REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggesstions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any oenalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
03-04-2013	Final Report	1-Jul-2002 - 30-Sep-2008

4. TITLE AND SUBTITLE
Science of Land Target Spectral Signatures

5a. CONTRACT NUMBER
DAAD19-02-2-0012

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
106011

6. AUTHORS
Dr. J. Michael Cathcart

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
Georgia Tech Research Corporation
Office of Sponsored Programs
505 Tenth Street NW
Atlanta, GA 30332 -0420

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
43578-EL-MUR.1

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
This MURI program progress report covers the fifteen months from 1 August 2007 to 30 September 2008. The overall purpose of this program is a study of the science underlying the signatures generated by land targets (both natural and manmade). Specifically, the objective is to understand the dependency of hyperspectral and hyperspectral/polarimetric signatures on variations in environmental conditions. An additional objective involves algorithm development and focuses on the identification of new target discriminants and the evaluation of their

15. SUBJECT TERMS
hyperspectral phenomenology, spectral signatures, landmines, spectral processing, detection algorithms, sensor fusion, spectral signature modeling

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			J. Cathcart
UU	UU	UU	UU		19b. TELEPHONE NUMBER: 404-407-6028

Report Title

Science of Land Target Spectral Signatures

ABSTRACT

This MURI program progress report covers the fifteen months from 1 August 2007 to 30 September 2008. The overall purpose of this program is a study of the science underlying the signatures generated by land targets (both natural and manmade). Specifically, the objective is to understand the dependency of hyperspectral and hyperspectral/polarimetric signatures on variations in environmental conditions. An additional objective involves algorithm development and focuses on the identification of new target discriminants and the evaluation of their utility for target detection and sensor fusion.

The phenomenology research continued to focus on spectroscopic soil measurements, optical property analyses, field data analysis, physics-based signature modeling, and synthetic scene generation for concealed targets. The algorithm work extended the signature-based detection research and conducted investigations into patterned based detection methods. Additional algorithm work exploiting local spatial-spectral background characteristics was conducted. Two endmember selection algorithms were developed based on a sparsity promoting approach. Fusion research continued to investigate methods for fusing SAR and hyperspectral data through the Choquet Integral approach. Target detection performance improved using this data fusion approach. Clutter complexity research continued to investigate metrics to tie image complexity to detection algorithm performance and continued work on methods for generating diverse sample images.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

J.B. Broadwater and R. Chellappa, "Hybrid Detectors for Subpixel Targets," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 29, No. 11, pp 1891-1903, November 2007.

Received Paper

J.B. Broadwater and R. Chellappa, "An Adaptive Threshold Method via Extreme Value Theory," IEEE Transactions on Signal Processing, May 2008.

A. Mendez-Vazquez, P. D. Gader, J. M. Keller, and K. Chamberlin, "Minimum Classification Error Training for Choquet Integrals with Applications to Landmine Detection," IEEE Transactions on Fuzzy Systems, Vol. 16, No. 1, pp 225-238, February 2008.

TOTAL:

A. Zare, J. Bolton, P. Gader, and M. Schatten, "Vegetation Mapping for Landmine Detection Using Long Wave Hyperspectral Imagery," IEEE Geoscience and Remote Sensing Letters, Vol. 46, No. 1, pp 172-178, January 2008.

A. Zare and P. Gader, "Hyperspectral Band Selection and Endmember Detection Using Sparsity Promoting Priors," IEEE Geoscience and Remote Sensing Letters, Vol 5, No. 2, pp 256-260, April 2008.

A. Zare and P. Gader, "PCE: Piece-wise Convex Endmember Detection," IEEE Trans. on Geoscience and Remote Sensing, (submitted).

J. Bolton, P. Gader, and J. Wilson, "Discrete Choquet Integral as a Distance Metric," IEEE Transactions on Fuzzy Systems, Vol. 14, No. 4, pp 1007-1110, August 2008.

J. Bolton and P. Gader, "The Benefits of Context Estimation Using the Random Set Framework for Target Spectra Detection in Hyperspectral Imagery," IEEE Transactions on Geoscience and Remote Sensing, (submitted).

O. O. Fadiran and P. Molnar, "Adaptive Sampling by Histogram Equalization," IEEE Transaction on Evolutionary Computation, (in preparation).

A.M. Thomas and J.M. Cathcart, "Applications of Grid Pattern Matching to the Detection of Buried Landmines," IEEE Transactions on Geoscience and Remote Sensing, (in progress).

Number of Papers published in peer-reviewed journals: 10.00

(b) Papers published in non-peer-reviewed journals (N/A for none)

J. Bolton and P. D. Gader, "Application of Context-Based Classifier to Hyperspectral Imagery for Mine Detection," Proceedings of the SPIE Conference on Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XIII, Orlando, FL, Vol. 6953, April 2008.

Received Paper

J. Bolton and P. D. Gader, "Random Set Model for Context-Based Classification," IEEE World Congress on Computational Intelligence, Hong Kong, pp 1999-2006, June 2008.

J. Bolton and P. D. Gader, "The Benefits of Context Estimation for Target Spectra Detection in Hyperspectral Imagery," Proceedings of IEEE Geoscience and Remote Sensing, Boston, MA, July 2008 (in press).

TOTAL:

A. Zare and P. D. Gader, "Endmember Detection using the Dirichlet Process," International Conference on Pattern Recognition, December 2008 (in press).

A.M. Thomas and J. M. Cathcart, "Adaptive Spatial Sampling Schemes for the Detection of Minefields in Hyperspectral Imagery," Proceedings of the SPIE Conference on Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XIII, vol. 6953, Orlando, FL, April 2008.

A.M. Thomas and J. M. Cathcart, "Automated Determination of Scale and Orientation of Mine Field Grid," Proceedings of the SPIE Conference on Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XIII, vol. 6953, Orlando, FL, April 2008.

S. Lane, J. M. Cathcart and J.T. Harrell "Exposure Effects on the Optical Properties of Building Materials," Proceedings of the SPIE Conference on Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XIII, vol. 6953, Orlando, FL, April 2008.

Number of Papers published in non peer-reviewed journals: 7.00

(c) Presentations

A.M. Thomas and J. M. Cathcart, "Spectral Signal Processing," 11th Annual Joint JUXOCO-ARO Landmine Detection Basic Research Review Meeting, Springfield, VA, January 2008.

J. Bolton, A. Zare, and P. D. Gader A.M. Thomas and J. M. Cathcart, "Stochastic Methods for Mine Detection with Hyperspectral Imagery," 11th Annual Joint JUXOCO-ARO Landmine Detection Basic Research Review Meeting, Springfield, VA, January 2008.

A.M. Thomas and J. M. Cathcart, "When Bad Results Happen to Good Methods: An examination of fundamental limitations to algorithms for the detection of landmines in hyperspectral Long-Wave IR imagery," Army Advanced Concept Workshop on Disturbed Soil Characterization and Exploitation, Atlanta, GA, January 2008.

J. Bolton, "Context-based Approach for the Recognition/ Discrimination of Disturbed Soil & Subsurface Targets," Army Advanced Concept Workshop on Disturbed Soil Characterization and Exploitation, Atlanta, GA, January 2008.

A. Zare, "False Alarm Reduction Using Vegetation Mapping for Disturbed Soil Detection in Long Wave Hyperspectral Imagery," Army Advanced Concept Workshop on Disturbed Soil Characterization and Exploitation, Atlanta, GA, January 2008.

Number of Presentations: 5.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received Paper

TOTAL:

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts): 0

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received Paper

TOTAL:

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts): 0

(d) Manuscripts

Received Paper

TOTAL:

Number of Manuscripts: 0.00

Books

Received Paper

TOTAL:

Patents Submitted

Patents Awarded

Awards

Graduate Students

NAME	PERCENT_SUPPORTED
Ken Camann	0.50
Brian Kocher	0.25
Joshua Broadwater	0.60
Hirsh Goldberg	0.50
Alina Zare	0.50
Jeremy Bolton	0.50
Andres Mendez-Vazquez	0.25
Alicia Paul	0.50
FTE Equivalent:	**3.60**
Total Number:	**8**

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
Oladipo Fadiran	0.50
FTE Equivalent:	**0.50**
Total Number:	**1**

Names of Faculty Supported

NAME	PERCENT_SUPPORTED	National Academy Member
Michael Cathcart	0.50	No
Rick Trebino	0.10	No
Rama Chellappa	0.10	No
Paul Gader	0.20	No
Peter Molnar	0.25	No
John Schott	0.02	No
Paul Lucey	0.02	No
FTE Equivalent:	**1.19**	
Total Number:	**7**	

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED	
Keith Prussing	0.25	No
FTE Equivalent:	**0.25**	
Total Number:	**1**	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period:	0.00
The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:.....	0.00
The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:.....	0.00
Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):	0.00
Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:	0.00
The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense	0.00
The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:	0.00

Names of Personnel receiving masters degrees

NAME
Total Number:

Names of personnel receiving PHDs

NAME
Alina Zare
Jeremy Bolton
Total Number: 2

Names of other research staff

NAME	PERCENT_SUPPORTED	
Alan Thomas	0.25	No
Sarah Lane	0.50	No
David Messinger	0.10	No
Tim Williams	0.20	No
Mark Wood	0.25	No
Scott Brown	0.10	No
David Pogorzala	0.10	No
FTE Equivalent:	**1.50**	
Total Number:	**7**	

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

See attachment.

Technology Transfer

Kernel Matched Signal Detectors for Hyperspectral Target Detection

Heesung Kwon

Nasser M. Nasrabadi

U.S. Army Research Laboratory, Attn: AMSRL-SE-SE

2800 Powder Mill Road, Adelphi, MD 20783, USA

Outline

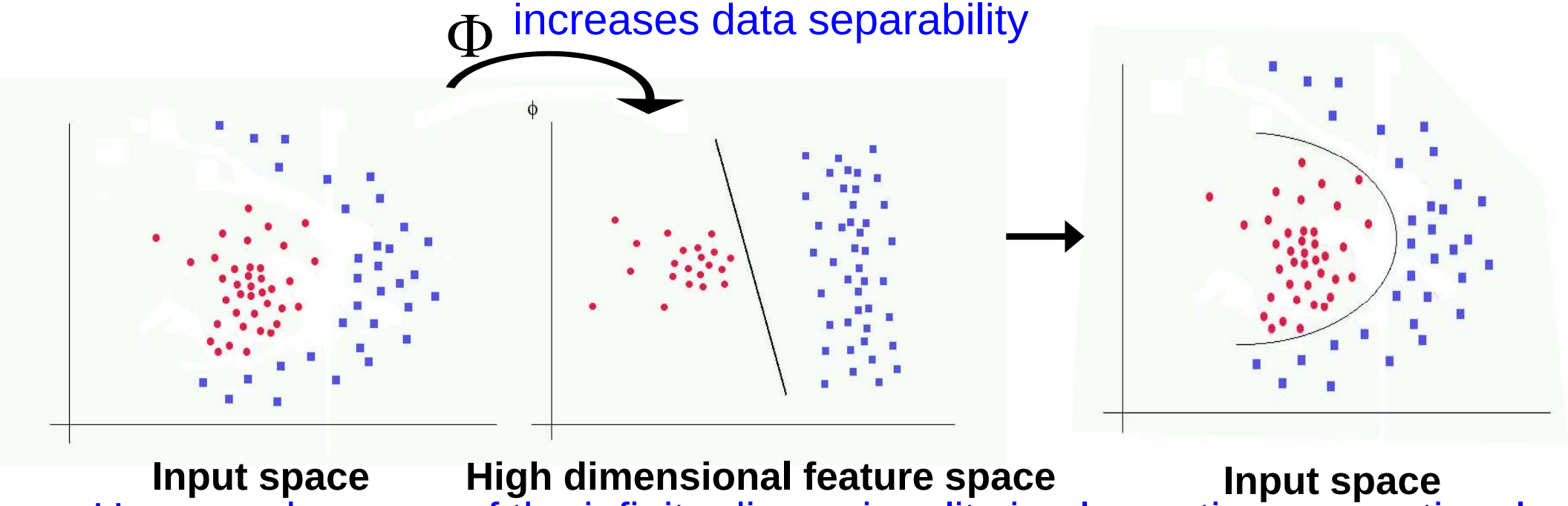
- Exploitation of Nonlinear Correlations Using Matched Filters
- Why Kernels
- Kernel Trick
- Conventional matched filters
- Kernel matched filters
- Detection results

Nonlinear Mapping of Data Exploitation of Nonlinear Correlations

- Nonlinear mapping Φ

$$\Phi : X \to F \qquad x \mapsto \Phi(x) = (\sqrt{\lambda_1}\psi_1(x), \sqrt{\lambda_2}\psi_2(x), \cdots)$$

- Statistical learning (VC): Mapping into a higher dimensional space increases data separability



Input space **High dimensional feature space** **Input space**

- However, because of the infinite dimensionality implementing conventional detectors in the feature space is not feasible using conventional methods
- Kernel trick : $k(x,y) = <\Phi(x), \Phi(y)>$
- Convert the detector expression into dot product forms →

Kernel-based nonlinear version of the conventional detector

Kernel Trick

$$k(\mathrm{x}, \mathrm{y}) = < \Phi(\mathrm{x}), \Phi(\mathrm{y}) >$$

- Consider 2-D input patterns $\mathrm{x} = (x_1, x_2)$, where $\mathrm{x} = (x_1, x_2) \in R^2$
- If a 2nd order monomial is used as the nonlinear mapping

$$\Phi : R^2 \rightarrow R^3, \Phi(\mathrm{x}) = \mathrm{x}^2$$

$$(x_1, x_2) \mapsto (z_1, z_2, z_3) := (x_1^2, \sqrt{2} x_1 x_2, x_2^2)$$

- Example of the kernel trick

$$< \Phi(\mathbf{x}), \Phi(\mathbf{y}) >= (x_1^2, \sqrt{2} x_1 x_2, x_2^2)(y_1^2, \sqrt{2} y_1 y_2, y_2^2)^T = x_1^2 y_1^2 + 2 x_1 x_2 y_1 y_2 + x_2^2 y_2^2$$

$$= ((x_1, x_2)(y_1, y_2)^T)^2 =< \mathbf{x}, \mathbf{y} >^2 := k(\mathbf{x}, \mathbf{y})$$

$$k(\mathrm{x}, \mathrm{y}) = < \Phi(\mathrm{x}), \Phi(\mathrm{y}) >, \; k : \text{kernel function}$$

- This property generalizes for $\mathbf{x}, \mathbf{y} \in R^N$ and $d \in R$

$$k(\mathrm{x}, \mathrm{y}) =< \mathrm{x}, \mathrm{y} >^d$$

Examples of Kernels

1. Gaussian RBF kernel: $k(\mathbf{x},\mathbf{y}) = \exp(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}) = \Phi(\mathrm{x})\Phi(\mathrm{y})$

Possible realization of

$$\Phi(\mathrm{x}) = (\sqrt{\lambda_1}\psi_1(\mathrm{x}), \sqrt{\lambda_2}\psi_2(\mathrm{x}), \cdots)$$

2. Inverse multiquadric kernel: $k(\mathbf{x},\mathbf{y}) = \frac{1}{\sqrt{\|\mathrm{x}-\mathrm{y}\|^2 + c^2}}$

3. Spectral angle-based kernel: $k(\mathbf{x},\mathbf{y}) = \frac{\mathrm{xy}}{\|\mathrm{x}\|\|\mathrm{y}\|}$

4. Polynomial kernel: $k(\mathbf{x},\mathbf{y}) = ((\mathbf{x}\cdot\mathbf{y}) + \theta)^d$

Matched Subspace Detection (MSD)

- Consider a linear mixed model:

$$H_0 : \mathbf{y} = \mathbf{B}\boldsymbol{\zeta} + \mathbf{n}, \qquad \text{Target absent} \approx N(\mathbf{B}\boldsymbol{\zeta}, \sigma^2\mathbf{I})$$

$$H_1 : \mathbf{y} = \mathbf{T}\boldsymbol{\theta} + \mathbf{B}\boldsymbol{\zeta} + \mathbf{n} \qquad \text{Target present} \approx N(\mathbf{T}\boldsymbol{\theta} + \mathbf{B}\boldsymbol{\zeta}, \sigma^2\mathbf{I})$$

- where $\mathbf{T}$ and $\mathbf{B}$ represent matrices whose column vectors span the target and the background subspaces
  $\boldsymbol{\zeta}$ and $\boldsymbol{\theta}$ are unknown vectors of coefficients, $\mathbf{n}$ is a Gaussian random noise distributed as $N(0, \sigma^2\mathbf{I})$
- The log Generalized likelihood ratio test (GLRT) is given by

$$L_2(\mathbf{y}) = \frac{\mathrm{p}(\mathrm{y} \mid \text{signal present})}{\mathrm{p}(\mathrm{y} \mid \text{signal absent})} = \frac{\mathbf{y}^T(\mathbf{I} - \mathbf{P}_\mathbf{B})\mathbf{y}}{\mathbf{y}^T(\mathbf{I} - \mathbf{P}_\mathbf{BT})\mathbf{y}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta$$

- where $\mathbf{P}_\mathbf{B} = \mathbf{B}\mathbf{B}^T$, $\mathbf{P}_\mathbf{TB} = [\mathbf{T}\ \mathbf{B}]\{[\mathbf{T}\ \mathbf{B}]^T[\mathbf{T}\ \mathbf{B}]\}^{-1}[\mathbf{T}\ \mathbf{B}]$

Kernel Matched Subspace Detection

- Define the matched subspace detector in the feature space
- To kernelize we use the kernel PCA, and kernel function properties as shown below

$$L_2(\Phi(y)) = \frac{\Phi(y)^T(I_\Phi - P_{B_\Phi})\Phi(y)}{\Phi(y)^T(I_\Phi - P_{B_\Phi T_\Phi})\Phi(y)} = \frac{\Phi(y)^T(I_\Phi - B_\Phi B_\Phi^T)(y)}{\Phi(y)^T\begin{bmatrix}T_\Phi & B_\Phi\end{bmatrix}\begin{bmatrix}T_\Phi^T T_\Phi & T_\Phi^T B_\Phi \\ B_\Phi^T T_\Phi & B_\Phi^T B_\Phi\end{bmatrix}^{-1}\begin{bmatrix}T_\Phi^T \\ B_\Phi^T\end{bmatrix}\Phi(y)}$$

$$B_\Phi = Z_{B_\Phi}\beta, T_\Phi = Z_{T_\Phi}\tau, \quad B_\Phi^T\Phi(y) = \beta^T k(Z_B, y), \text{ and } T_\Phi^T\Phi(y) = \tau^T k(Z_T, y)$$

$$\therefore \Phi(y)^T B_\Phi B_\Phi \Phi(y) = k(Z_B, y)^T \beta\beta^T k(Z_B, y)$$

$$L_{2k} = \frac{k(y, y) - k(Z_B, y)^T \beta\beta^T k(Z_B, y)}{k(y, y) - \begin{bmatrix}\tau^T k(Z_B, y) & \beta^T k(Z_B, y)\end{bmatrix}^T \Lambda_1^{-1}\begin{bmatrix}\tau^T k(Z_B, y) \\ \beta^T k(Z_B, y)\end{bmatrix}}$$

$$\Lambda_1 = \begin{bmatrix}\tau^T K(Z_T, Z_T)\tau & \tau^T K(Z_T, Z_B)\beta \\ \beta^T K(Z_B, Z_T)\tau & \beta^T K(Z_B, Z_B)\beta\end{bmatrix}$$

MSD vs. Kernel MSD

- GLRT for the MSD:

$$L_2(\mathbf{y}) = \frac{\mathbf{y}^T(\mathbf{I} - \mathbf{P_B})\mathbf{y}}{\mathbf{y}^T(\mathbf{I} - \mathbf{P_{BT}})\mathbf{y}}$$

- Nonlinear GLRT for the MSD in feature space:

$$L_2(\Phi(\mathbf{y})) = \frac{\Phi(\mathbf{y})^T(\mathbf{I}_\Phi - \mathbf{P}_{\mathbf{B}_\Phi})\Phi(\mathbf{y})}{\Phi(\mathbf{y})^T(\mathbf{I}_\Phi - \mathbf{P}_{\mathbf{B}_\Phi\mathbf{T}_\Phi})\Phi(\mathbf{y})}$$

- Kernelized GLRT for the kernel MSD:

$$L_{2k} = \frac{k(y,y) - k(Z_B,y)^T \beta\beta^T k(Z_B,y)}{k(y,y) - \begin{bmatrix}\tau^T k(Z_B,y) & \beta^T k(Z_B,y)\end{bmatrix}^T \Lambda_1^{-1} \begin{bmatrix}\tau^T k(Z_B,y) \\ \beta^T k(Z_B,y)\end{bmatrix}}$$

Orthogonal Subspace Projector vs. Kernel OSP

- The model in the nonlinear feature space is

$$H_{0_\Phi} : \Phi(\mathbf{y}) = \mathbf{B}_\Phi \zeta_\Phi + \mathbf{n}_\Phi, \quad \text{Target absent}$$
$$H_{1_\Phi} : \Phi(\mathbf{y}) = \mathrm{s}_\Phi \mu_\Phi + \mathbf{B}_\Phi \zeta_\Phi + \mathbf{n}_\Phi \quad \text{Target present}$$

- The MLE for μ_ϕ in feature space is given as

$$\mu_\Phi = \frac{\Phi(\mathrm{s})^T (\mathbf{I}_\Phi - \mathbf{P}_{\mathbf{B}_\Phi}) \Phi(\mathbf{y})}{\Phi(\mathrm{s})^T (\mathbf{I}_\Phi - \mathbf{P}_{\mathbf{B}_\Phi}) \Phi(\mathrm{s})} \underset{H_{0\Phi}}{\overset{H_{1\Phi}}{\gtrless}} \eta$$

- The kernel version of μ_ϕ is given as

$$\mu_k = \frac{\mathrm{k}(\mathrm{s},\mathrm{y}) - \mathrm{k}(Z_B,\mathrm{s})^T \beta\beta^T \mathrm{k}(Z_B,\mathrm{y})}{\mathrm{k}(\mathrm{s},\mathrm{s}) - \mathrm{k}(Z_B,\mathrm{s})^T \beta\beta^T \mathrm{k}(Z_B,\mathrm{s})}$$

Linear Spectral Matched Filter & Nonlinear Spectral Matched Filter

- Spectral signal model

$$H_0 : \mathbf{x} = \mathbf{n}, \qquad a = 0: \text{ no target,}$$
$$H_1 : \mathbf{x} = a\mathbf{s} + \mathbf{n} \qquad a > 0: \text{ target present}$$

$\mathbf{n}$: background clutter noise
$\mathbf{s}$: target spectral signature,

- Linear matched filter is given as:

$$y(\mathrm{x}) = \mathrm{w}^{\mathrm{T}}\mathrm{x} = \frac{\mathrm{s}^{\mathrm{T}}\mathrm{C}^{-1}\mathrm{x}}{\mathrm{s}^{\mathrm{T}}\mathrm{C}^{-1}\mathrm{s}},$$

- In the feature space, the equivalent signal model

$$H_0 : \Phi(\mathrm{x}) = \mathrm{n}_\Phi, \qquad \text{No target}$$
$$H_1 : \Phi(\mathrm{x}) = a_\Phi \Phi(\mathrm{s}) + \mathrm{n}_\Phi \qquad \text{Target present}$$

- Output of the matched filter in the feature space

$$y(\Phi(\mathrm{x})) = \mathrm{w}_\Phi^{\mathrm{T}}\Phi(\mathrm{x}) = \frac{\Phi(\mathrm{s})\mathrm{C}_\Phi^{-1}\Phi(\mathrm{x})}{\Phi(\mathrm{s})\mathrm{C}_\Phi^{-1}\Phi(\mathrm{s})}$$

Kernelization of Spectral Matched Filter in Feature space

- Using the following properties of PCA and Kernel PCA

$$\mathbf{C}_{\Phi}^{-1} = \mathbf{V}_{\Phi}\Lambda^{-1}\mathbf{V}_{\Phi}^{\mathrm{T}}, \qquad \mathbf{V}_{\Phi} = [\mathbf{v}_{\Phi}^{1}, \mathbf{v}_{\Phi}^{2}, \cdots, \mathbf{v}_{\Phi}^{M}]$$

- Each eigenvector can be represented in terms of the input data

$$\mathbf{V}_{\Phi} = \mathbf{X}_{\Phi}\mathbf{B}, \qquad \mathbf{B} = [\mathbf{b}^{1}, \mathbf{b}^{2}, \cdots, \mathbf{b}^{M}]$$

- Inverse Covariance matrix is now

$$\mathbf{C}_{\Phi}^{-1} = \mathbf{X}_{\Phi}\mathbf{B}\Lambda^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{X}_{\Phi}^{\mathrm{T}}$$

- Kernel matrix, $\mathbf{K}$, spectral decomposition (kernel PCA)

$$\mathbf{K}^{-1} = \frac{1}{M}\mathbf{B}\Lambda^{-1}\mathbf{B}^{T}, \quad \text{where } \mathbf{K}(\mathbf{X},\mathbf{X}) = \mathbf{K}_{ij} = k(\mathbf{x}_i,\mathbf{x}_j) \quad \mathbf{x}_i,\mathbf{x}_j \in \mathbf{X}$$

$$y(\Phi(\mathrm{x})) = \frac{\Phi(\mathrm{s})^{\mathrm{T}}\mathrm{X}_{\Phi}\mathrm{B}\Lambda^{-1}\mathrm{B}^{\mathrm{T}}\mathrm{X}_{\Phi}^{\mathrm{T}}\Phi(\mathrm{x})}{\Phi(\mathrm{s})\mathrm{X}_{\Phi}\mathrm{B}\Lambda^{-1}\mathrm{B}^{\mathrm{T}}\mathrm{X}_{\Phi}^{\mathrm{T}}\Phi(\mathrm{s})},$$

- The kernelized version of matched filter

$$y(\mathrm{k}_{\mathrm{x}}) = \frac{\mathrm{k}(\mathrm{X},\mathrm{s})\mathrm{K}^{-1}\mathrm{k}(\mathrm{X},\mathrm{x})}{\mathrm{k}(\mathrm{X},\mathrm{s})\mathrm{K}^{-1}\mathrm{k}(\mathrm{X},\mathrm{s})},$$

$$\mathrm{k}(\mathrm{X},\mathrm{s}) = (k(\mathrm{x}_1,\mathrm{s}), k(\mathrm{x}_2,\mathrm{s}), \cdots, k(\mathrm{x}_{\mathrm{N}},\mathrm{s}))^{\mathrm{T}}$$
$$\mathrm{k}(\mathrm{X},\mathrm{x}) = (k(\mathrm{x}_1,\mathrm{x}), k(\mathrm{x}_2,\mathrm{x}), \cdots, k(\mathrm{x}_{\mathrm{N}},\mathrm{x}))^{\mathrm{T}}$$

Conventional MF vs. Kernel MF

- Conventional spectral matched filter

$$y(x) = w^T x = \frac{s^T C^{-1} x}{s^T C^{-1} s},$$

- Nonlinear matched filter

$$y(\Phi(x)) = w_\Phi^T \Phi(x) = \frac{\Phi(s) C_\Phi^{-1} \Phi(x)}{\Phi(s) C_\Phi^{-1} \Phi(s)}$$

- Kernel matched filter

$$y(k_x) = \frac{k(X, s) K^{-1} k(X, x)}{k(X, s) K^{-1} k(X, s)}$$

Adaptive Subspace Detection (ASD) & Nonlinear ASD

- Consider a linear mixed model:

$$H_0 : \mathbf{r} = \mathbf{n}, \qquad \text{Target absent} \approx (0, \mathbf{C})$$

$$H_1 : \mathbf{r} = \mathbf{U}\theta + \sigma\,\mathbf{n} \qquad \text{Target present} \approx (\mathbf{U}\theta, \sigma^2\mathbf{C})$$

where $\mathbf{U}$ represent the target subspace and $\mathbf{C}$ is the background covariance.

- The ASD is given by

$$D_{ASD}(\mathbf{r}) = \frac{\mathbf{r}^T \hat{\mathbf{C}}^{-1}\mathbf{U}(\mathbf{U}^T\hat{\mathbf{C}}^{-1}\mathbf{U})^{-1}\mathbf{U}^T\hat{\mathbf{C}}^{-1}\mathbf{r}}{\mathbf{r}^T\hat{\mathbf{C}}^{-1}\mathbf{r}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta$$

- The model in the nonlinear feature space is

$$H_{0_\Phi} : \Phi(\mathbf{r}) = \mathbf{n}_\Phi, \qquad \text{Target absent}$$

$$H_{1_\Phi} : \Phi(\mathbf{r}) = \mathbf{U}_\Phi\theta_\Phi + \sigma\,\mathbf{n}_\Phi \qquad \text{Target present}$$

- The ASD in feature space is given as

$$D_{ASD}(\Phi(\mathbf{r})) = \frac{\Phi(\mathbf{r})^T\hat{\mathbf{C}}_\Phi^{-1}\mathbf{U}_\Phi(\mathbf{U}_\Phi^T\hat{\mathbf{C}}_\Phi^{-1}\mathbf{U}_\Phi)^{-1}\mathbf{U}_\Phi^T\hat{\mathbf{C}}_\Phi^{-1}\Phi(\mathbf{r})}{\Phi(\mathbf{r})^T\hat{\mathbf{C}}^{-1}\Phi(\mathbf{r})} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{ASD}$$

ASD vs. Kernel ASD

- GLRT for the ASD:

$$D_{ASD}(\mathrm{r}) = \frac{\mathrm{r}^{\mathrm{T}}\,\hat{\mathbf{C}}^{-1}\mathbf{U}(\mathbf{U}^{\mathrm{T}}\hat{\mathbf{C}}^{-1}\mathbf{U})^{-1}\,\mathbf{U}^{\mathrm{T}}\hat{\mathbf{C}}^{-1}\,\mathrm{r}}{\mathrm{r}^{\mathrm{T}}\,\hat{\mathbf{C}}^{-1}\,\mathrm{r}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta$$

- Nonlinear GLRT for the ASD in feature space:

$$D_{ASD}(\Phi(\mathrm{r})) = \frac{\Phi(\mathrm{r})^{\mathrm{T}}\,\hat{\mathbf{C}}_{\Phi}^{-1}\mathbf{U}_{\Phi}(\mathbf{U}_{\Phi}^{\mathrm{T}}\hat{\mathbf{C}}_{\Phi}^{-1}\mathbf{U}_{\Phi})^{-1}\,\mathbf{U}_{\Phi}^{\mathrm{T}}\hat{\mathbf{C}}_{\Phi}^{-1}\,\Phi(\mathrm{r})}{\Phi(\mathrm{r})^{\mathrm{T}}\,\hat{\mathbf{C}}^{-1}\,\Phi(\mathrm{r})} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{ASD}$$

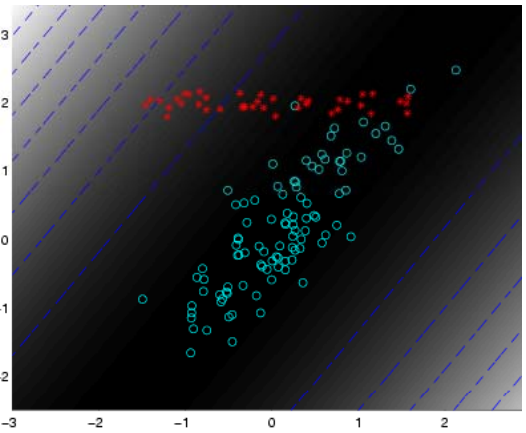
- Kernelized GLRT for the kernel ASD:

$$D_{KASD}(\mathrm{r}) = \frac{\mathrm{K_r}\,[\tau^{\mathrm{T}}\mathrm{K}(\mathrm{X},\mathrm{Y})^{\mathrm{T}}\mathrm{K_b}(\mathrm{X},\mathrm{X})^{-1}\mathrm{K}(\mathrm{X},\mathrm{Y})\tau\,]^{-1}\,\mathrm{K_r^T}}{\mathrm{k}(\mathrm{r},\mathrm{X})^{\mathrm{T}}\mathrm{K_b}(\mathrm{X},\mathrm{X})^{-1}\mathrm{k}(\mathrm{r},\mathrm{X})}$$

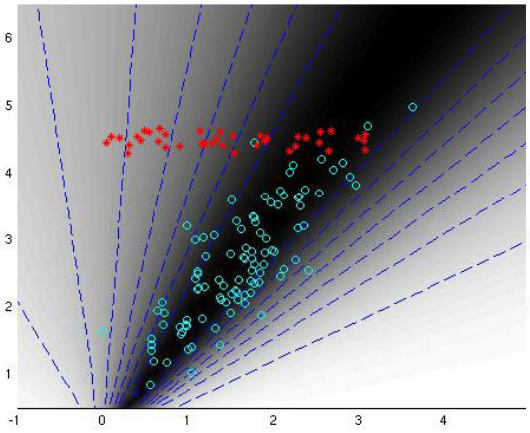
τ corresponds to eigenvectors of kernel matrix K(Y,Y)

A 2-D Gaussian Toy Example

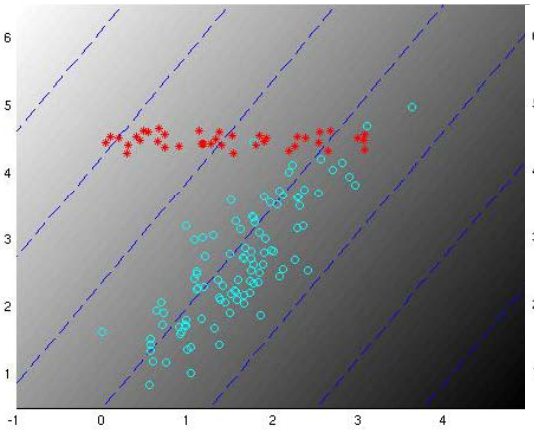
- Red dots belong to class H_1, blue dots belong to H_0



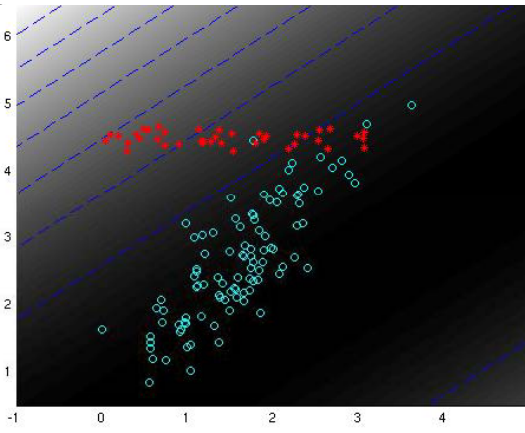
(a) MSD



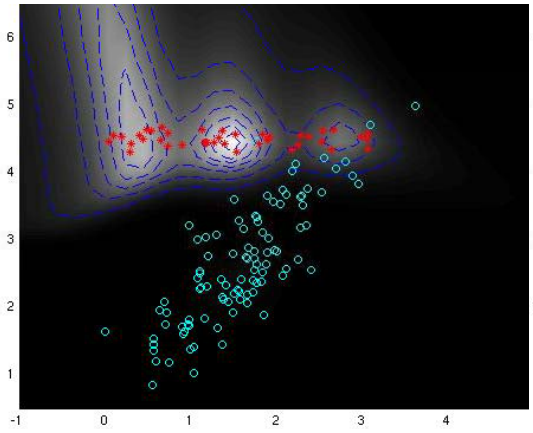
(c) ASD



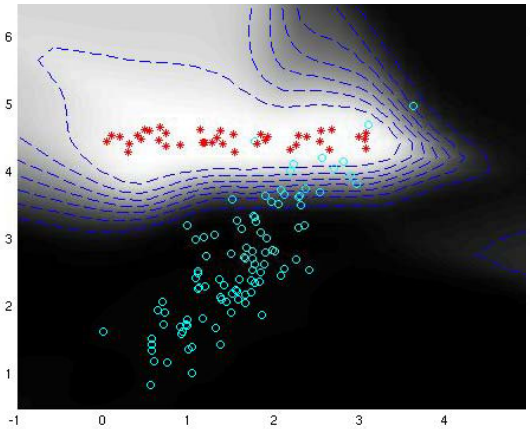
(e) OSP



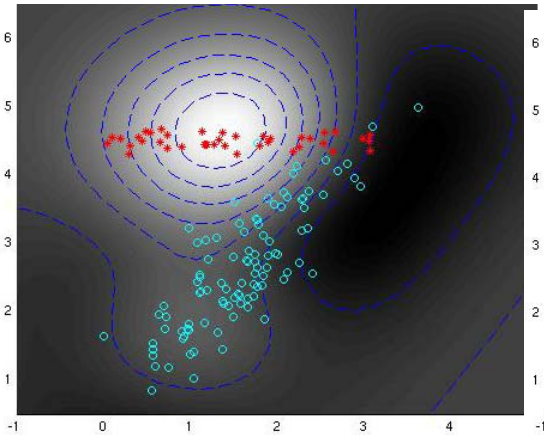
(g) SMF



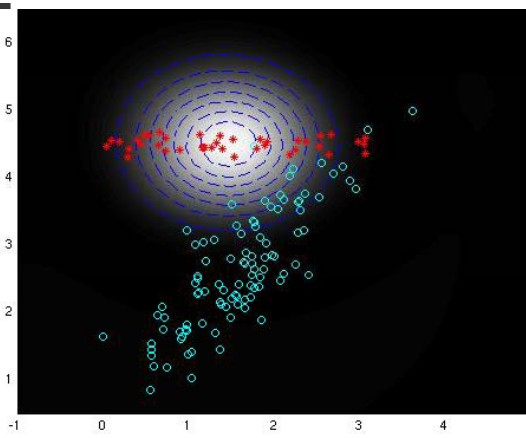
(b) KMSD



(d) KASD



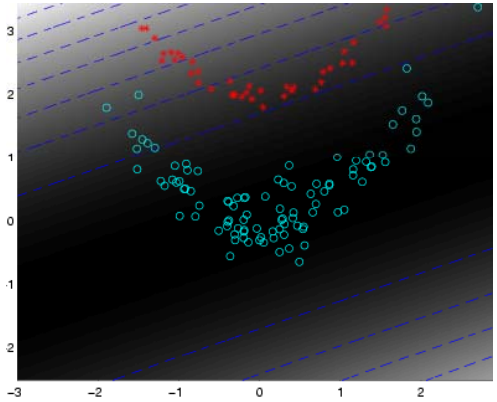
(f) KOSP



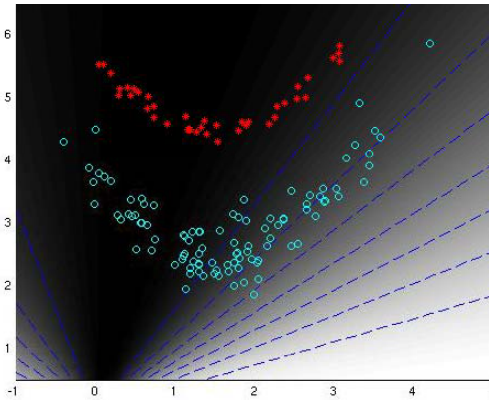
(h) KSMF

A 2-D Toy Example

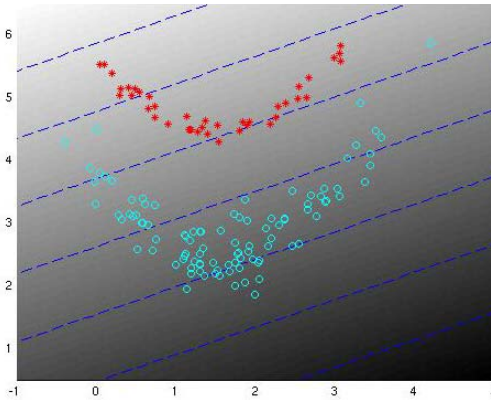
- Red dots belong to class H_1, blue dots belong to H_0



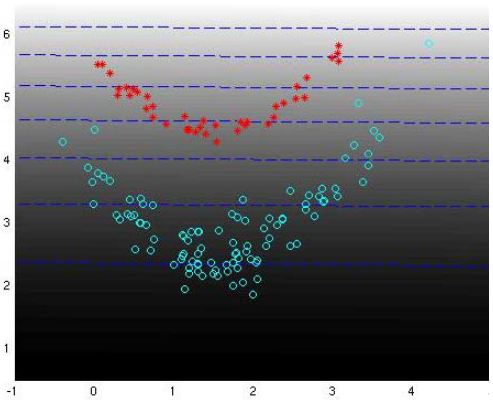
(a) MSD



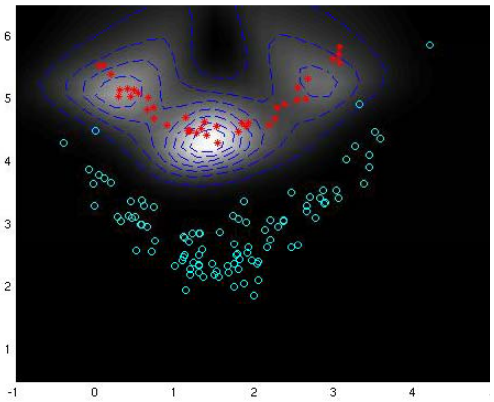
(c) ASD



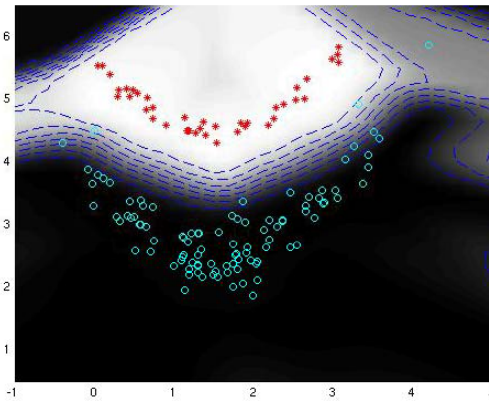
(e) OSP



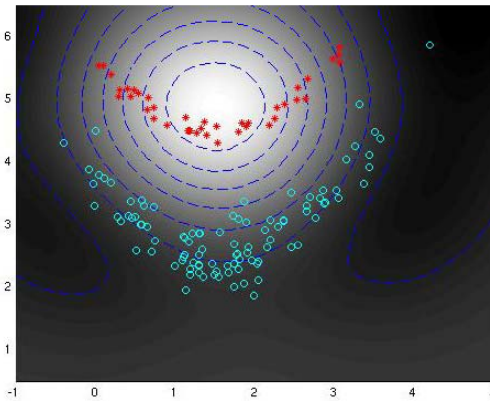
(g) SMF



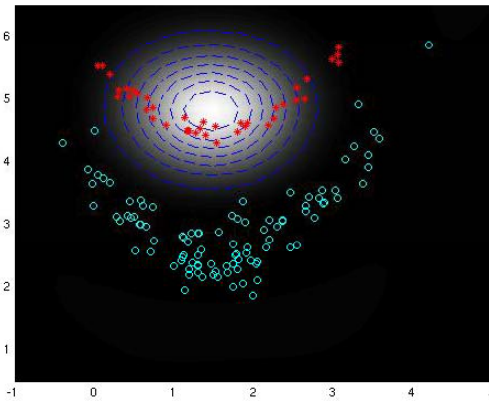
(b) KMSD



(d) KASD

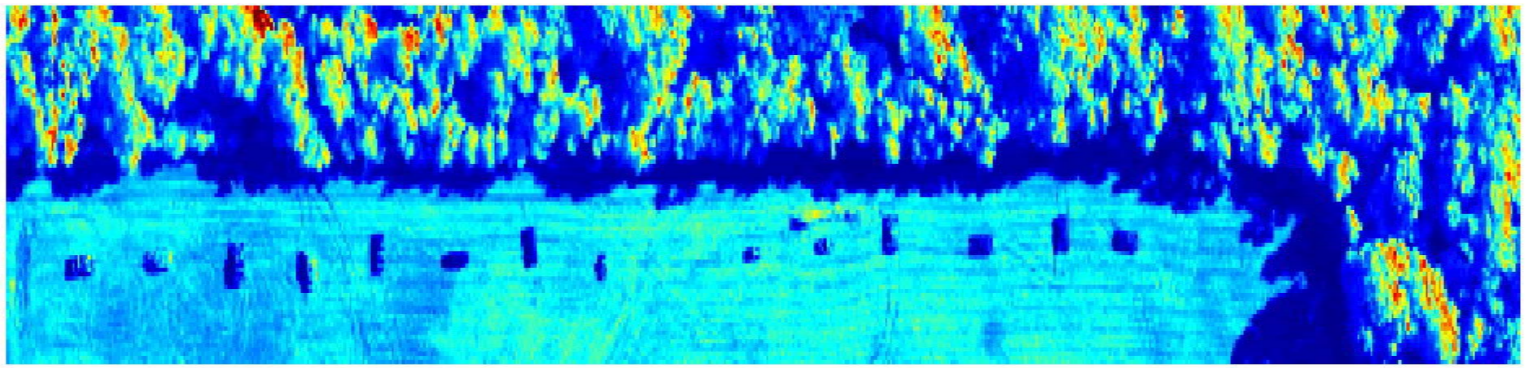


(f) KOSP

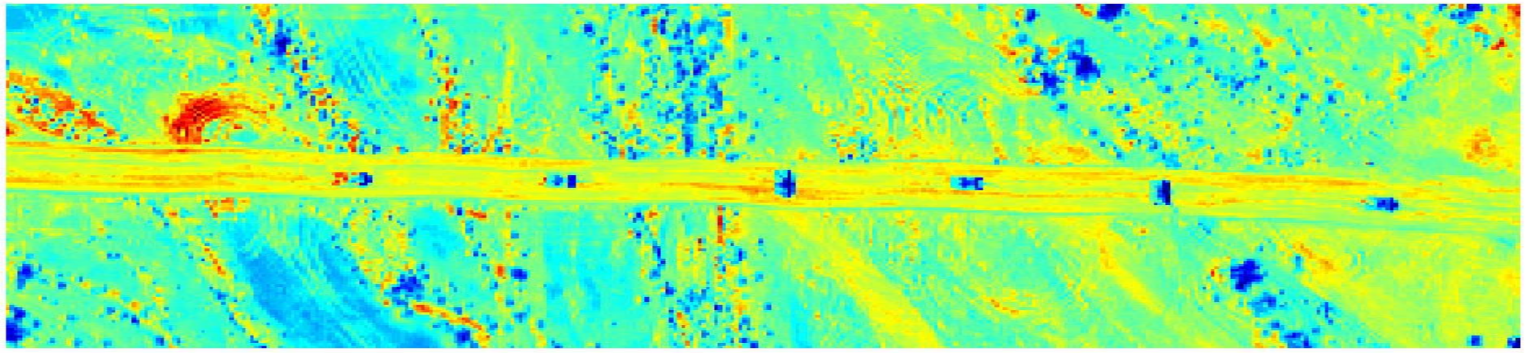


(h) KSMF

Test Images

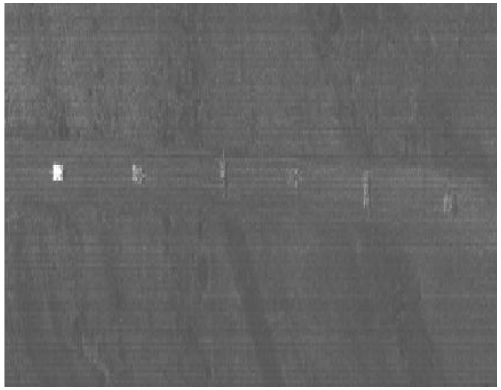


Forest Radiance I



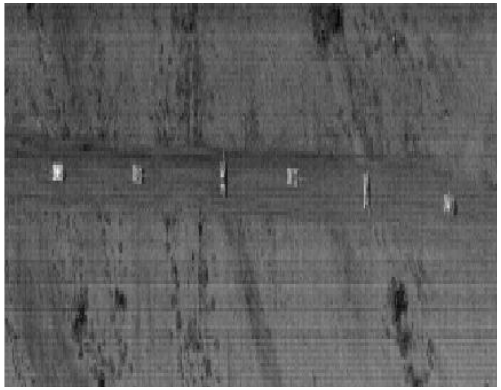
Desert Radiance II

Results for DR-II Image



(a) MSD

(c) ASD



(e) OSP

(g) SMF

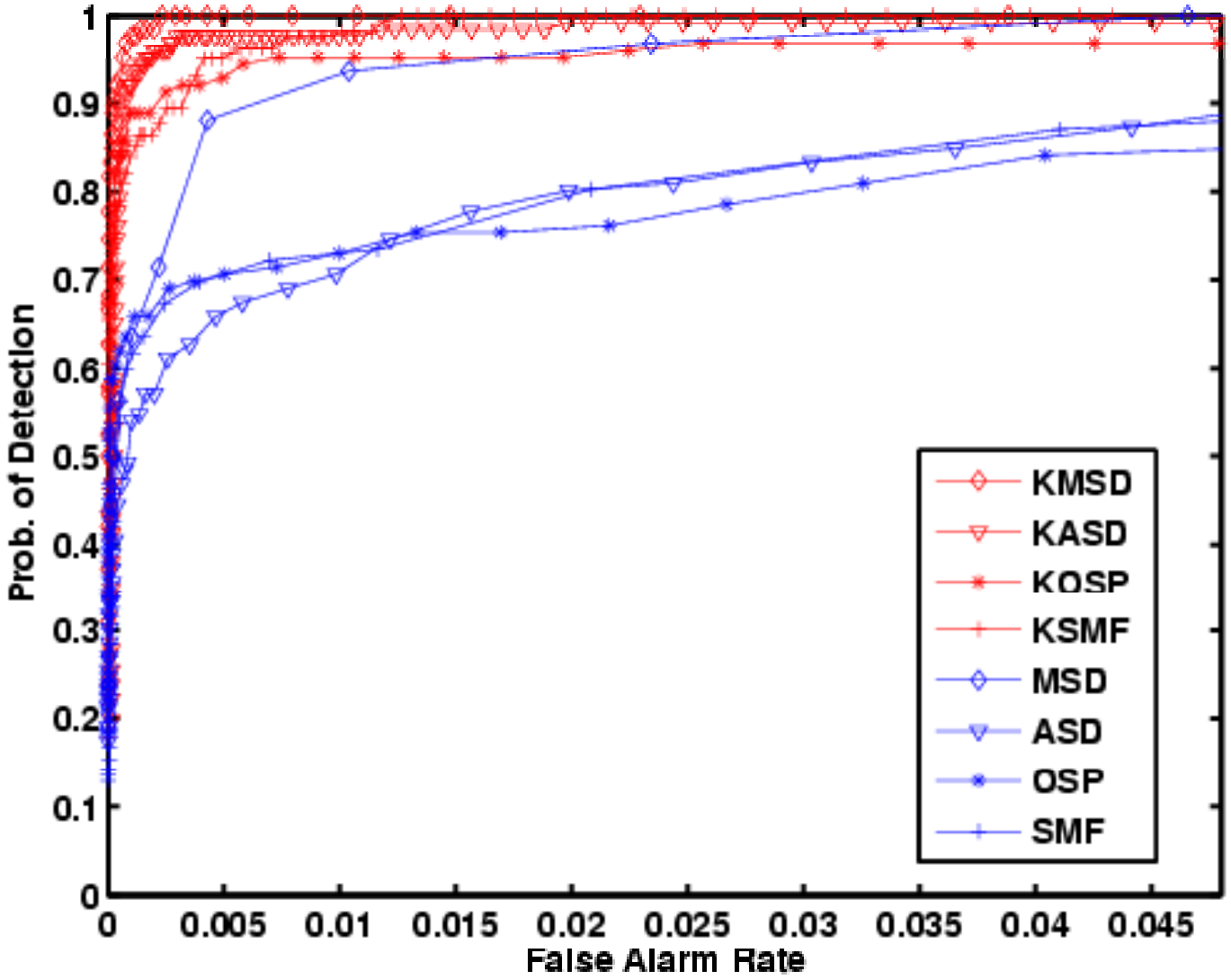
(b) KMSD

(d) KASD

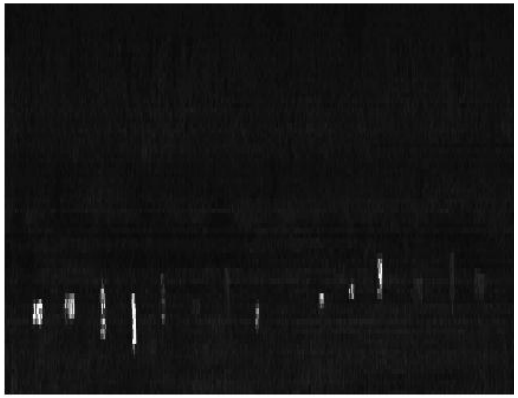
(f) KOSP

(h) KSMF

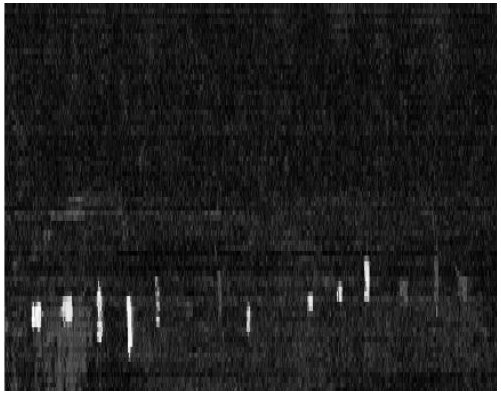
ROC Curves for DR-II Image



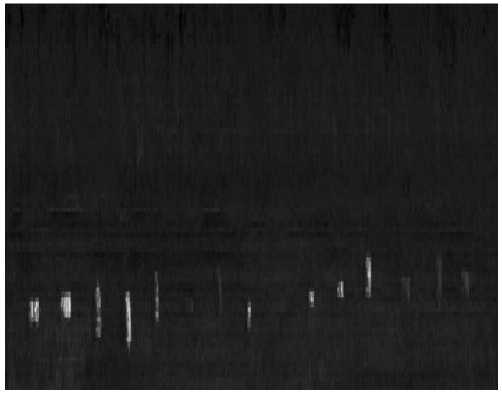
Results for FR-II Image



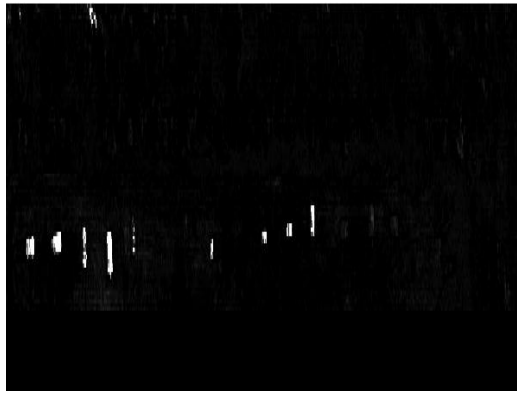
(a) MSD



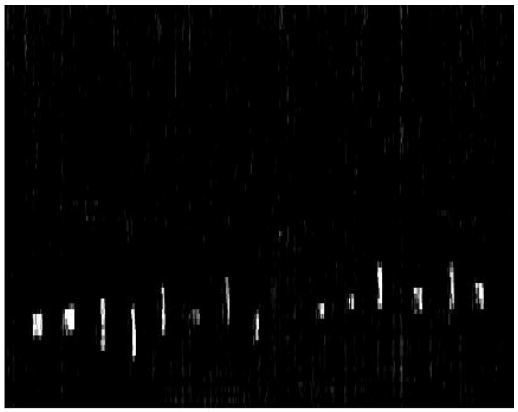
(c) ASD



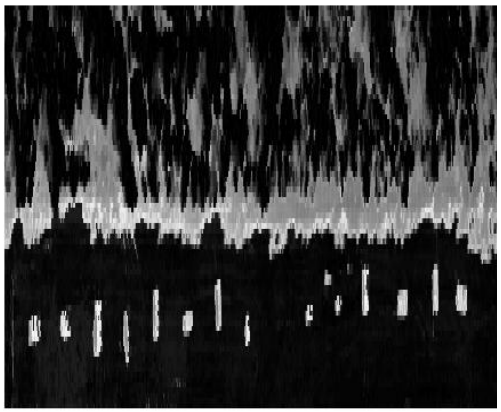
(e) OSP



(g) SMF

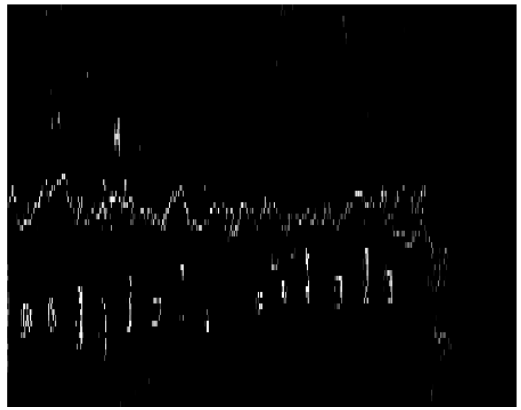


(b) KMSD



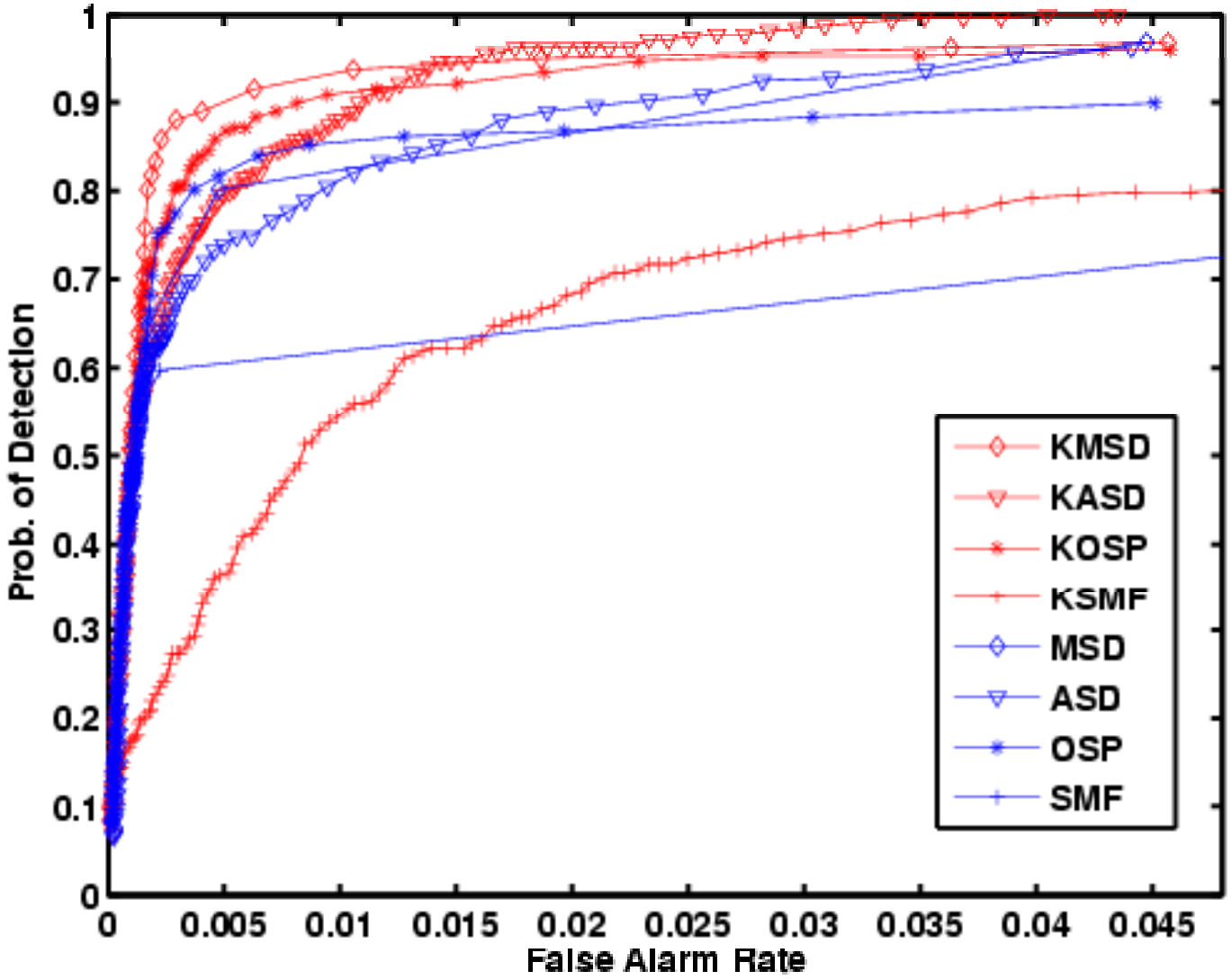
(d) KASD

(f) KOSP



(h) KSMF

ROC Curves for FR-II Image



Scene Anomaly Detection @ Ground Level Using HSI

Dalton Rosario
Army Research Laboratory

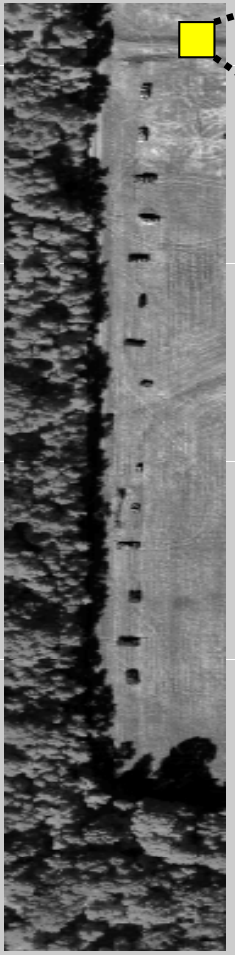
Rama Chellappa
University of Maryland

Outline

- Motivation/Idea
- New Family of Anomaly Detectors for HSI
  - Ground Vehicle Detection –Top View
  - Scene Anomaly Detection @ Ground Level
- Final Remarks

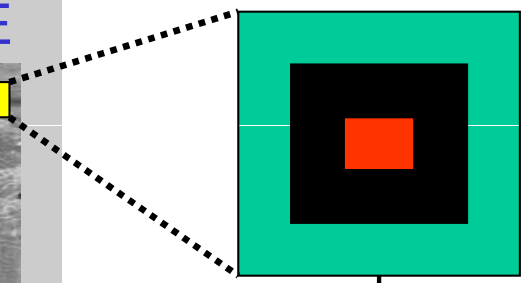
Physical Motivation

HYDICE



Aberdeen, MD

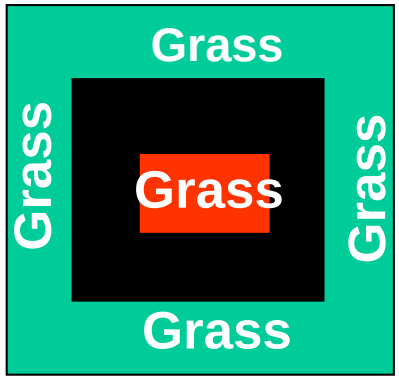
Visible to SWIR Hyperspectral Data



Inside-Outside Windows

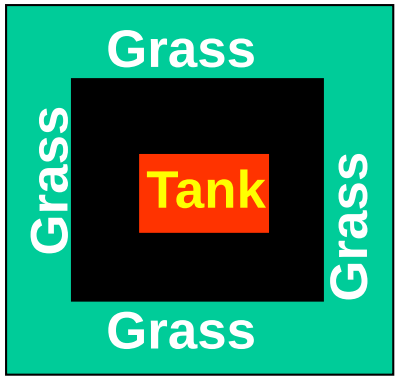
Local Anomaly Detector (e.g., FLD)

Case 1



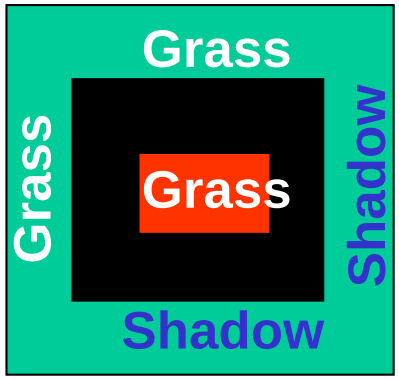
Non-Anomaly

Case 2

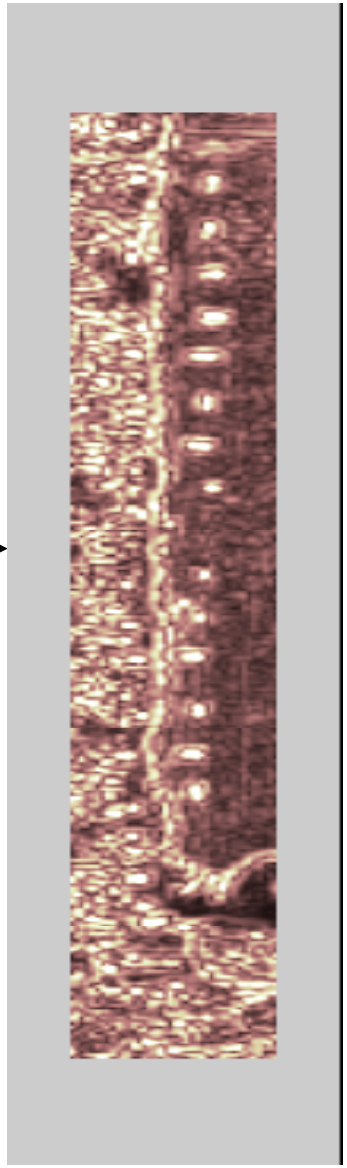


Anomaly

Case 3



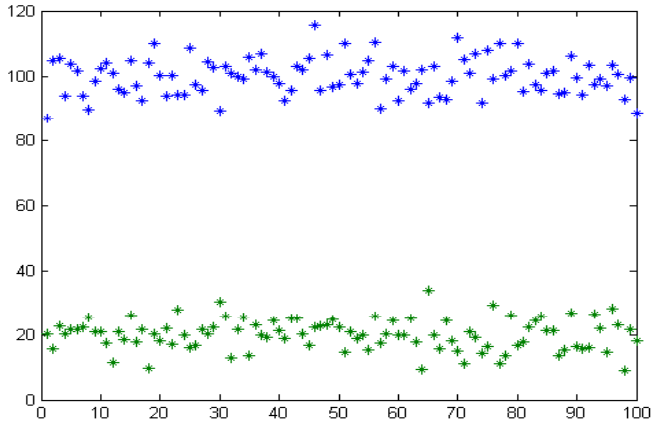
Anomaly

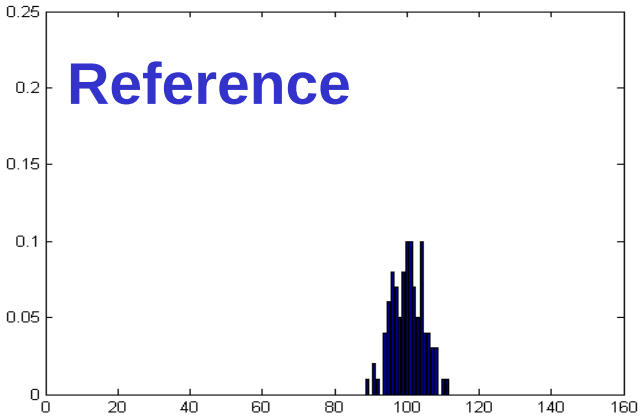


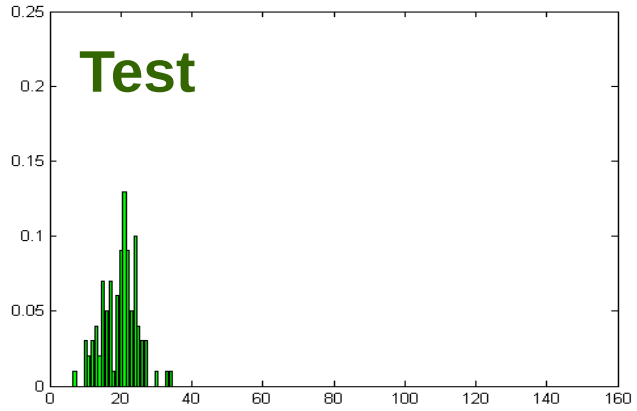
Output Surface

Statistical Motivation

Case 2

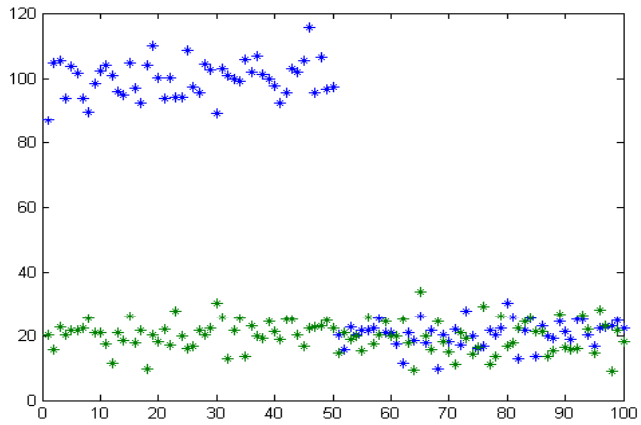


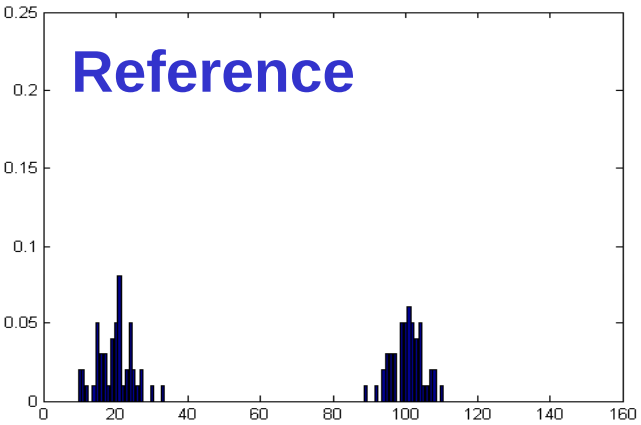


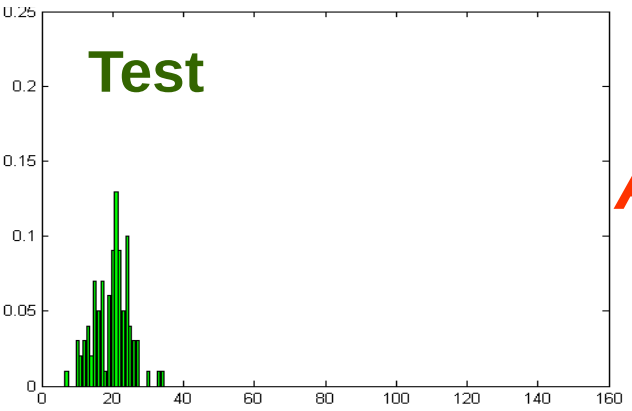


Anomaly

Case 3



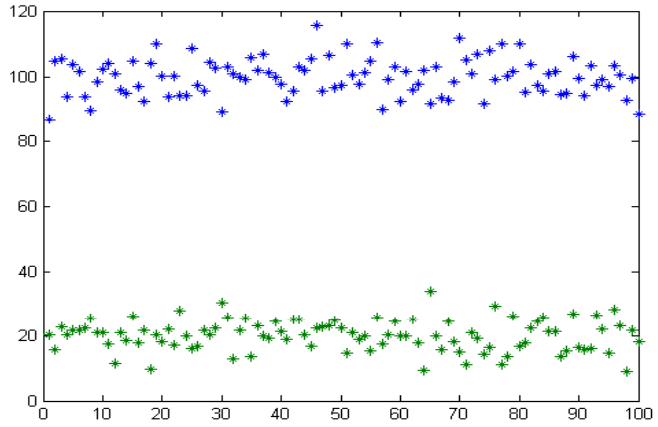


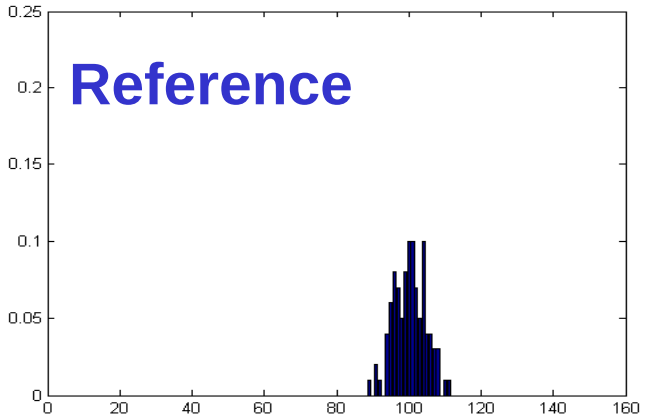


Anomaly

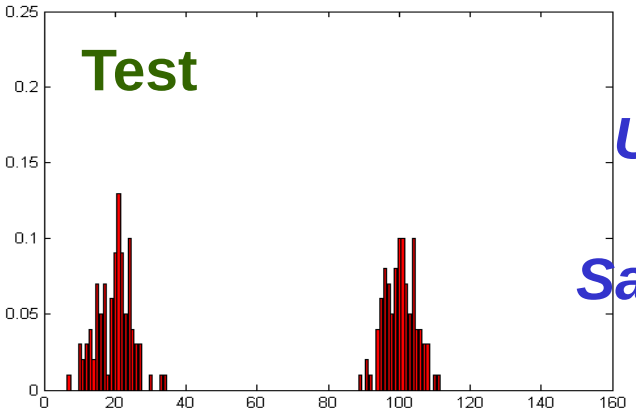
Indirect Comparison: *Combine & Compare*

Case 2



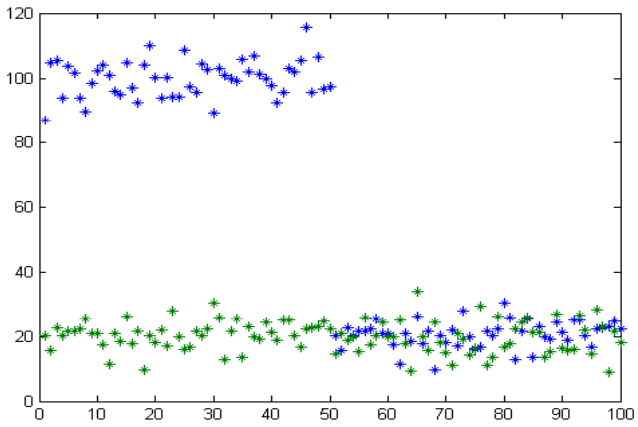


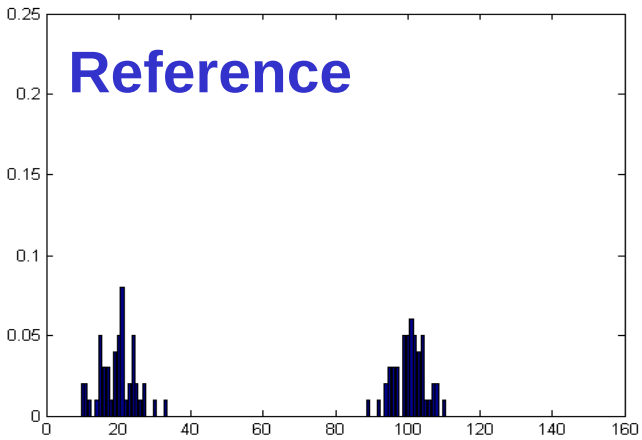
Anomaly

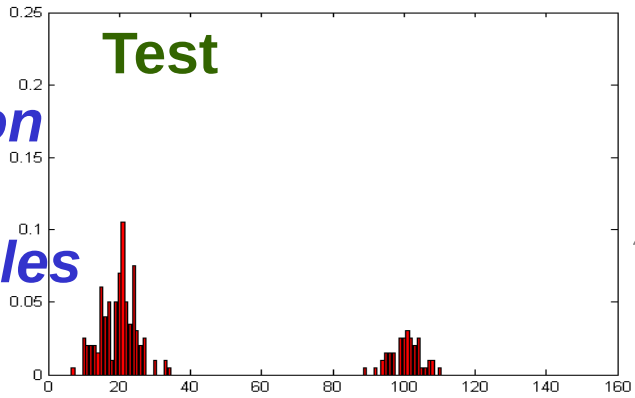


Union of Samples

Case 3



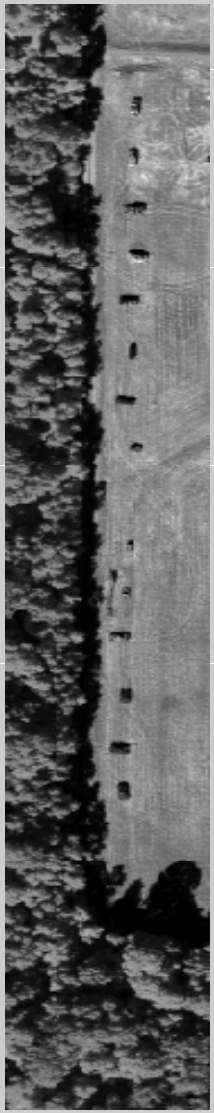




Softer Anomaly

New Family of Anomaly Detectors

Scene



V-SWIR

SemiP Detector

$$\eta_1 = \rho_1 \hat{\beta}^2 \left[\hat{E}(\mathbf{t}^2) - \hat{E}^2(\mathbf{t}) \right] \Rightarrow \chi_1^2$$

CFT Detector

$$\eta_3 = \rho_3 \left(\hat{\beta}_1^2 / S_1^2 \right) \Big/ \left(\hat{\beta}_2^2 / S_2^2 \right) \Rightarrow F_{1,1}$$

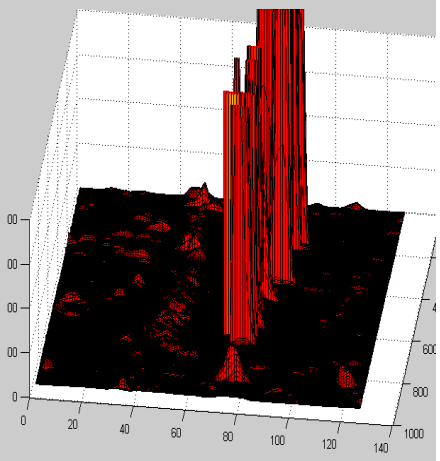
AsemiP Detector

$$\eta_2 = \frac{\hat{\tilde{\beta}}^2}{\left(\frac{1}{n_0} + \frac{1}{n_1} \right)} \frac{S_t^2}{S^4} \Rightarrow \chi_1^2$$

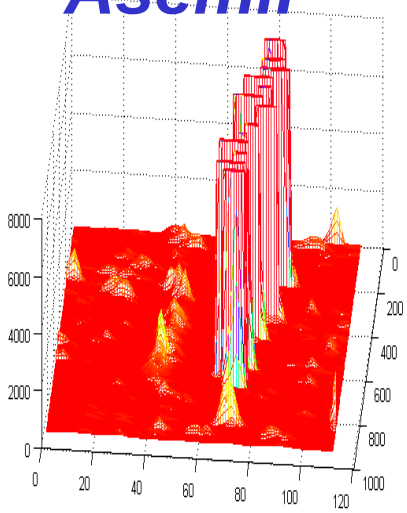
ANOVA Detector

$$\eta_4 = \frac{\sum_{i=1}^{2} n_i \left(\bar{x}_i - \bar{\bar{x}} \right)^2}{\frac{1}{N-2} \sum_{i=1}^{2} (n_i - 1) s_i^2} \sim F_{1,N-2}$$

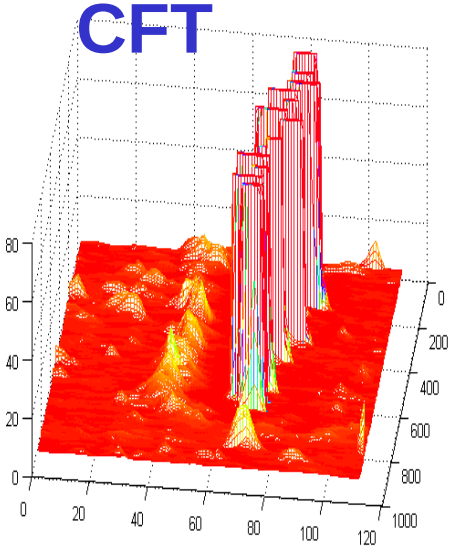
SemiP



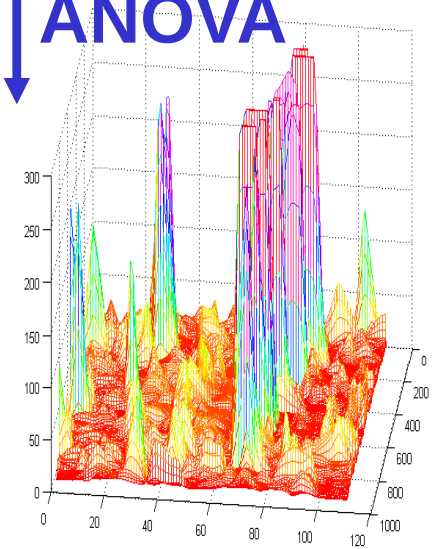
AsemiP

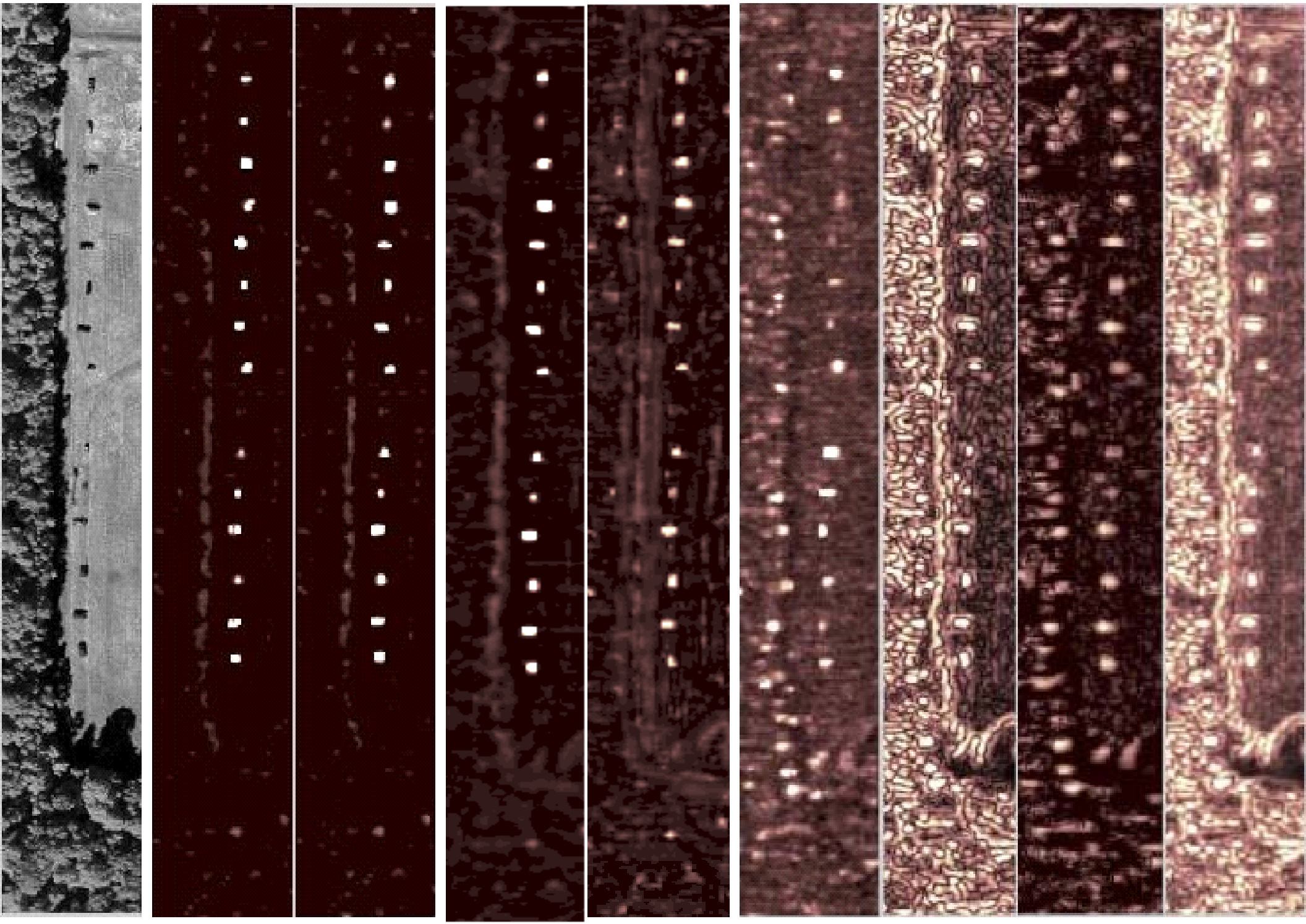


CFT



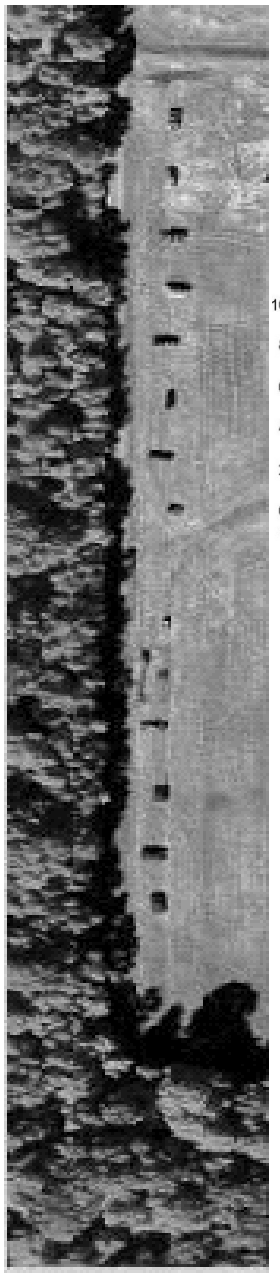
ANOVA



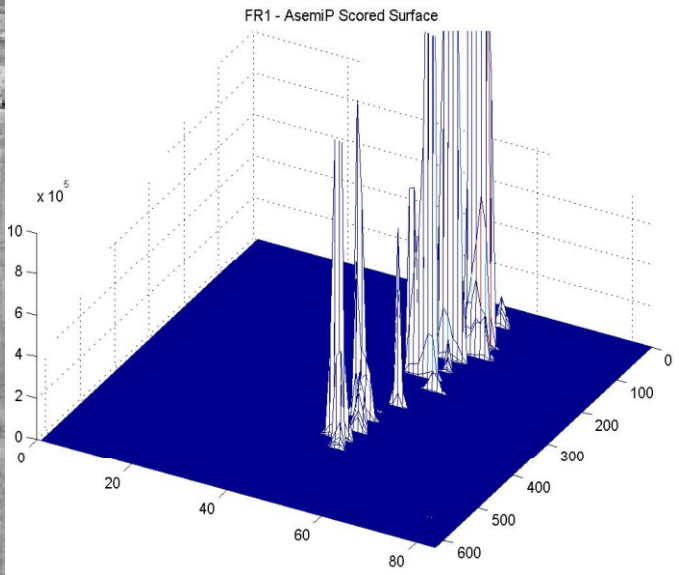


Scene *SemiP* *AsemiP* CFT ANOVA RX PCA EST FLD

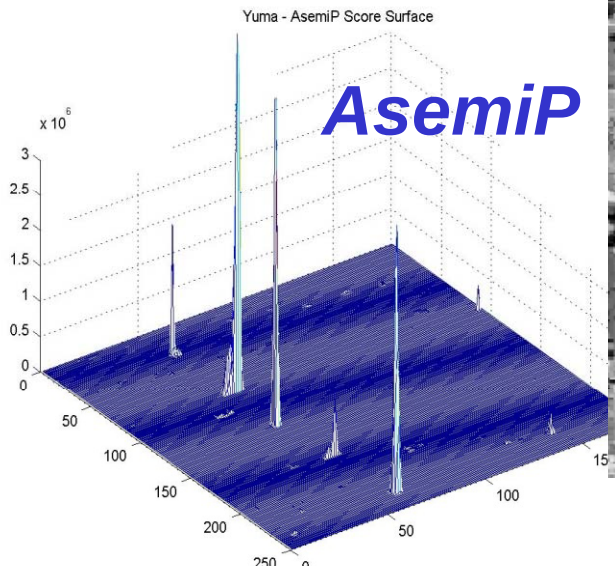
AsemiP Anomaly Detector



AsemiP



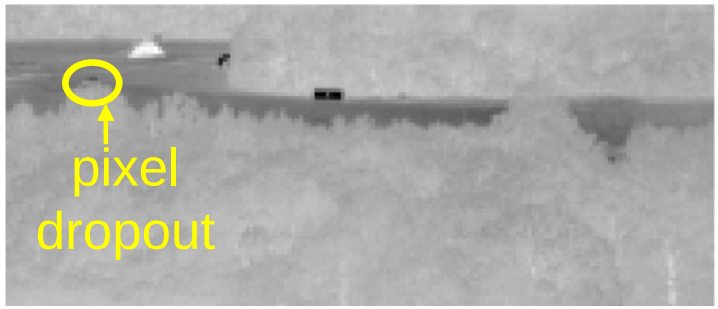
Desert Radiance



AsemiP

V-SWIR

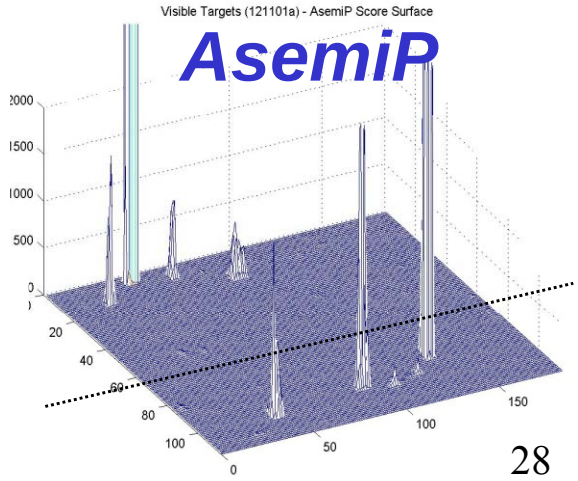
← **Forest Radiance** →



LWIR

No Targets

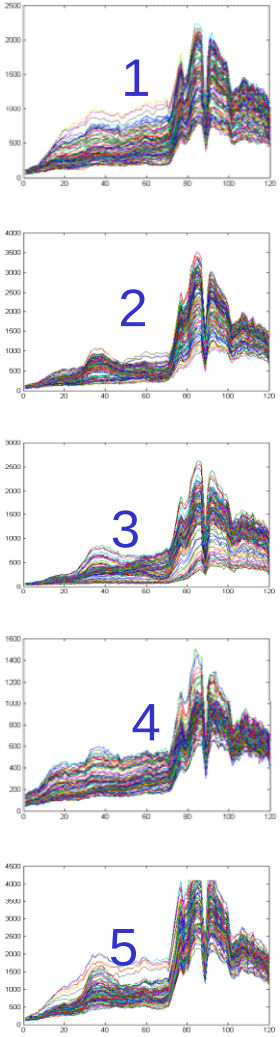
V-SWIR



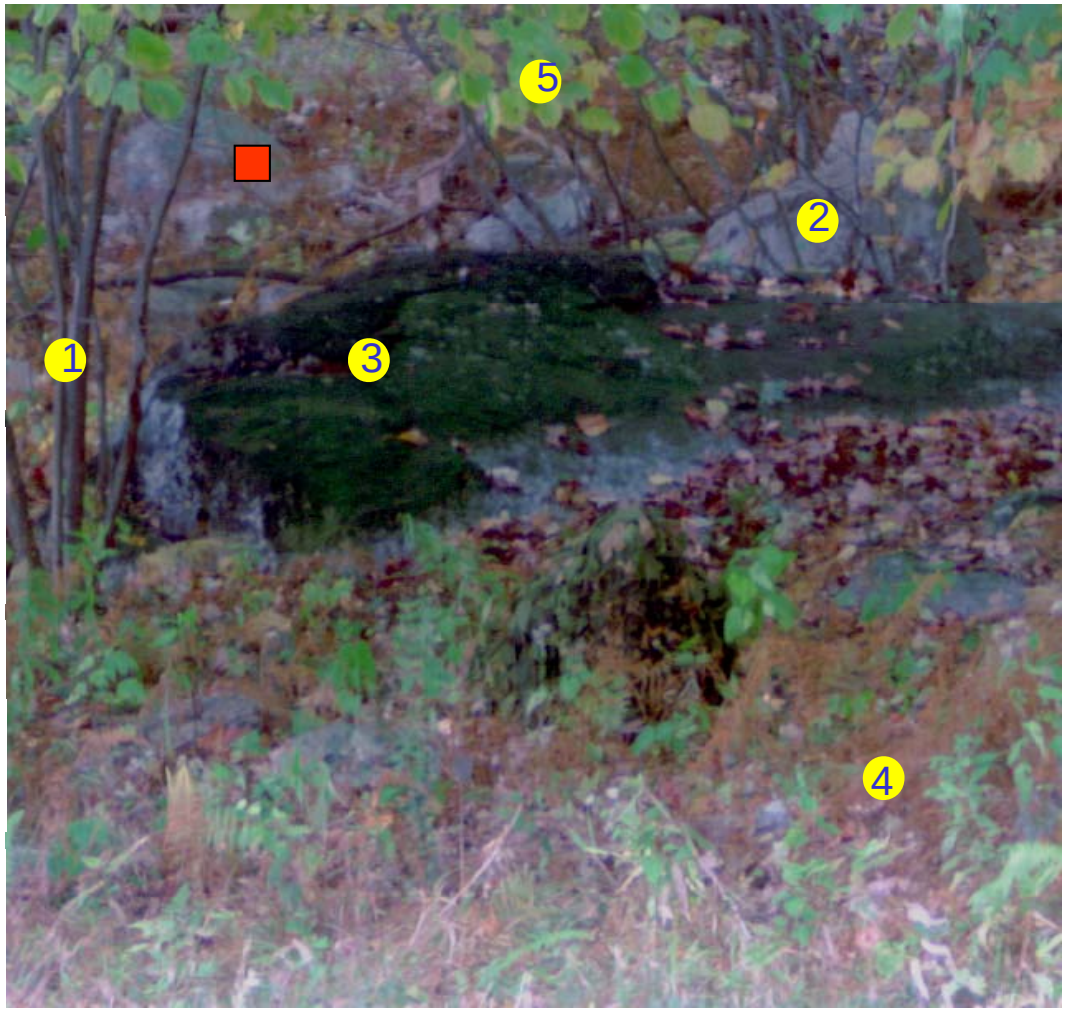
Scene Anomaly Detection @ Ground Level

***AsemiP* Algorithm**
Multi-Sample Extension

***10x10* area**

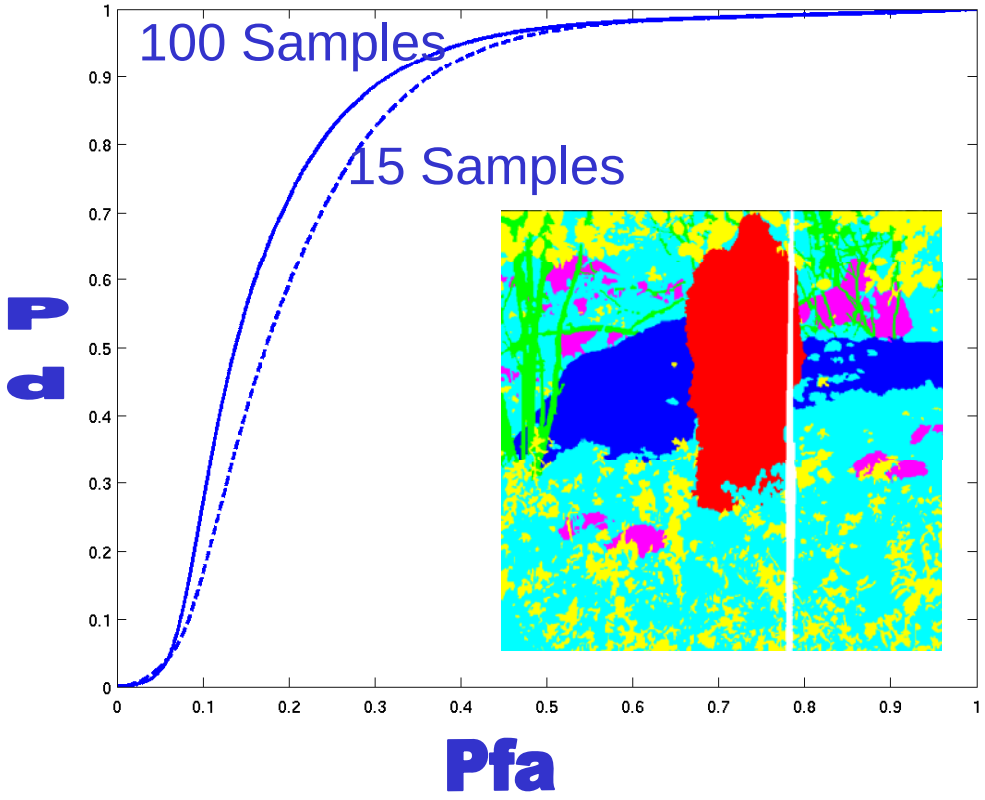


All Samples: 0.12% (500 / 373,321)

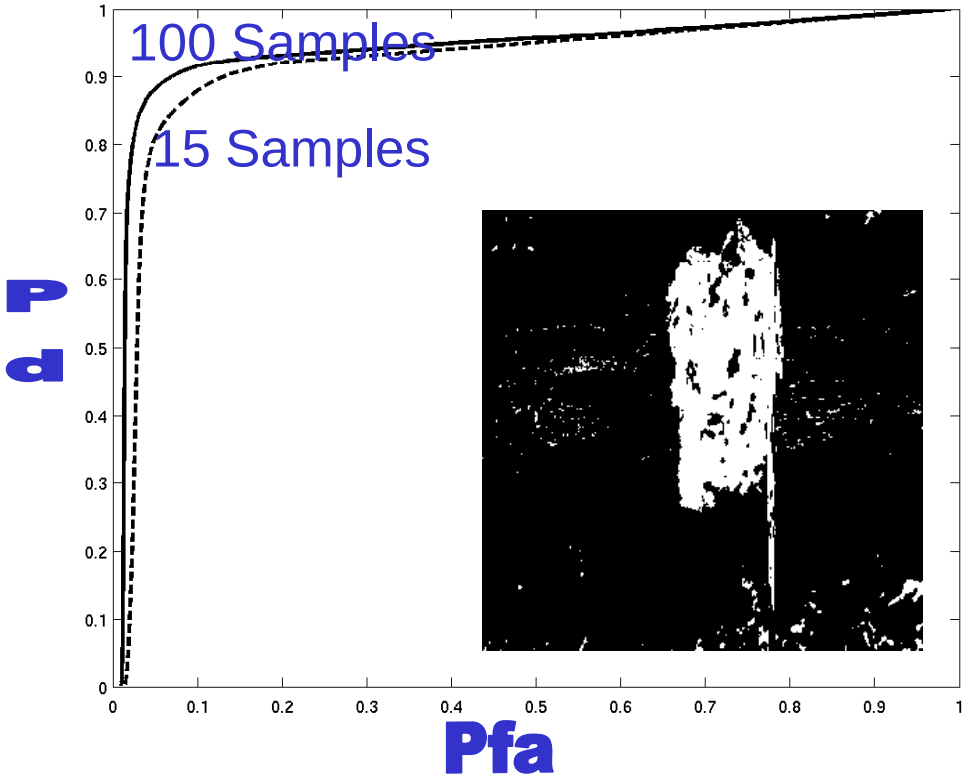


Six-Class Anomaly Detection

Supervised Learning Approach



Unsupervised Learning Approach



Artificial Neural Network

AsemiP Anomaly Detector

Final Remarks

- Statistical-Motivated Idea
- New Family of Anomaly Detectors
- Many Applications

Follow Up

- Auto Sampling
- Unsupervised Learning
  - Target Detection/Classification

MURI
Science of Land Target Spectral Signatures

MURI Program

- Spectral Signatures of Land Based Targets MURI Awarded 1 July 2002
  - Physics of Hyperspectral/Polarimetric Signatures
  - Fusion of Sensor Modalities
- Target Classes
  - Surface and buried land mines
  - Obscured Ground vehicles (CC&D)
  - Underground Facilities
  - Chemical/biological weapons

MURI Objectives

- **SPECTRAL SIGNATURES**: Understand and exploit optical signatures of land targets in complex environmental and terrain conditions
- **PHYSICS-BASED MODELING**: understanding and prediction of 'full' optical signatures (spectral, polarimetric)
- **ATR**: Identify and exploit features in the physical signatures to increase detection and decrease FA
- **FUSION**: Devise algorithms to exploit signature information over a wide spectral range; utilize ancillary information

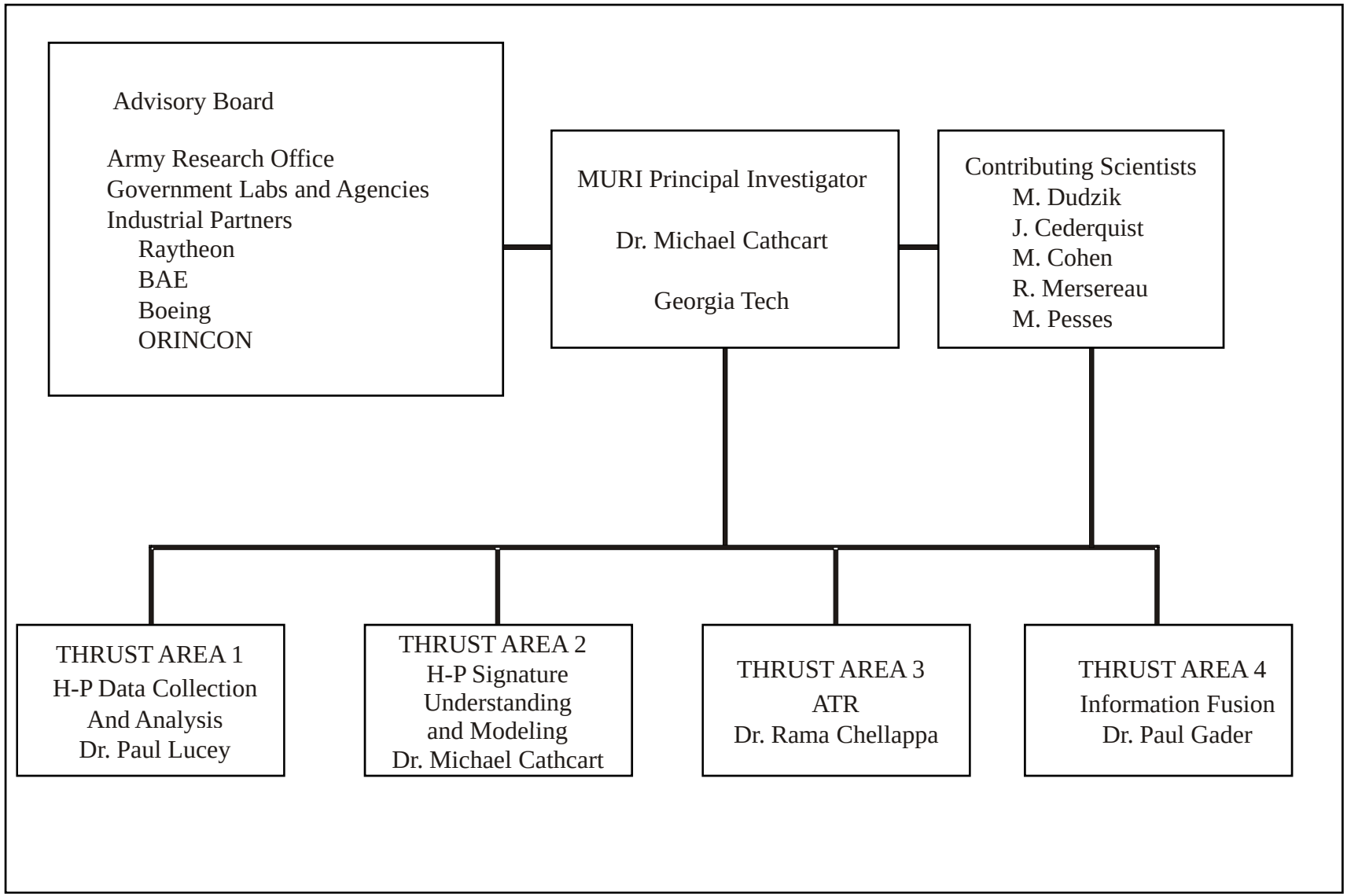
MURI Program Collaborators

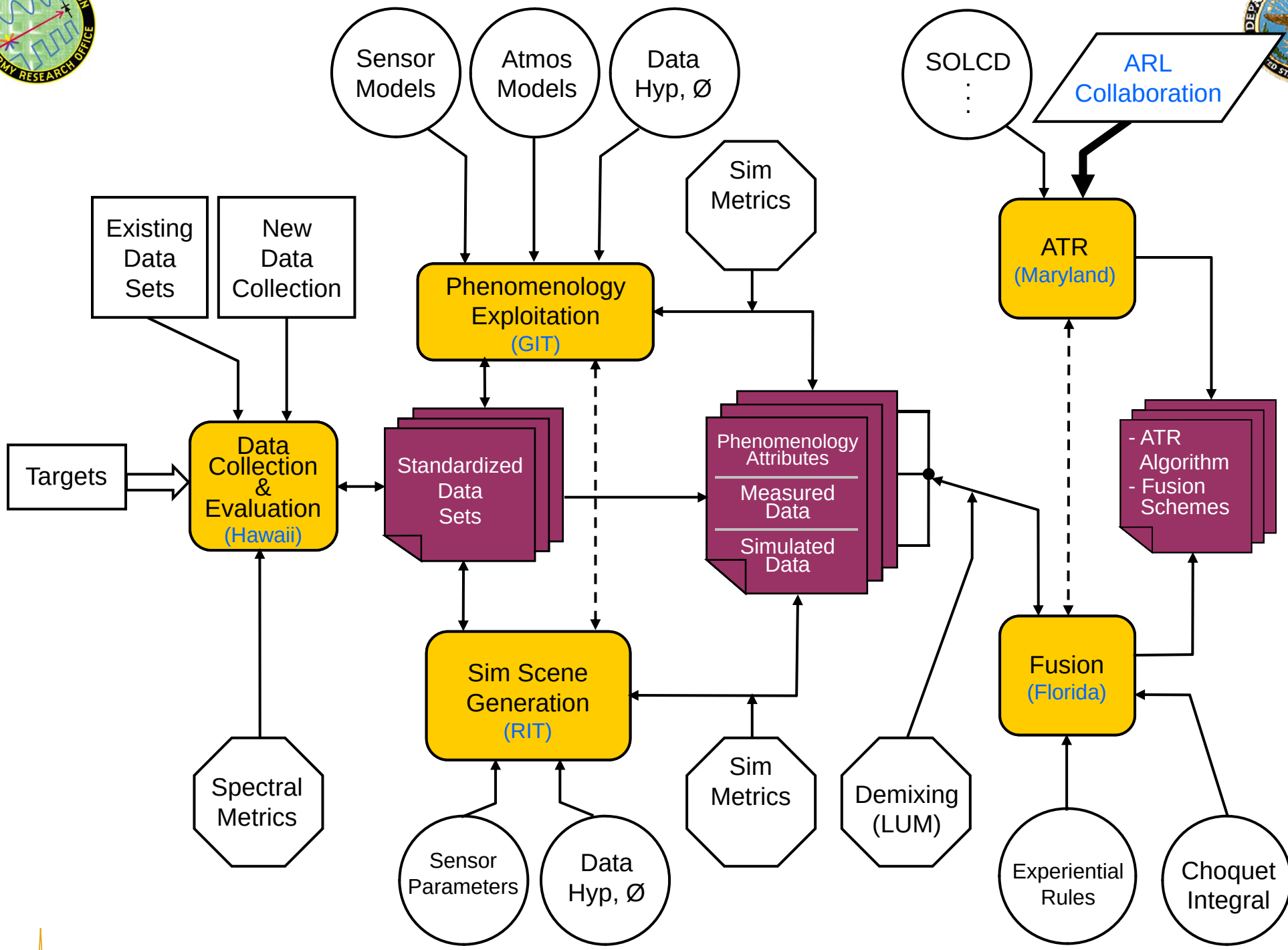
- Georgia Institute of Technology
  - University of Hawaii
  - University of Florida
  - University of Maryland
  - Rochester Institute of Technology
  - Clark Atlanta University
- Army Research Laboratory

Interested Government Agencies

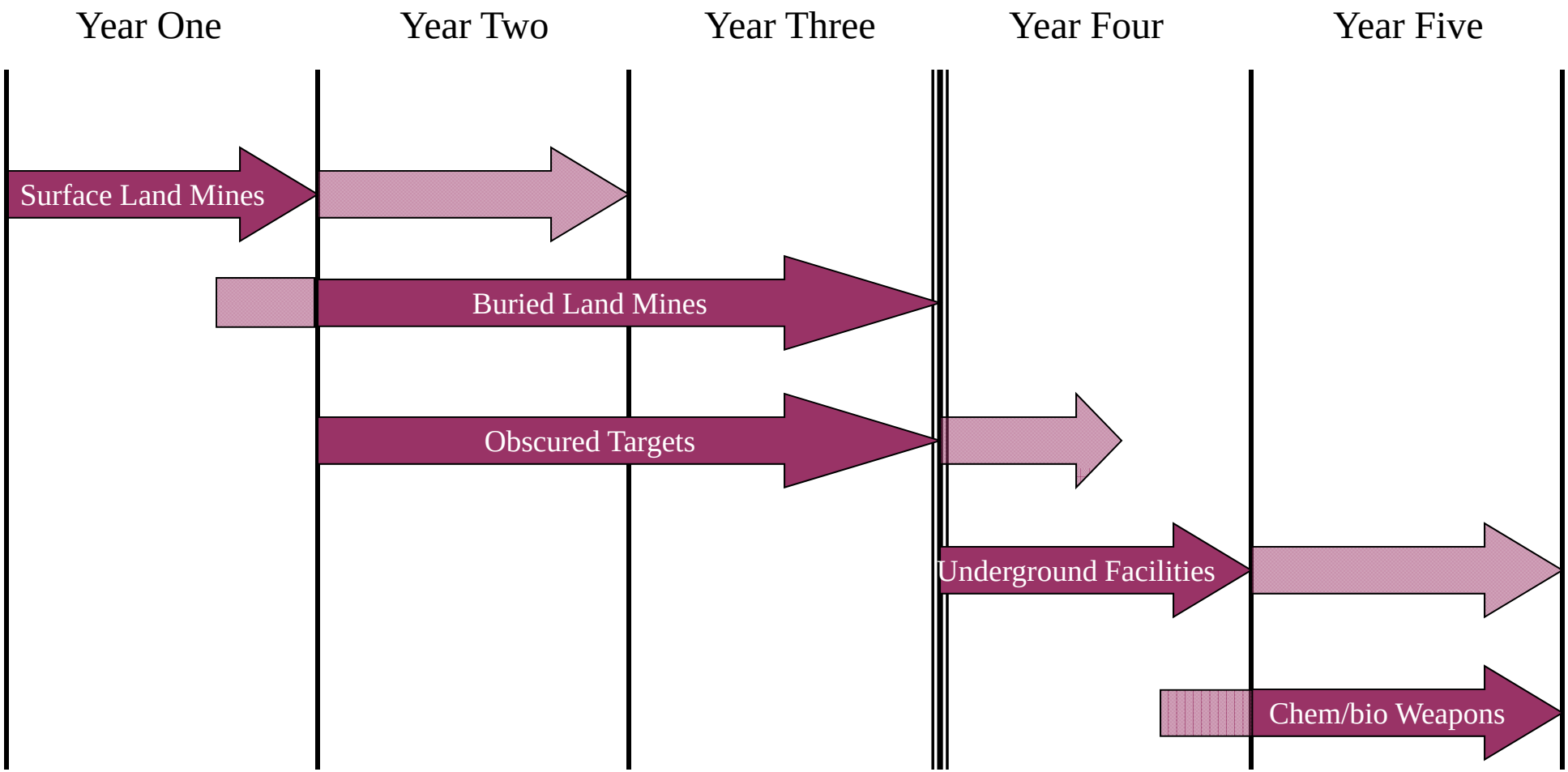
- ARO
- DARPA
- NVESD
- Army Space
- BMDO
- Army - Huntsville
- NRO
- NSSA
- NIMA
- ARL
- NSWC – Crane
- Navy – SPAWAR
- NSCS – Panama City
- AFTAC
- AFRL (Dayton, Hanscom, Albuquerque)
- National Guard Bureau / Counter Drug
- NRL

MURI Technical Management

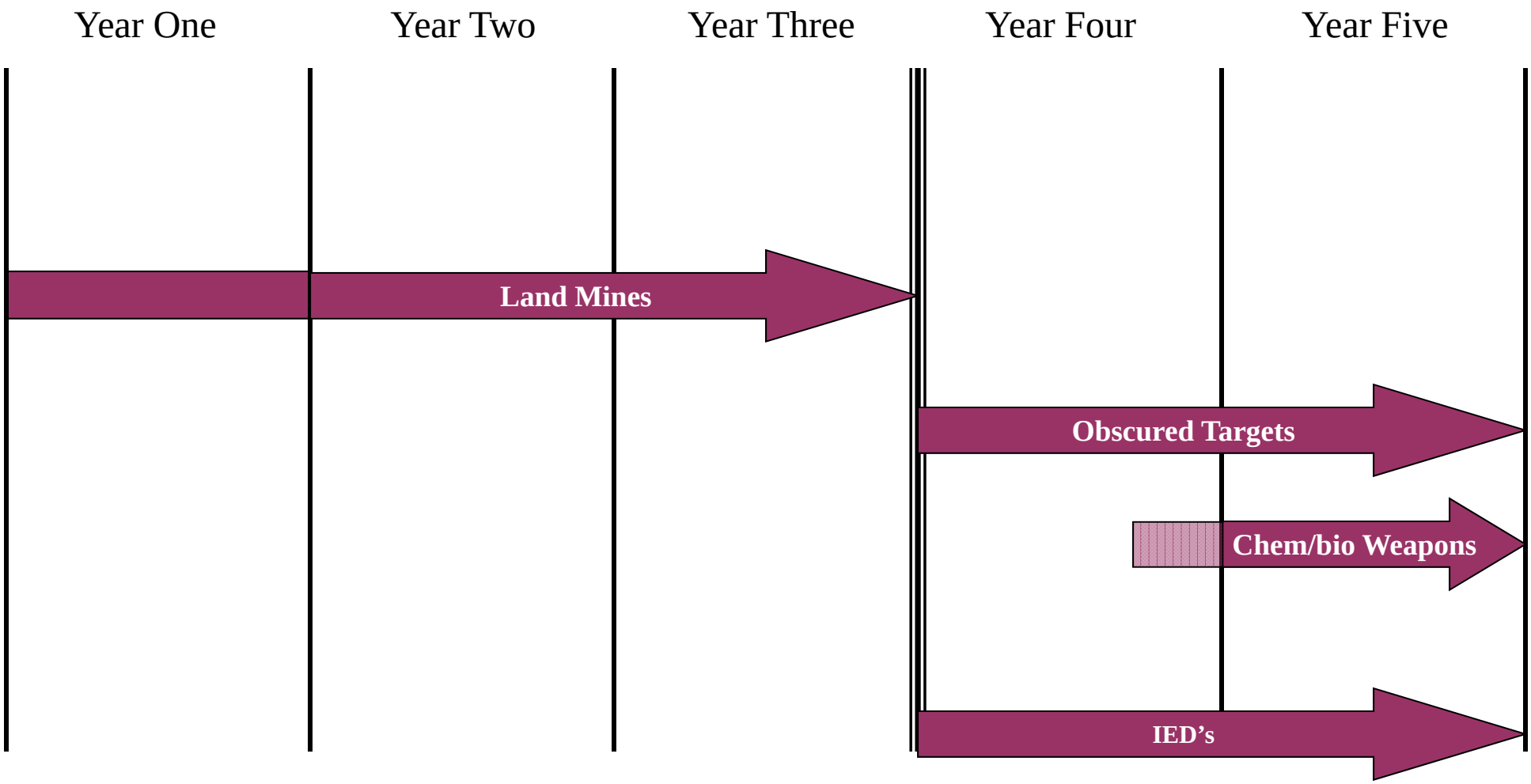



Sensor Models
Atmos Models
Data Hyp, Ø
SOLCD
ARL Collaboration
Sim Metrics
Existing Data Sets
New Data Collection
Phenomenology Exploitation (GIT)
ATR (Maryland)
Targets
Data Collection & Evaluation (Hawaii)
Standardized Data Sets
Phenomenology Attributes
Measured Data
Simulated Data
- ATR Algorithm
- Fusion Schemes
Sim Scene Generation (RIT)
Fusion (Florida)
Spectral Metrics
Sim Metrics
Demixing (LUM)
Sensor Parameters
Data Hyp, Ø
Experiential Rules
Choquet Integral

MURI Application Focus - Initial



MURI Application Focus – Mid-Program



Thrust Area 1: Hyperspectral-Polarimetric Data Collection & Analysis

- Thrust Director: Paul Lucey (University of Hawaii)

- Objectives
  - Evaluate existing hyperspectral imagery & data
  - Provide legacy data sets if available & applicable
  - Create quantitative data sets of applicable classes of targets and backgrounds
  - Develop spectral metrics for data characterization
  - Conduct special HSI “gap-filler” data collections as required

Thrust Area 2: Hyperspectral-Polarimetric Signature Understanding & Modeling

- Thrust Director: Michael Cathcart (Georgia Tech)
- Objectives
  - Develop detailed understanding of the phenomenology & discriminants which form the basis of hyperspectral-polarimetric signatures
  - Develop hyperspectral-polarimetric signature and scene models for targets and backgrounds
  - Validate modeling and phenomenology paradigms
- Collaborator: Rochester Institute of Technology

Thrust Area 3: Automatic Target Recognition

- Thrust Director: Rama Chellappa (University of Maryland)

- Objectives
  - Maximize discrimination through physics-based algorithms
  - Development of phenomenologically-based subpixel target detection algorithms
  - Development of structured and non-structured ATR algorithms
  - Development of clutter complexity measures for HSI
  - Exploration of utility of non-Gaussian models for detection and recognition

- Collaborator: Clark Atlanta University, Army Research Laboratory

Thrust Area 4: Information Fusion

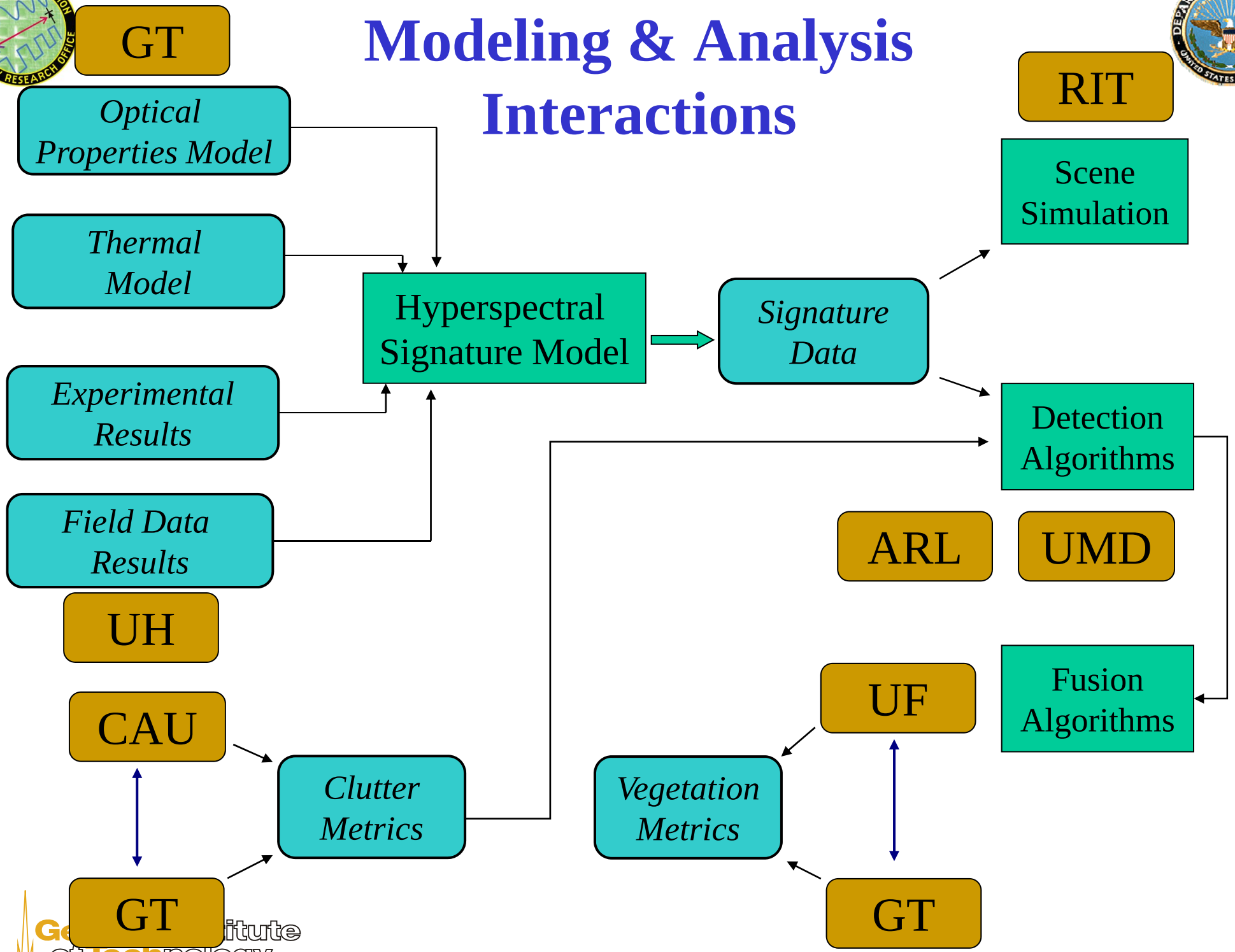
- Thrust Director: Paul Gader (University of Florida)

- Objectives
  - Fusion of multi-sensor data with HSI to leverage additional modality discriminants (e.g., GPR)
  - Development of algorithms that fuse the individual ATR outputs (Multi-INT Fusion)
  - Incorporate physics-based approaches into fusion architecture and algorithm development

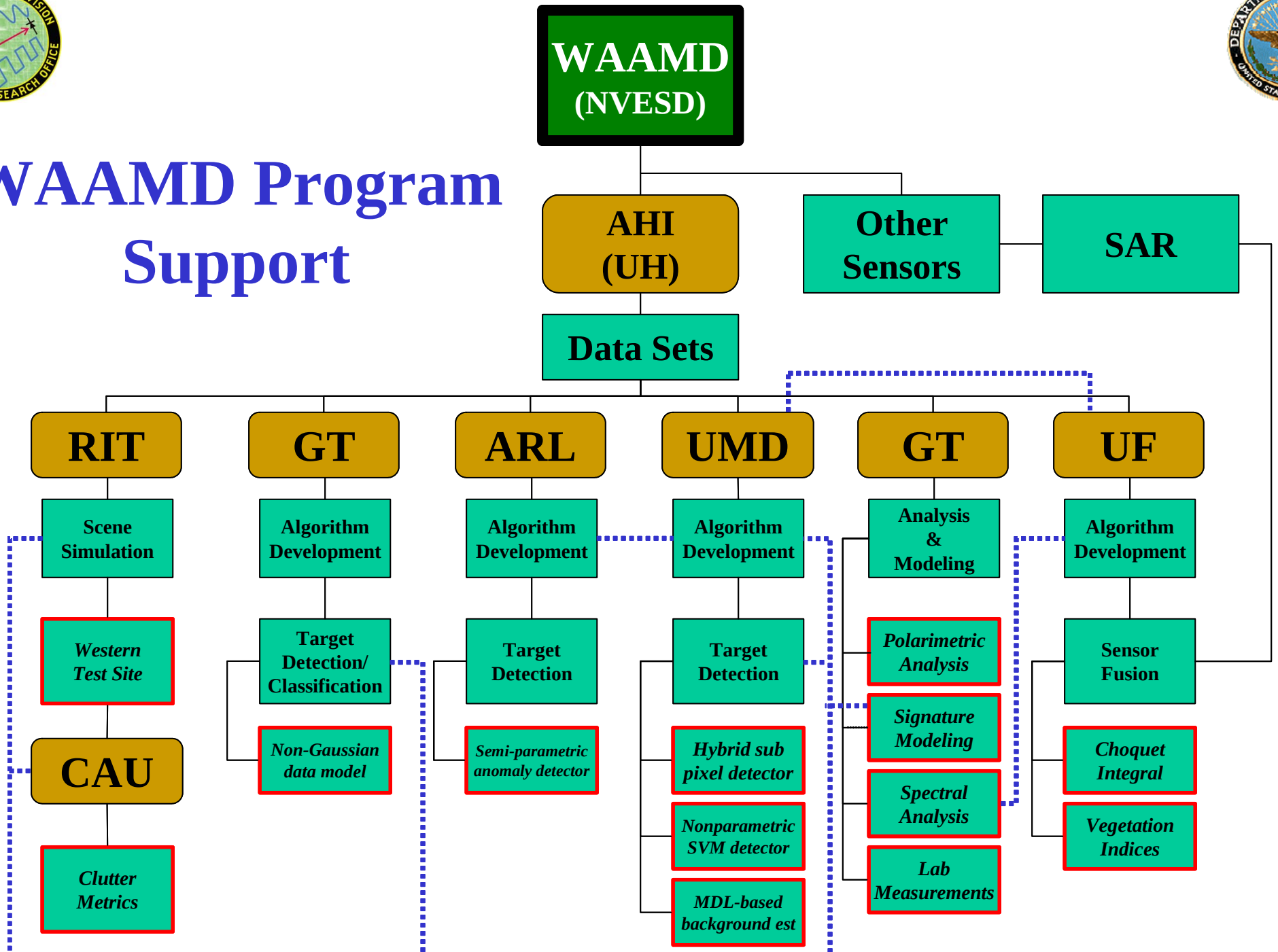
Government Agencies

- ARO
- DARPA
- NVESD
- Army Space
- MDA
- Army - Huntsville
- NRO
- NSSA
- NIMA
- TEC
- National Guard Bureau / Counter Drug
- ARL
- NSWC – Crane
- Navy – SPAWAR
- NSCSS – Panama City
- AFTAC
- AFRL (Dayton, Hanscom, Albuquerque, Eglin)
- NRL
- ERDC
- LLNL
- EPA
- ONR
- DIA (MASINT)

Modeling & Analysis Interactions



WAAMD Program Support



Impact of MURI Research

- Hyperspectral Data Collection
  - Defines the requirements for hyperspectral data collection efforts to support current and future military system development
  - Provides well-defined & characterized data sets (laboratory & field) for current and future research
- Phenomenology Understanding
  - Identifies phenomenological basis for improvement of algorithm operation
  - Provides data and models on variation of spectral signatures under realistic environmental conditions
  - Aids in the definition of sensor requirements for future detection systems
- Algorithm Development
  - Development of algorithms incorporating phenomenological results to provide improved detection performance and false alarm reduction of hyperspectral sensors
- Sensor Fusion
  - Development of fusion concepts and approaches which maximize performance of hyperspectral sensors through band selection, correlation with other sensor data (SAR), physics-based signatures, etc.

MURI Research Collaborations - Examples

- WAAMD
  - Continuing data analysis and evaluation
  - Participation in Level I evaluation
- ARL
  - University of Maryland collaboration (Hirsch Goldberg)
- JHU – APL
  - Algorithm studies (Dr. Amit Banerjee)
- ONR
  - Continuing participation in Counter-IED Program (3 year program, 6.1)
- NVESD – Forward Looking Program
  - Continuing participation (Processing)
- DIA – spectral sensor program

Document Contents

- Georgia Institute of Technology
- University of Maryland
- University of Florida
- Army Research Laboratory
- Clark Atlanta University
- University of Hawaii
- Rochester Institute of Technology
- PhD Theses

Hyperspectral-Polarimetric Signature Understanding & Modeling

Georgia Institute of Technology

Research Summary

Participants

- Dr. Michael Cathcart – Principal Investigator
- Dr. Boris Mizaikoff – Principal Investigator
- Dr. Thomas Orlando
- Dr. Alan Thomas
- Dr. Robert Bock
- Mr. Ricardo Campbell
- Mr. Bryce Remesch
- Mr. Michael Dudzik
- Dr. Russell Mersereau
- Mr. Rafael Love
- Ms. Sarah Lane
- Ms. Sarah Greenwood
- Mr. Brett Mauro
- Dr. Mark Pesses
- Mr. Manfred Karlowatz
- Mr. Aleksandr Alexandrov
- Ms. Kelly O'Connor
- Mr. Michael Farrell
- Mr. Robert Worrall

Research Objectives

- **Develop detailed understanding of the phenomenology & discriminants which form the basis of hyperspectral-polarimetric signatures**
  - Understand and exploit optical signatures of land targets in complex environmental and terrain conditions
- **Develop hyperspectral-polarimetric signature models for targets and backgrounds**
  - Develop an understanding of 'full' optical signatures (spectral, polarimetric)
  - Develop a prediction methodology for 'full' optical signatures (spectral, polarimetric)
- **Validate modeling and phenomenology paradigms**
- **Identify and exploit features in the physical signatures to increase detection and decrease false alarms**

Research Activities

1. ATR Measurements of particulate minerals
   a. Berreman effect
2. LWIR polarization in soils
   a. Spectral & angle effects
3. Soil optical properties
4. Minefield detection algorithms
   a. Adaptive sampling
   b. Scale & orientation of minefields

5. LWIR HS signature processing
   a. Algorithm critique
   b. Statistical spatial
   c. End member techniques
6. Hyperspectral LWIR model
   a. Landmine, soils, false targets
   b. Environmental effects
7. Environmental exposure measurements on materials
8. Disturbed Soil Characterization Workshop

Technology Transfer-related Activities

- Army Research Office (tunnel detection workshop, disturbed soil workshop)
- NVESD Wide Area Airborne Minefield Detection Program
- US Army Engineer Research and Development Command (WES, CRREL)
- US Army Research Laboratory
- Defense Intelligence Agency
- Joint ARO-ERDC 'Battlespace Environments' Basis Research Review Meeting
- NATO Advanced Study Institute
- NAVSEA Coastal Systems Station
- National Geospatial – Intelligence Agency
- JIEDDO
- Office of Naval Research
- US Air Force – reconnaissance wing
- US Department of Energy
- SMDC
- National Guard

School of Chemistry

PI: Boris Mizaikoff

Atlanta, September 2005

Attenuated Total Reflection (ATR) Studies of Particulate Minerals (Annual MURI Review Meeting)

Alexandra Molinelli, Manfred Karlowatz, Alexandr Aleksandrov, Thomas Orlando, and Boris Mizaikoff

Introduction

Remote Landmine Detection: The Disturbed Soil Approach

- Field Measurements: disturbed soil visible in spectra as[1]:

  (a) change in spectral contrast

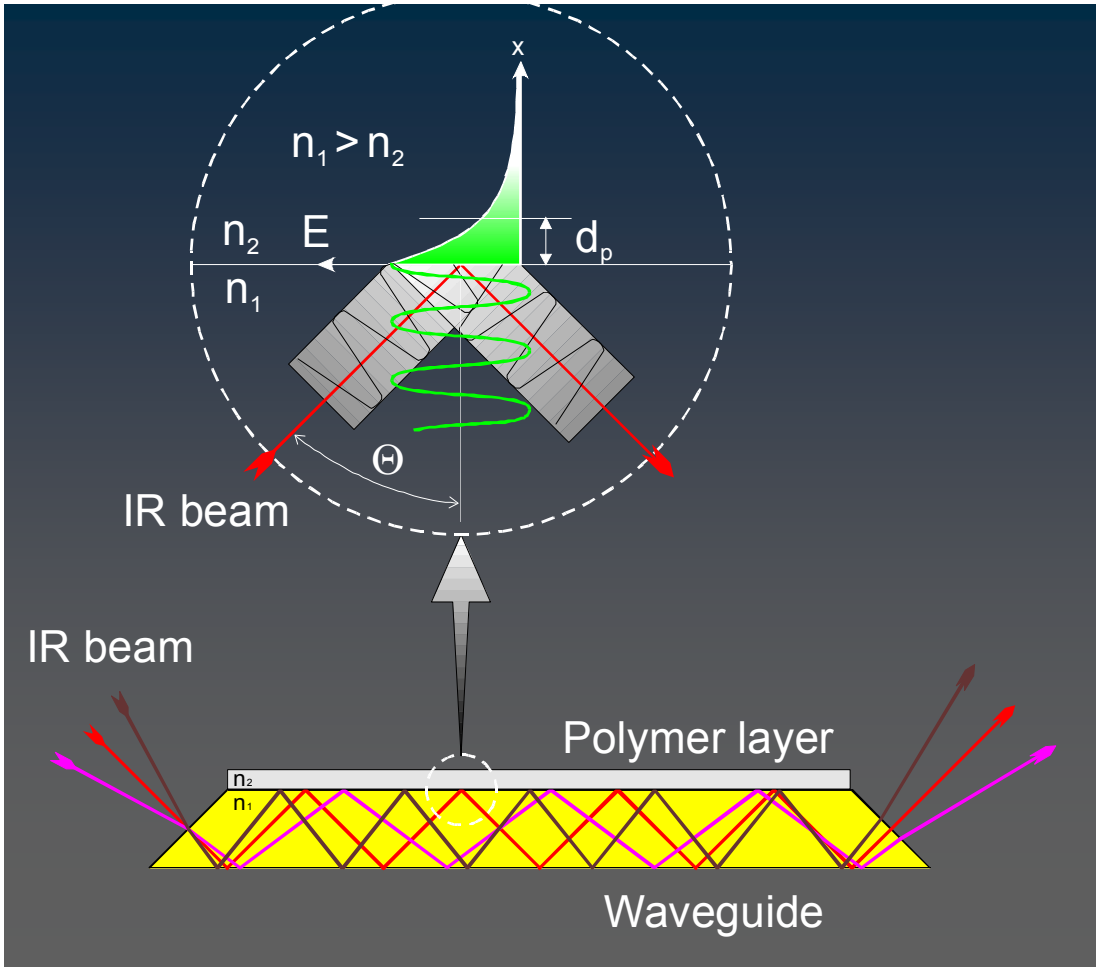
  (b) spectral shifts in the region in the region of 9.2 μm (~1080 cm^{-1})

- Data evaluation difficult due to complexity of spectroscopic signatures

Fundamental infrared spectroscopic studies on minerals at controlled laboratory conditions for improved understanding of remote sensing and hyperspectral imaging data.

[1] J. R. Johnson et al., *Remote Sens. Environ.* 1998, 64, 34-46

ATR-Principle



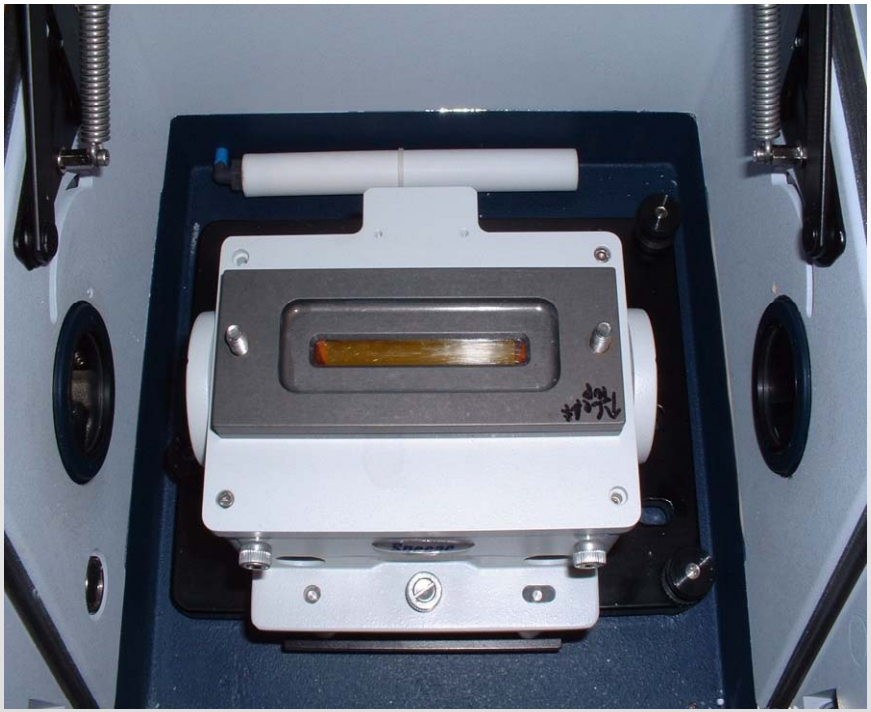
PENETRATION DEPTH

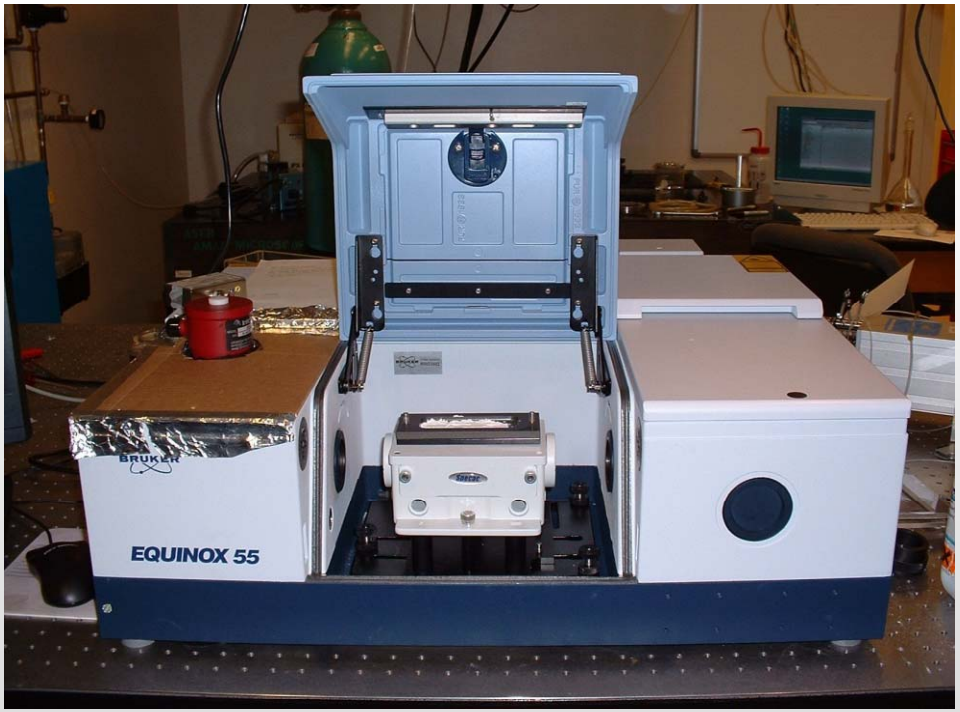
$$d_p = \frac{\lambda}{2\pi n_1 \sqrt{\sin^2\theta - \left(\frac{n_2}{n_1}\right)^2}}$$

EVANESCENT FIELD INTENSITY

$$E = E_0 \cdot e^{\left(\frac{-z}{d_p}\right)}$$

ATR-Setup





ATR Parameters

ATR Parameters

- Waveguide material: ZnSe
- Number of internal reflections: 12
- Refractive index: ~ 2.4

Measurement Parameters

- Recorded spectral range: 6000 - 400 cm^{-1}
- Spectral resolution: 1 cm^{-1}
- Averaged scans: 100

Introduction

S-polarization: $E = E_y = E_0$

Transverse optic modes (TO)

P-polarization: $E = (E_0.\cos\vartheta, 0, E_0.\sin\vartheta)$

Transverse optic modes (TO)

Longitudinal optic modes (LO)

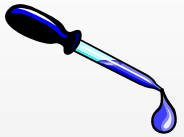
Wetting / Drying Studies

Measurement Procedure

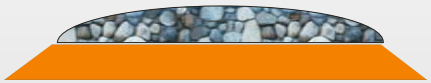
- Empty crystal; reference spectrum

- Application of sample; sample spectrum

- Addition of water



- Drying process; higher packed sample



- Dried spectrum

- Disturbing event (spatula)

- Disturbed spectrum

Investigated Samples To Date

- Quartz; SiO_2 (Fluka): > 62 µm, < 62 µm

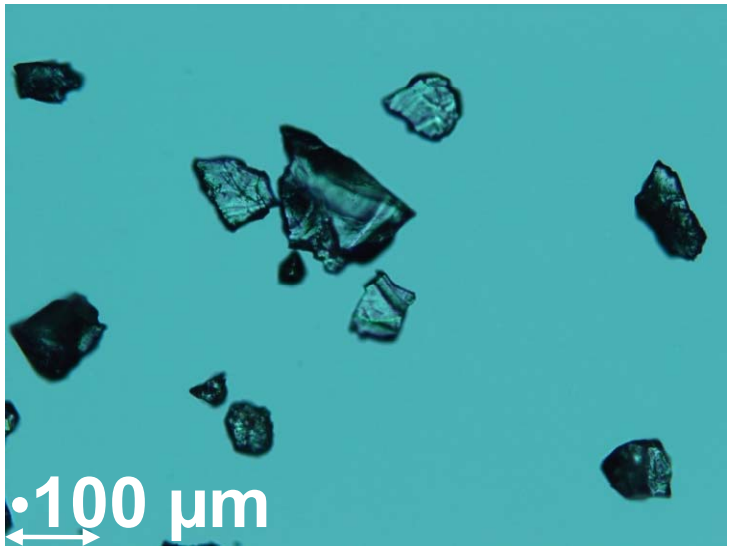
- Soda lime glass microspheres: 1-3 µm, 4-10 µm
  (MO-SCI, Rolla, MO)

- General purpose glass microspheres: 1700-1800 µm, 400-425 µm,
  (Whitehouse Scientific, Chester, UK) 112-125 µm, 25-32 µm

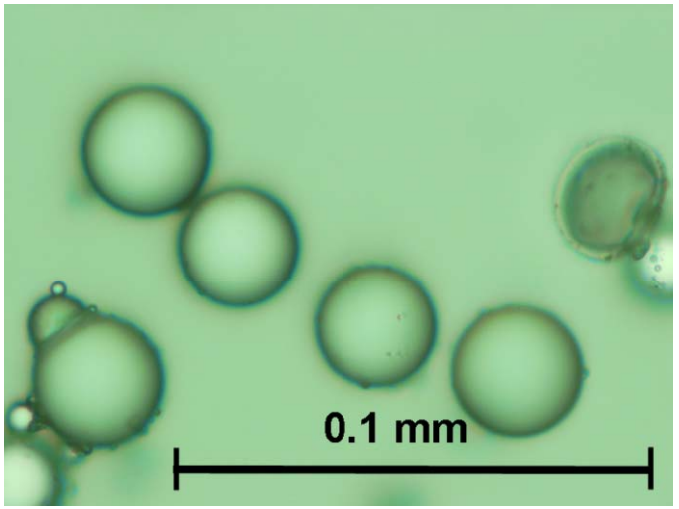
- Plain silica beads: 10 µm, 3 µm, and 200 nm
  (Kisker Biotech, Germany)

- Silica micro- and nanospheres: 3 µm, 100 nm,
  (C.P. Wong, Georgia Tech) 2 µm, 7 µm, 15 µ

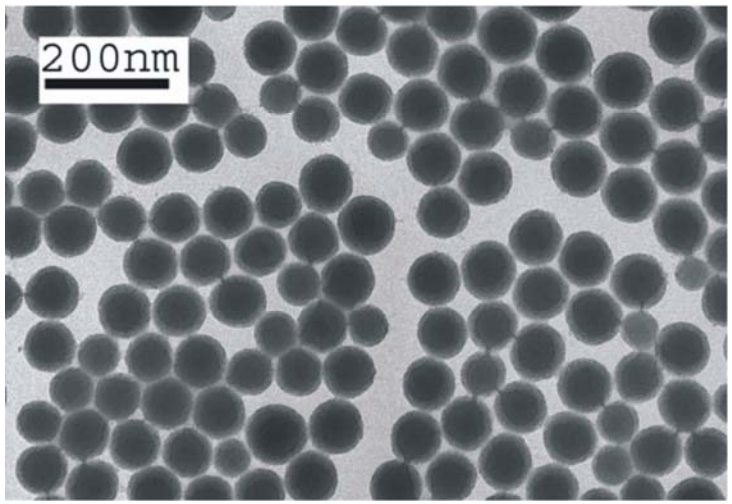
Samples



Quartz



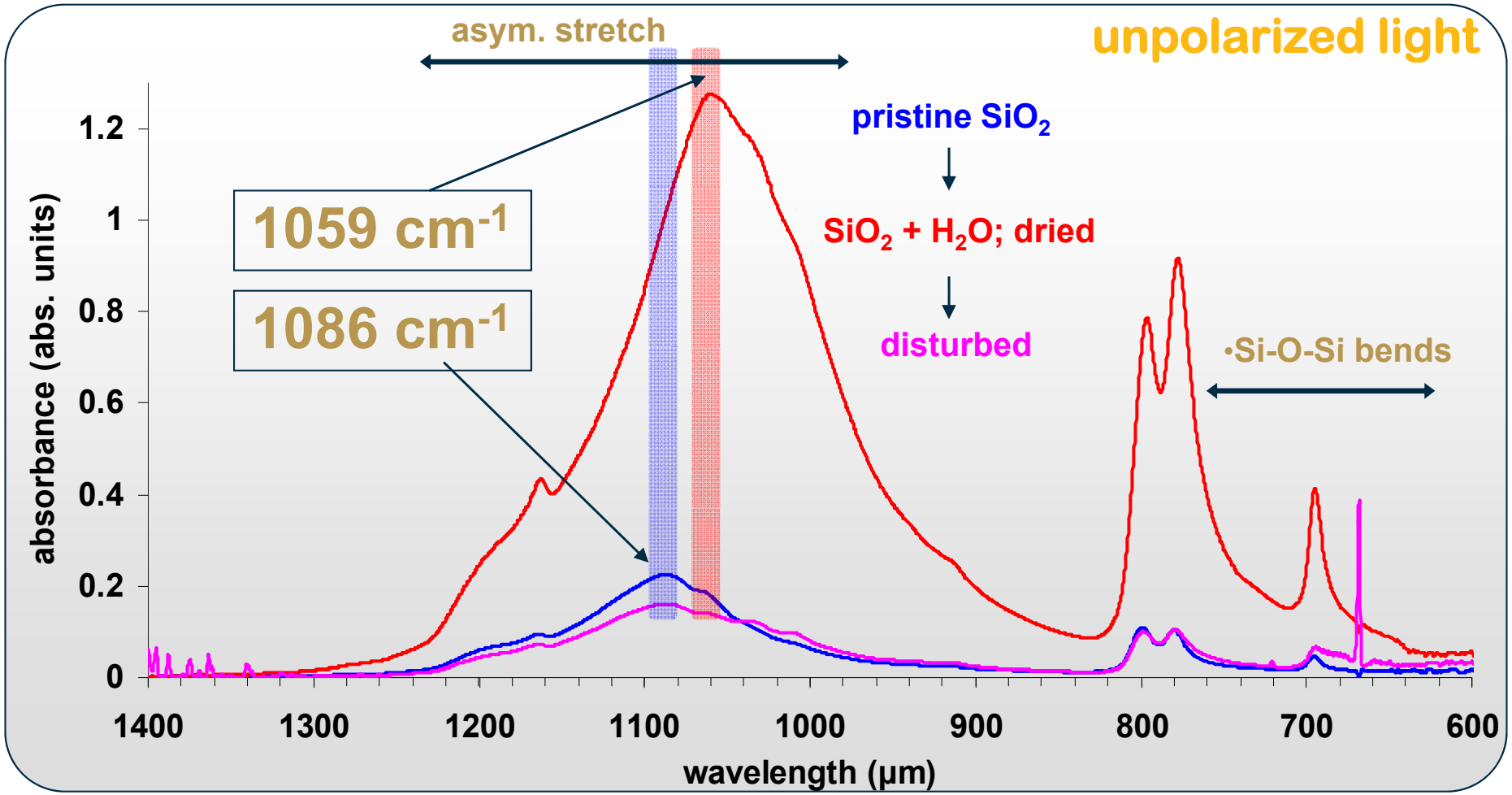
Soda lime glass spheres (25-32 µm)



Silica nanospheres (100 nm, TEM)

Wetting/Drying Studies with Quartz

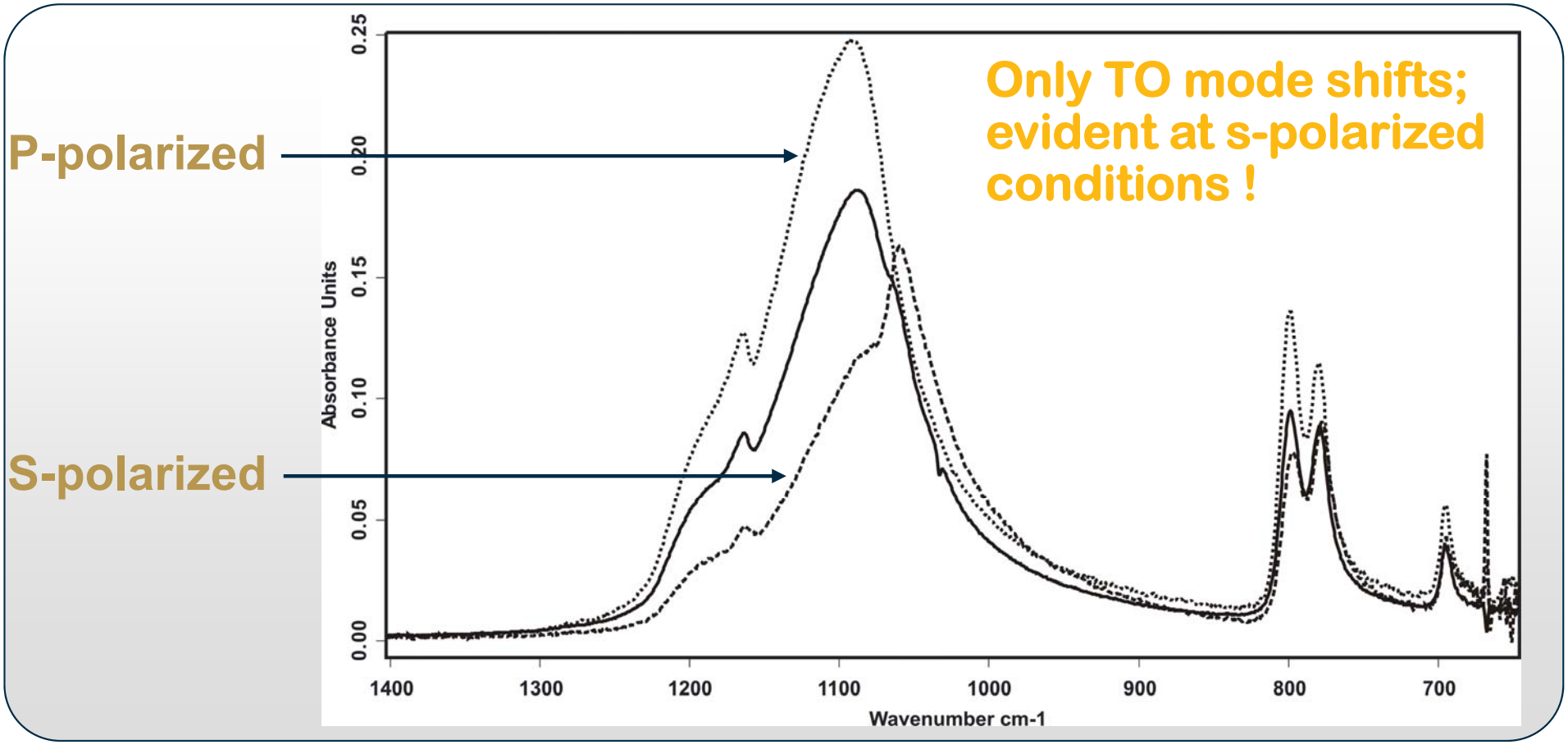
ATR – Quartz; SiO_2 (polydisperse particles)



Manuscript in preparation to Phys. Chem. B , 2005

Wetting/Drying Studies with Quartz

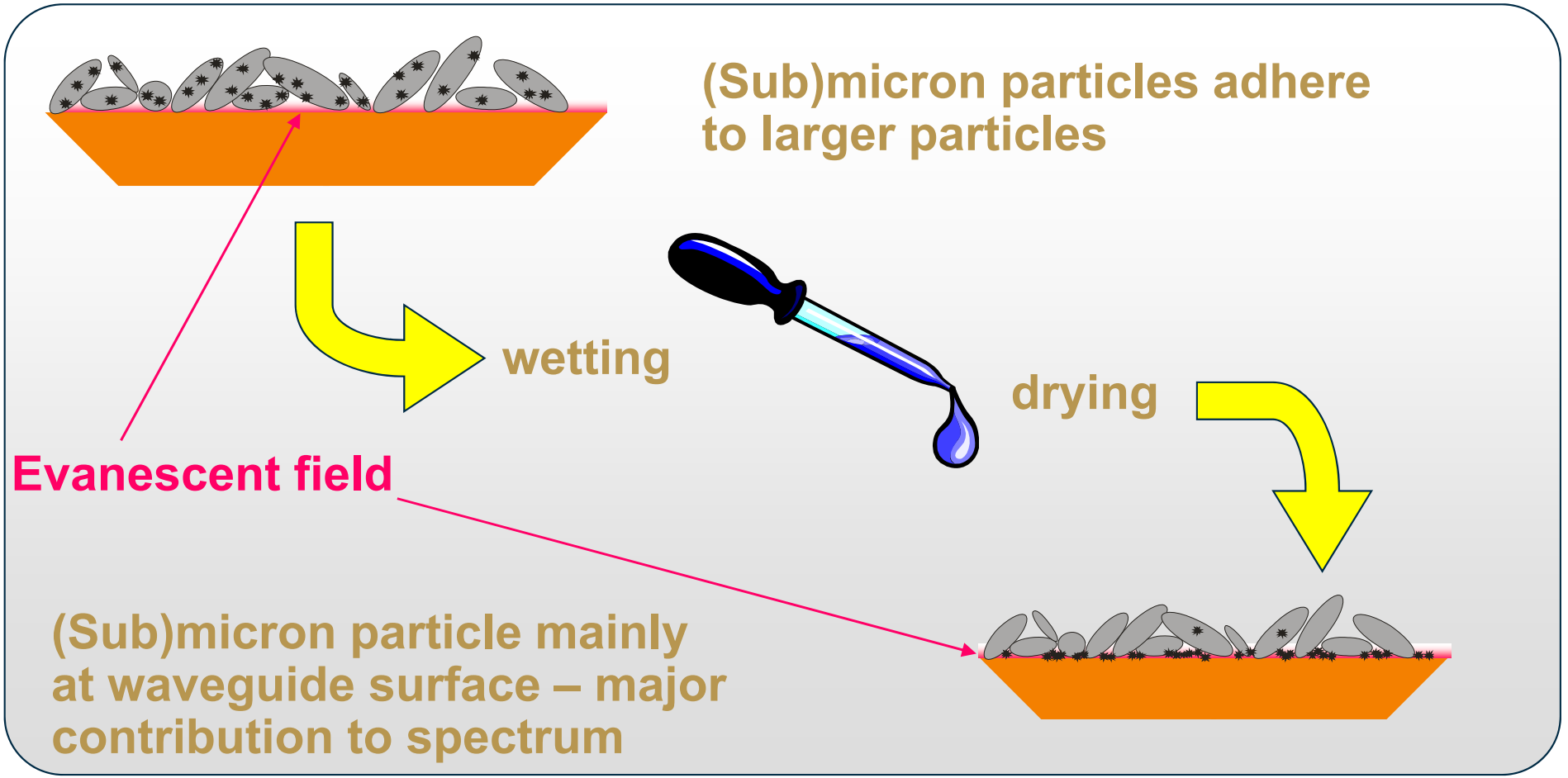
ATR – Quartz; SiO_2 (polydisperse particles) – Polarized light



Manuscript in preparation to Phys. Chem. B , 2005

Wetting/Drying Studies with Quartz

Hypothesis for Hyperspectral and Laboratory Data



Conclusion: ATR Studies of Quartz

Wetting / Drying Studies of Quartz

- Significant shift of the asymmetric Si-O bands around 1080 cm^{-1}
- Shift is related to re-organization of particles leading to a higher density of smaller particles close to the waveguide surface during the wetting/drying process
- Change in contrast and spectral shift was reversible after disturbing event
- Experiments with polarized light show strong LO / TO mode splitting (**suspicion: Berreman effect!**); only the TO mode shifts
- Hypothesis: first experimental confirmation of the Berreman effect for particle films

Berreman Effect

- 1963, D. W. Berreman[1]: specific boundary conditions in thin layers of cubic crystals causes longitudinal (LO) and transversal (TO) mode splitting in IR absorption features (see Density Functional Perturbation Theory)

- Berreman effect shown for:

  - Thin films of crystalline structures[1] (oxides)

  - Bulk crystalline structures[2] (oxides)

  - Bulk glass[3]

  - Particle films have not been investigated yet

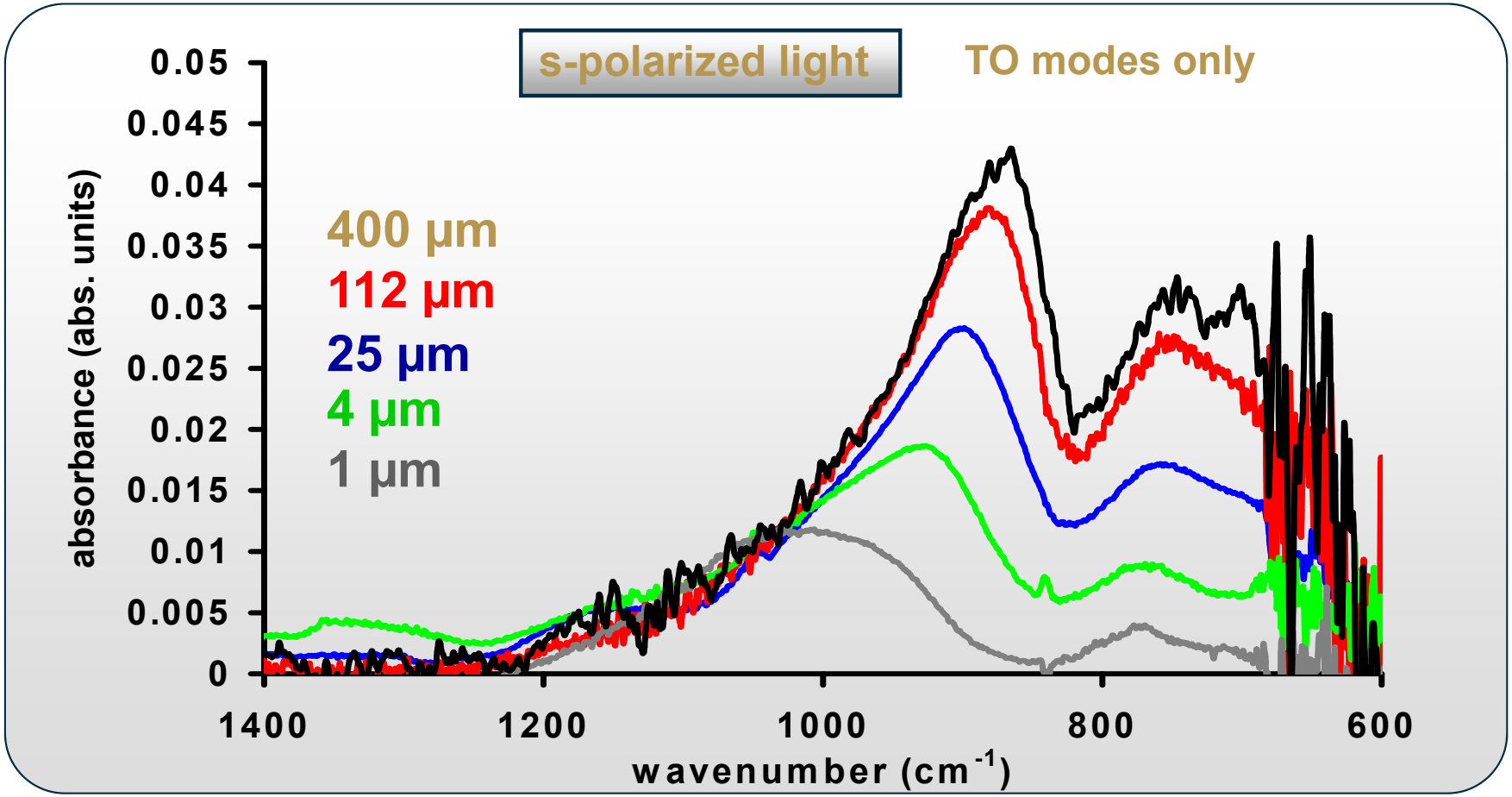
[1] D. W. Berreman; Phys. Rev., (1963), 130 (6), 2193-2198

[2] O. E. Piro, et al; Phys. Rev. B, (1988), 38 (12), 8437-8443

[3] M. Almeida; Phys. Rev. B., (1992), 45 (1), 161-170

Wetting/Drying Studies with Soda Lime Spheres

Normalized ATR Spectra - Nonbridging Oxygen Groups due to high cation content



Manuscript in preparation to Phys. Chem. B , 2005

Conclusion: ATR Studies of Soda Lime Glass Spheres

Soda Lime Glass Microspheres

chemical composition	
NaO	- 13.7%
CaO	- 9.8%
MgO	- 3.3%
Al_2O_3	- 0.4%
FeO, $Fe2O_3$	- 0.2%
K_2O	- 0.1%

- Problem for fundamental studies: variation of chemical composition

- Batch to batch variations

Conclusion: ATR Studies of Soda Lime Glass Spheres

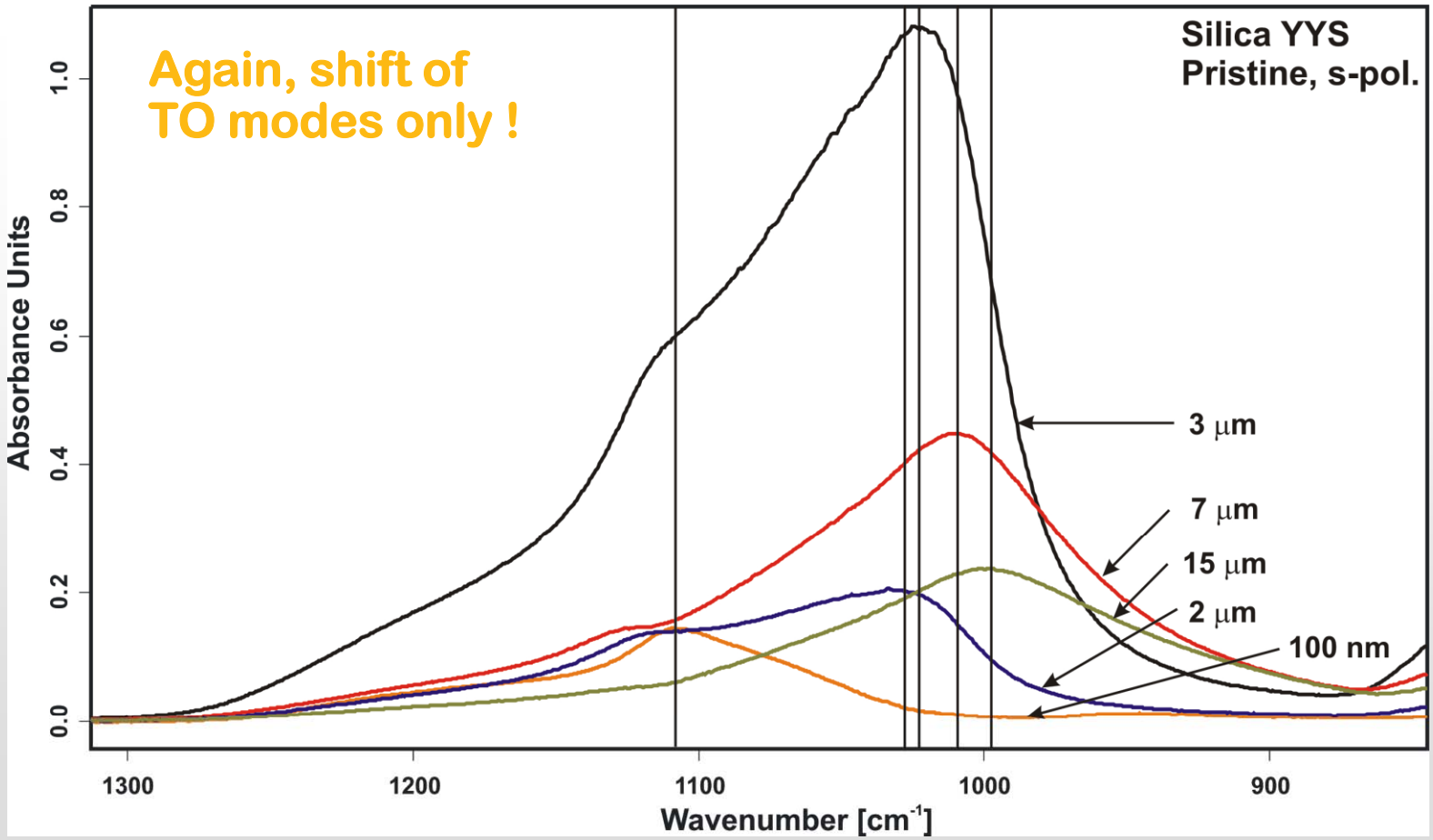
- **Wetting/drying studies of mono-disperse samples show no spectral shift and no changes in intensity**

→ **Changes in disturbed / undisturbed soil spectra have to be particle size related !**

- Significant shifts and changes of relative intensities of other absorption bands
- $Si\text{-}O^-$ vibrational band increases in intensity in relation to other spectral features with increasing particle size
- Experiments with polarized light show again strong LO-TO mode splitting (Berreman effect) of the broad absorption features
- Only TO bands show significant particle size related changes
- LO bands do not show significant particle size

Silica Micro- and Nanospheres

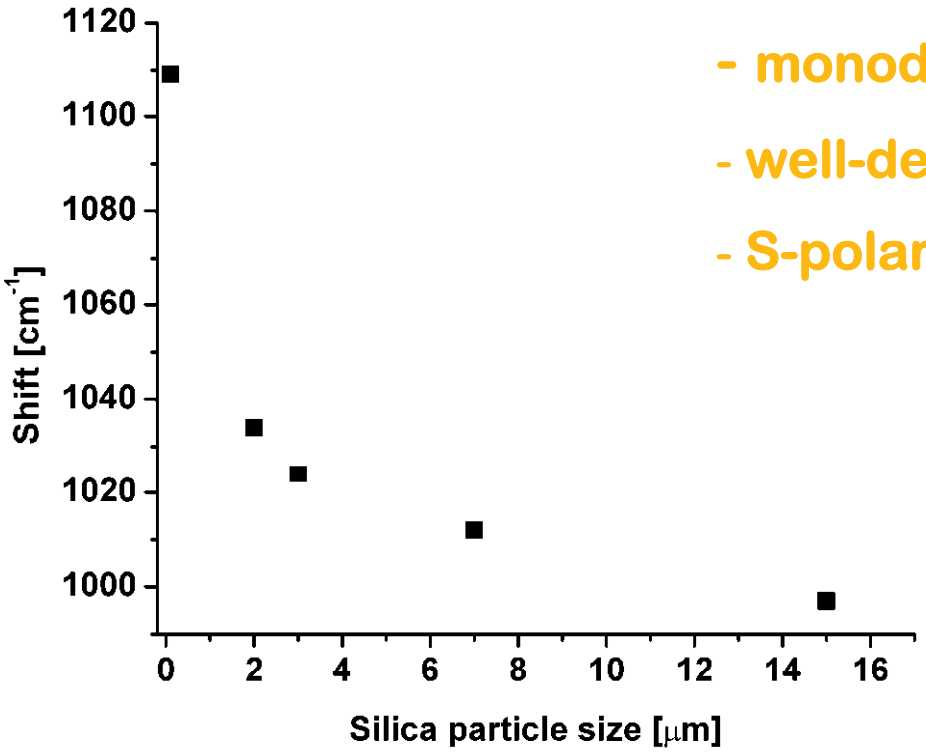
ATR Spectra – C. P. Wong's Silica Spheres – s-Polarization



Manuscript in preparation to Phys. Chem. B , 2005

Silica Micro- and Nanospheres

ATR Spectra – C. P. Wong's Silica Spheres – Shift of TO modes



- monodisperse pristine spheres
- well-defined material
- S-polarized light

Manuscript in preparation to Phys. Chem. B , 2005

Conclusion: ATR Studies of Silica Spheres

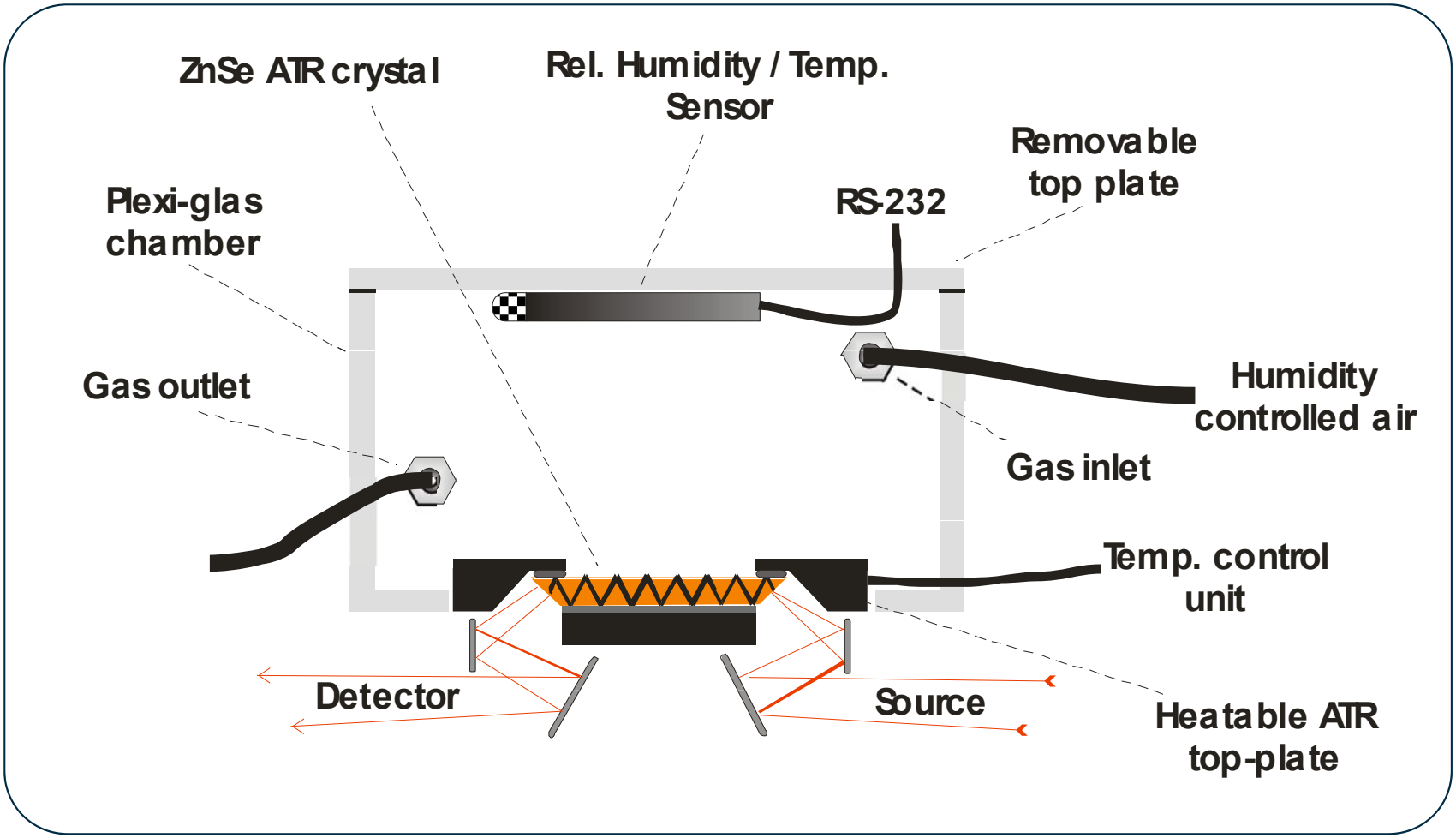
- Wetting/drying studies of silica samples show only spectral changes in intensity
- Wetting/drying studies of silica samples show no significant shift of absorption bands
- **BUT: significant particle size related spectral shifts with increasing particle size**
- Experiments with polarized light show strong LO-TO mode splitting (Berreman effect) of the broad absorption features with substantial shift of TO mode(s) in dependence on particle size

Outlook

- Continue studies with silica micro- and nanospheres of different particle size (higher µm range, lower nm range)
- Comparison with bulk spectral behavior of same material
- Studies under controlled environments (humidity, temperature)
- Investigation of real world samples (e.g. Calcite, Kaolinite, etc.) and samples from field sites (e.g. AHI/tower experiments)
- Fractionation of natural samples with mechanical sieve shaker
- Diffuse reflectance studies (ideally simultaneous with ATR); system needs to be developed
- Optical simulations of Berreman effect (SCOUT/SPRAY)
- Implementation of these concepts into modeling efforts of this MURI

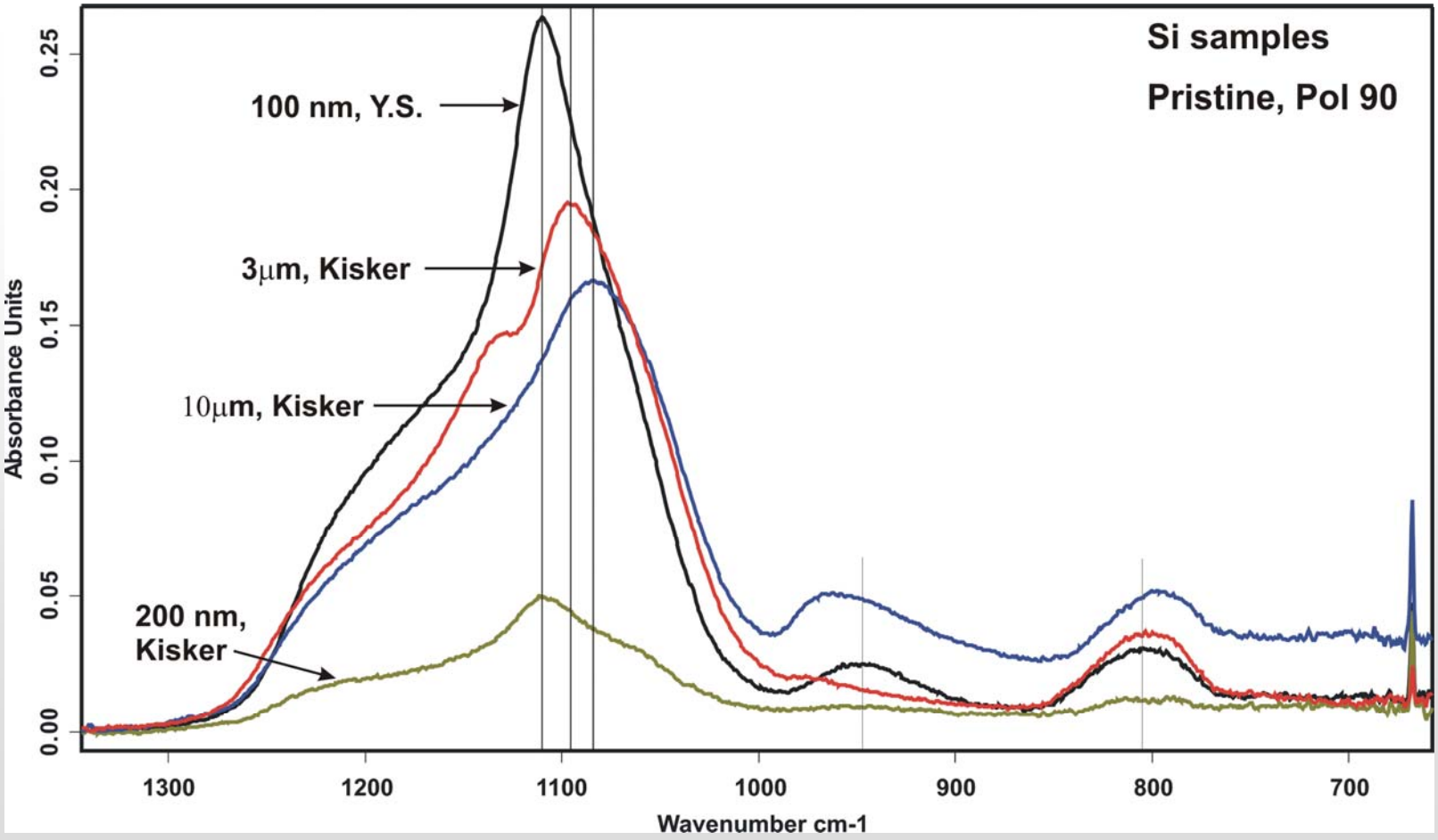
Outlook

Measurements in Controlled Environment



Wetting/Drying Studies with Silica Micro- and Nanospheres

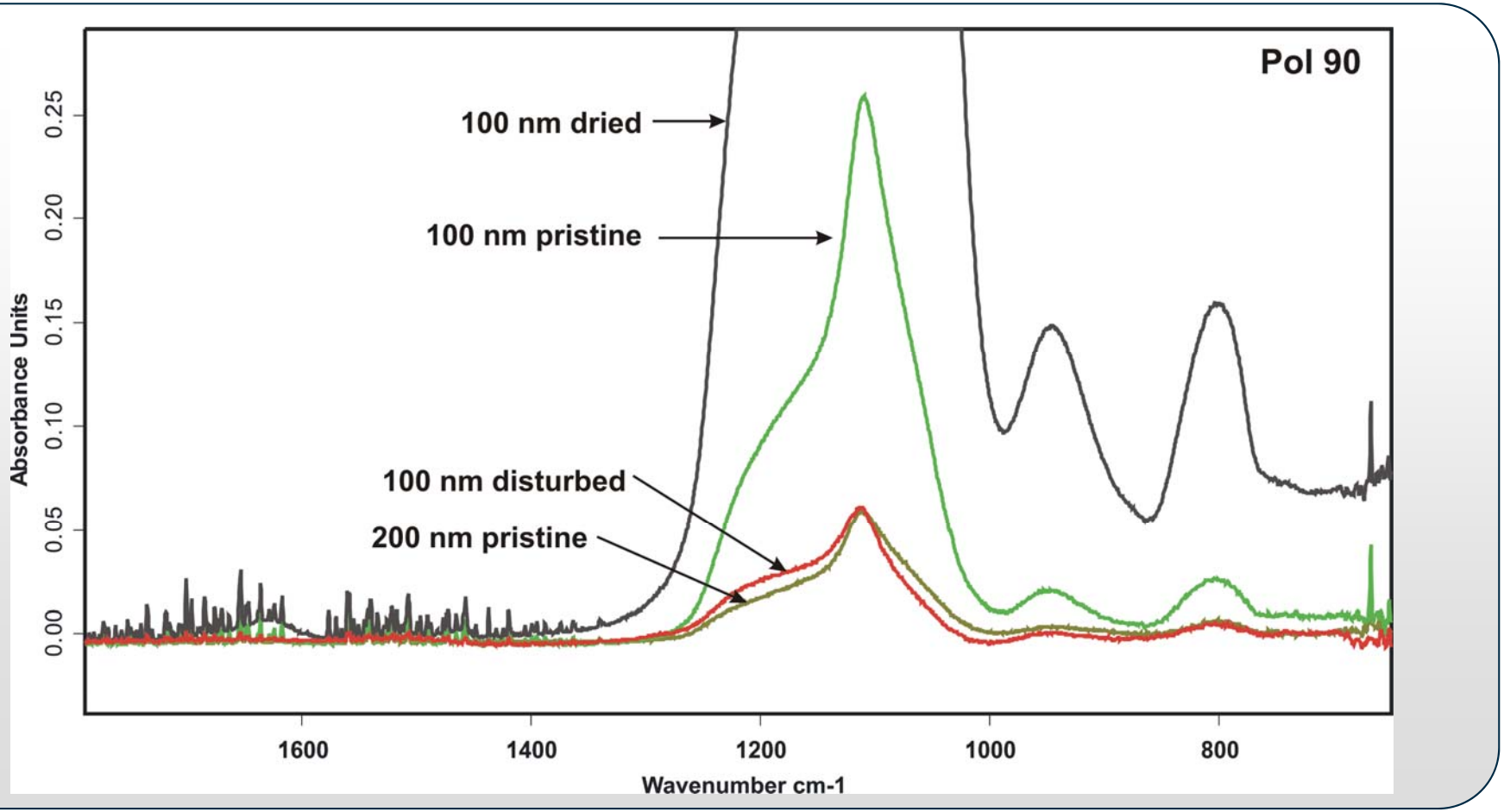
ATR Spectra of Silica Micro- and Nanopheres, p-Polarization



Manuscript in preparation to Phys. Chem. B , 2005

Wetting/Drying Studies with Silica Nano- and Microspheres

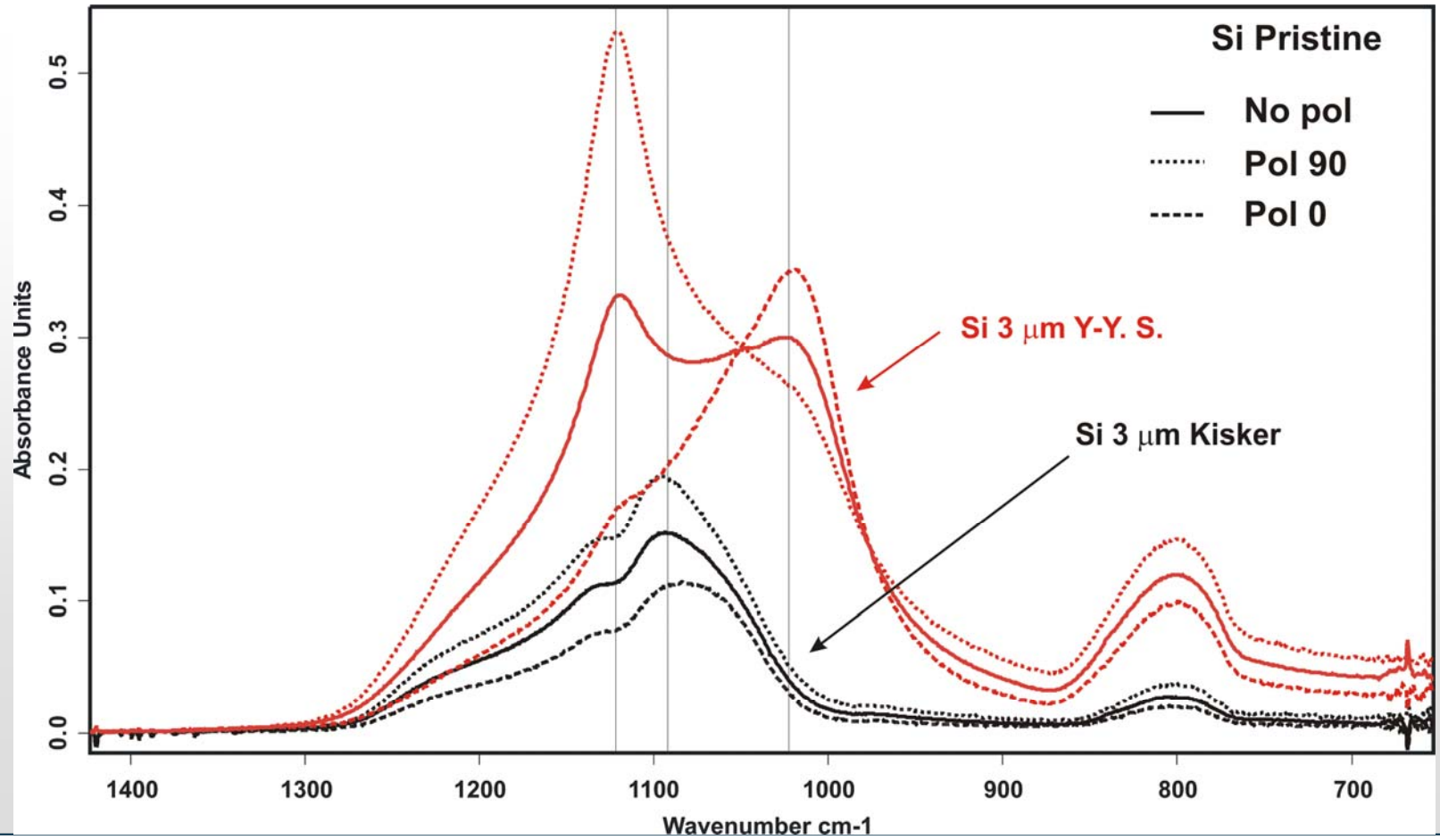
ATR Spectra – Kisker vs. C. P. Wong's Silica Nanospheres



Manuscript in preparation to Phys. Chem. B , 2005

Silica Nano- and Microspheres

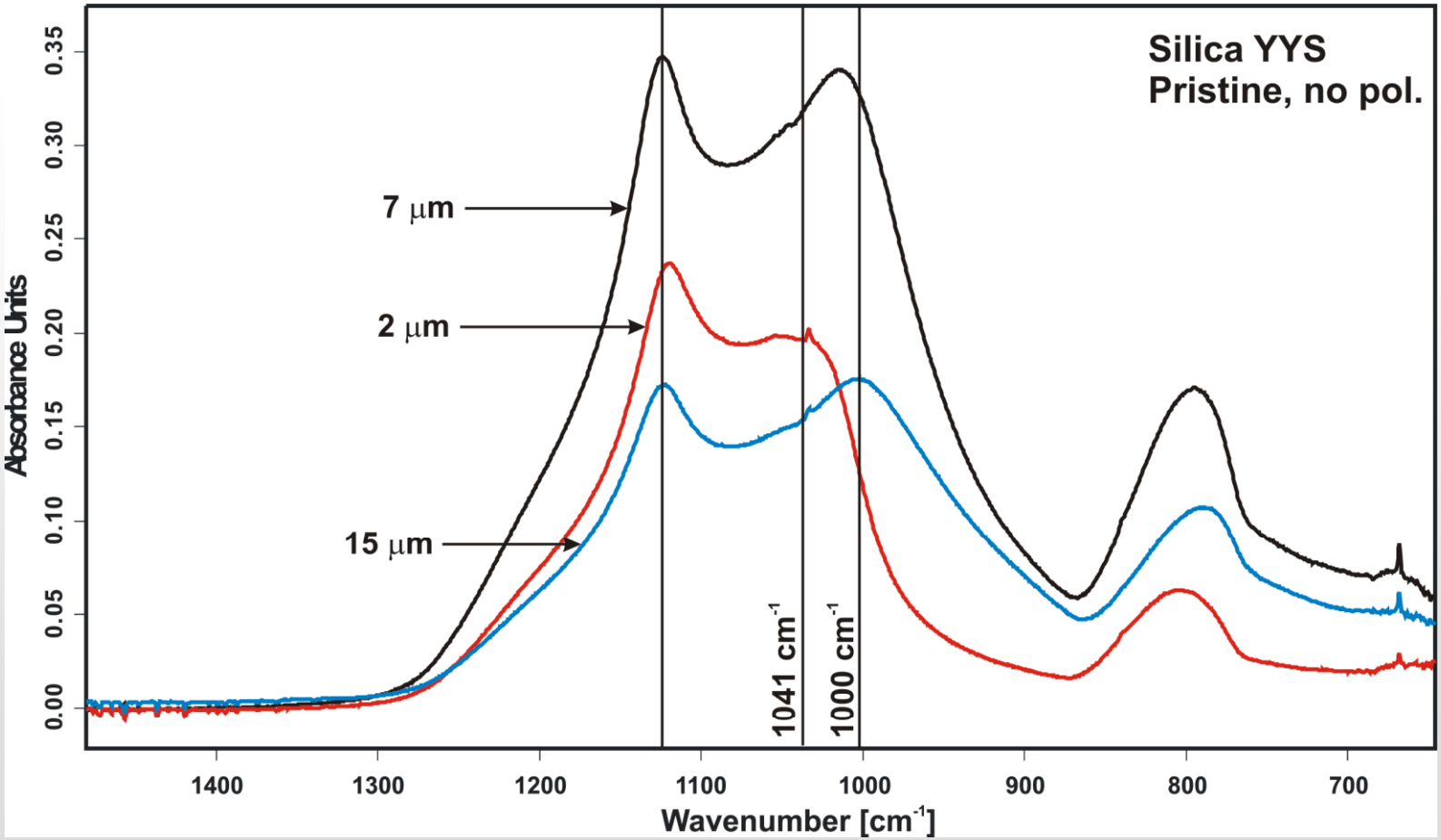
ATR Spectra – Kisker vs. C. P. Wong's 3 µm Silica Spheres



Manuscript in preparation to Phys. Chem. B , 2005

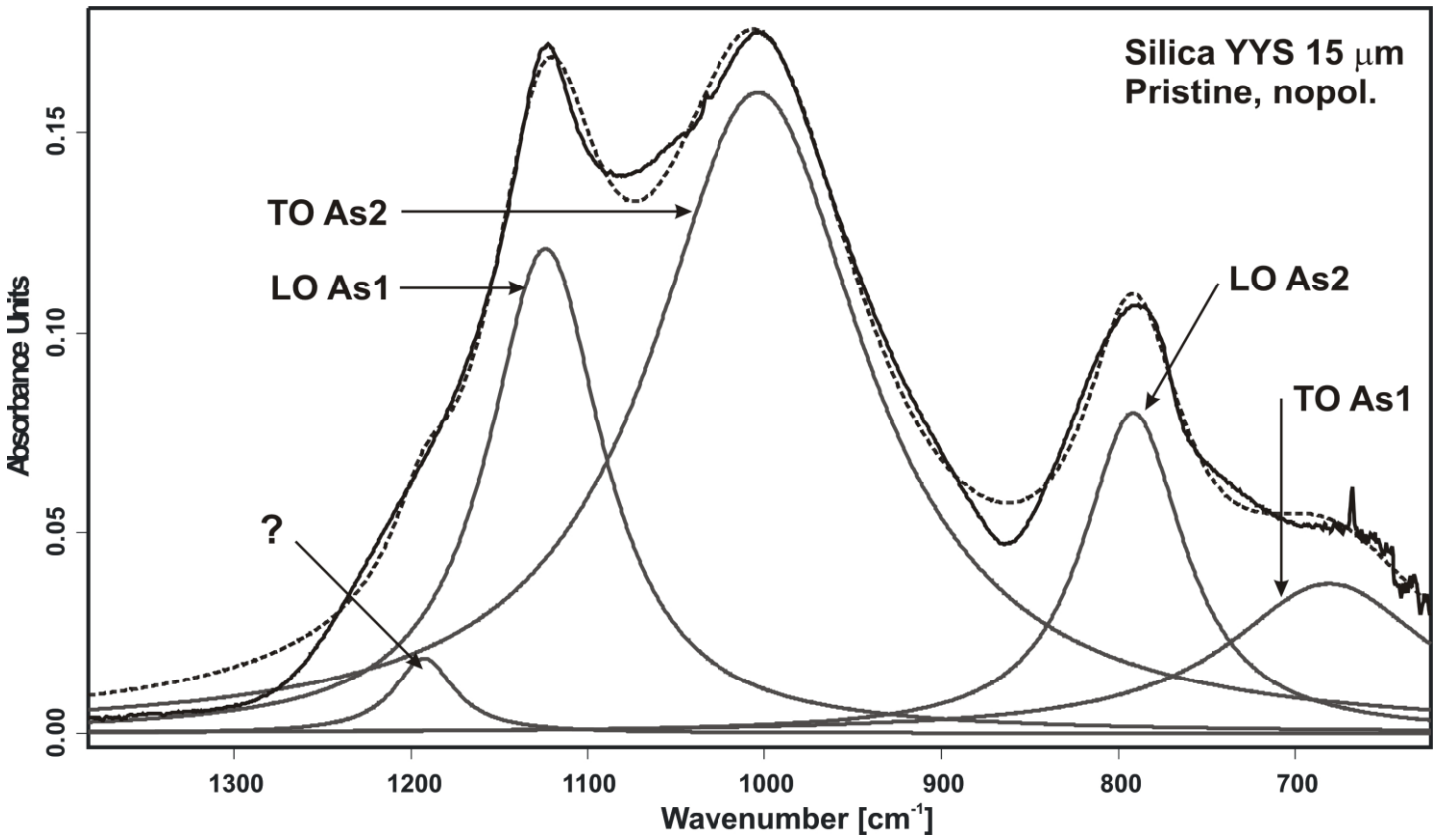
Silica Nano- and Microspheres

ATR Spectra – C. P. Wong‘ s Silica Spheres



Manuscript in preparation to Phys. Chem. B , 2005

Silica Nano- and Microspheres



Levenberg-Marquardt Curve fitting

AS_1: Asymmetric O-Si-O stretch in-phase

SS: Bending

AS_2: Asymmetric O-Si-O stretch out-of-phase

R: Rocking motions

Conclusion: ATR Studies of Silica Spheres

- Differences between Kisker & C. P. Wong's silica microspheres contributing to changes in LO-TO mode splitting?

  1) Fumed and fused silica

  2) Drying methods (oven, sublimation at 4 ˚C)

  May also contribute to differences in particle agglomeration → observed changes in intensity

- Can the TO shifts of soda lime spheres and silica spheres be related?

Acknowledgements

- Colleagues at ASL and EPIC laboratories especially
  - Manfred Karlowatz
  - Alexandr Aleksandrov
  - Thomas Orlando
- Michael Cathcart and all other members of the MURI program for fruitful discussions
- MURI for funding

Soil LWIR Polarization Effect

PI: Michael Cathcart & Robert Bock

Overview

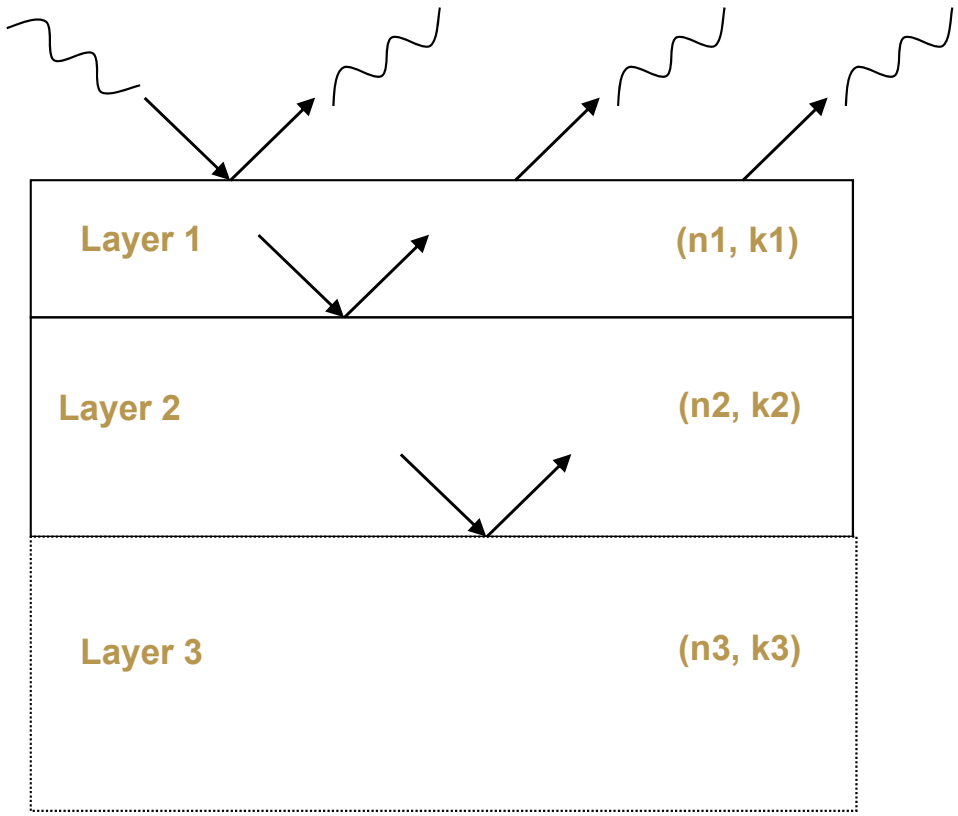
- Participants – School of Physics; Electro-Optics Laboratory
- Objectives
  - Conduct first-principles, physics-based theoretical investigations of the polarization signatures of land mines and soils
  - Provide theoretical analysis to support measurement and thermal modeling efforts
- Analysis Tools
  - MATLAB, ENVI

Analysis Tasks

- Develop a realistic first-principles polarization model for buried land mines
- Examine dependency of polarimetric properties on realistic factors
- Investigate diurnal variations of polarization signatures
- Examine the effect of the atmosphere on polarization signatures
- Examine the influence of particle sizes on the polarization metrics
- Investigate the source of spectral and polarimetric differences
- Incorporate thermal mine model results into analysis

‘Dirty’ Landmine – Reflection Model

- Classical electromagnetic modeling approach
- Fresnel equations for s- & p-reflection coefficients
- Multi-layer models:
  - 2 layer: Soil – Metal
  - 3 layer: Soil – Paint – Metal

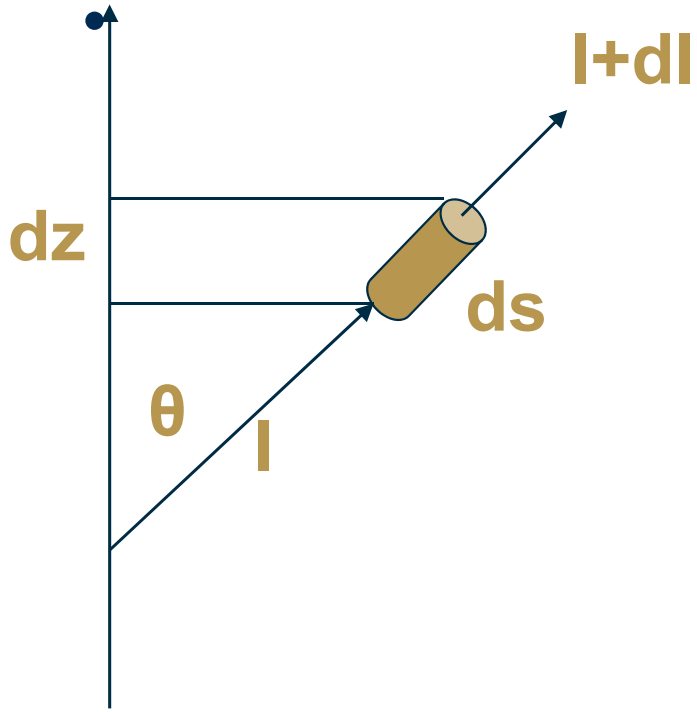


Why Radiative Transfer Modeling?

- Analytical and modeling results disagree with some field measurements
  - Classical Fresnel results indicate virtually no polarization in LWIR
  - Field measurements have shown some polarization effects (surface mines)
- Soil dominates the optical properties of buried and flush land mines – for all layers greater than micron depths
- A higher fidelity model of the optical properties of soil – both disturbed and undisturbed – and the effects of moisture and temperature was indicated
- All models of sand and the influence of moisture (primarily in the microwave) are semi-empirical
- No first principles digital model for IR optical properties of sand exists which include polarization

Radiative Transfer Overview

- The equation of radiative transfer describes the changes dI that occur in the beam of radiant energy as it traverses the cylinder of length ds
- Analytical solutions do not exist
- Particulate surface issues:
  - Previous approaches based on semi-empirical approach
  - Powerful radiative transfer models for the scattering of light by particulate surfaces have recently emerged
  - Numerical models required for realistic scenarios (e.g., radiation from particulate surfaces)



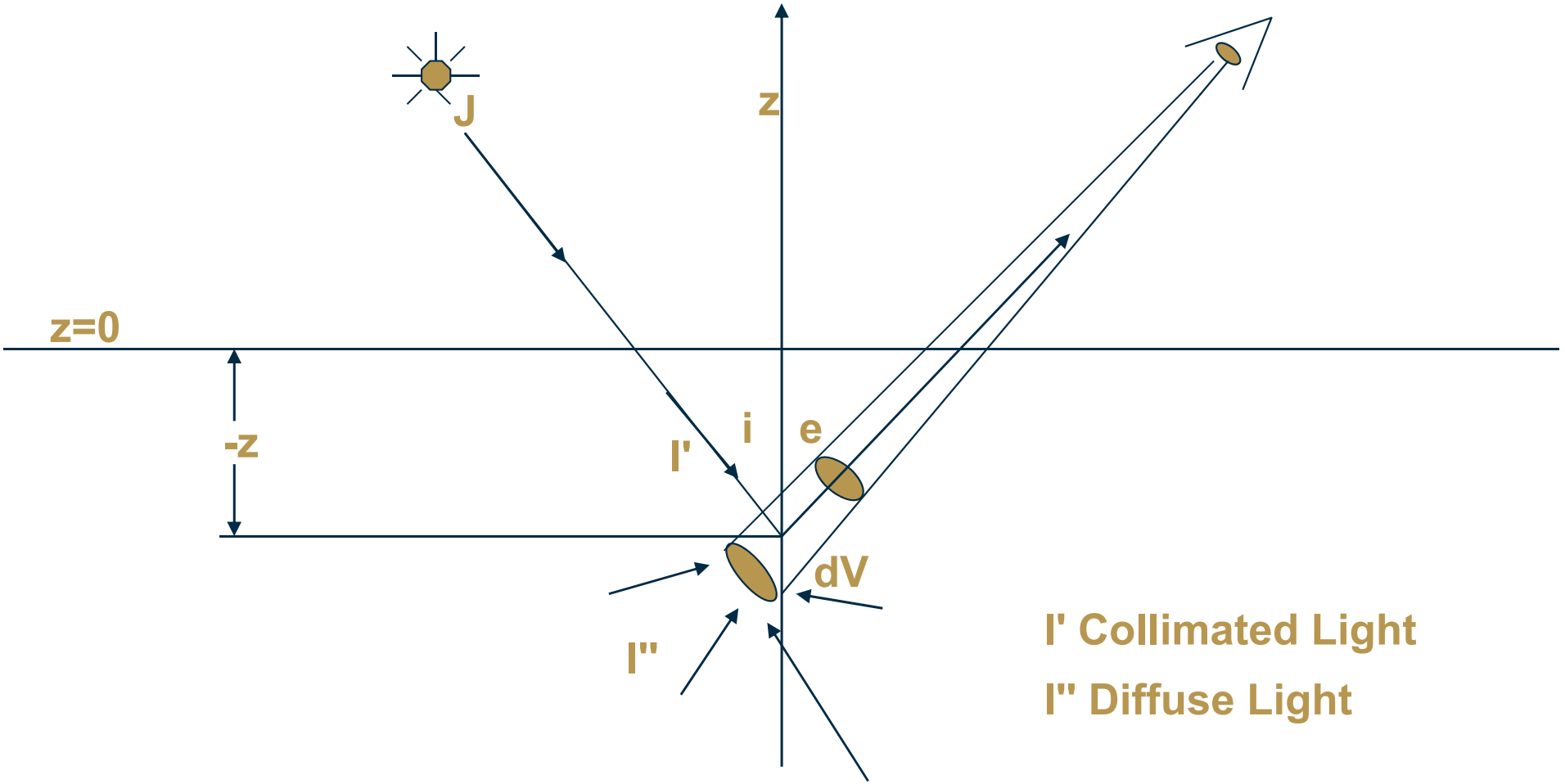
Hapke Reflectance Theory

- First analytic physical theory for the scattering of light by a surface composed of particles (Hapke, 1981)
- Only includes reflected light
- Assumes particles are large compared with the wavelength of light (2-3x)
- Allows for irregular particles
- Allows for close-packing (multiple particle scattering)
- Does not account for polarization or large-scale surface roughness

Combined Theory of Reflectance and Emittance Spectroscopy

- Hapke (1993)
- Modification of previous theory to include emission effects due to physical processes
- Same assumptions as before:
  - Assumes particles are large compared to the wavelength
  - Allows for irregular particles
  - Allows for close-packing
- Also does not account for polarization or large-scale surface roughness

The Geometry of the Model



Diffuse Light is scattered one or more times

Physical Model of Hapke Emission + Reflection Theory

- Consider an increment of volume dV located a depth z below the surface a distance r from the detector
- There are three contributions to the radiance received by the detector from dV:

$$I = I_1 + I_2 + I_3$$

I_1 • Radiation from the source (sun) scattered once by the particles in dV into the direction toward the detector

I_2 • Radiation thermally emitted by the particles in dV toward the detector

I_3 • Radiation that has been emitted or scattered at least once, impinging on the particles in dV and being scattered toward the detector

Method of Solution

- The radiance is the integral of I over all depths:

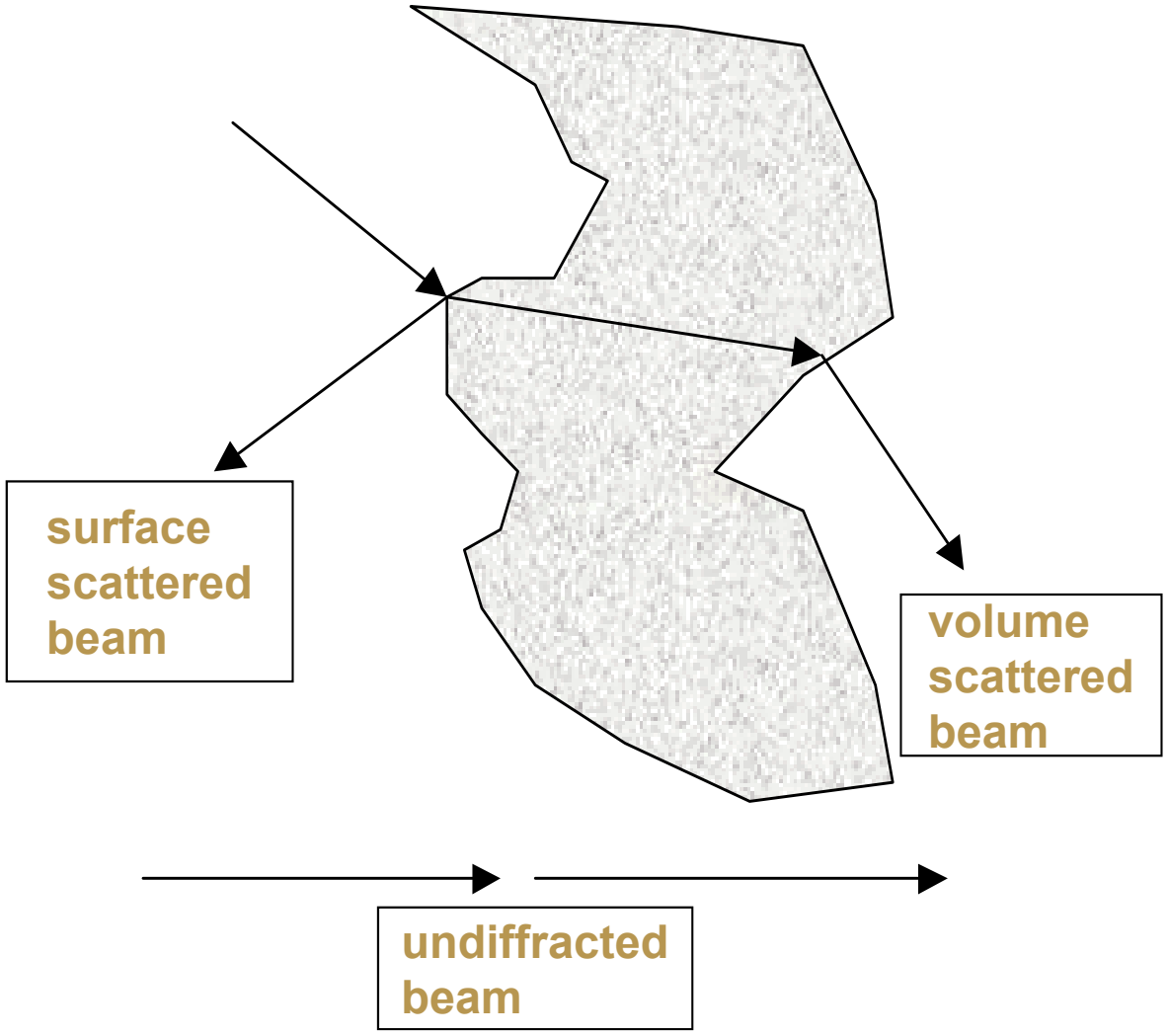
$$I = \frac{1}{4\pi}\int_{-\infty}^{0}\left[Jwp(g)e^{\frac{u}{\mu_0}} + 4\gamma^2 B(T) + w\int I(u,\Omega')\,p(g')\,d\Omega' \right] e^{\frac{\mu}{\mu_0}}\,\frac{du}{\mu}$$

- The first two terms can be evaluated directly
- The last term (multiple-scattered radiation) is evaluated using the equation of radiative transfer (with the two-stream approx.):

$$\frac{dI(z,\Omega)}{ds} = -EI(z,\Omega) + \frac{1}{4\pi}\int I(z,\Omega')G(g')\,d\Omega' + Je^{-E|z|/\mu_0}\,\frac{G(g)}{4\pi} + \frac{F}{\pi}B(z)$$

Particle Scattering Assumption

- In order to evaluate I, one must consider in detail how a large, irregular particle refracts light
- Hapke theory does not include diffraction
  - geometrical optics
- Scattering from a large particle takes place by two processes:
  - External – off the surface of the particle
  - Internal – volume scattering of rays which have been refracted into the interior of the grain and scattered or refracted back out
- Under these assumptions, the following is calculated:
  - External scattering coefficient
  - Internal scattering coefficient
  - Scattering efficiency
  - Single-scattering albedo



Polarization Extension

- Hapke theory does not include polarization
- First-order polarization effects are included through the following assumptions:
  - Incident radiation is unpolarized
  - Radiation that has been emitted or scattered at least once impinging on the particles in dV is unpolarized – becomes polarized under scattering toward detector
  - Volume-scattering contribution unpolarized
    - Assume that for an irregular particle the orientation of the scattering plane of the refracted light is randomized so that:

$$\omega_{\perp} = R_{\perp} + \frac{\left(w - S_E\right)}{2} \qquad \omega_{\parallel} = R_{\parallel} + \frac{\left(w - S_E\right)}{2}$$

  - Intensity equations derived for each field component
  - Polarization metrics computed

Intensity Equations

- **Emissive Component**

$$I_{\perp}^{(T)}(i,e) = \left[\frac{B_0}{\pi} \gamma_{\perp} H(w_{\perp}, \mu) + \frac{B_1}{\pi} \frac{L}{L+\mu} \gamma_{\perp}^2 H(w_{\perp}, L) H(w_{\perp}, \mu) \right]$$

$$I_{\parallel}^{(T)}(i,e) = \left[\frac{B_0}{\pi} \gamma_{\parallel} H(w_{\parallel}, \mu) + \frac{B_1}{\pi} \frac{L}{L+\mu} \gamma_{\parallel}^2 H(w_{\parallel}, L) H(w_{\parallel}, \mu) \right]$$

- **Reflective Component**

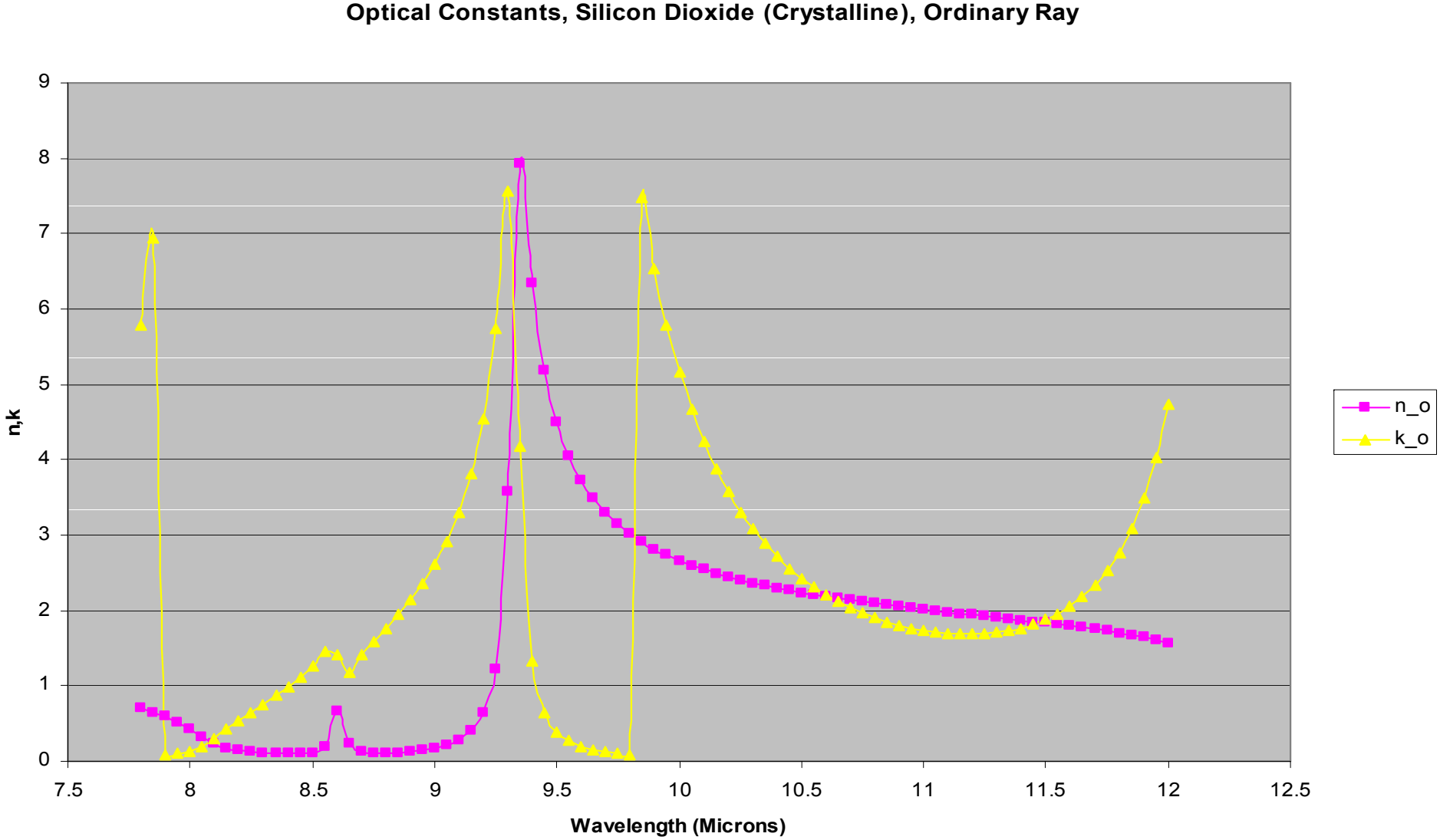
$$I_{\perp}^{(R)}(i,e,g) = \frac{J}{2} \frac{1}{4\pi} \frac{\mu_0}{\mu_0 + \mu} \left\{ \left[wp(g) \right]_{\perp} + \left(H(w,\mu_0) H(w,\mu) - 1 \right) \right\}$$

$$I_{\parallel}^{(R)}(i,e,g) = \frac{J}{2} \frac{1}{4\pi} \frac{\mu_0}{\mu_0 + \mu} \left\{ \left[wp(g) \right]_{\parallel} + \left(H(w,\mu_0) H(w,\mu) - 1 \right) \right\}$$

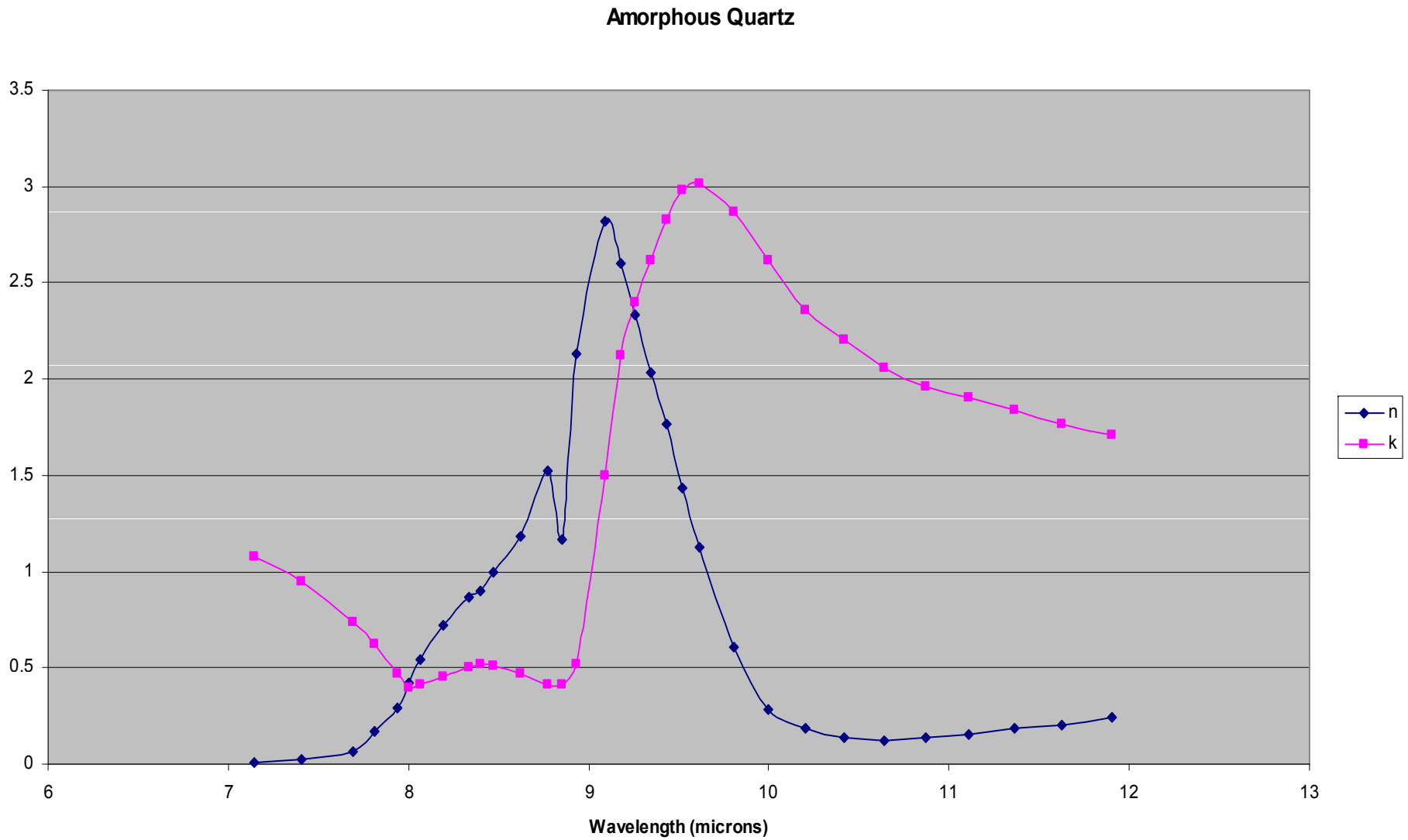
Soil Composition

- Solid phase of soil composed of primary and secondary minerals
- Three main particle size fractions:
  - Sand (2000 to 200 μm)
  - Silt (200 to 2 μm)
  - Clay (<2 μm)
- Primary Minerals
  - Quartz, feldspar, orthoclase, plaglioclase
- Secondary Minerals
  - Dependent on weathering stage
  - Aluminosilicates, aluminum and iron oxides, hydroxides, carbonates
- Disturbed vs. Undisturbed Soil
  - Disturbed: ~ 10 μm
  - Undisturbed: >100 μm

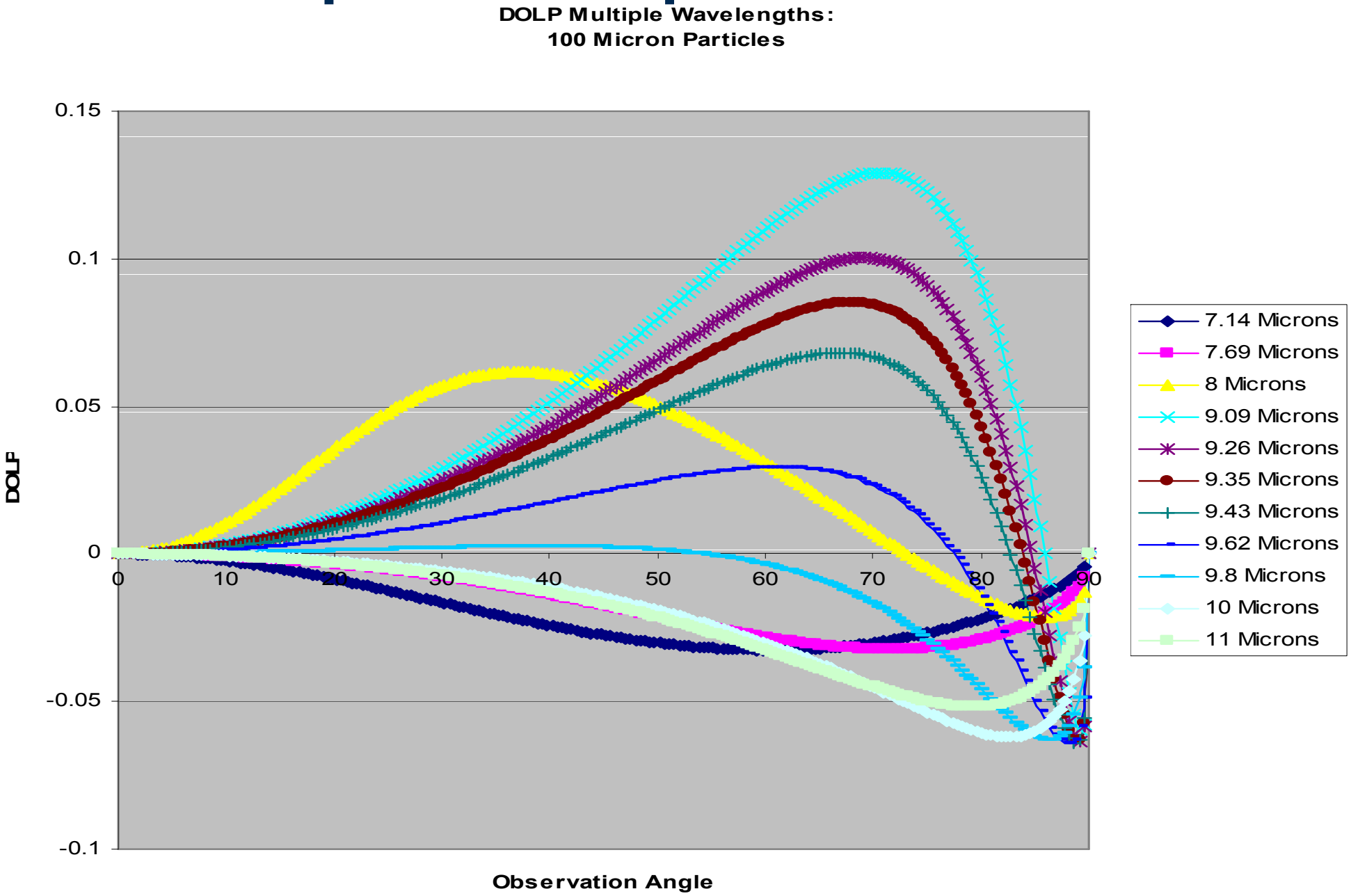
Crystalline Quartz



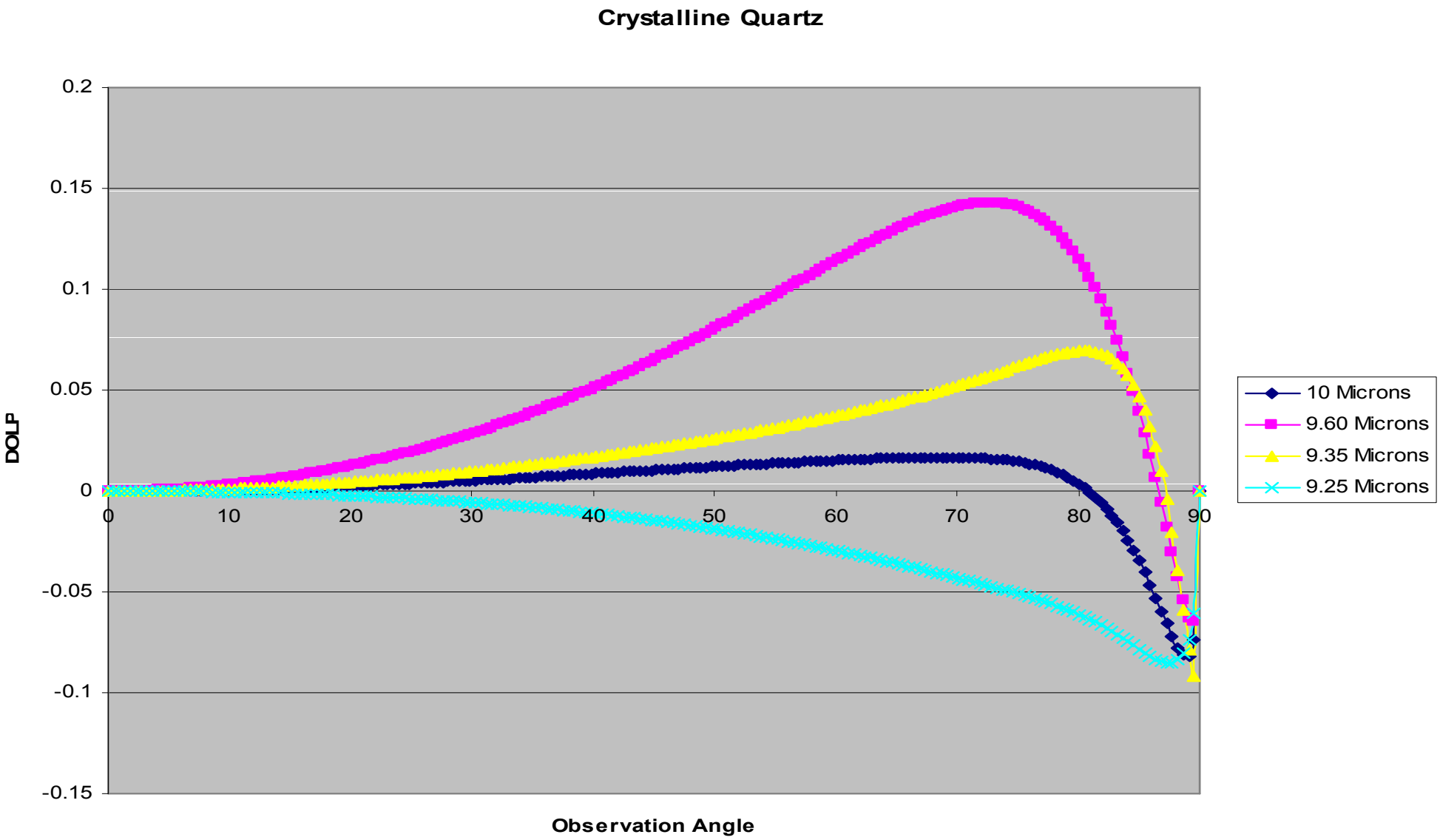
Amorphous Quartz



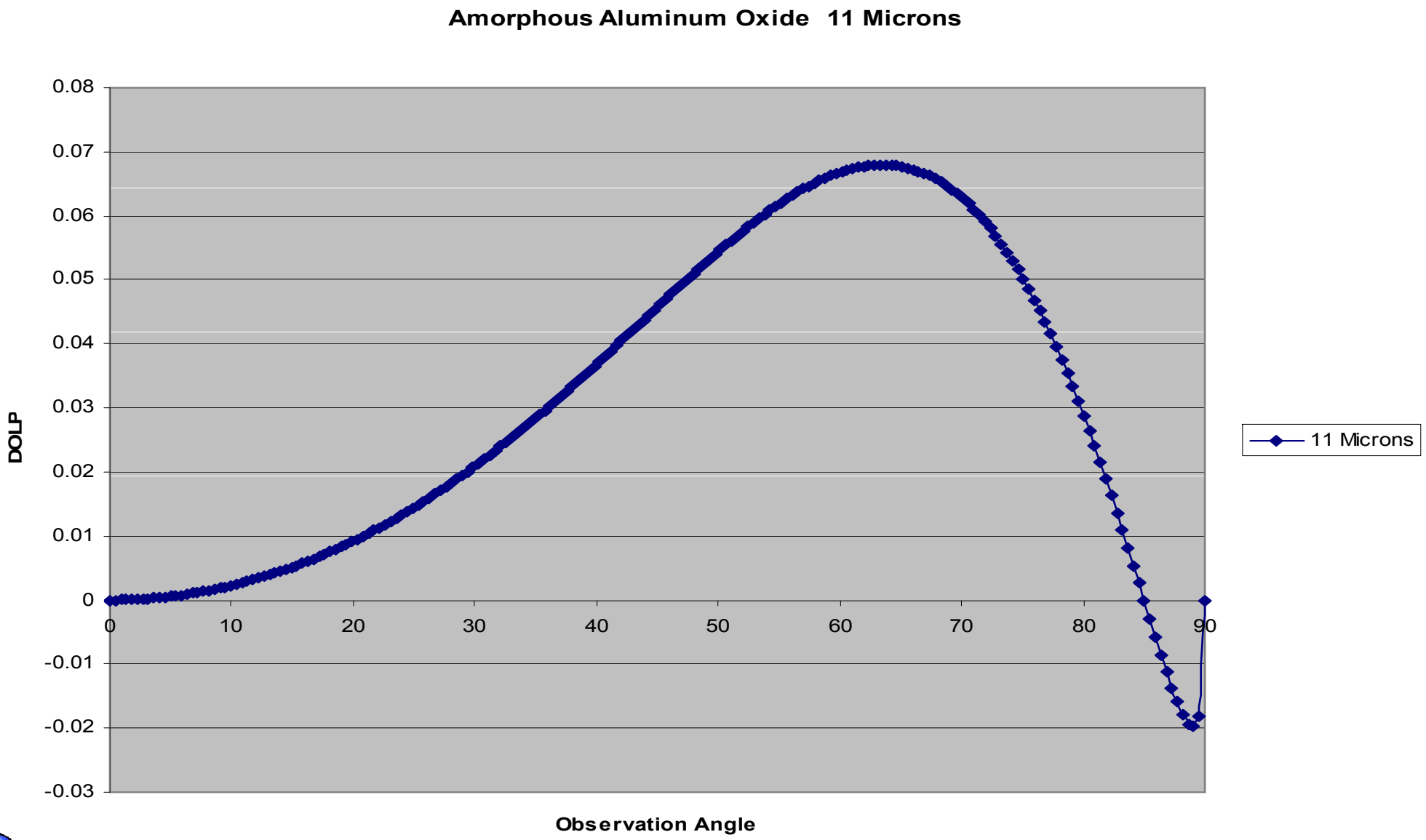
Hapke: Amorphous Quartz



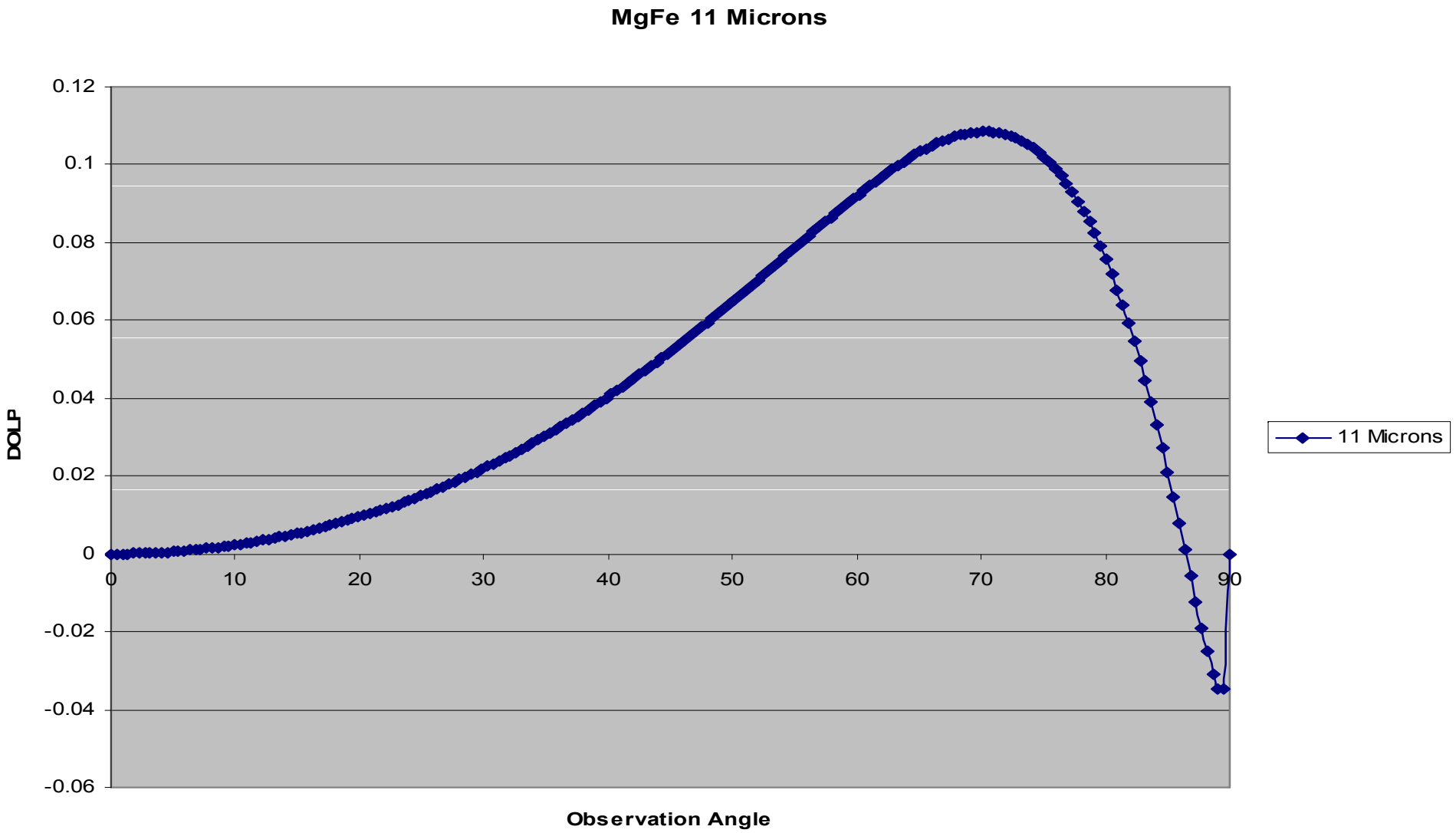
Hapke: Crystalline Quartz



Hapke: Amorphous Aluminum Oxide



Hapke: Magnesium Ferrite



Particle Size Effects

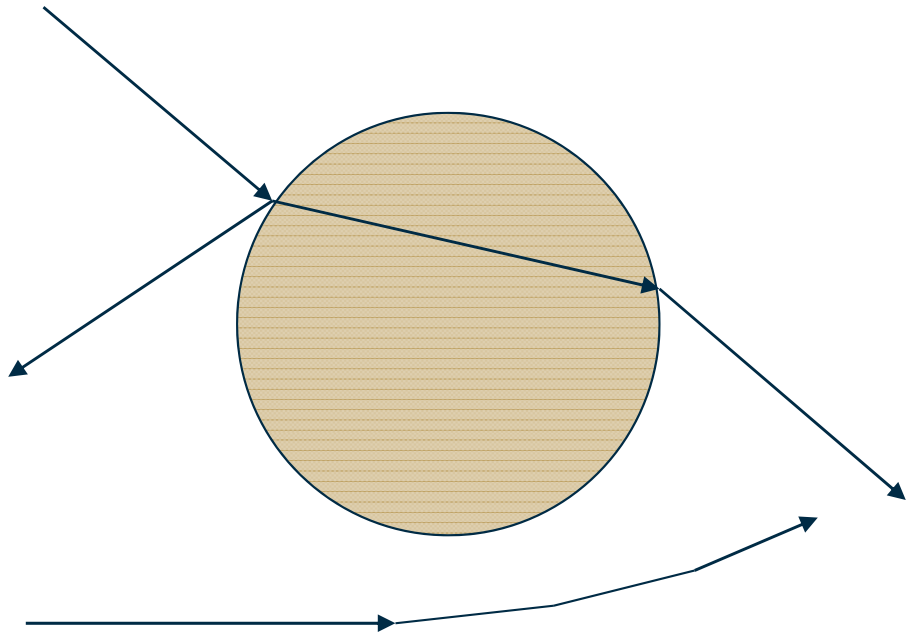
- Hapke Theory only valid for particles large compared to the wavelength of light
- Need to extend polarization model to include effects that are likely to be important when the wavelength and the particle size are similar
- Mie Theory provides a tool for understanding the effects of decreasing particle size
- Hapke/Mie Hybrid
  - Proposed by Moersch and Christensen (1995)
  - Used the Mie-derived single-scattering albedo, corrected for close-packing (Wald subtraction), in the Hapke multiple-scattering theory
  - Provides best fit of observed emissivity measurements of quartz when compared to other models

Mie Scattering Approach

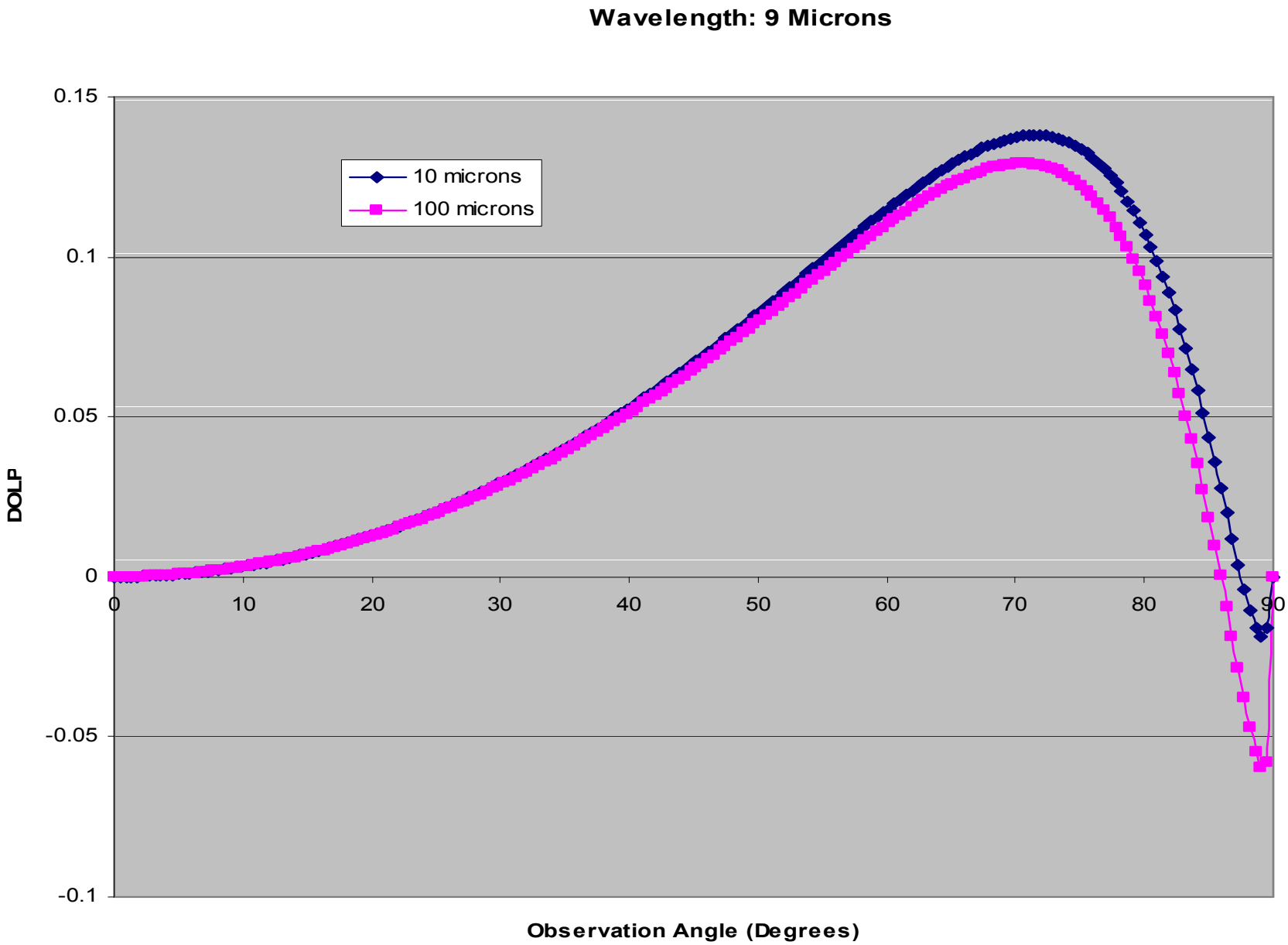
- Assumes spherical particles
- Assumes particles well-separated (by at least 3 radii)
- Allows for diffraction
- Two parameters control scattering
  - Ratio of particle size to wavelength:

$$X = \frac{\pi D}{\lambda}$$

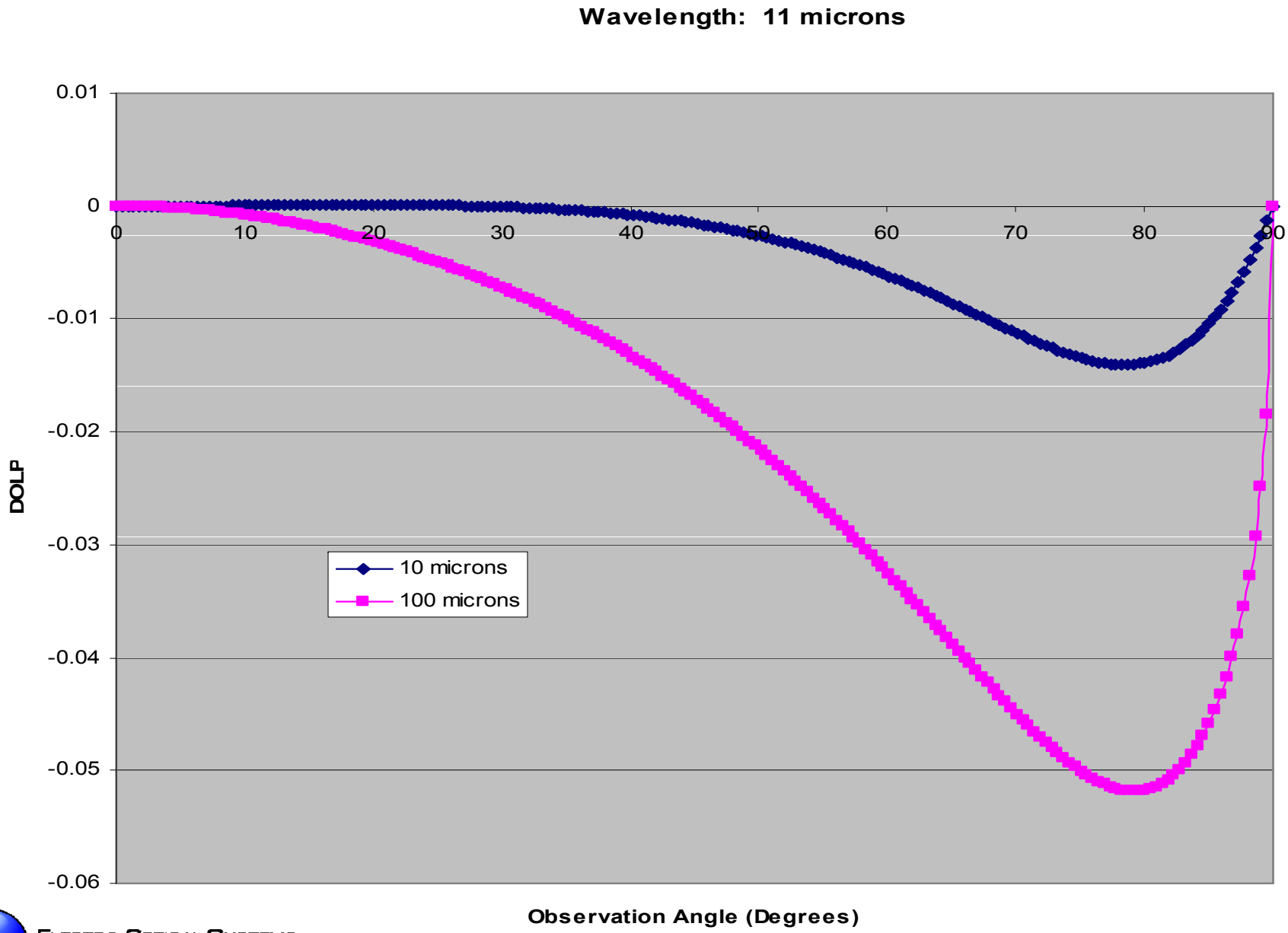
  - Relative refractive index: N = n - ik



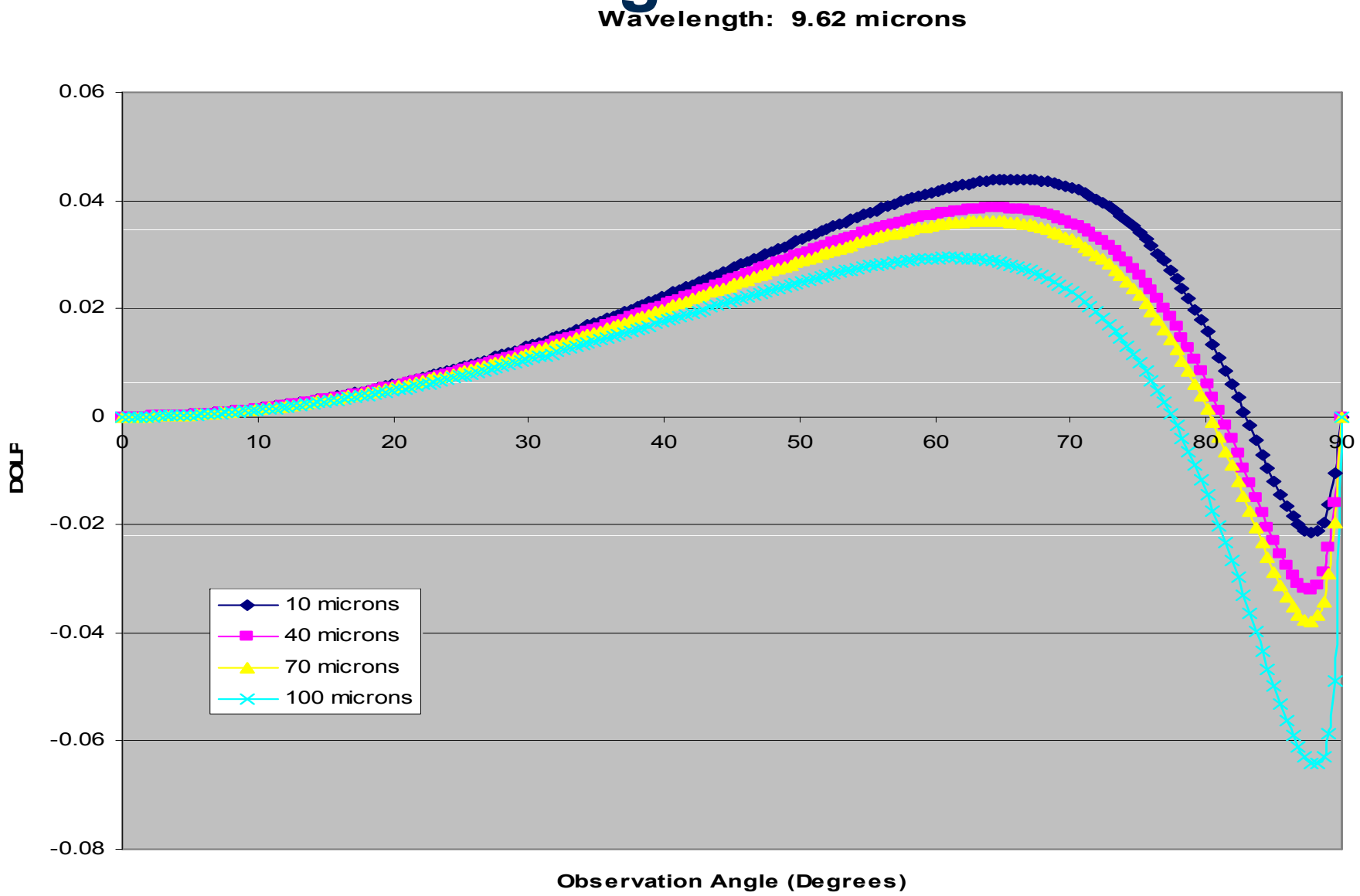
Particle Size Impact: 9 microns



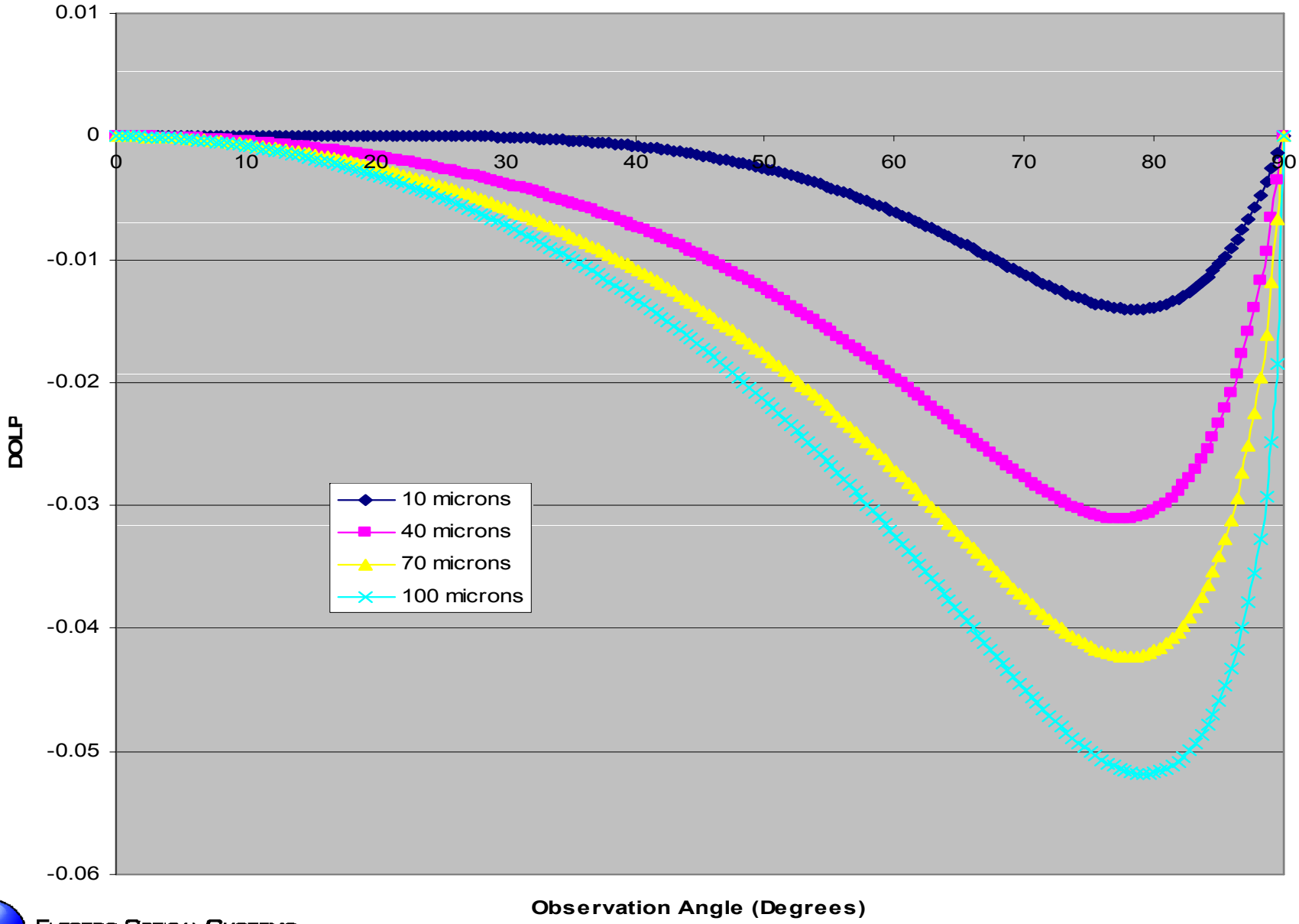
Particle Size Impact: 11 microns



Variation of Particle Size Wavelength: 9.62 microns



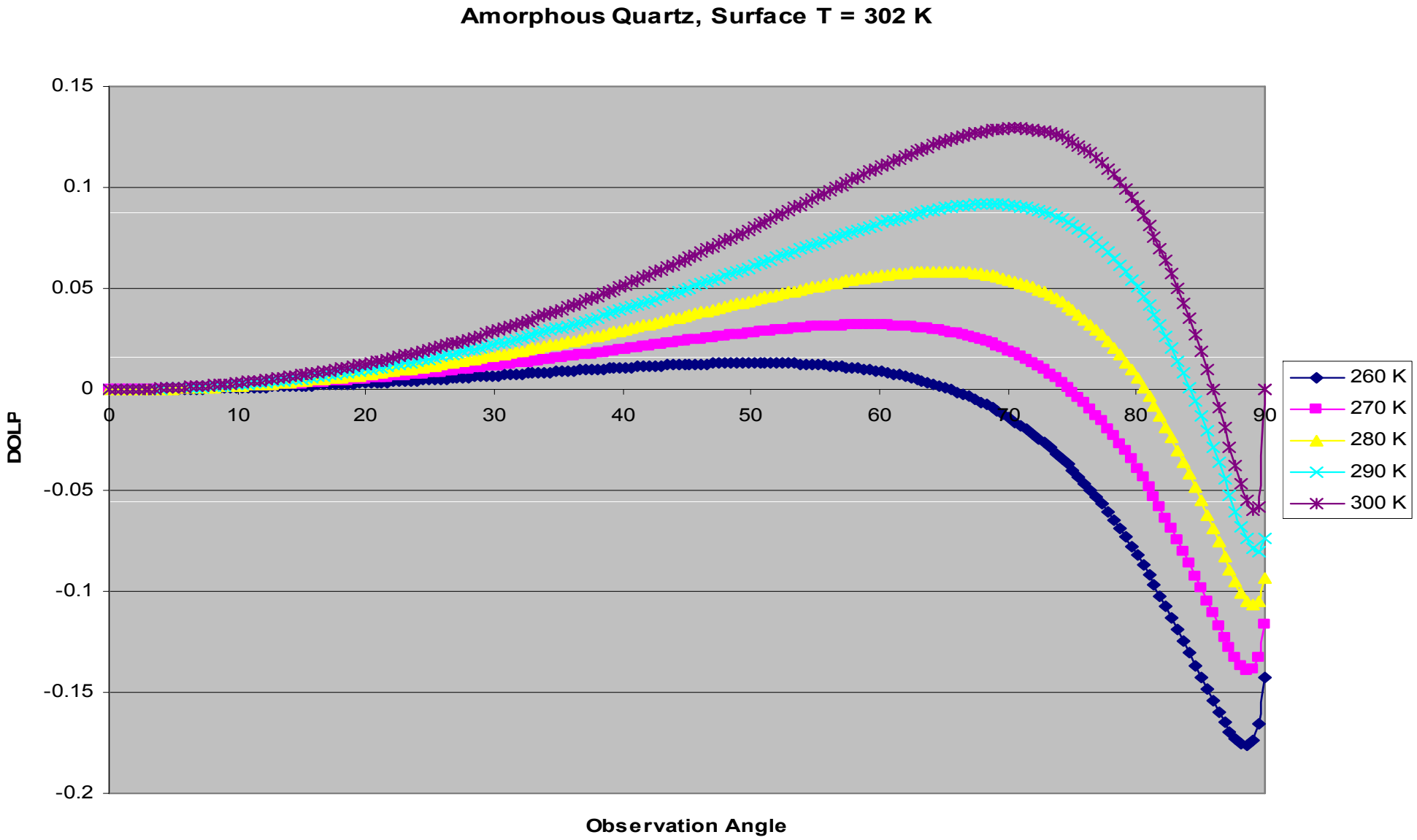
Variation of Particle Size Wavelength: 11 microns



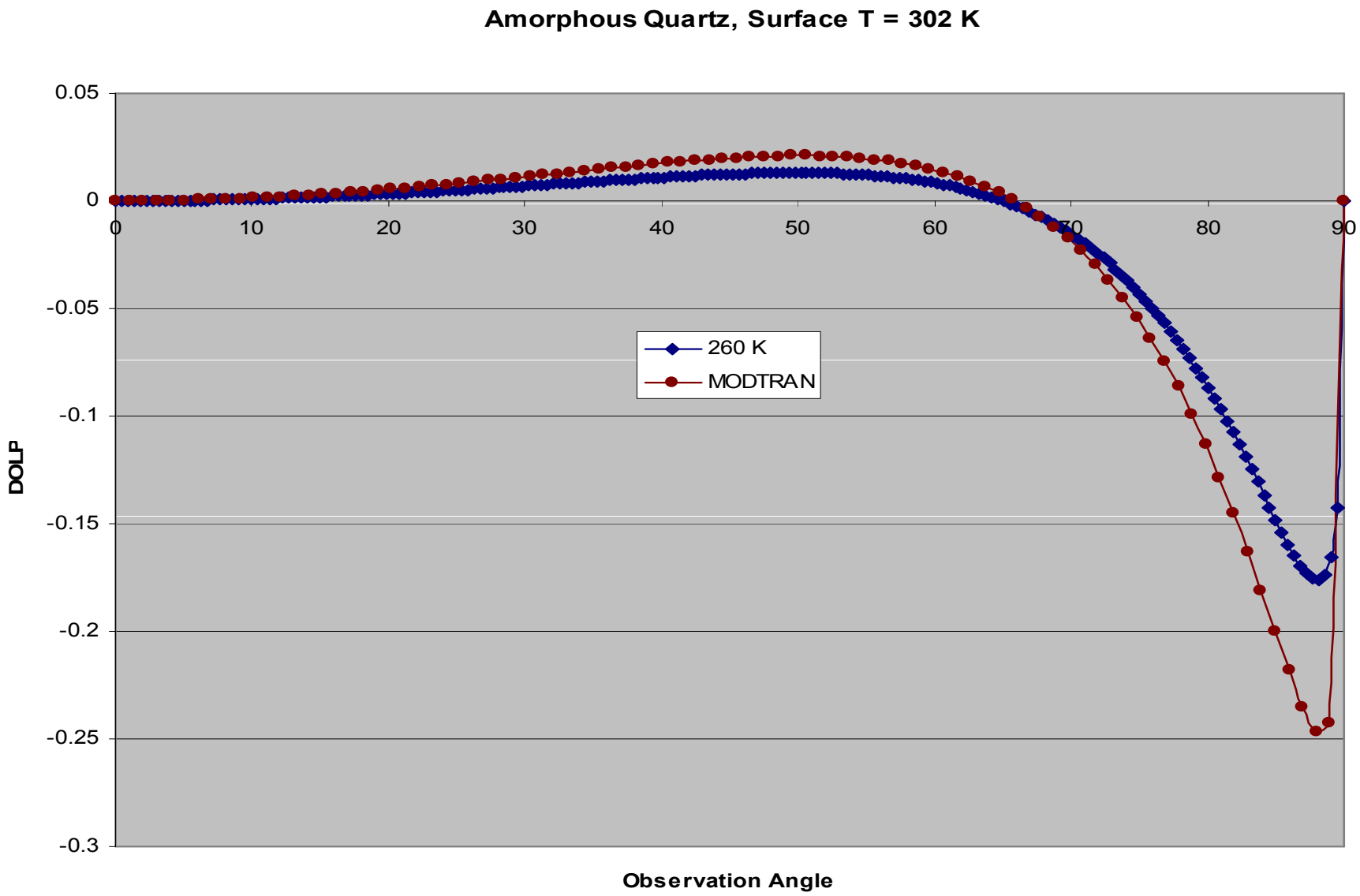
Atmospheric Impact

- Atmosphere primary source of reflective component in the LWIR
- Initial Calculations
  - Atmosphere treated as a black-body source at an effective temperature
- Realistic atmosphere
  - MODTRAN 4.0

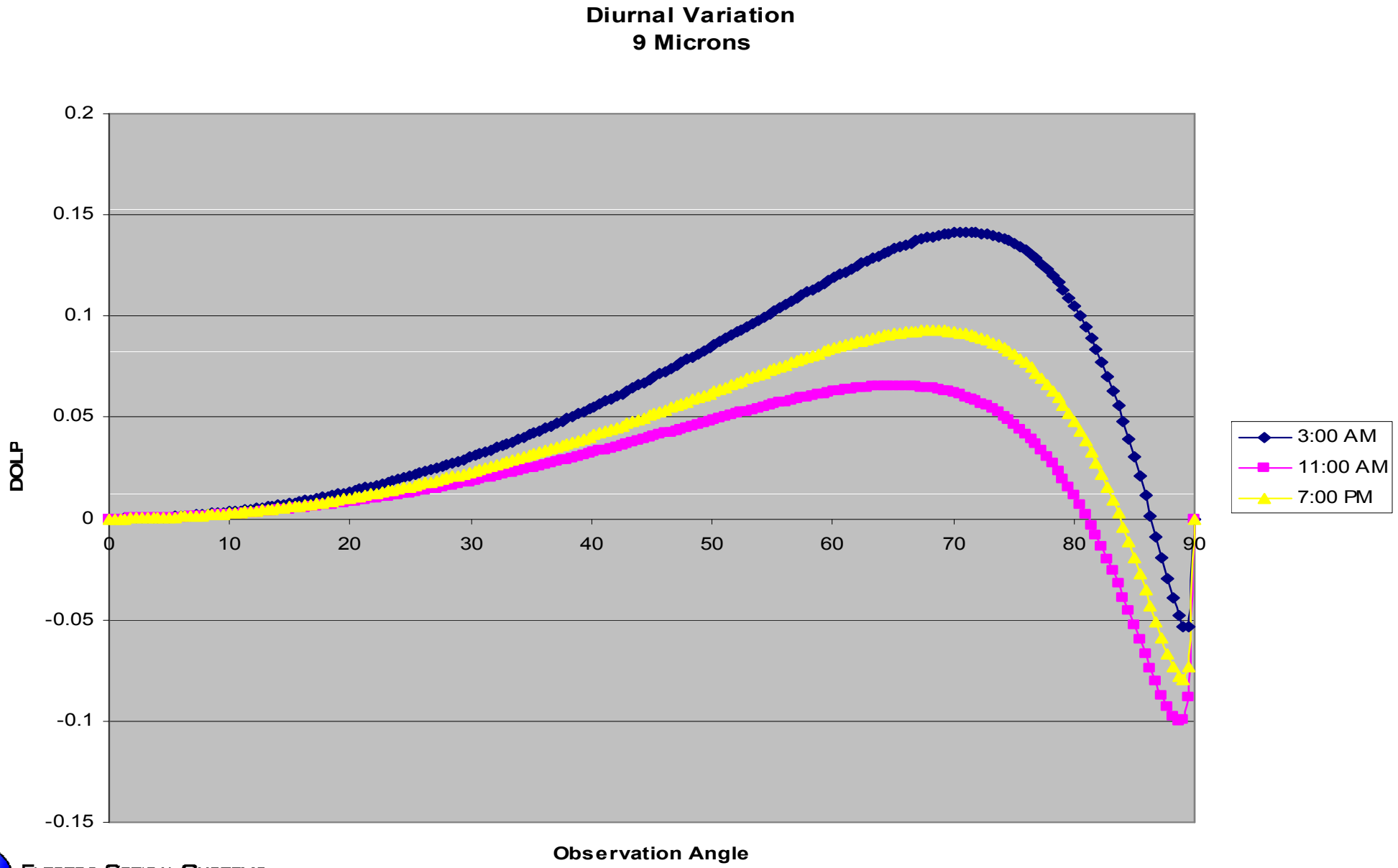
Atmospheric Impact: 9 Microns



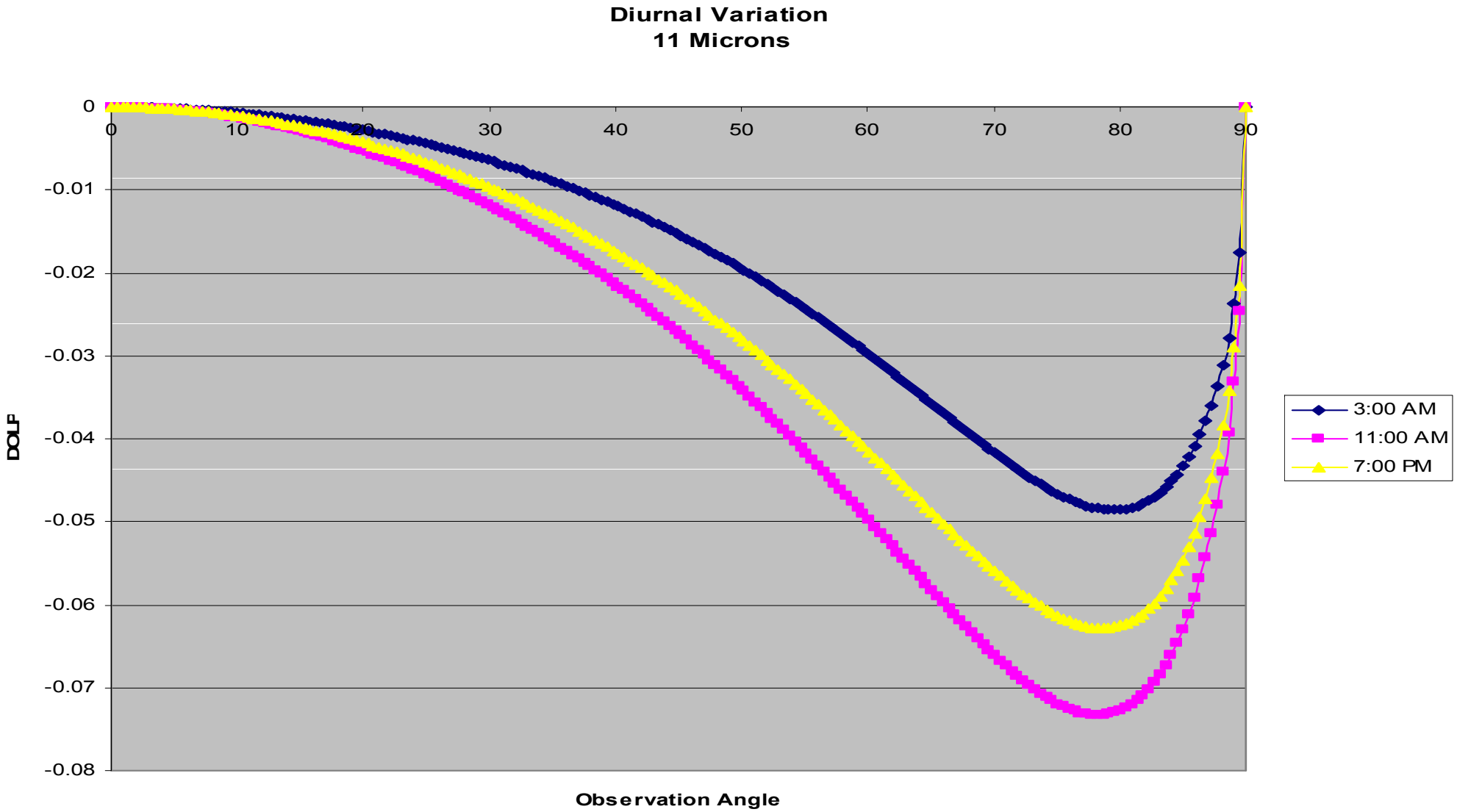
Atmospheric Impact



Diurnal Impact



Diurnal Impact



Summary

- DOLP predicted by Hapke theory differs substantially from that predicted by a simple Fresnel model.
- In the LWIR the Hapke/Mie hybrid theory predicts a particle-size dependence.
- DOLP sensitive to optical properties of material (n,k)
- DOLP sensitive to atmospheric conditions as evidenced by change in DOLP with time of day.
- Research continuation
  - Extend model to particles << λ
  - Extend calculations to particle size distribution
  - Minerals study (discrimination)
- Laboratory and field data needed

Soil Optical Properties

PI: Michael Cathcart

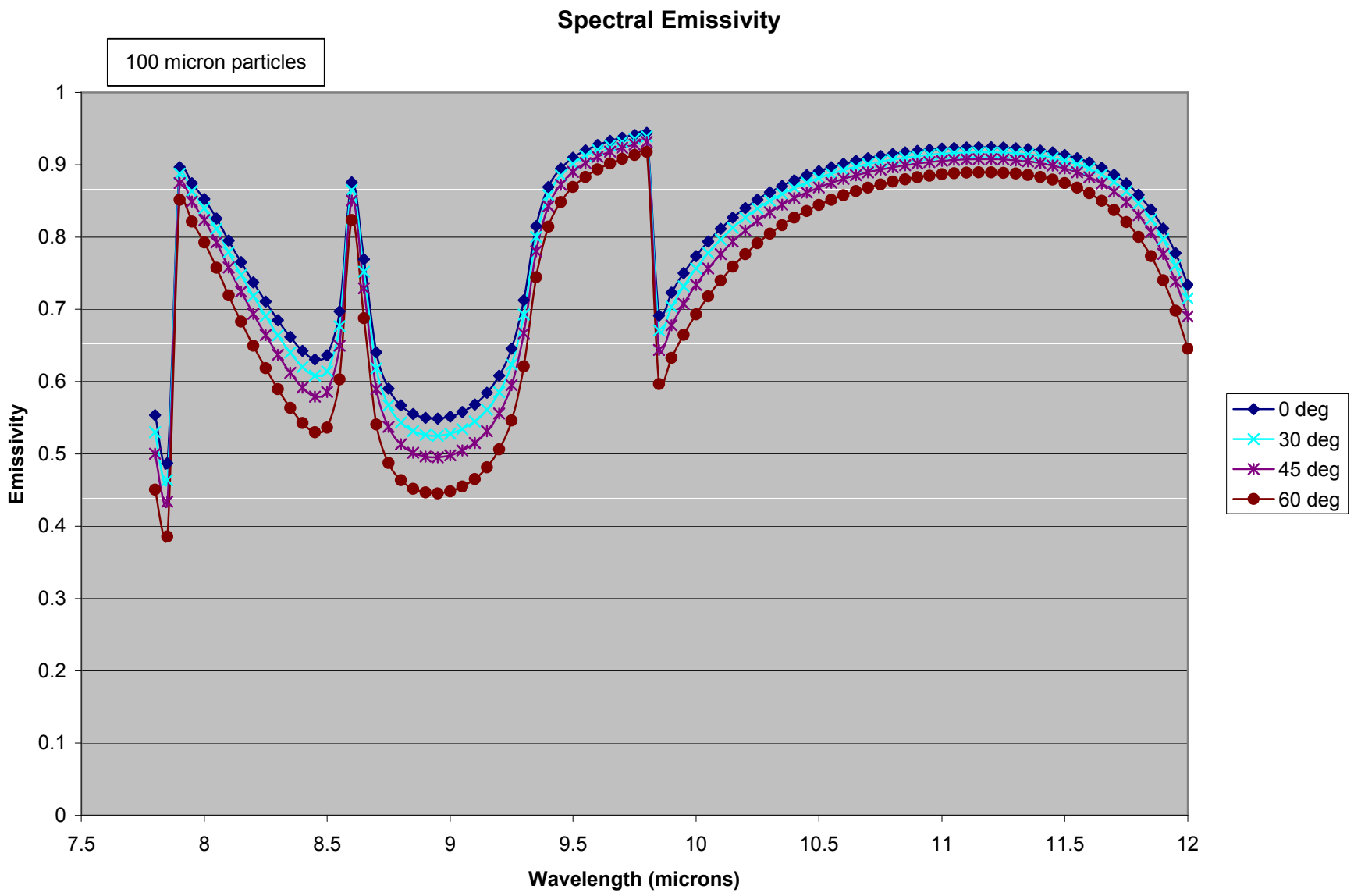
Overview

- Participants – School of Physics; Electro-Optics Laboratory
- Objectives
  - Conduct first-principles, physics-based theoretical investigations of optical property computation of optical properties of soils
  - Develop physics-based models of optical properties of quartz-laden soils
  - Employ GT-modified version of Hapke reflectance / emittance theory as the basis for the analysis

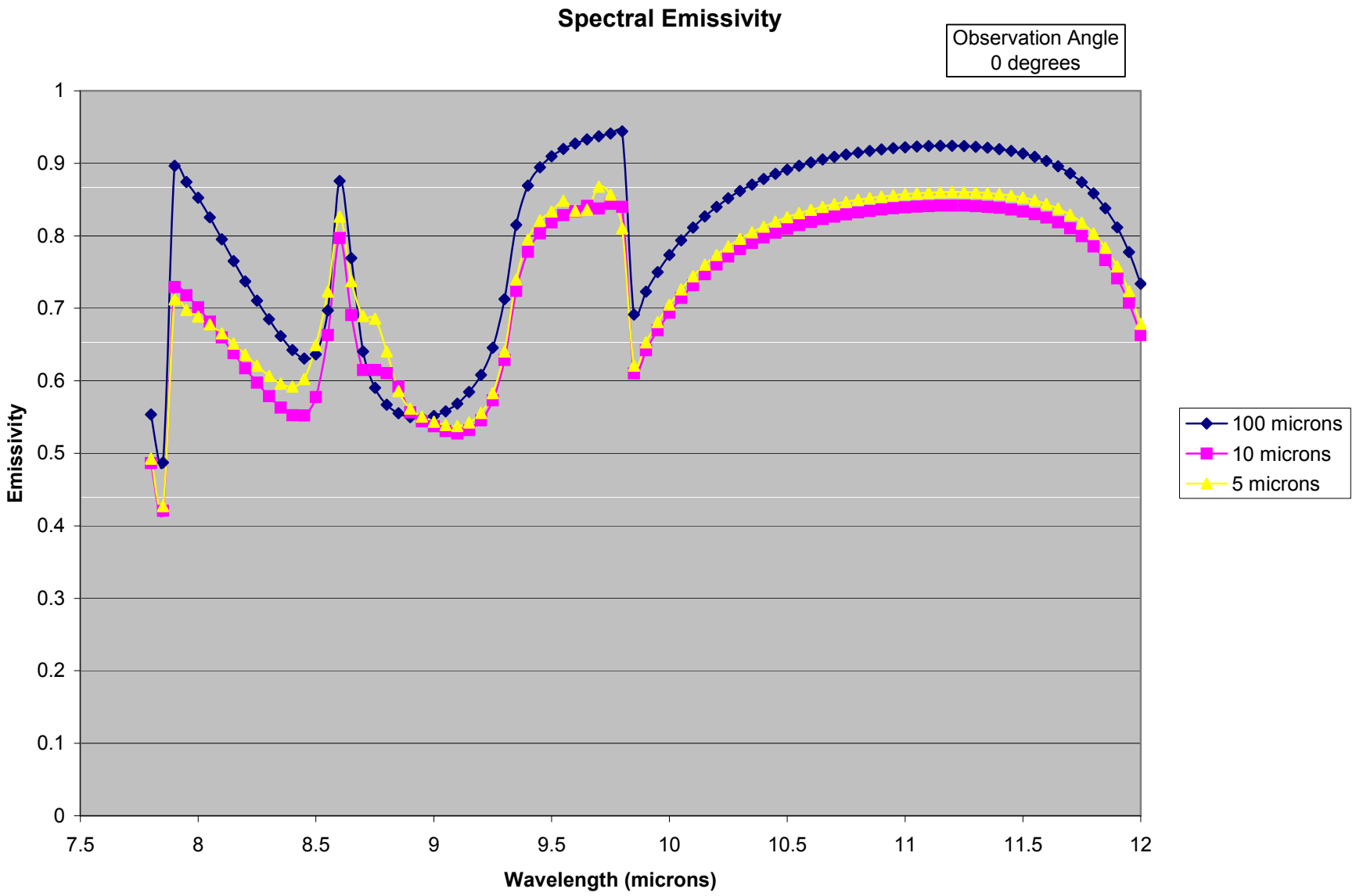
Combined Theory of Reflectance and Emittance Spectroscopy

- Hapke (1993)
- Modification of previous theory to include emission effects due to physical processes
- Same assumptions as before:
  - Assumes particles are large compared to the wavelength
  - Allows for irregular particles
  - Allows for close-packing
- Also does not account for polarization or large-scale surface roughness

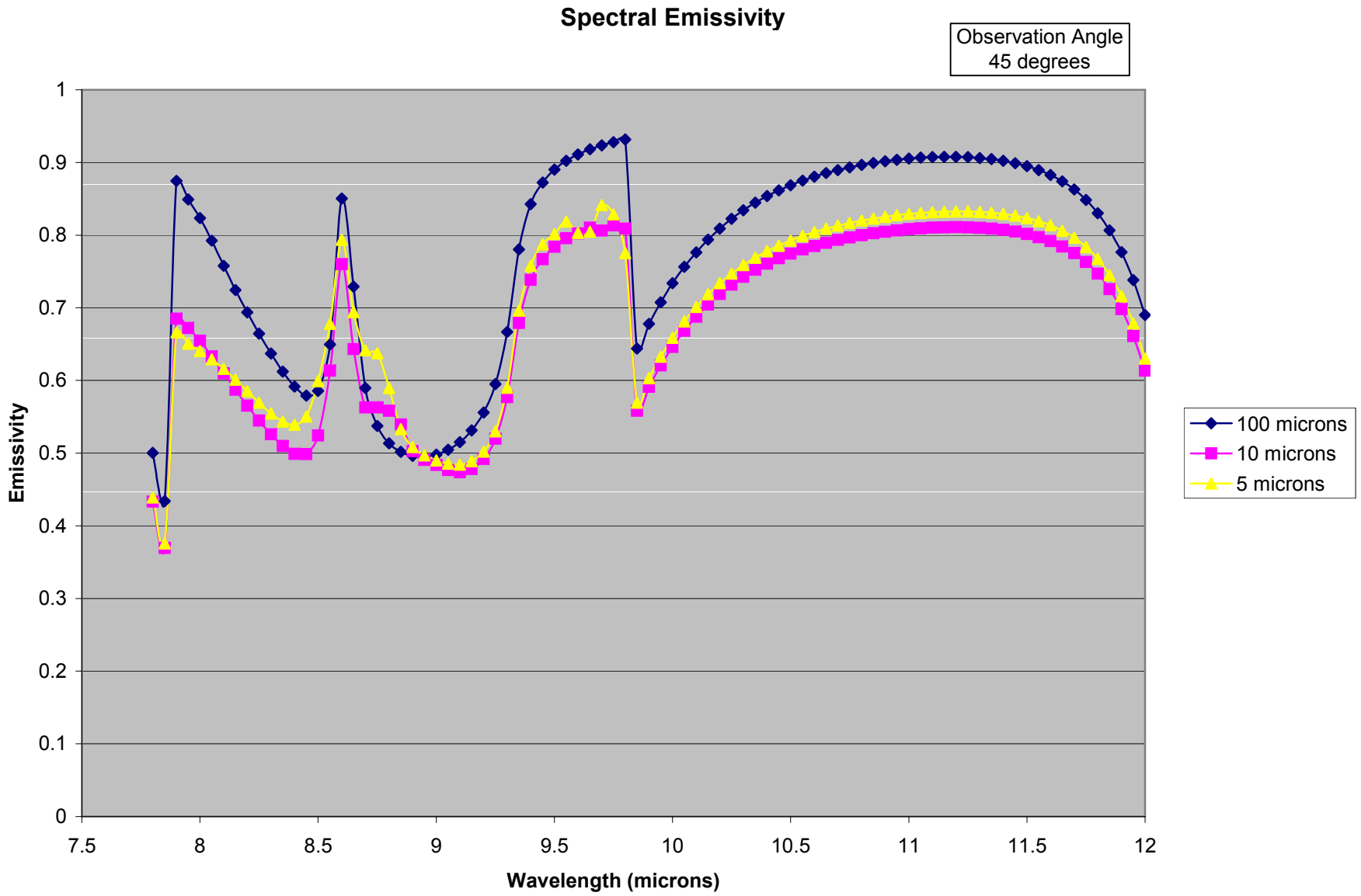
Spectral Emissivity: Quartz Observation Angle



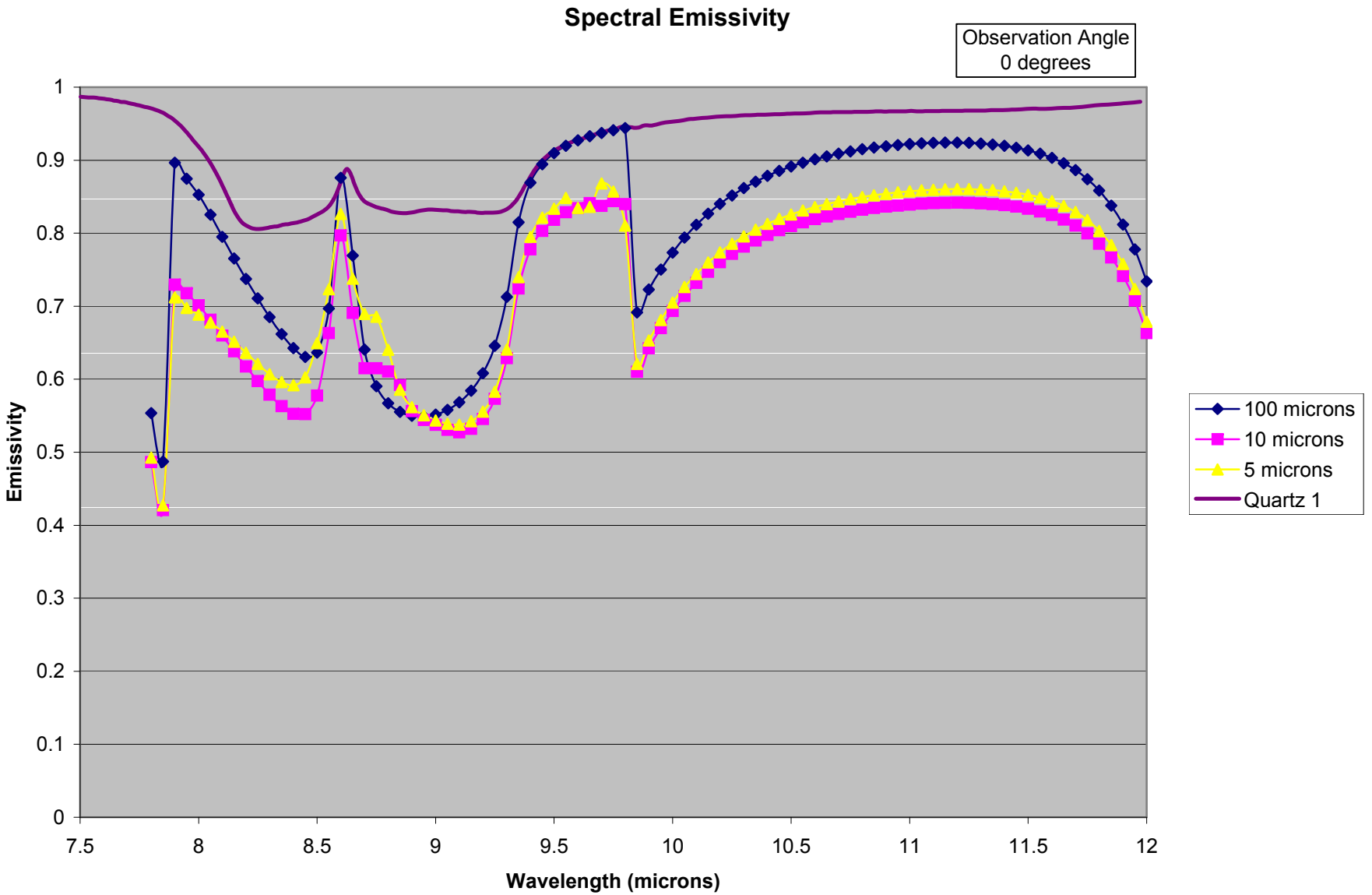
Spectral Emissivity: Quartz Particle Size Effects



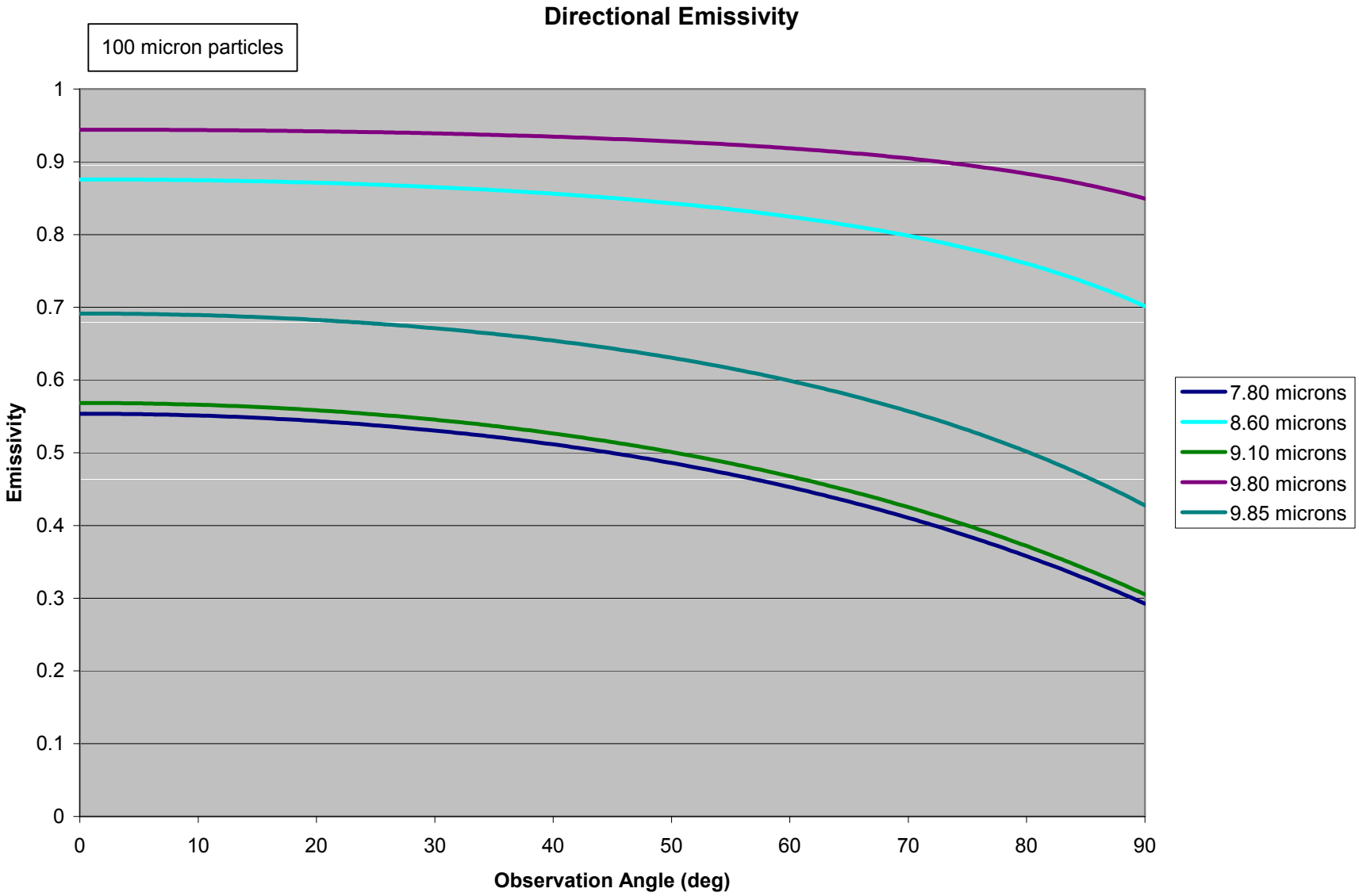
Spectral Emissivity: Quartz Particle Size Effects (2)



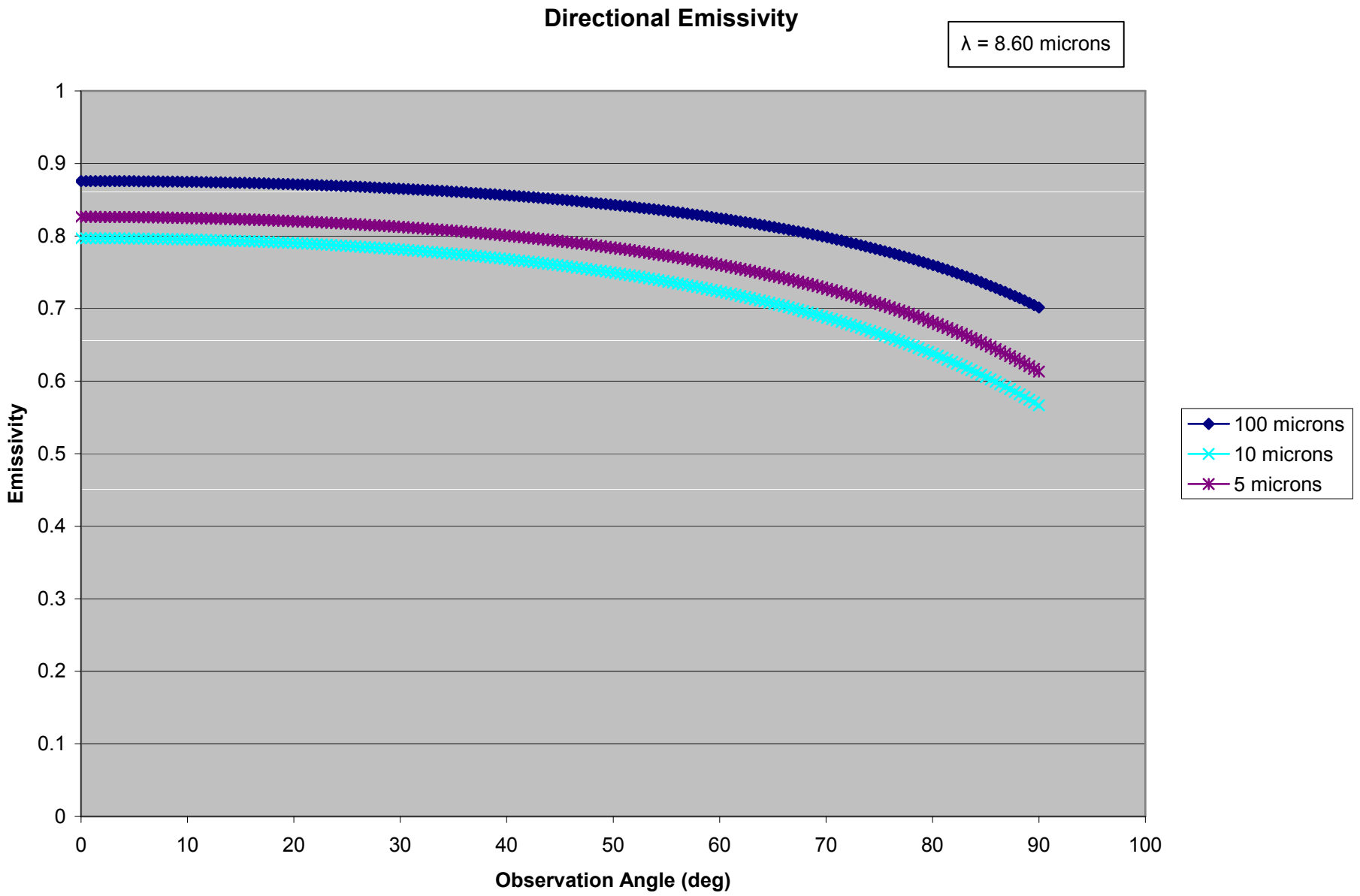
Spectral Emissivity Comparison



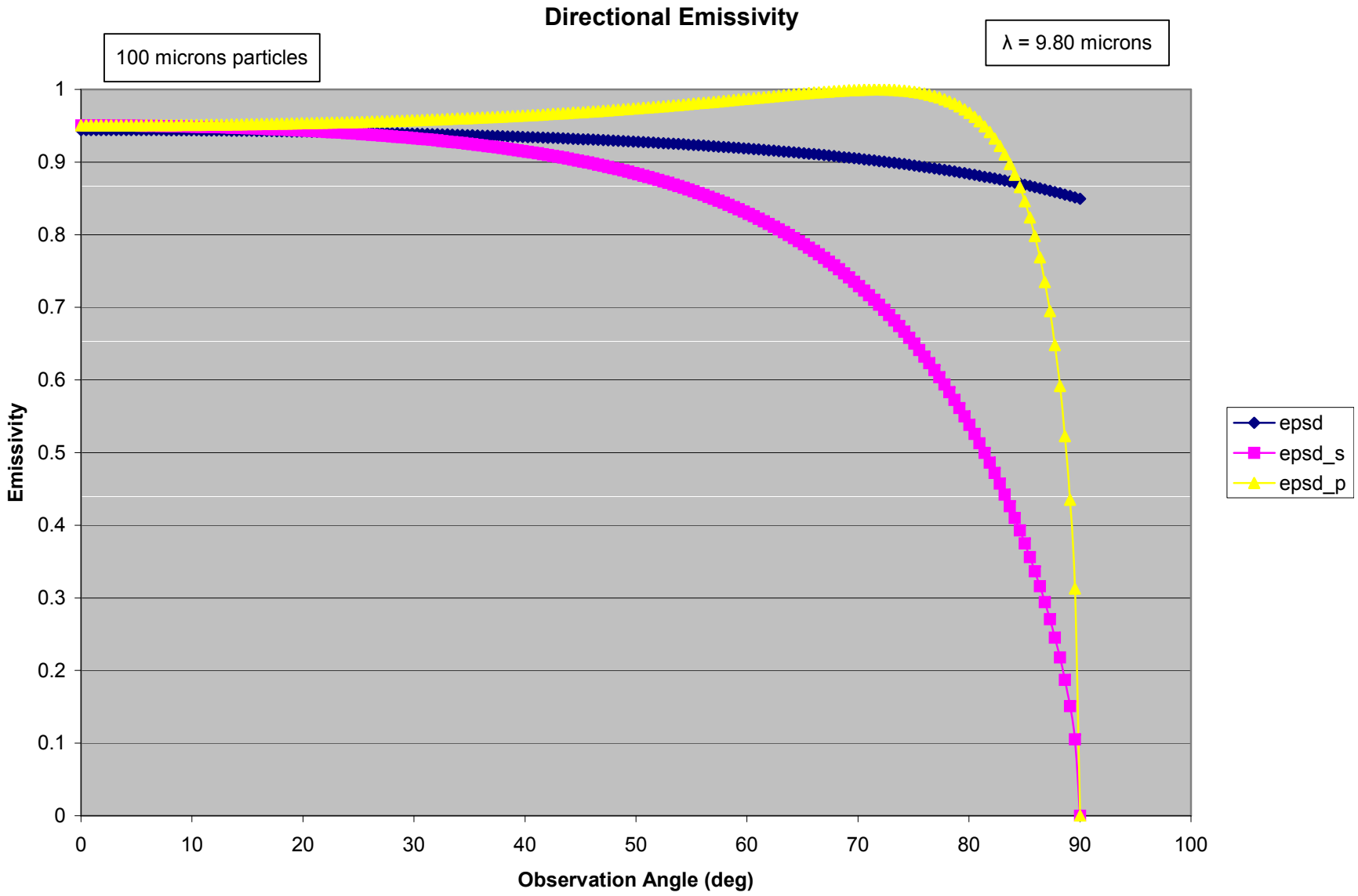
Directional Emissivity: Wavelength Comparison



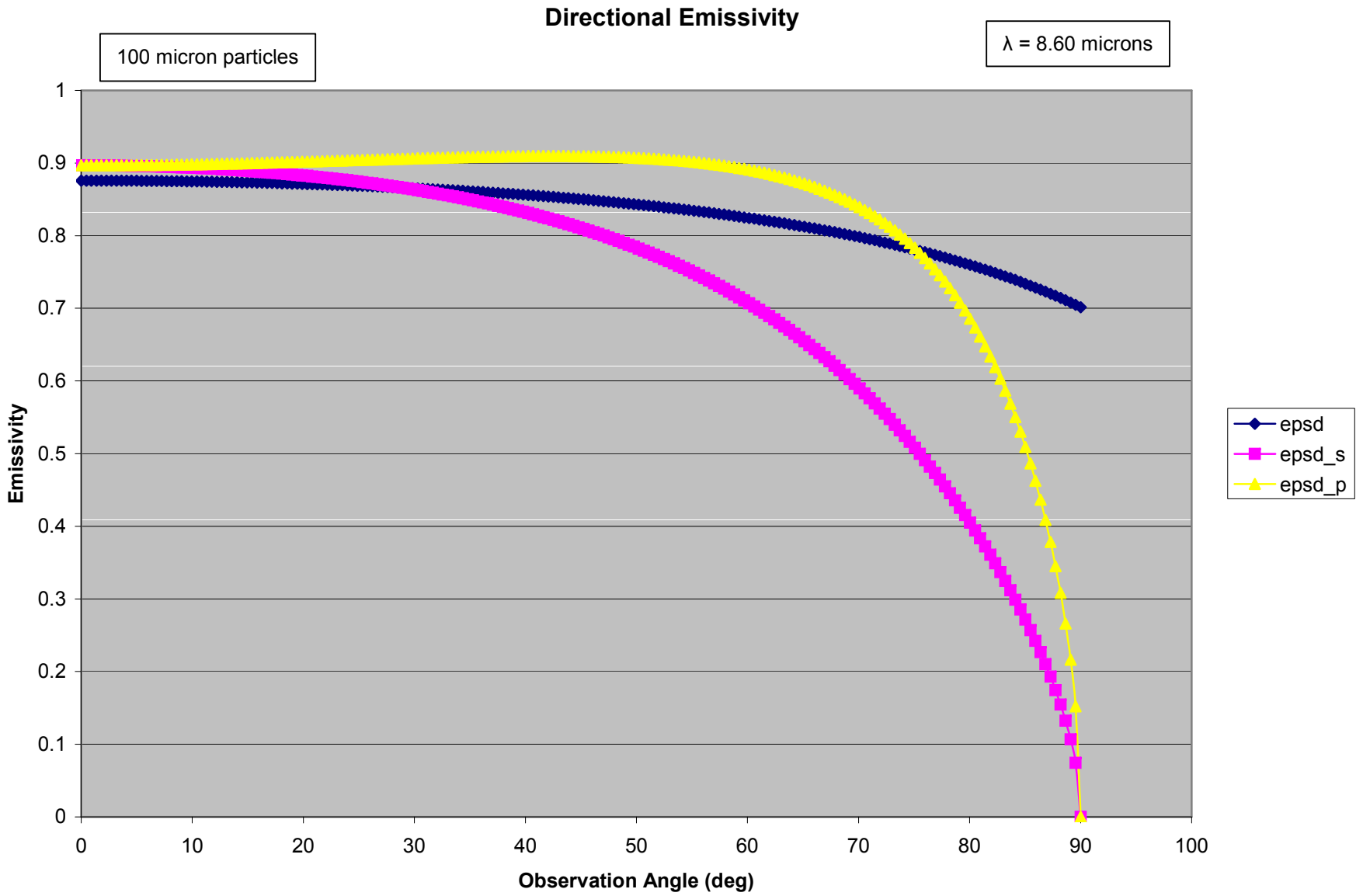
Directional Emissivity: Particle Size Effects



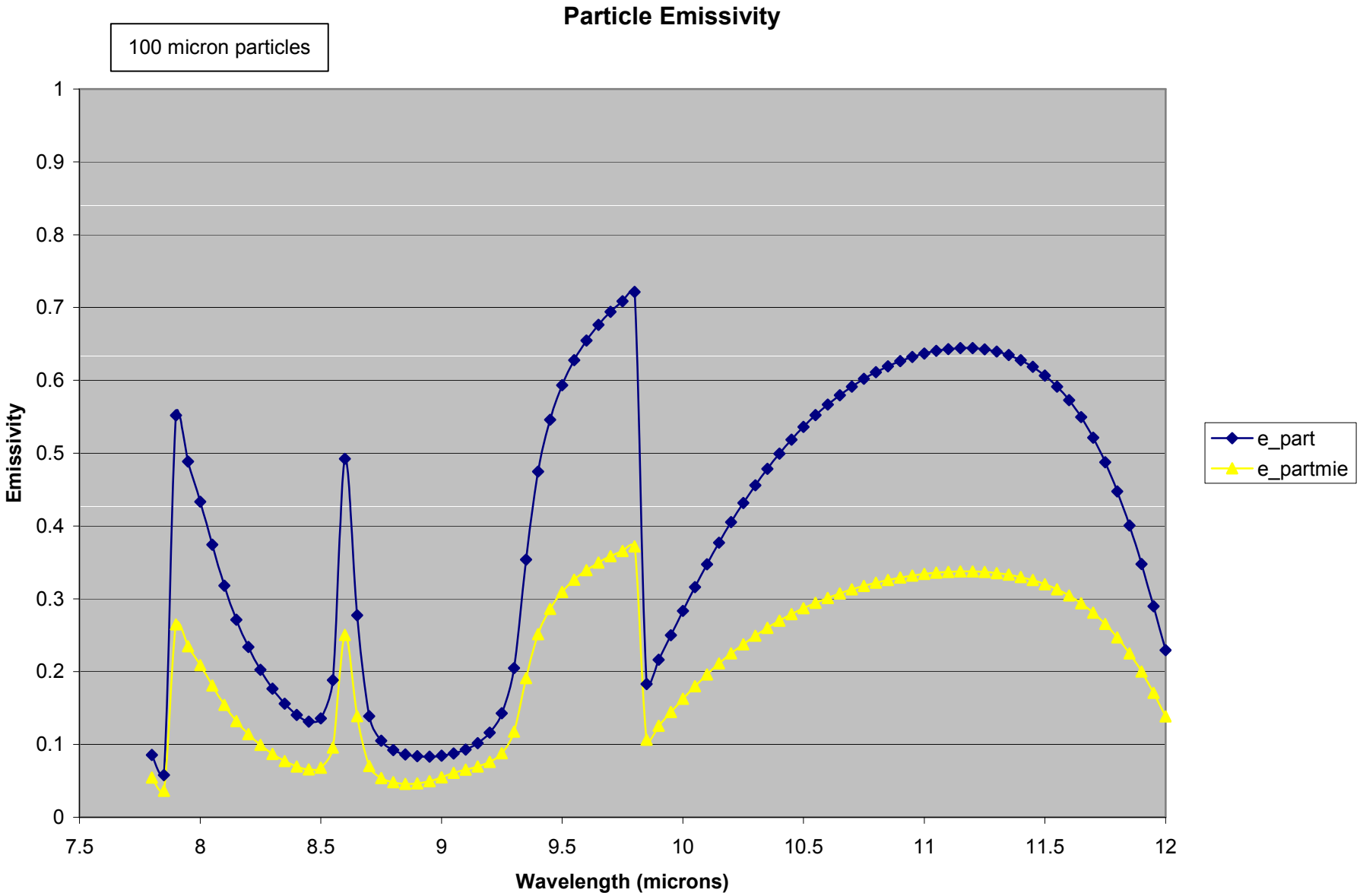
Directional Emissivity – Polarization



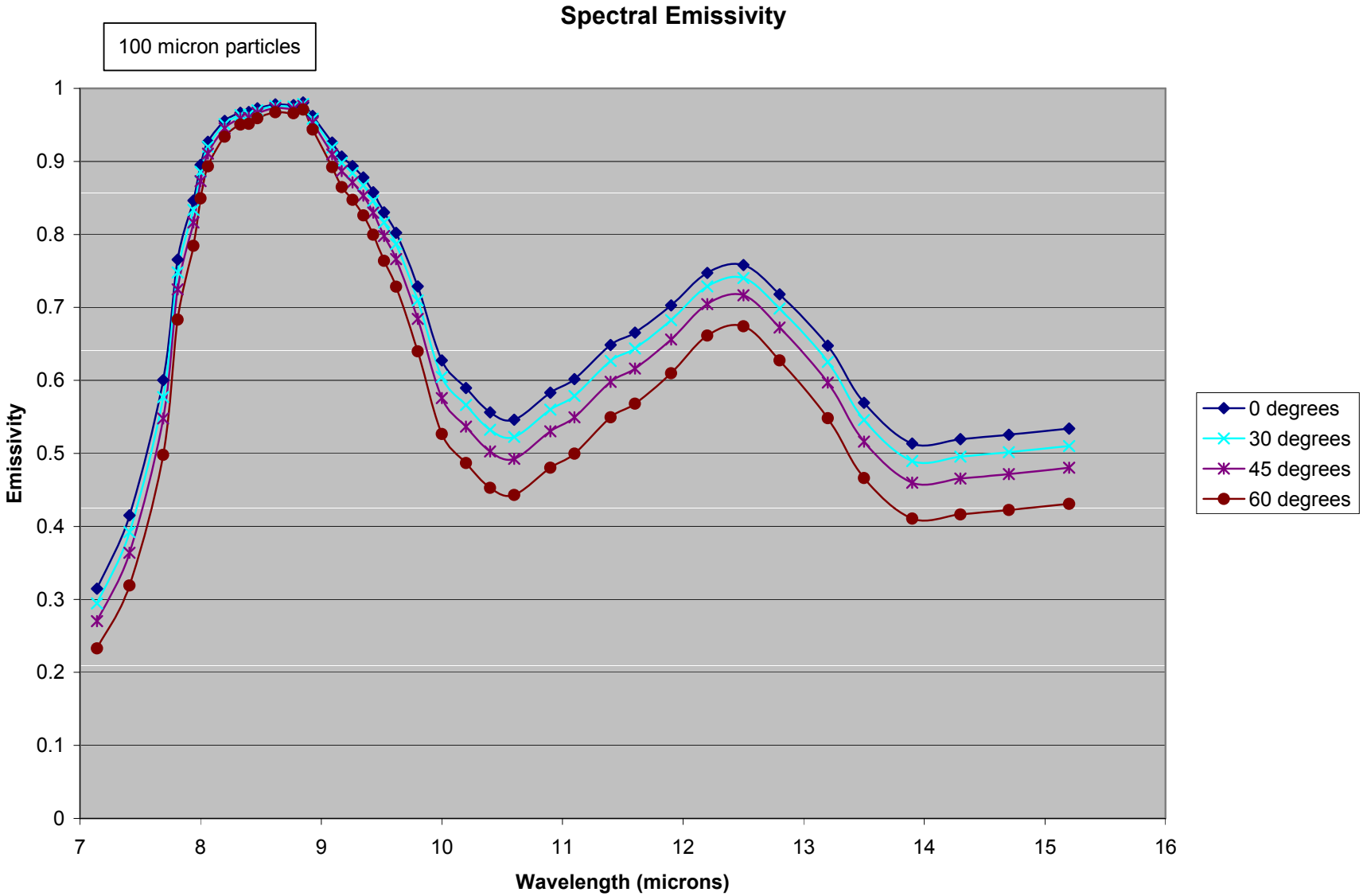
Directional Emissivity – Polarization (2)



Particle Emissivity: Quartz



Spectral Emissivity: Amorphous Quartz



Minefield Detection Using LWIR Hyperspectral Imagery

PI: Alan Thomas & Michael Cathcart

Adaptive Spatial Sampling Schemes for the Detection of Minefields in Hyperspectral Imagery

LWIR Mine Detection

- The principle challenge in LWIR hyperspectral land mine detection is discriminating land mines from highly varying a priori unknown backgrounds.
- Many soil and vegetation regions may exist within the background of a single minefield.
- Mine signatures may often times show only small spectral differences from any one background.

Algorithms vs. Analysts

- It is highly desirable that mine detection be automated.
- In a study, Reddy et al.[1] observed a much lower false alarm rate for mine detection in observers that had a mine field focus over observers with an individual mine focus.
- This motivates us to exploit mine field pattern information

[1] Reddy, M., Agarwal, S., Hall, R., Brown, J., Woodard, T., and Trang, A., "Warfighter-in-the-loop: mental models in airborne minefield detection," Proc. of SPIE Detection and Remediation Technologies for Mines and Minelike Targets X 5794(1), 1050–1059 (2005).

Big Idea

- Find spectral anomalies
- Look for a grid pattern in the anomalies
- Use the determined pattern to predict locations for additional anomalies
- Perform a subsequent anomaly search in those locations to determine if an anomaly truly is present
- Fuse the results of both anomaly searches

Table of Contents

- Analysis presented in two parts
- Part One: We will look at exploiting the mine field grid pattern through subsequent sampling
  - (Assumes the pattern has already been determined)
- Part Two: We will look at a method for finding the mine field grid orientation and spacing

Part 1

Exploiting the Mine Field Grid Through Adaptive Sampling

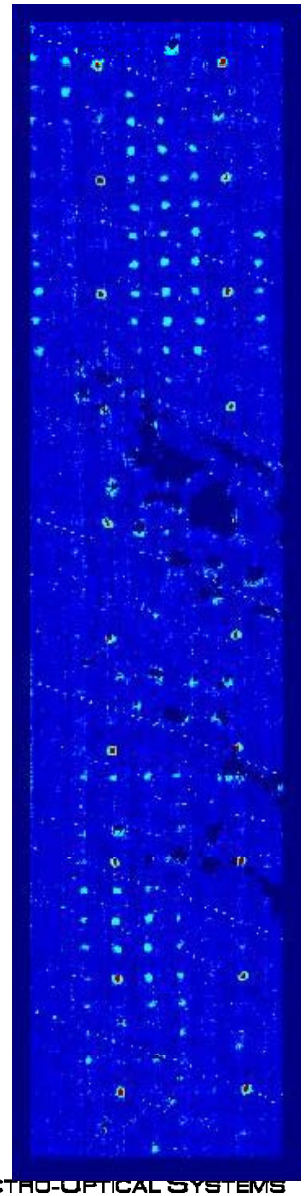
The Example Data

- 70 Band LWIR radiance data from a desert region in the western United States
- Multiple target types
- Multiple false target types
- Varying background
- Large vegetation regions
- Natural anomalies

Vegetation Mask

- We will use a vegetation mask to eliminate false alarms.
- Use K-means clustering to find 6 pixel clusters.
- Pick the cluster with the highest mean emissivity. These pixels are vegetation
- Form vegetation mask

RX Anomaly Detection



- Run local RX with buffer region and vegetation masking
- The RX detector value is the covariance normalized difference between the value of the test pixel and the mean of the background .

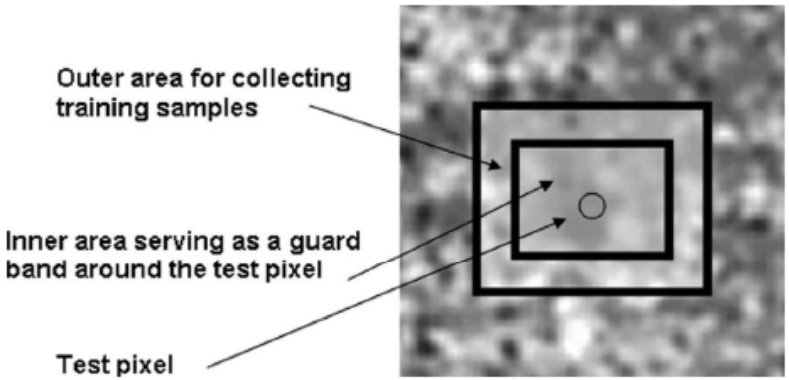


Illustration on right taken from :

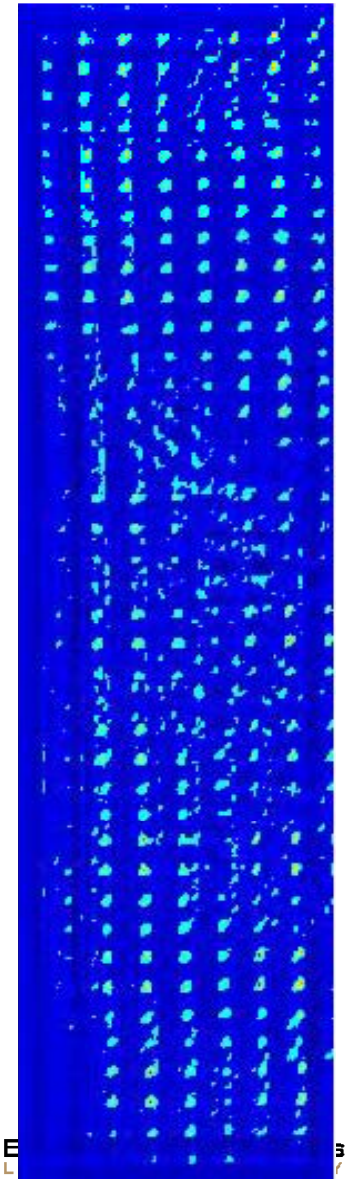
[4] A. Banerjee, P. Burlina, C. Diehl. A Support Vector Method for Anomaly Detection in Hyperspectral Imagery. *IEEE Trans. Geosci. Remote Sens.* 44(8), pp2282-2291, 2006.

Record Blobs

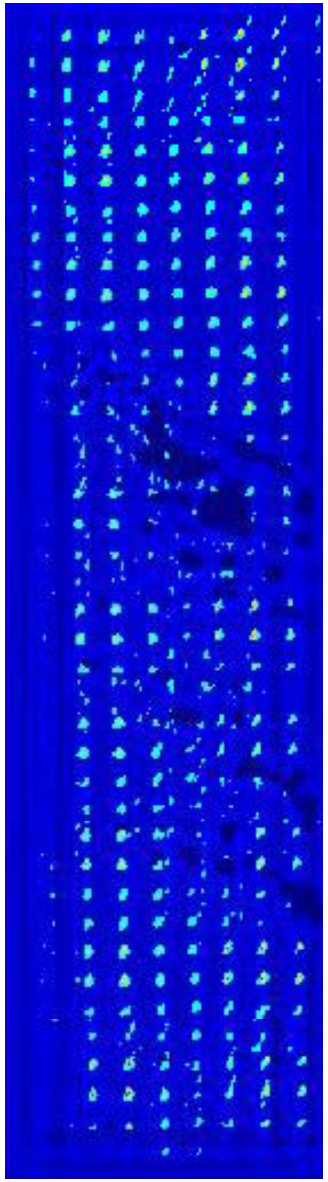
- Apply threshold to the image then search for blobs.
- Eliminate Blobs with less than 6 pixels.
- Number of blobs reduced from 1248 to 158.
- Record the blob information for later use.

Template Matching

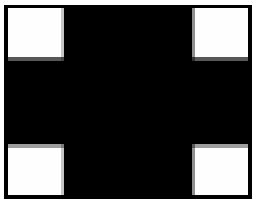
Before



After



- Run field pattern template matching to produce an image with search regions
- The image is derived by projecting the detector value multiplied by the template pattern back onto the image
- In this example, we again apply the vegetation mask, however this may be neglected if one wishes to predict mine locations beneath vegetation.



Pattern template

Blob Analysis

Before

After

- Identify connected blob objects using threshold of 6
- Eliminate blobs that contain less than 6 pixels.
- Number of blobs is reduced from 612 to 401.

Blob Analysis

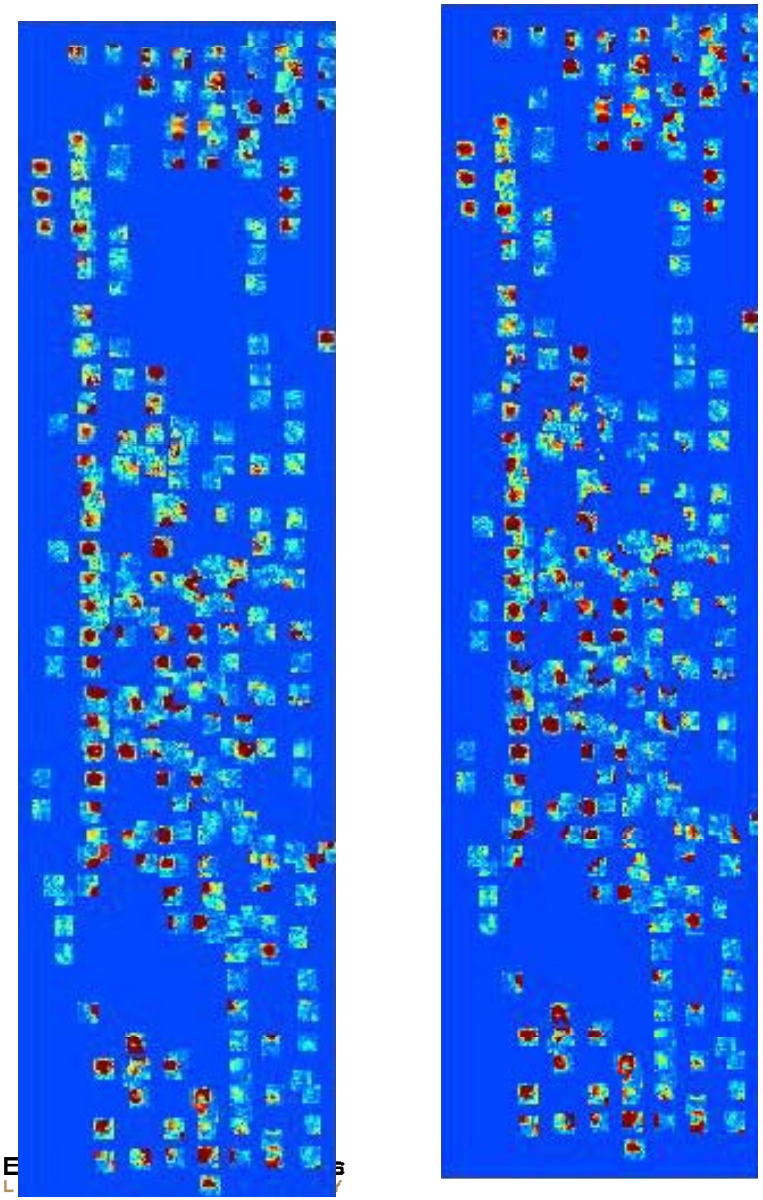
Before

For Removal

After

- Eliminate blobs that were encountered immediately after RX anomaly detection.
- Number of blobs is reduced from 401 to 276.

Focused Anomaly Detection

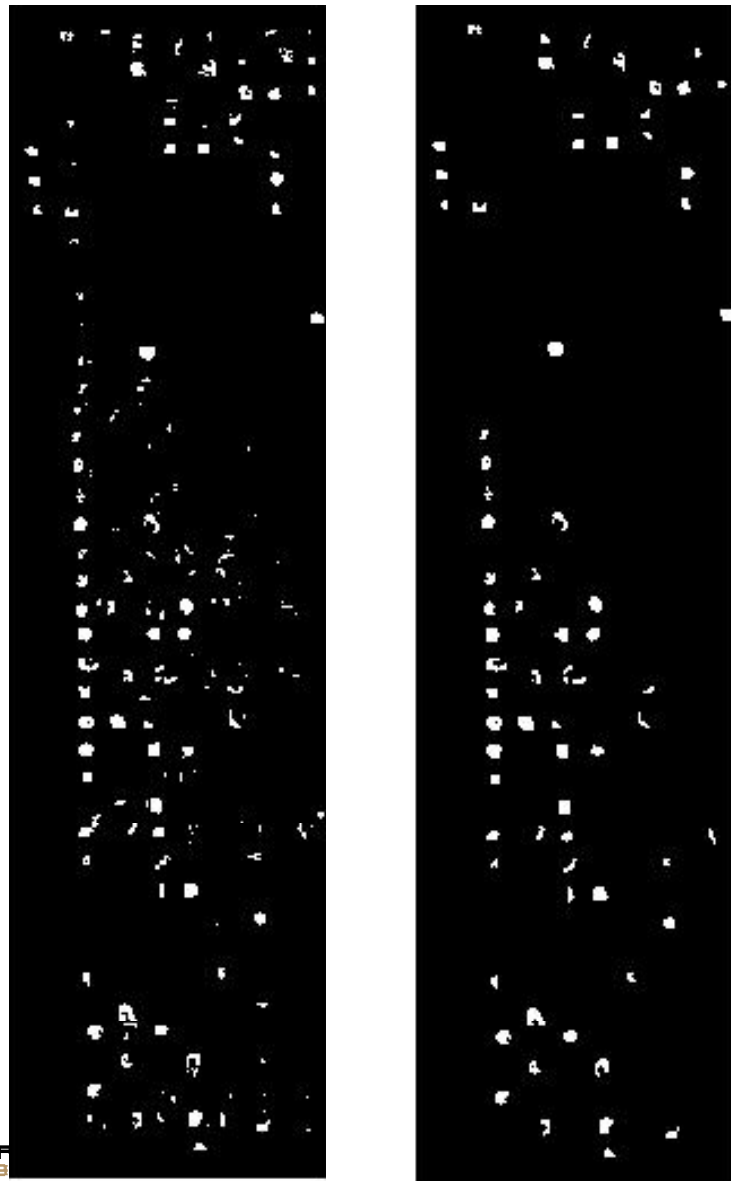


- Apply Dual Window-based Eigen Separation Transform (DWEST) anomaly detector to the focus regions

$$DWEST_{i,j} = \left| \sum_{n=1}^{3} v_{i,j}^{n} \cdot \left(\mu_{i,j}^{\text{out}} - \mu_{i,j}^{\text{in}} \right) \right|$$

- Reapply Vegetation masking
- Find the blobs

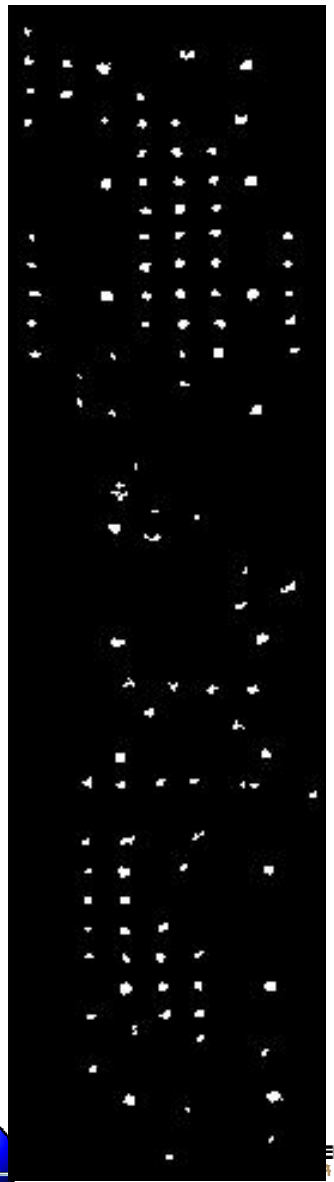
Focused Anomaly Detection



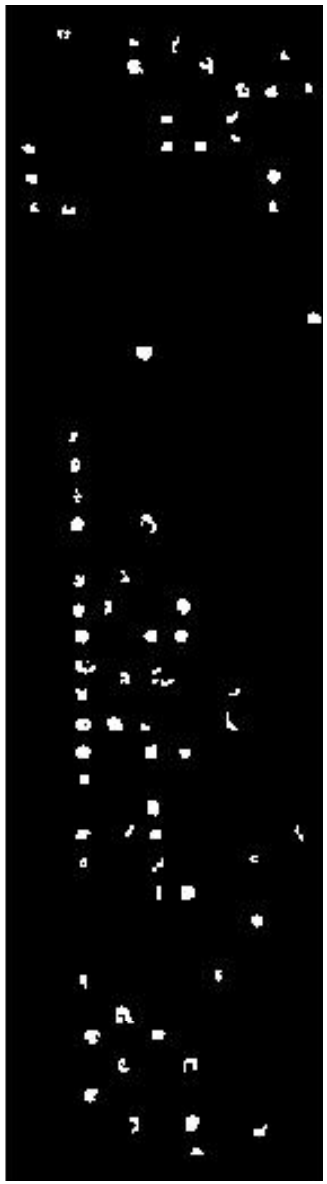
- Find the blobs
- Eliminate Blobs with less than 20 pixels
- Number of blobs reduced from 262 to 72

Fusion of Anomalies

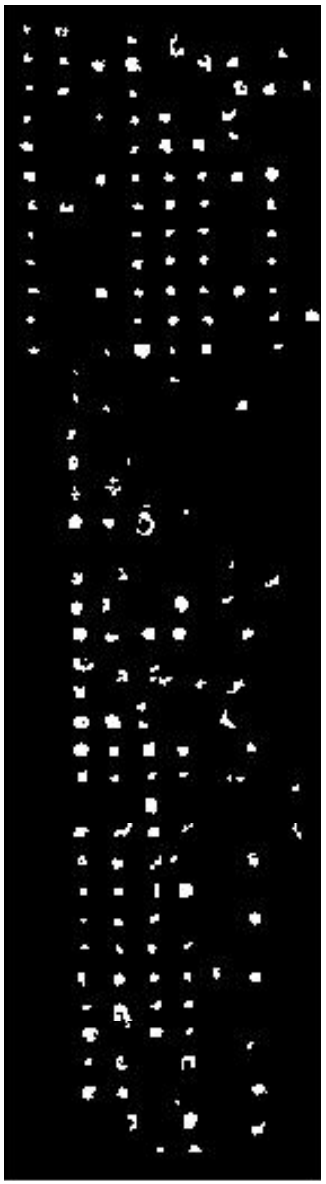
Original



Focused



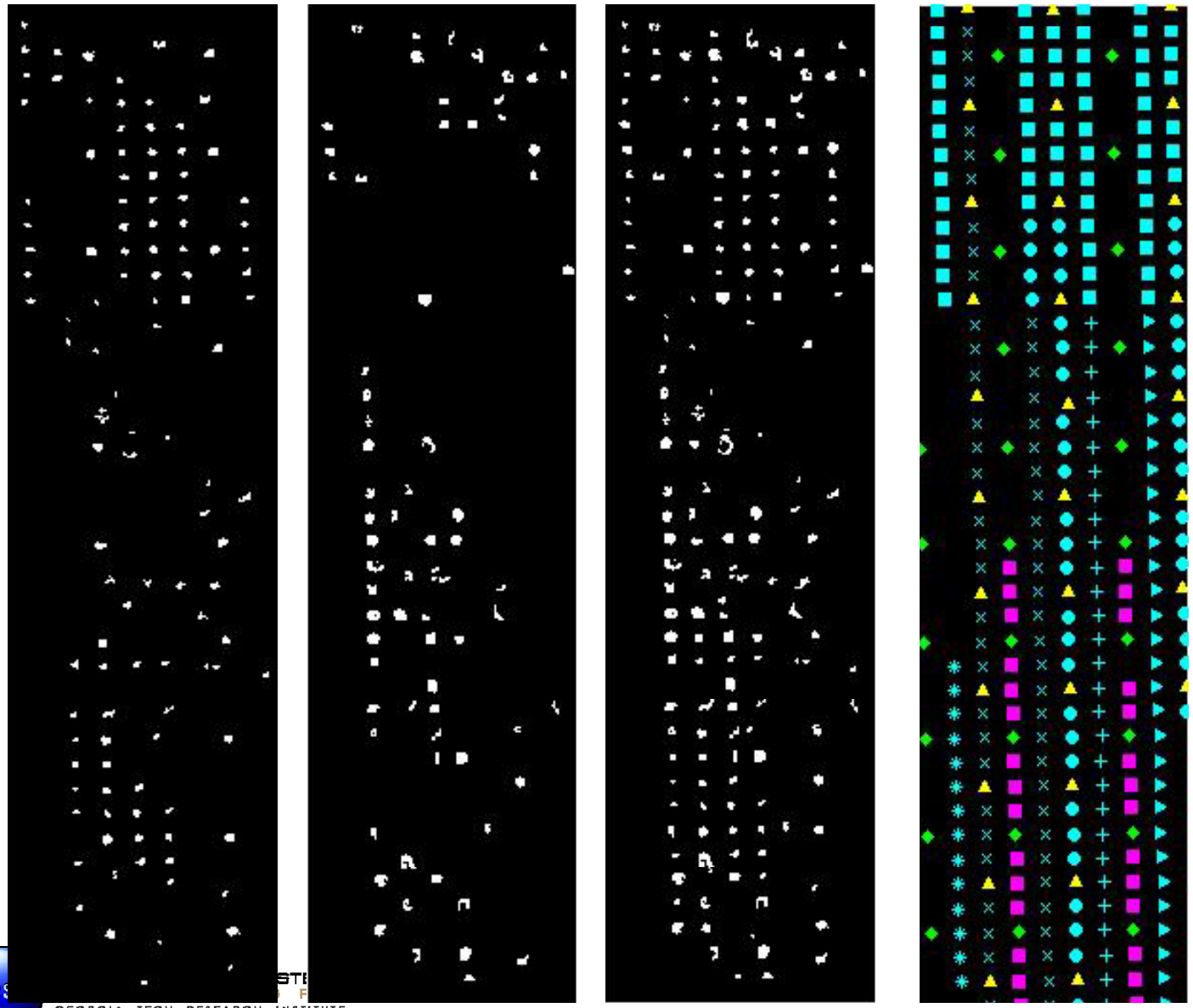
Fused



- Eliminate Blobs with less than 10 pixels in original RX
- Number of blobs reduced to 118
- Add the results of the original and focused anomaly searches

Fusion of Anomalies

Original Focused Fused Ground Truth



Future Research

- Automated parameter selection e.g. determination of thresholds.
- Evaluate usefulness for different situations. Quantify the improvement.
- Generalize to a fusion procedure for the output of many anomaly detection algorithms.

Part 2

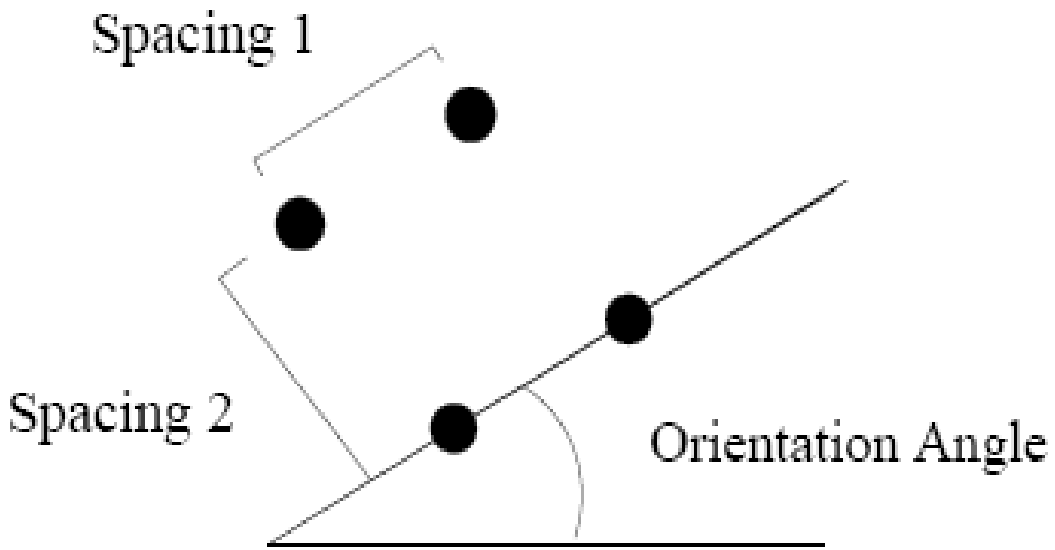
Determination of the Grid Orientation and Spacing Parameters

Automated Determination of Scale and Orientation of Mine Field Grid

Big Idea

- Find spectral anomalies
- Look for a grid pattern in the anomalies
- Use the determined pattern to predict locations for additional anomalies
- Perform a subsequent anomaly search in those locations to determine if an anomaly truly is present
- Fuse the results of both anomaly searches

Pattern Parameters

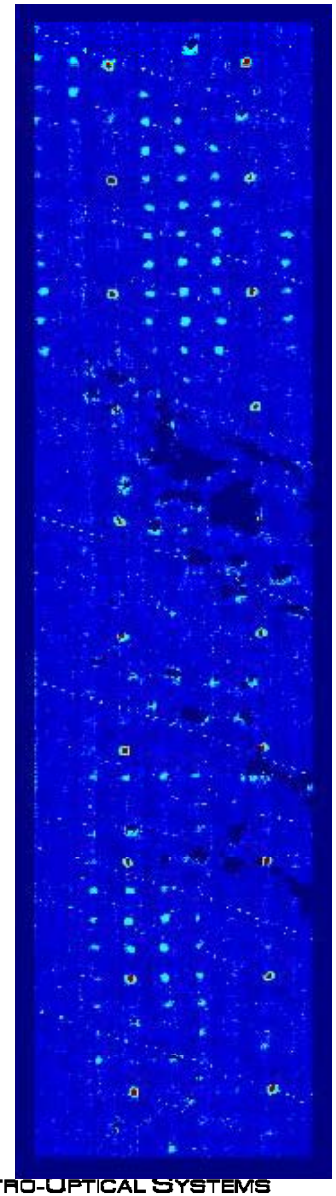


- Three parameters to be determined: an orientation angle and two spacing parameters.

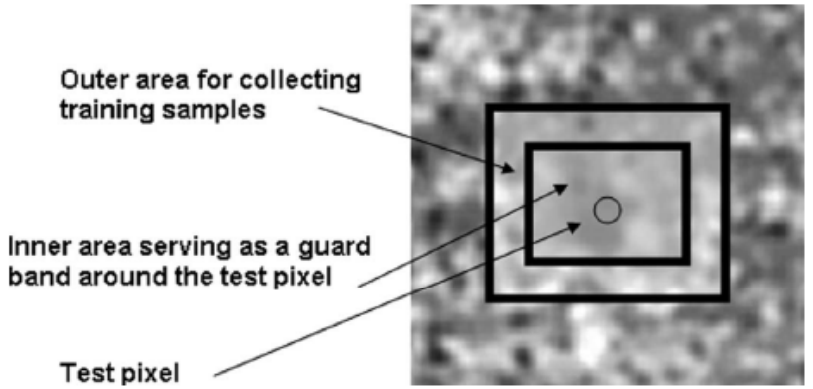
Vegetation Mask

- We will use a vegetation mask to eliminate false alarms.
- Use K-means clustering to find 6 pixel clusters.
- Pick the cluster with the highest mean emissivity. These pixels are vegetation
- Form vegetation mask

RX Anomaly Detection



- Run local RX with buffer region and vegetation masking
- The RX detector value is the covariance normalized difference between the value of the test pixel and the mean of the background .



•Illustration on right taken from :

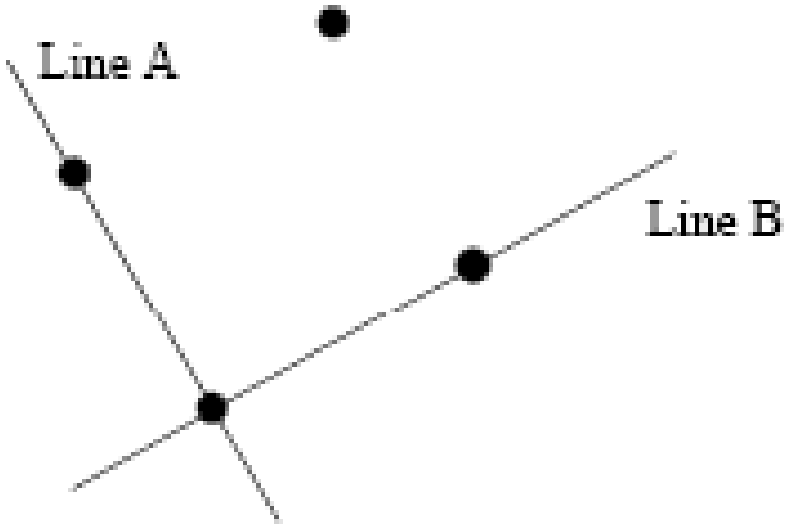
•A. Banerjee, P. Burlina, C. Diehl. A Support Vector Method for Anomaly Detection in Hyperspectral Imagery. *IEEE Trans. Geosci. Remote Sens.* 44(8), pp2282-2291, 2006.

Eliminate Blobs

- Apply threshold to the image then search for blobs.
- Eliminate Blobs with less than 6 pixels.
- Number of blobs reduced from 1248 to 158.
- Record the blob information for later use.

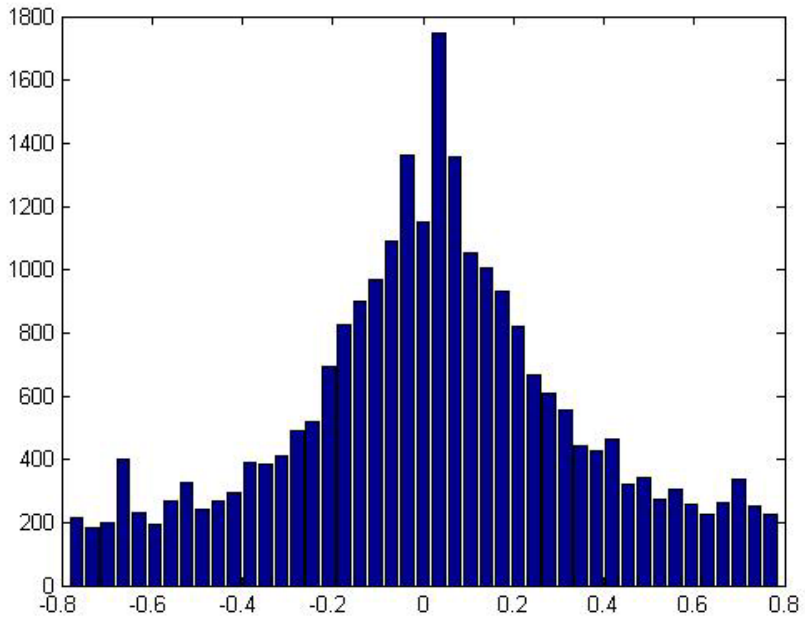
Grid Orientation

- Calculate the distance and slope angle between every pair of distinct blobs.
- We then associate with each angle an equivalence class in the modulus pi/2 arithmetic.



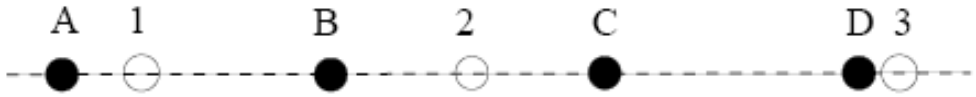
- In this way, perpendicular lines are associated with the same orientation angle

Grid Orientation



- Sort the slopes into 45 equally space bins (i.e. accuracy to within 2 degrees).
- Find bin with max entries
- In this case we obtained, .0351 radians or 2 degrees.
- We call this Angle 1 and the perpendicular direction Angle 2.

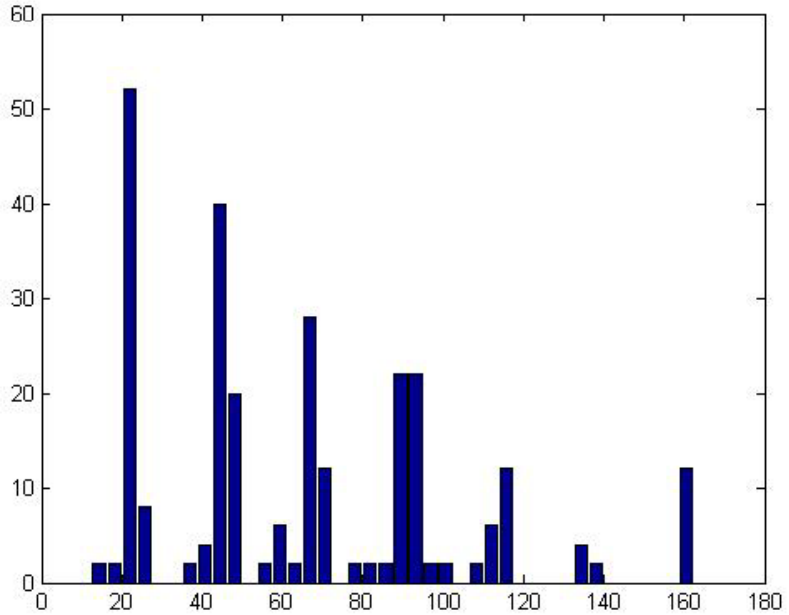
Grid Spacing



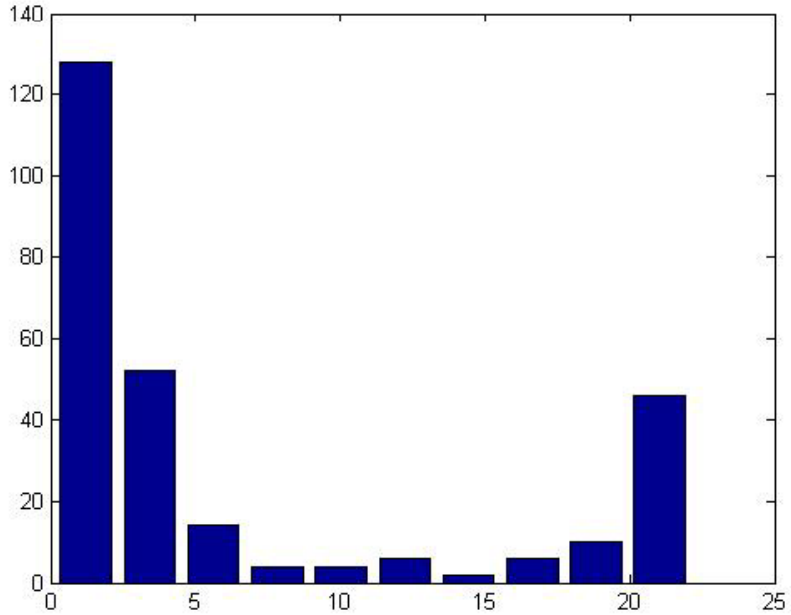
First Anomaly	Second Anomaly	Distance	Distance (mod 13)
A	B	13	0
A	C	26	0
A	D	39	0
A	1	3	3
A	2	20	7
A	3	41	2
B	C	13	0
B	D	26	0
B	1	9	9
B	2	7	7
B	3	28	2
C	D	13	0
C	1	22	9
C	2	6	6
C	3	15	2
D	1	35	9
D	2	19	6
D	3	2	2
1	2	16	3
1	3	27	1
2	3	21	8

- **Use modular arithmetic to find the most likely distance in the slope direction**
- **The modulus that yields the most zero must be the true spacing distance (within some bounds).**

Grid Spacing



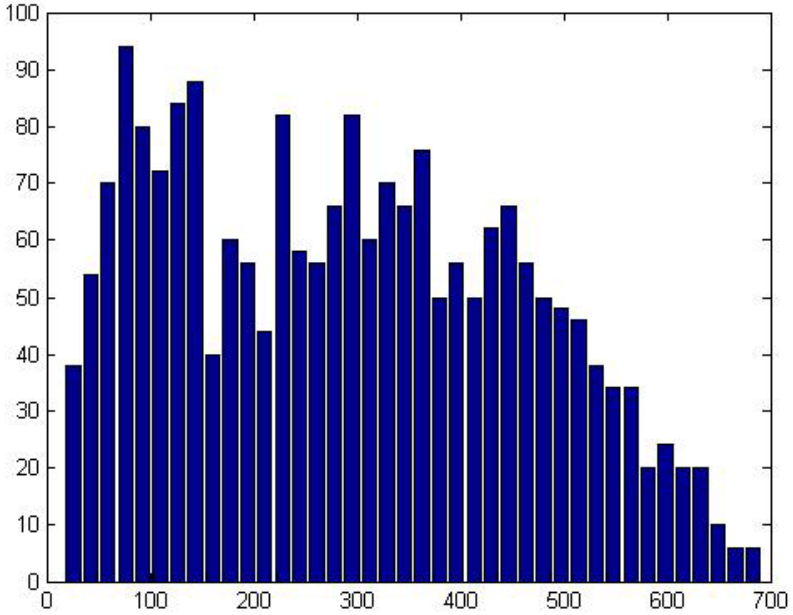
Histogram of Angle 1 Distances in 40 bins



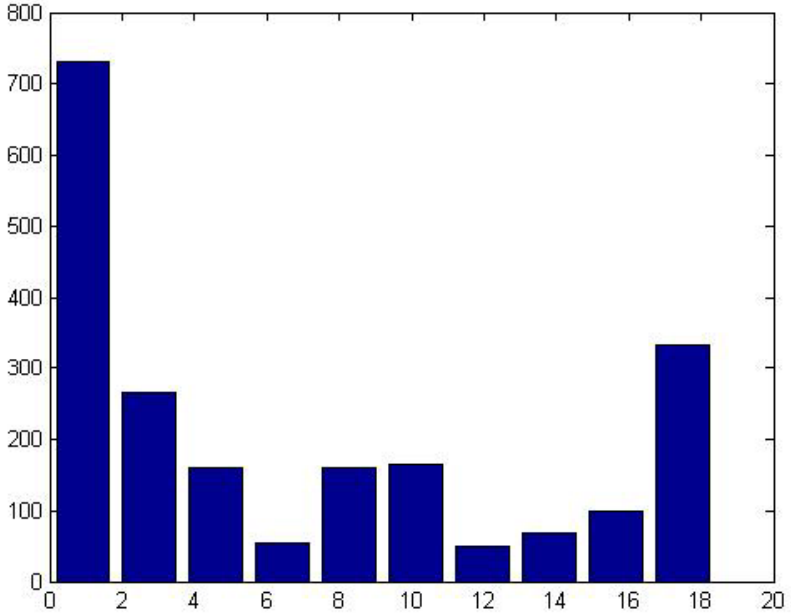
Histogram of Angle 1 Distances in modulo 22.35 in 10 bins

The modulus that maximizes the number of entries in the first bin is determined to be the Angle 1 spacing.

Grid Spacing



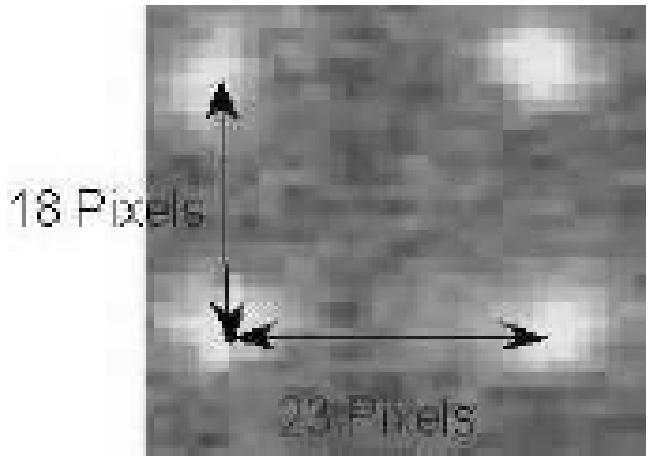
Histogram of Angle 1 Distances in 40 bins



Histogram of Angle 1 Distances in modulo 18.40 in 10 bins

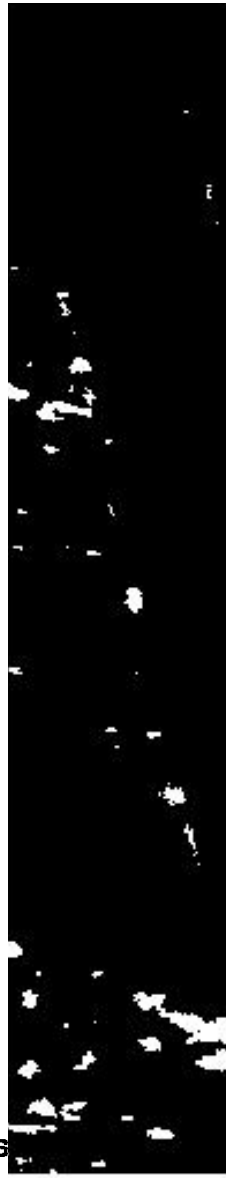
Example 1 - Results

- Angle 1 = .0351 radians or about 2 degrees
- Scale 1 = 22.35 pixels
- Scale 2 = 18.40 pixels



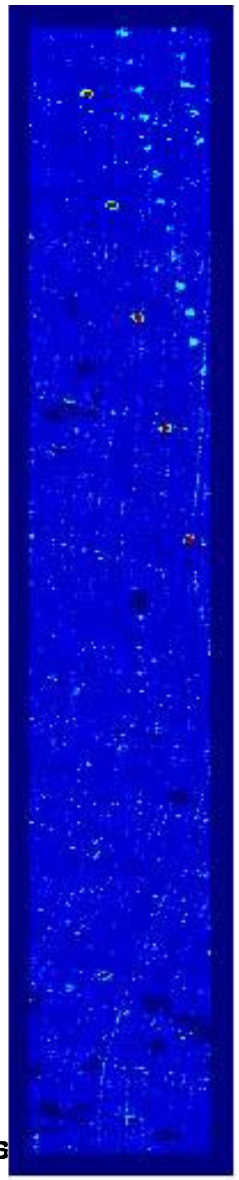
Visual verification of parameters in broadband image

Example 2 – Vegetation Mask



- Use K-means clustering to find
- 8 pixel clusters. Note that the k-means algorithm did not converge for 6 clusters and 7 clusters
- Pick the cluster with the highest mean emissivity. These pixels are vegetation
- Form vegetation mask

Example 2 – RX Detection

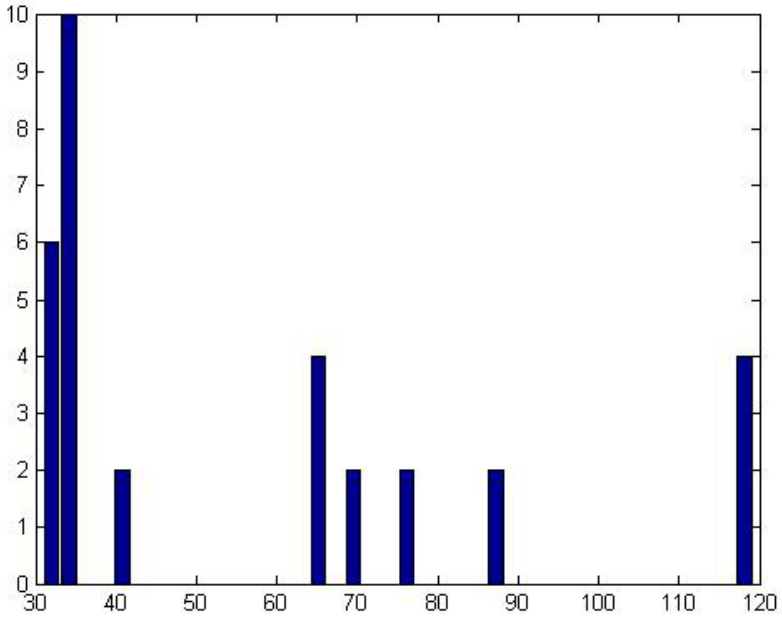


- Run local RX with buffer region and vegetation masking

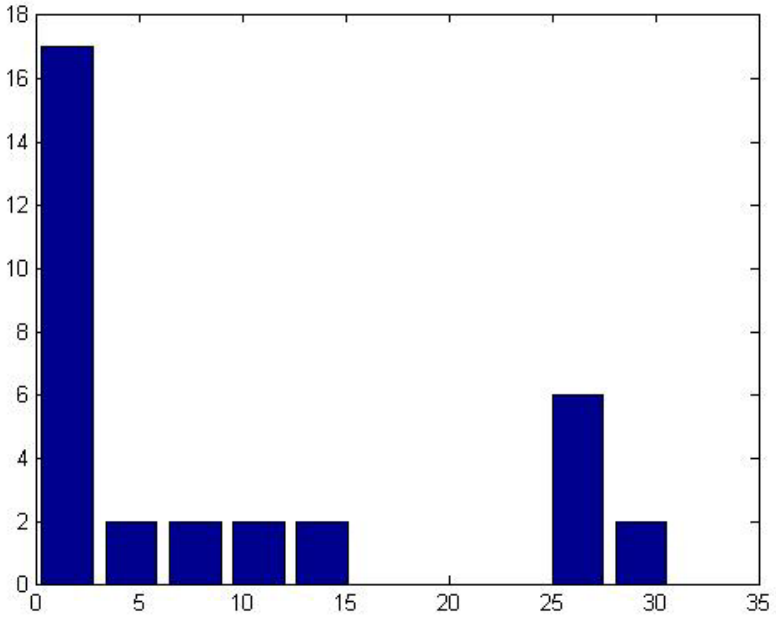
Example 2 – Eliminate Blobs

- Apply threshold to the image then search for blobs.
- Eliminate Blobs with less than 6 pixels.

Grid Spacing

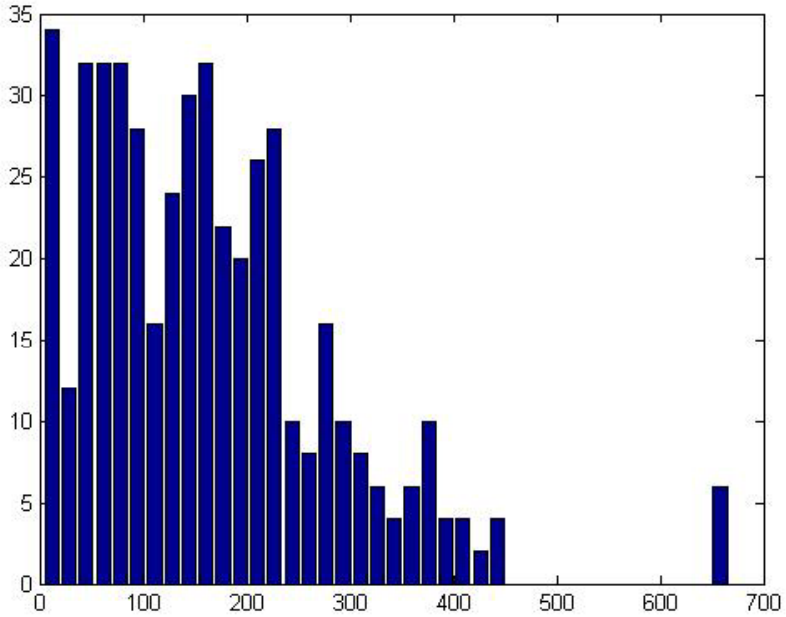


Histogram of Angle 1 Distances in 40 bins

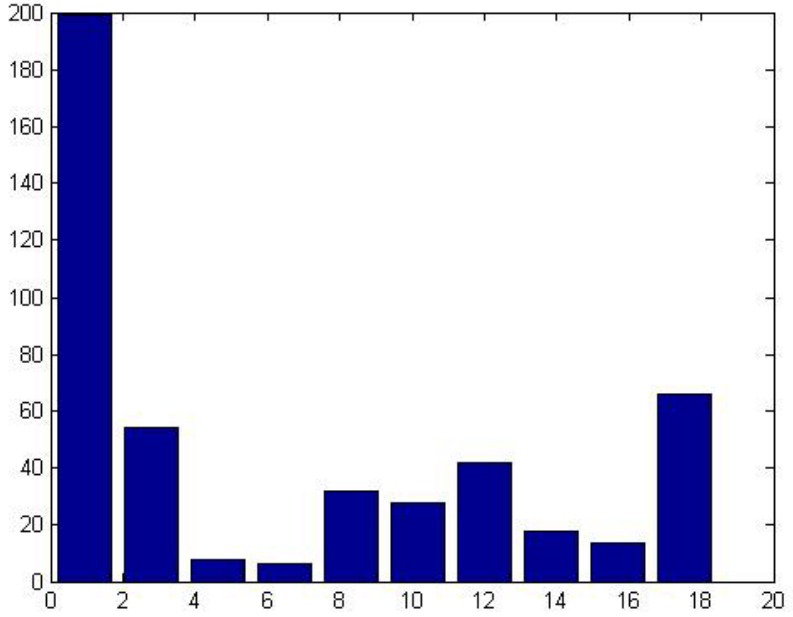


Histogram of Angle 1 Distances in modulo 30.90 in 10 bins

Grid Spacing



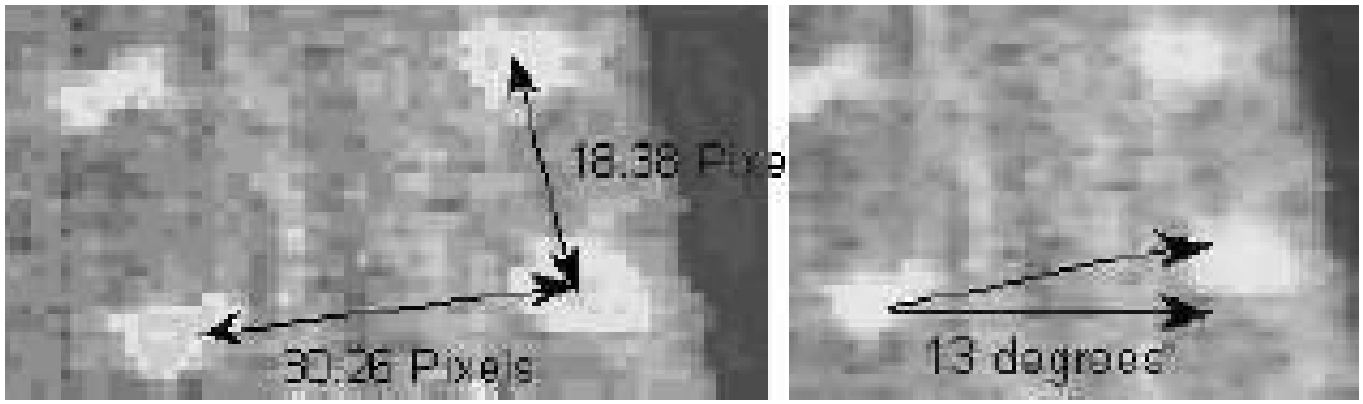
Histogram of Angle 1 Distances in 40 bins



Histogram of Angle 1 Distances in modulo 18.40 in 10 bins

Example 2 - Results

- Angle 1 = .2262 radians or about 13 degrees
- Scale 1 = 30.9 pixels
- Scale 2 = 18.5 pixels



Future Research

- Identification of nonlinear patterns such as skewed grids, radially symmetric patterns.
- Identification of patterns of mines laid in relation to naturally occurring contours such as roads.
- The ability to determine a pattern may provide a means to perform automated threshold selection for the outputs of anomaly detectors.

Adaptive Sampling & Spatial Pattern
– Overview Slide –

Patterned Based Accentuation of Mines and Minefields in Hyperspectral LWIR Imagery

Objective:

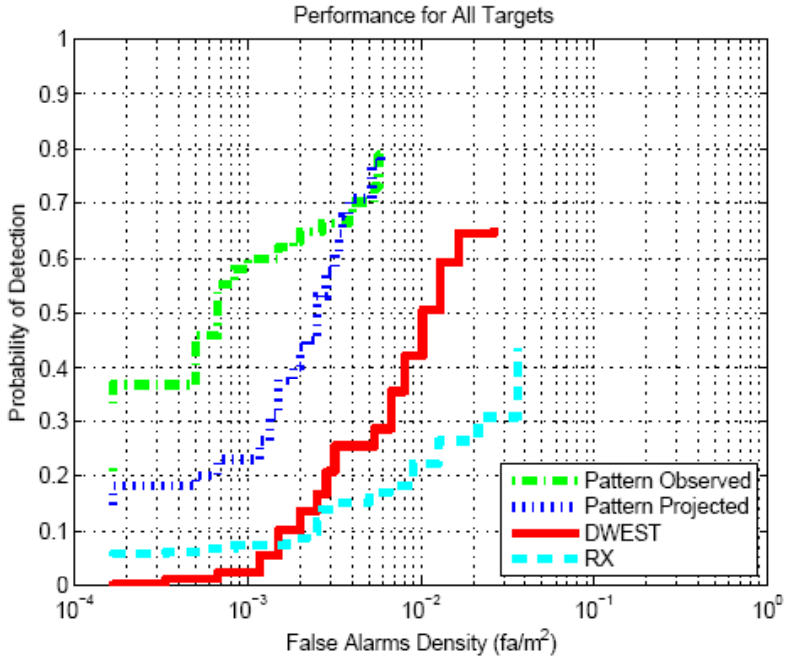
The goal was to exploit the fact that mines are typically related to each other by a spatial pattern for the purposes of improving the detection of individual mines in minefields and the detection of minefields as a whole. One should note that the LWIR spectral signatures of the buried mines are not a priori known because they vary locally with the soil composition.

Approach:

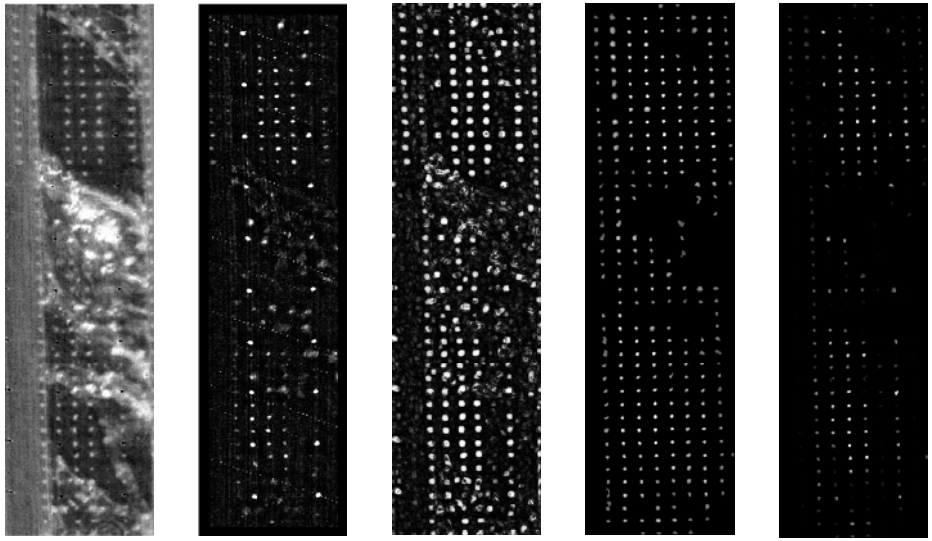
The approach is to first use the DWEST local anomaly detector to specify points of interest in the hyperspectral imagery. Second we look for a dominant grid pattern in those points of interest. The "*Pattern Projection*" is the likely location of mines based upon the local strength of the previously observed pattern. This approach allows for some prediction of obstructed mine locations. The *"Pattern Observed"* output roughly gives those locations where the pattern predicts a mine and a local anomaly has independently been observed.

Results:

The figures to the left show the results from one of the data sets used. The DWEST anomaly detector and the well known RX anomaly detector are shown here for comparison purposes. The two pattern based methods exhibit higher probability of detection at all false alarm rates than do the anomaly detectors. In all of the data sets we considered, the *"Pattern Observed"* method showed particularly good results at low false alarm rates.



The above figure shows ROC curves plotted on a log scale for two anomaly detectors (RX, DWEST) and the two developed pattern based methods.



Broadband | RX detector | DWEST detector | Pattern Projected | Pattern Observed

LWIR Hyperspectral Signature Processing

PI: Bryce Remesch, Michael Cathcart, & Alan Thomas

Summary

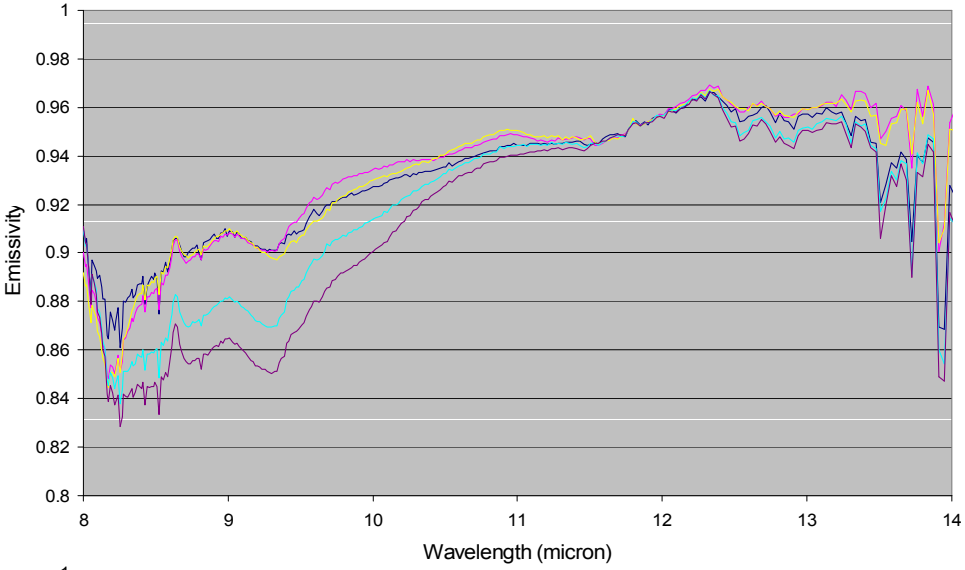
- Data collection summary
  - Yuma collection
    - Aerial data with AHI sensor
    - Ground truth
- Critique of common hyperspectral algorithms
  - Spectral matching
  - Anomaly detection
  - Model-based methods
- Spectral Analysis
  - Statistical Spatial Analysis
  - End member techniques
  - Dimension reduction

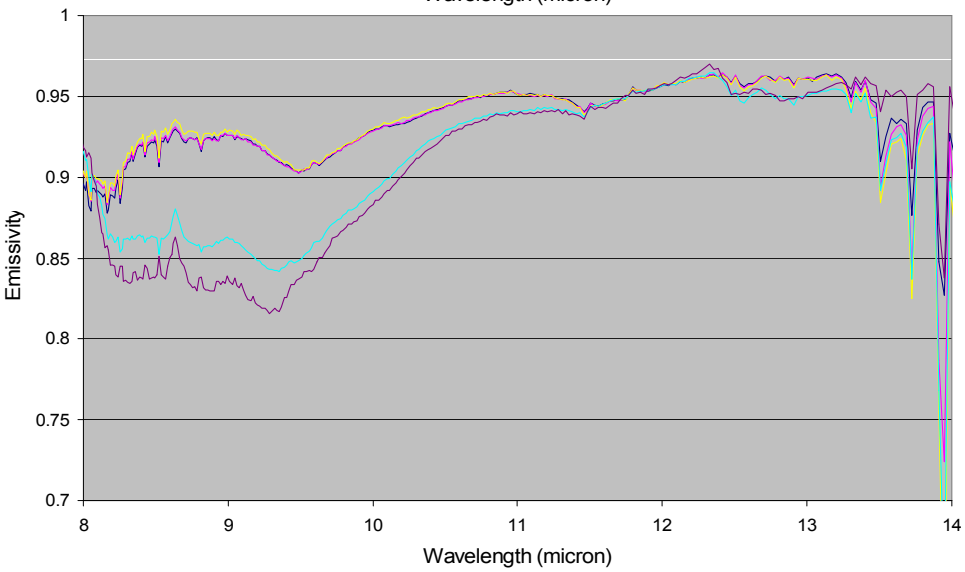
Soils Signature Data

- **WAAMD Measurements**
  - UH AHI instrument data
  - Ground truth data
- **MURI measurements**
  - 4 soil, 7 road, 6 pipe, and 8 other samples for a total of 73 measurements
  - Measured with a D&P Instruments Turbo FT portable FT-IR spectrometer
  - 2-16 micron wavelength range
  - Each signature the result of 4,200 co-added spectra
  - UH AHI instrument data

Soils Signature Database

Disturbed Dirt Emissivity Spectra from two locations





When Bad Results Happen to Good Methods

- Thermal IR multi-spectral imagery presents some unique challenges for automated mine detection.
- In this research, we will not show methods that slightly improve results through incorporation of a priori information.
- Instead, we will show how methods fail when subtle assumptions are violated.
- We will point out the limitations of common trends in algorithms and discuss means for addressing those limitations.

The Methods

- Algorithms for detecting targets in hyperspectral data generally can be categorized as either spectral matchers or anomaly detectors [1].
- For this reason, consideration will be given to two basic methods representing each of these two categories:

  a) A spectral matcher based upon a Fisher Linear Discriminant Analysis.

  b) The RX anomaly detector.

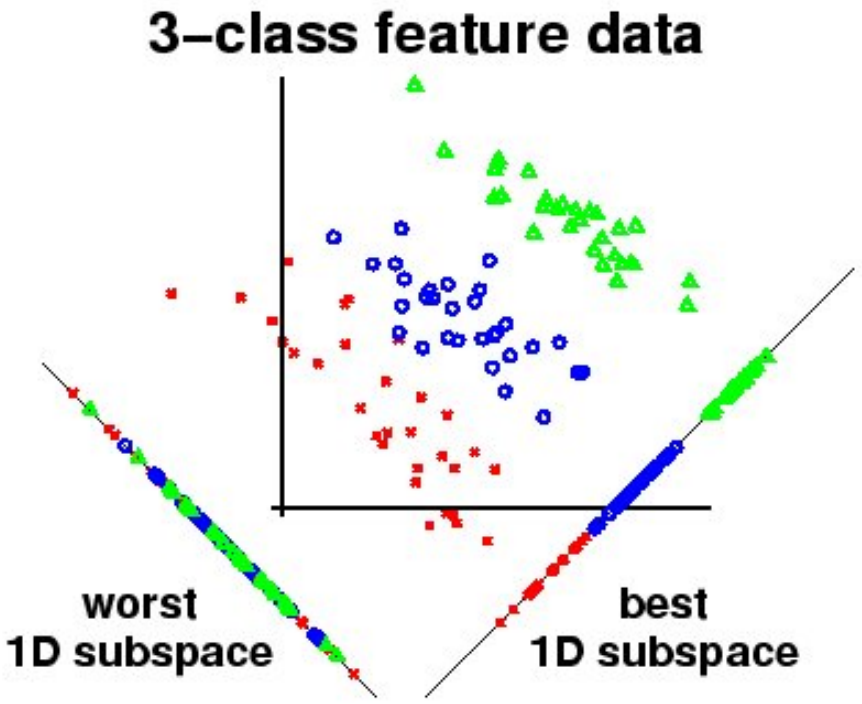
[1] H. Kwon and N. Nasrabadi. Adaptive anomaly detection using subspace separation for hyperspectral imagery. *Optical Engineering*, 42(11), pp 3342-3351, 2003.

The Data

- 70 Band LWIR radiance data from a desert region in the western United States
- Multiple target types
- Multiple false target types
- Varying background
- Large vegetation regions
- Natural anomalies

Spectral Matching

Fisher Linear Discriminant

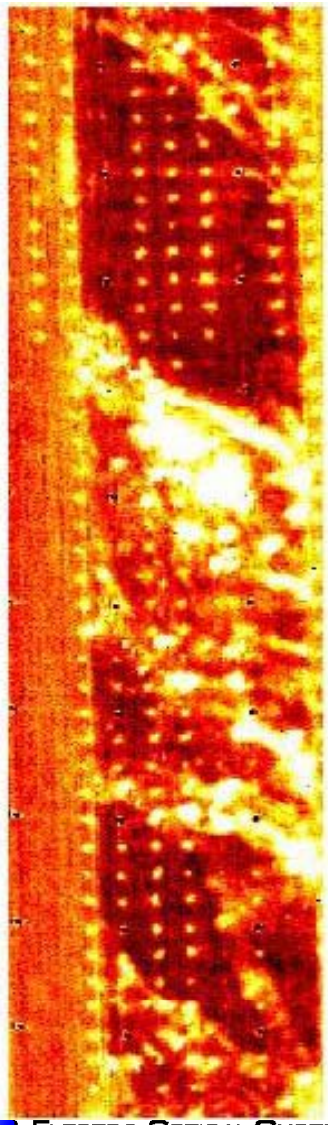


- Use a set of target and background pixels for training.
- Fisher Linear Discriminant: produces a projection vector that maximizes the ratio of between-class variance to within-class variance.
- Project data onto this vector.
- Note: many more advance methods exist for trained classification e.g. Kernel Fisher, and Support Vector Machines.

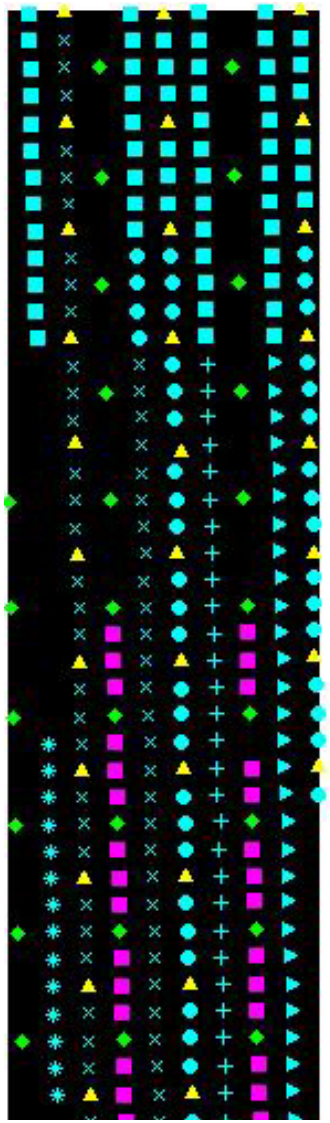
Illustration taken from : http://www.dtreg.com/lda.htm

Spectral Matching — Data Set 1

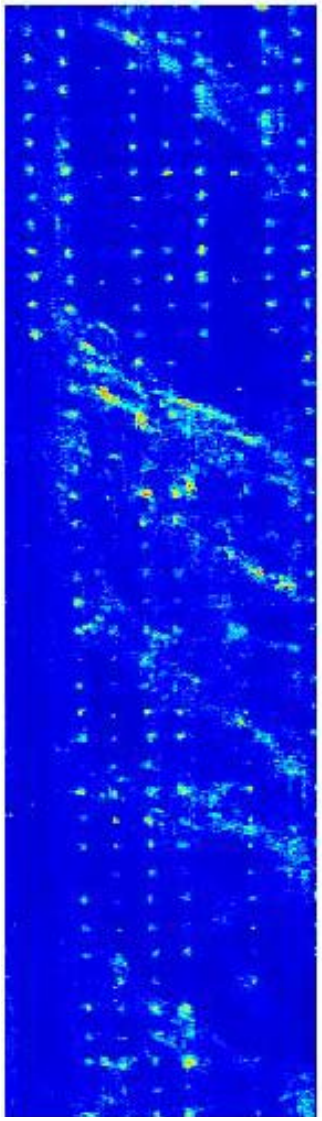
Broadband



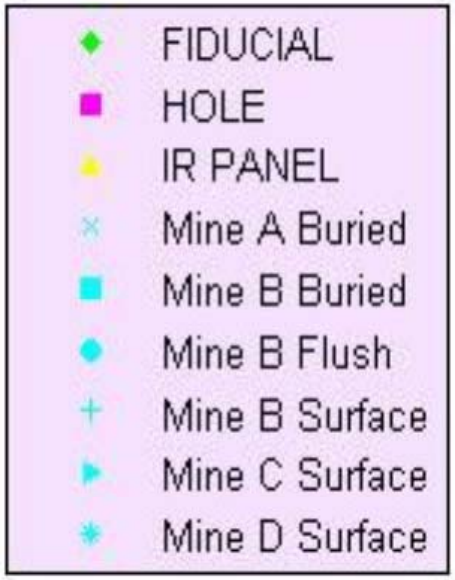
Ground Truth



Spectral Matcher

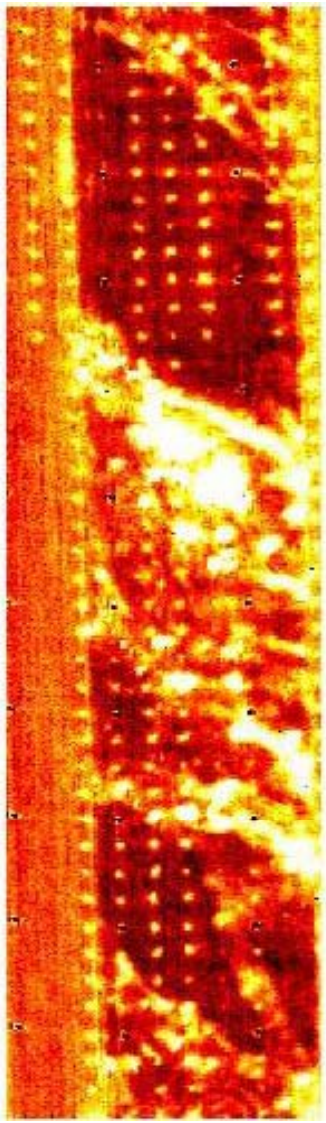


- Broadband Image
- Ground Truth
- Spectral Matcher Trained from Data Set 1

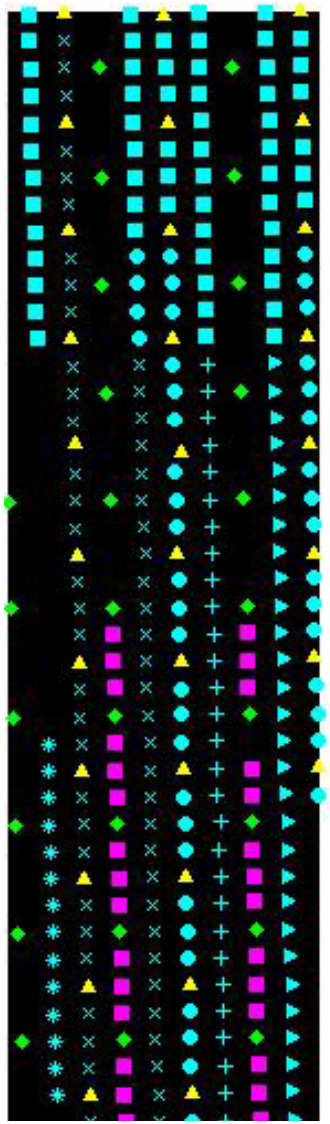


Spectral Matching — Data Set 1

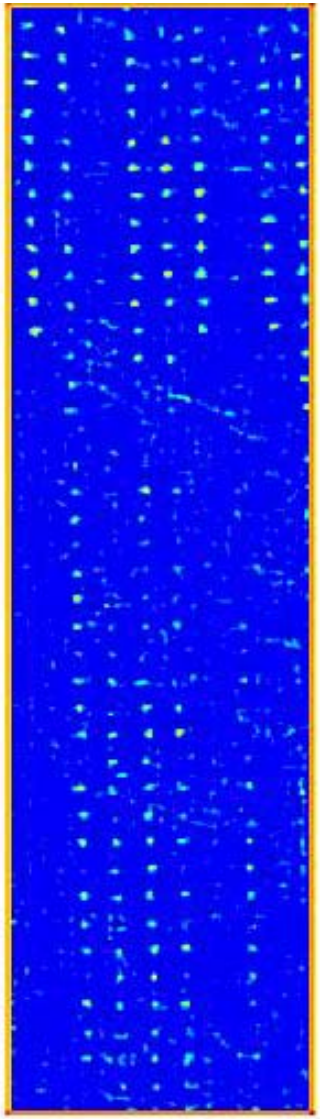
Broadband



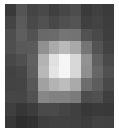
Ground Truth



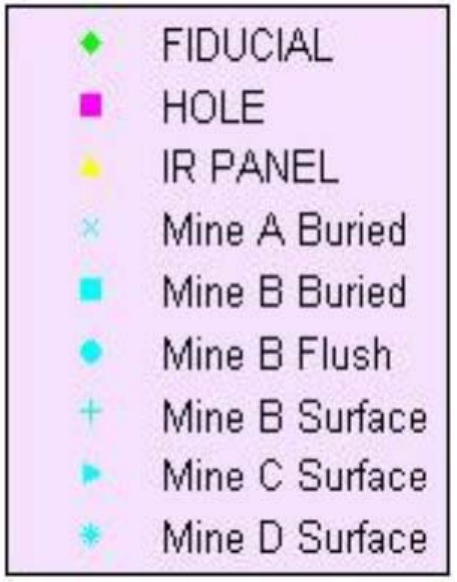
Spectral Matcher



- **Results of Template Matching applied to Spectral ATR**

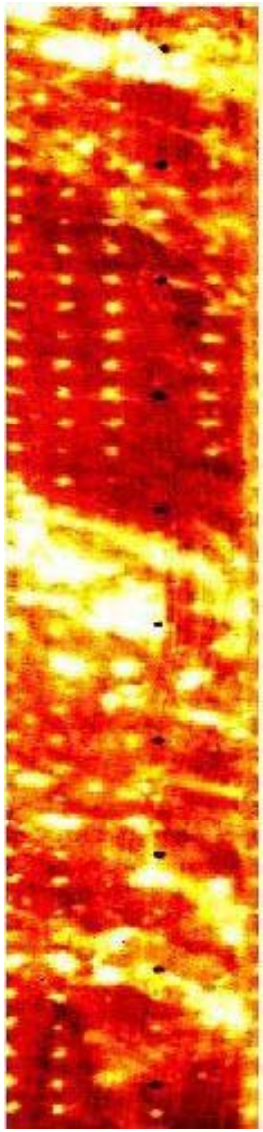


Blown up image of template used to process the above image

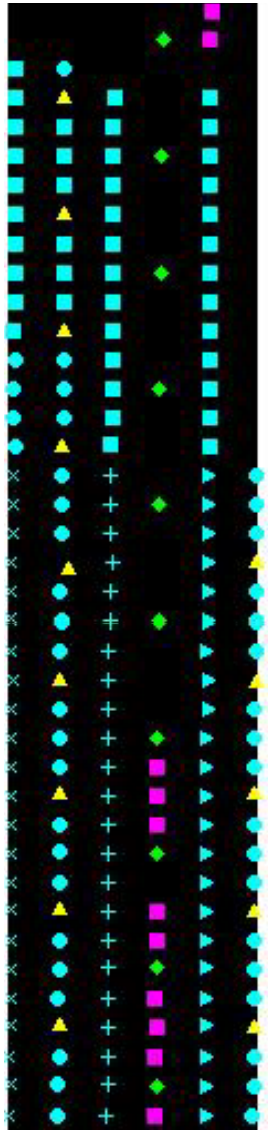


Spectral Matching — Data Set 2

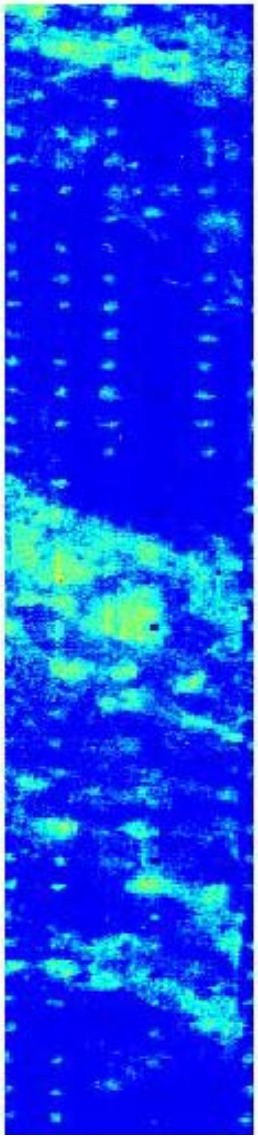
Broadband



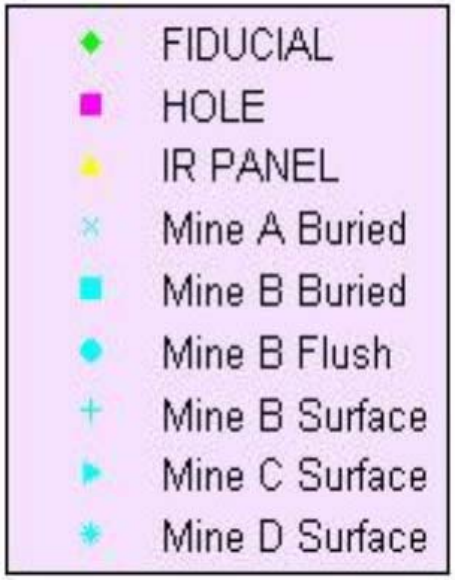
Ground Truth



Spectral Matcher

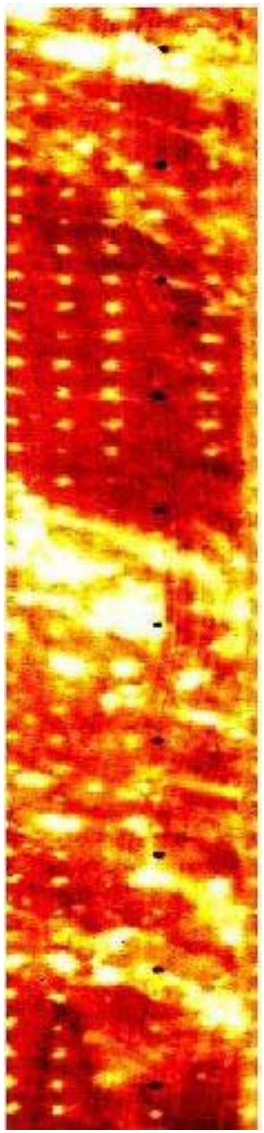


- Broadband Image
- Ground Truth
- Spectral Matcher Trained from Data Set 1

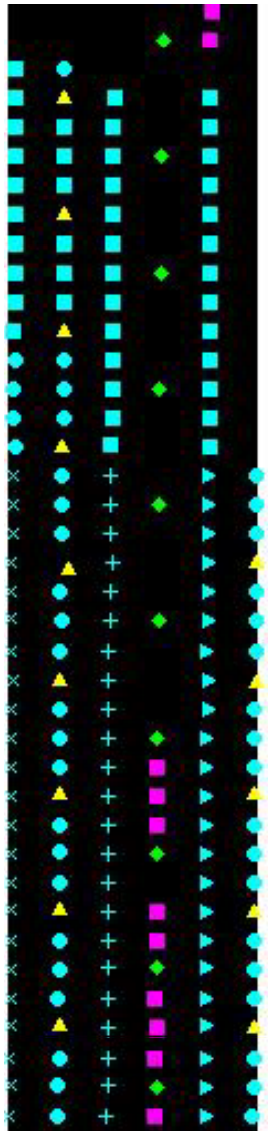


Spectral Matching — Data Set 2

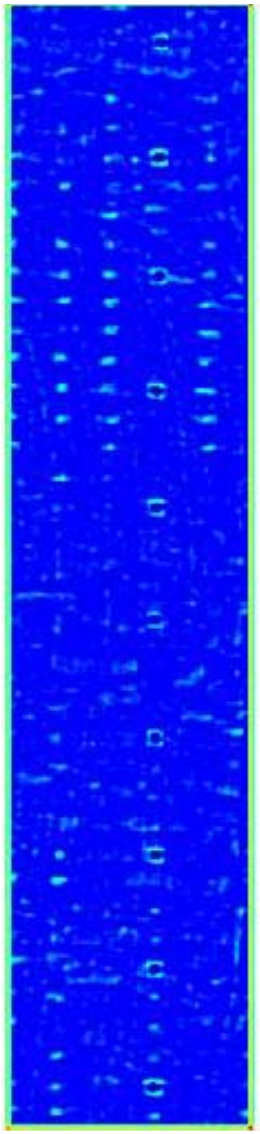
Broadband



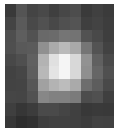
Ground Truth



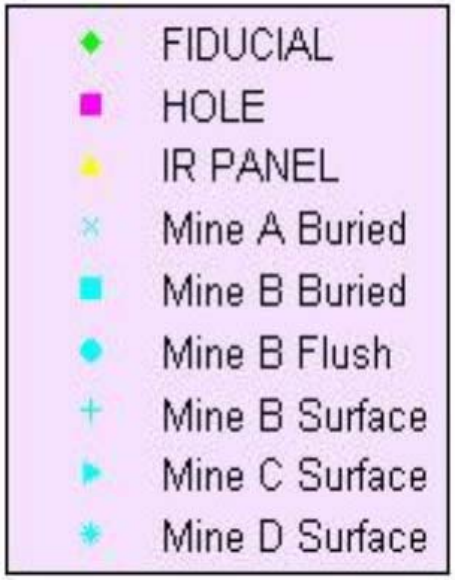
Spectral Matcher



- **Results of Template Matching applied to Spectral ATR**

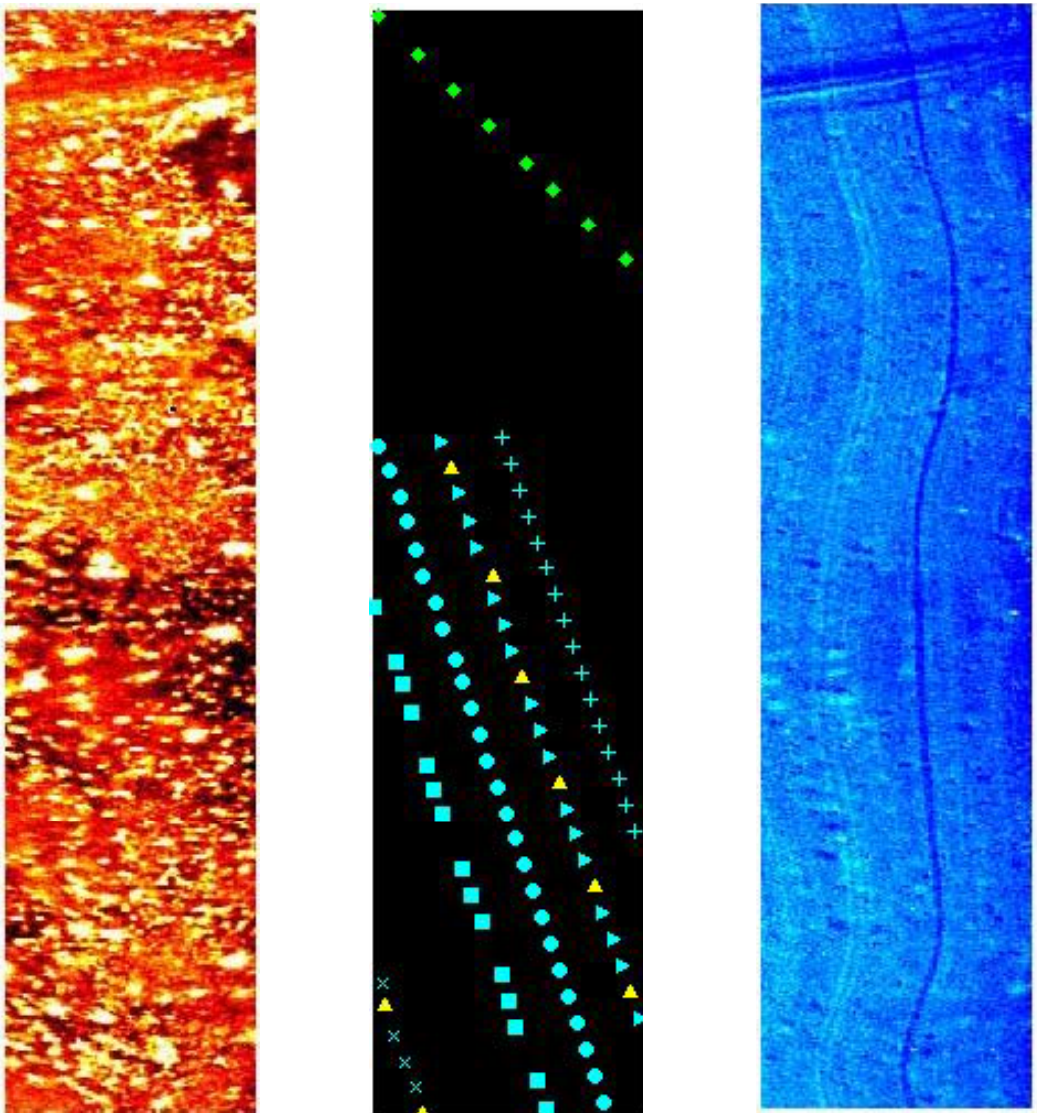


Blown up image of template used to process the above image

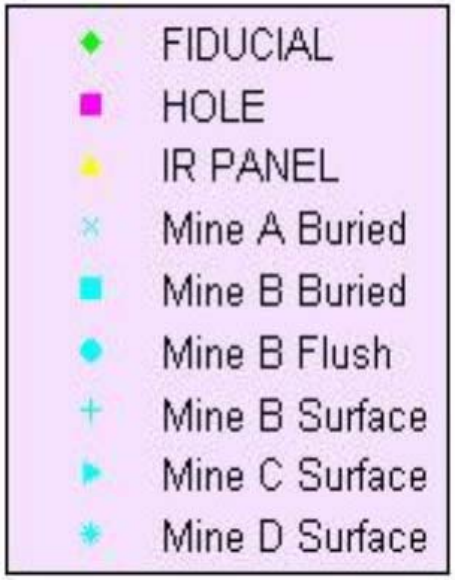


Spectral Matching — Data Set 3

Broadband | Ground Truth | Spectral Matcher

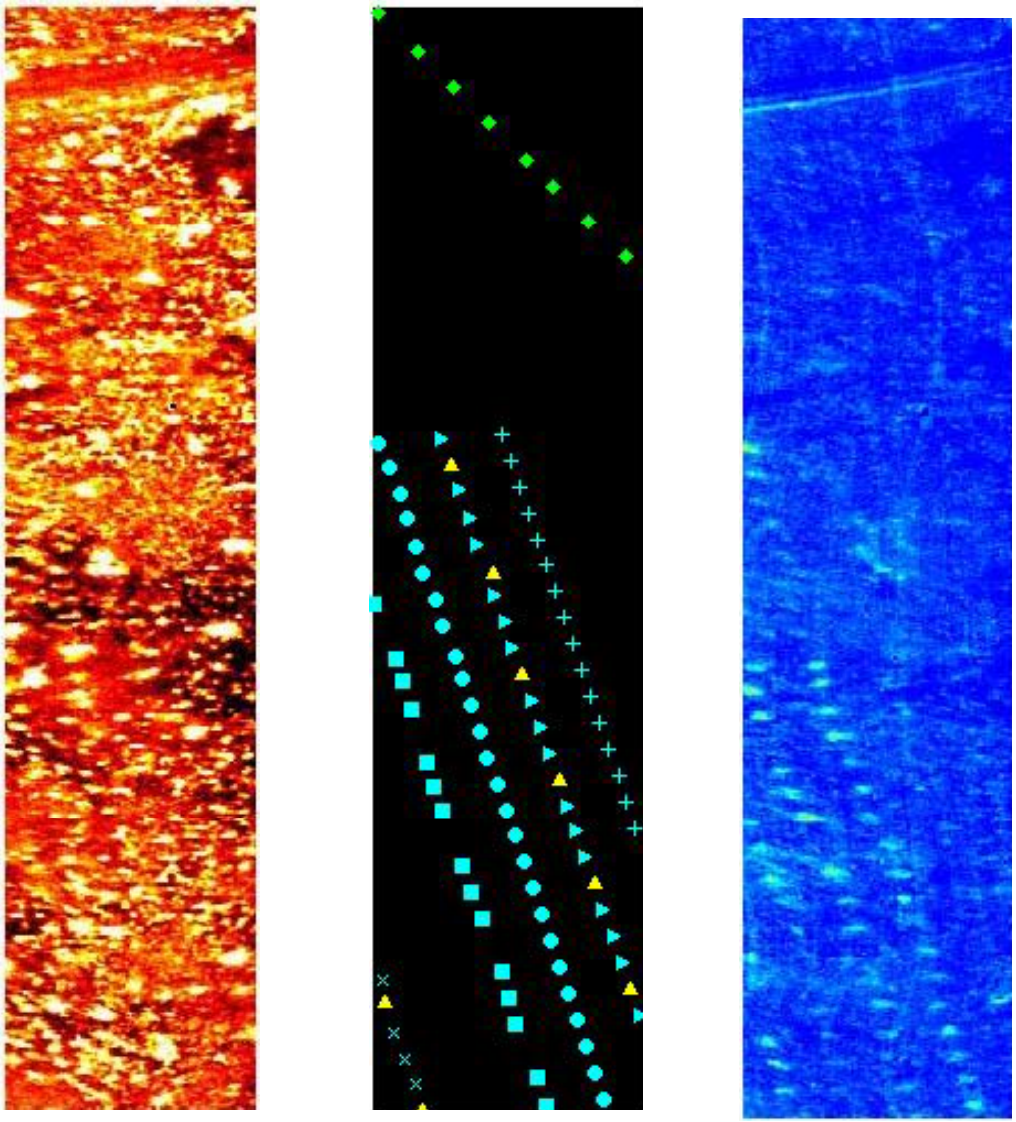


- Broadband Image
- Ground Truth
- Spectral Matcher Trained from Data Set 1



Spectral Matching — Data Set 3

Broadband | Ground Truth | Spectral Matcher

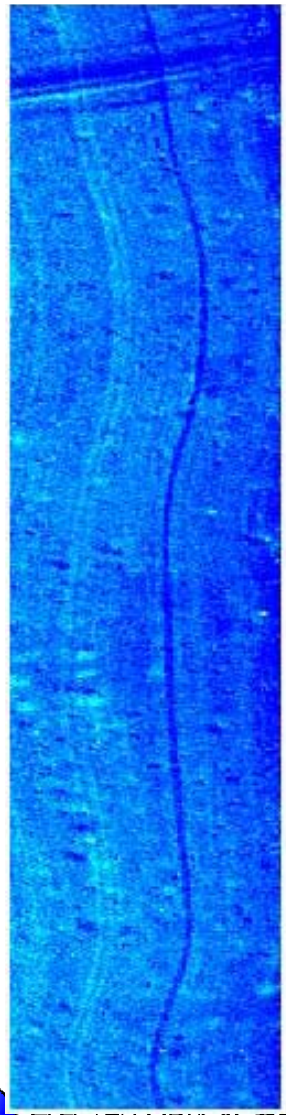


- Broadband Image
- Ground Truth
- Spectral Matcher Trained from Data Set 3

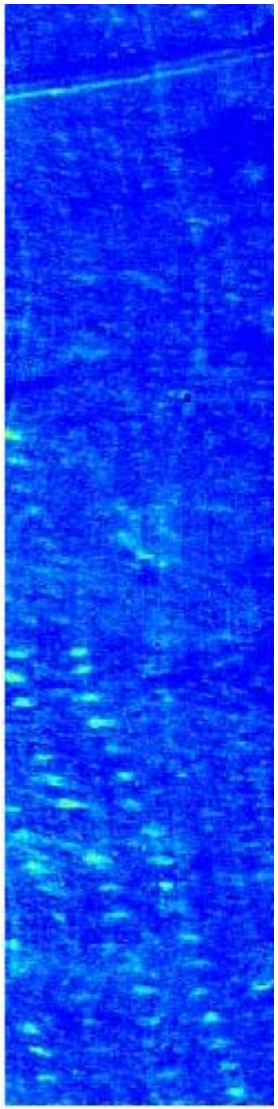
FIDUCIAL
HOLE
IR PANEL
Mine A Buried
Mine B Buried
Mine B Flush
Mine B Surface
Mine C Surface
Mine D Surface

Spectral Matching — Data Set 3

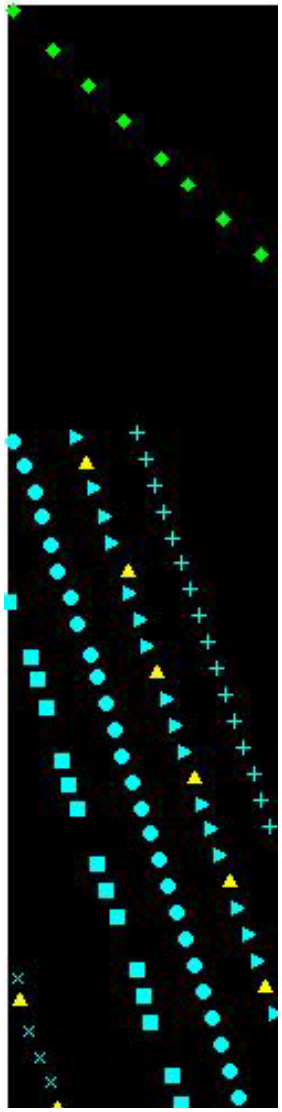
Trained Set 1



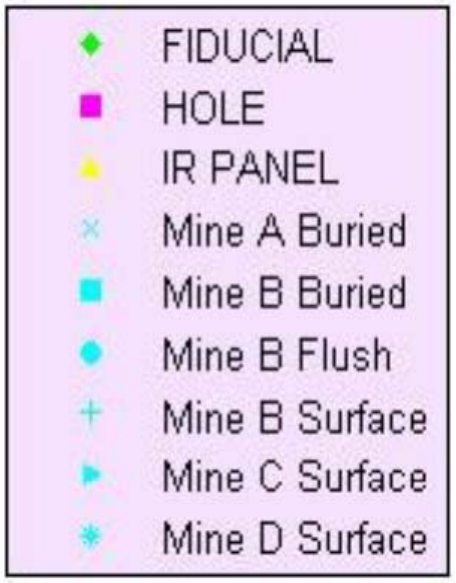
Trained Set 3



Ground Truth

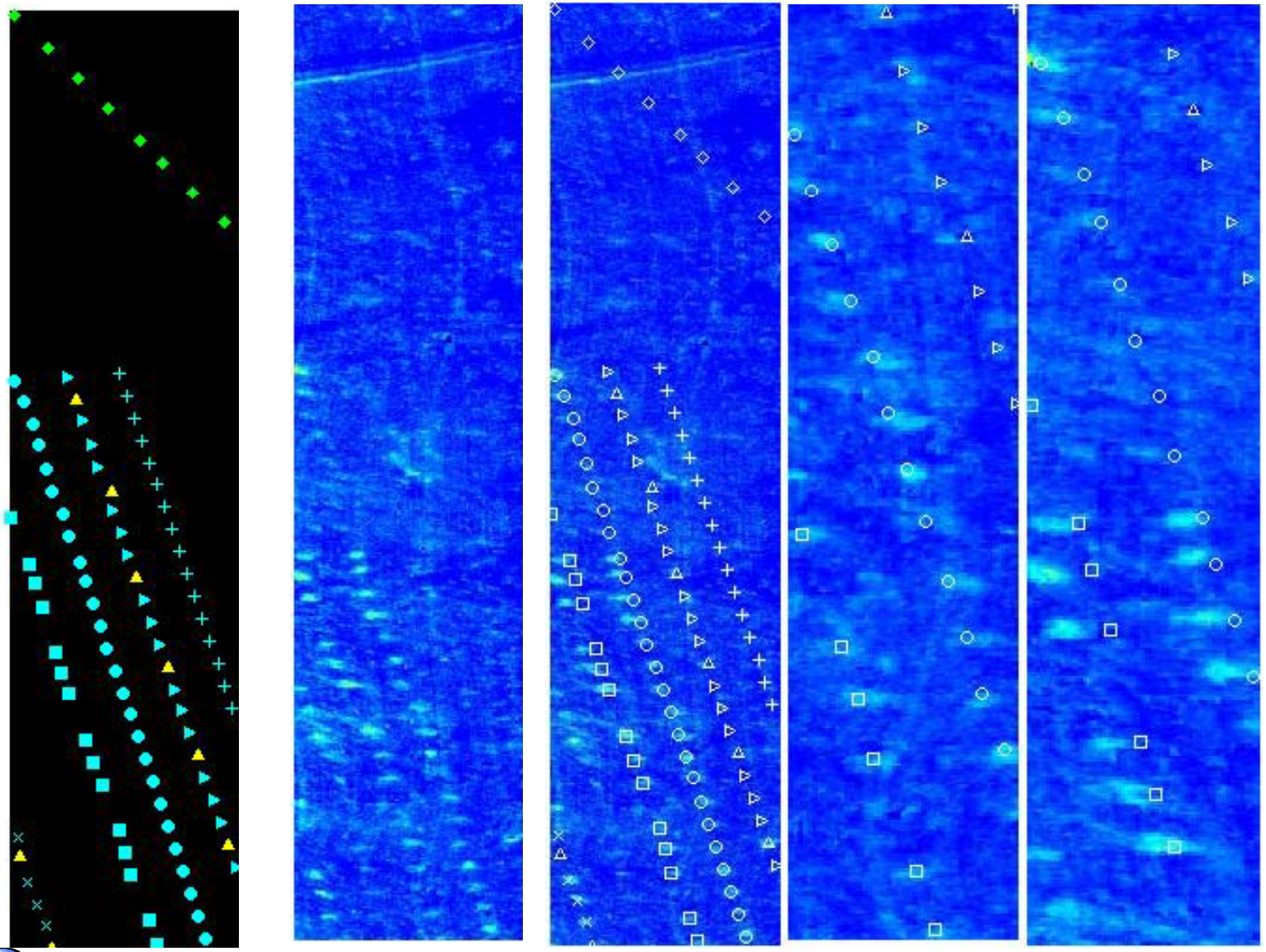


- Spectral ATR Trained from Data Set 1
- Spectral ATR Trained from Data Set 3, i.e. Self Trained
- Ground Truth



Spectral Matching — Data Set 3

Ground Truth | Trained Set 1 | Overlay | Zoom 1 | Zoom 2



- Spectral ATR Trained from Data Set 3, i.e. Self Trained
- Ground Truth Overlay
- Enlarged Sub Regions

Thoughts on Spectral Matching

- Knowing both target and background signatures is powerful !
- Disturbed soil signatures can change dramatically within a geographic region.
- Trained methods are limited by the validity of their training data. More advanced learning methods are still only as good as the training data.
- Testing on one data set is not sufficient !
  - Testing should be done on many data sets in areas with different soil types.

RX Anomaly Detection

- The RX Anomaly Detection Algorithm was developed by Irving Reed and Xiaoli Yu circa 1990 [2].
- Attempts to find pixels that are spectrally different from the rest of the scene.
- Is the foundation for many hyperspectral anomaly detection algorithms.

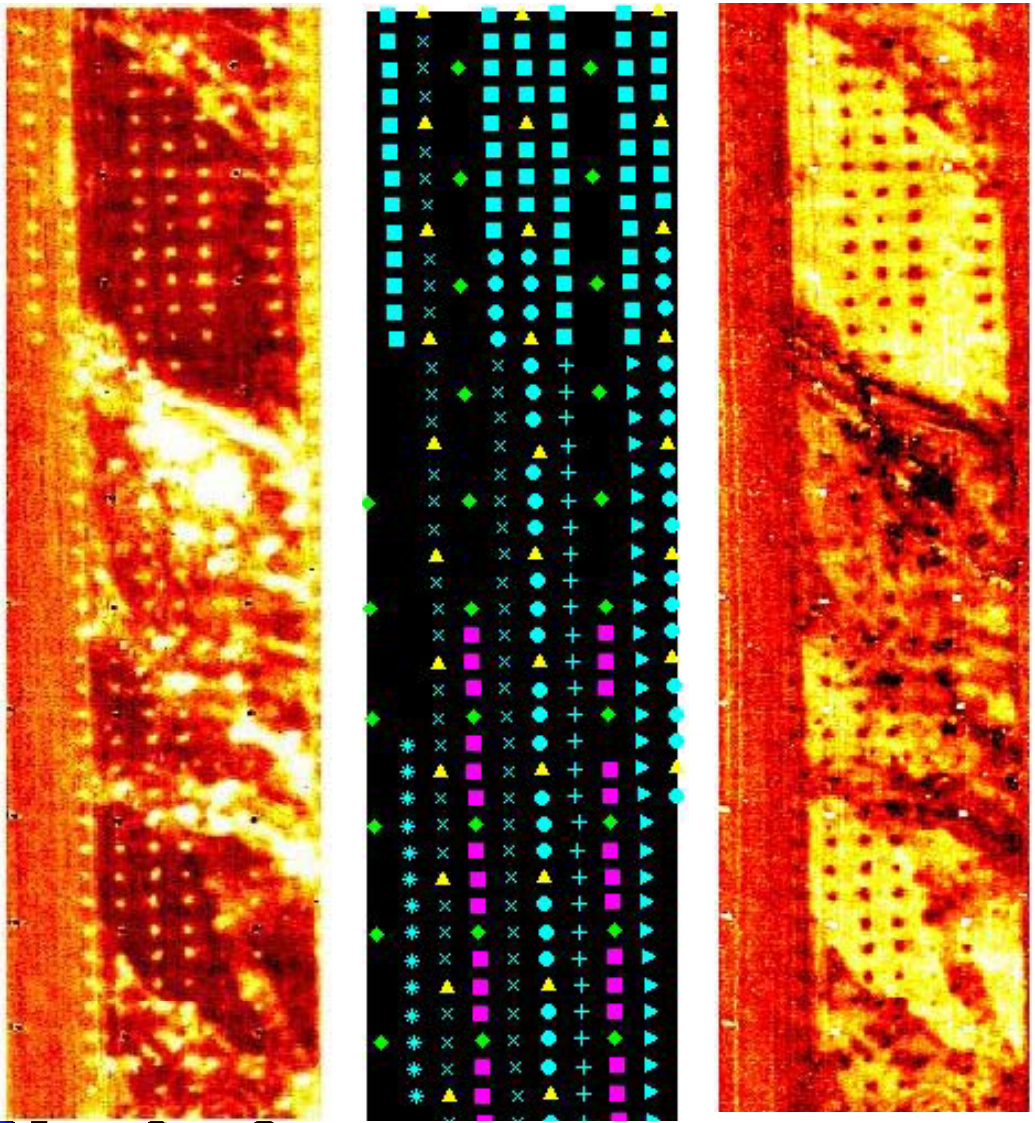
[2] I. Reed and X. Yu. Adaptive Multi-Band CFAR Detection of an Optical Pattern with Unknown Spectral Distribution, *IEEE Transactions on Acoustics, Speech, and Signal Processing,* 38(10), pp 1760-1770, 1990.

RX Anomaly Detection

- The RX algorithms detects anomalies based upon the calculation of the covariance normalized distance between a pixel vector and the mean of the background vectors.
- It is dependent upon the in scene estimation of the mean vector and covariance matrix for the hyperspectral data.
- This estimation can be done using the entire image or can be done using local sub regions.

Global RX – Data Set 1

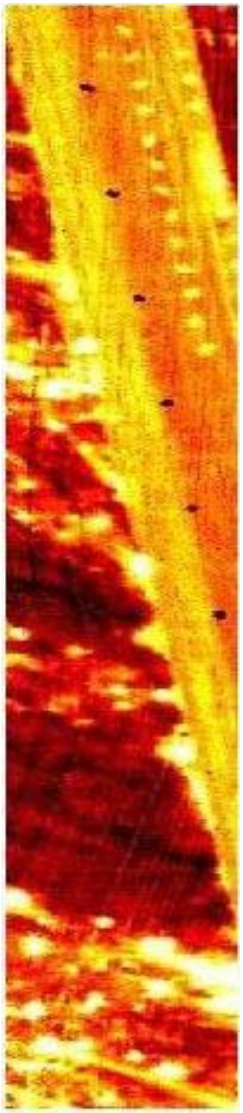
Broadband | Ground Truth | Global RX



- Covariance matrix is estimated using all of the pixels in the image.
- Mines actually have a relatively low detector value.
- Not Good!

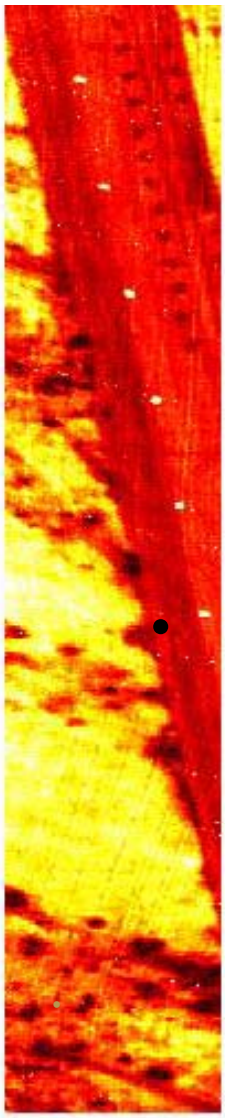
Global RX – Data Set 4

Broadband



Ground Truth

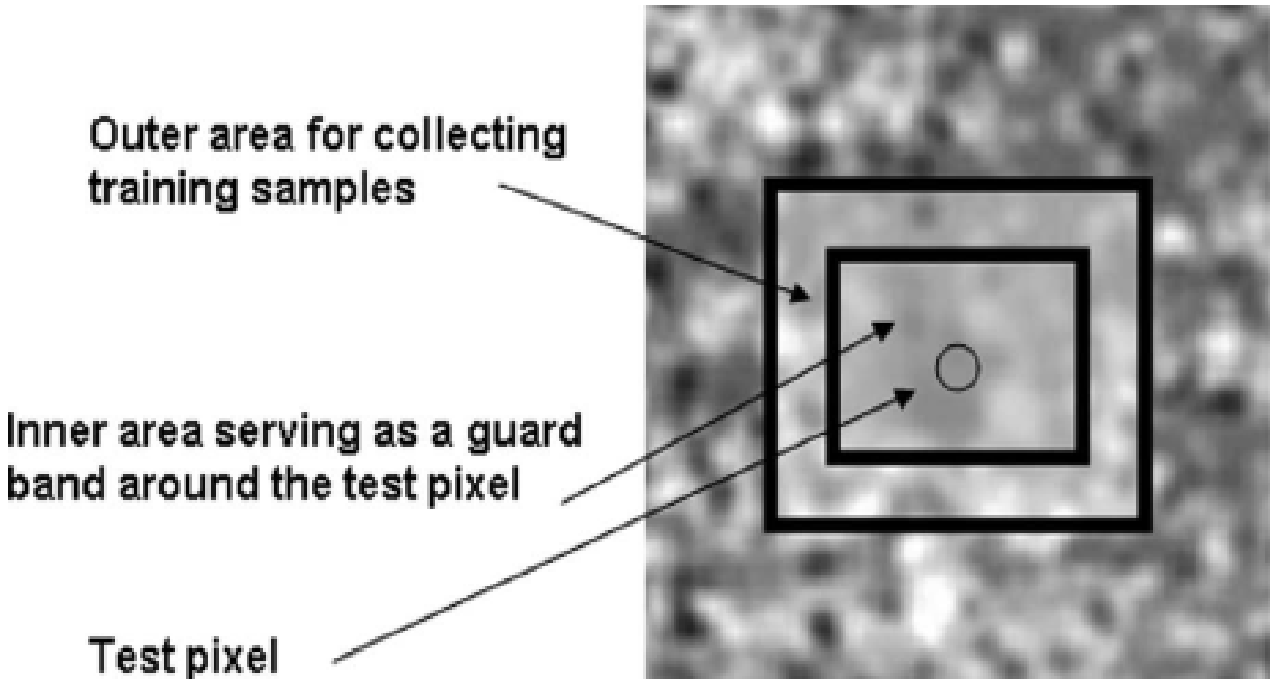
Global RX



- Again, mines actually have a relatively low detector value.
- Why? ...Spectral Variations due to changes in soil type, presence of vegetation and clutter is much greater than the variation due to the presence of a mine.
- This is a well known problem [3].

[3] D. Stein, S. Beaven, L. Hoff, E. Winter, A. Schuam, and A. Stocker. Anomaly detection from hyperspectral imagery, *IEEE Signal Process. Mag., 19(1), pp 58-69, 2002.*

Local RX



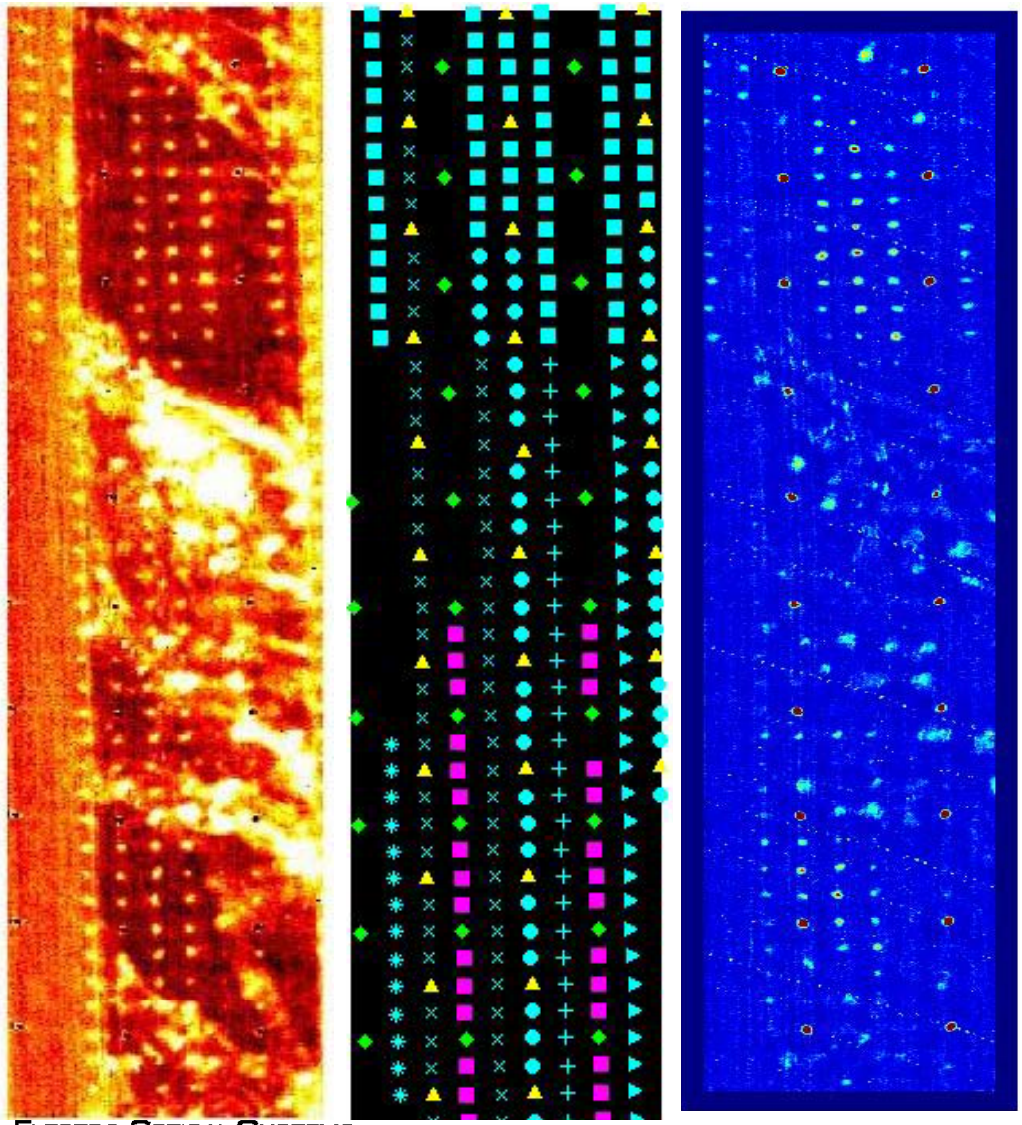
- Uses local data for estimation of mean and covariance.
- The approach shown here utilizes a buffer region between the test pixel and the data used for estimation.

Illustration taken from :

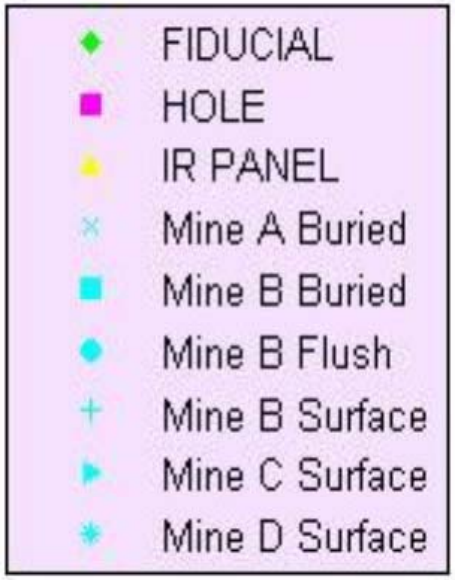
[4] A. Banerjee, P. Burlina, C. Diehl. A Support Vector Method for Anomaly Detection in Hyperspectral Imagery. *IEEE Trans. Geosci. Remote Sens.* 44(8), pp 2282-2291, 2006.

Local RX – Data Set 1

Broadband Ground Truth Local RX

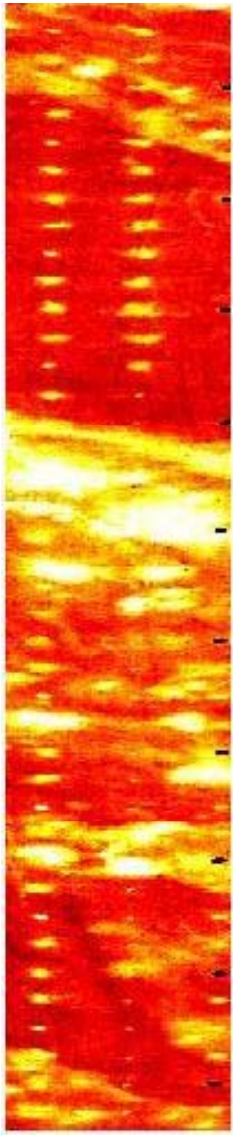


- Broadband Image
- Ground Truth
- Local RX detector statistic

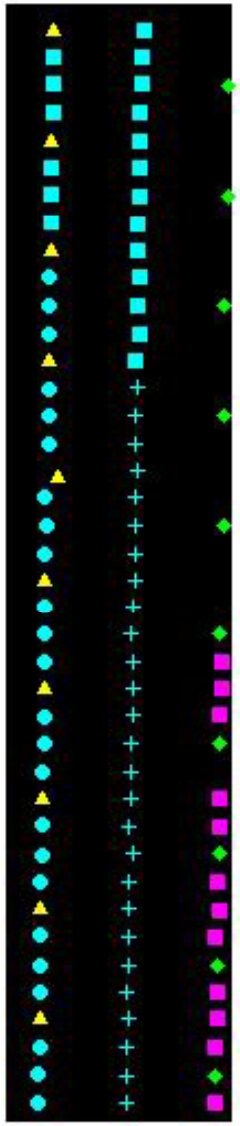


Local RX – Data Set 5

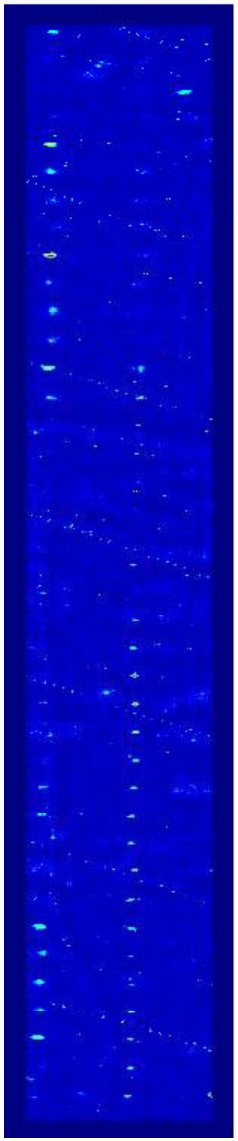
Broadband



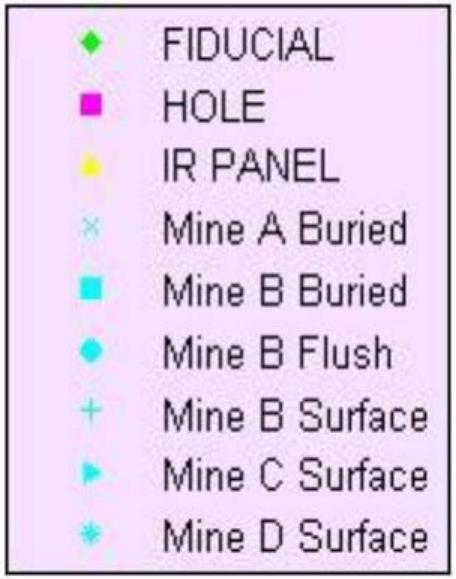
Ground Truth



Local RX



- Broadband Image
- Ground Truth
- Local RX detector statistic



Thoughts on Anomaly Detection

- Anomaly detector still rely on a priori information in the form of subtle spatial assumptions.
- Knowing the target spatial distribution is powerful ! More advanced methods further exploit spatial information for improved results.
- The spatial distribution of a target signature can change with soil type (perhaps also with depth, time of day, et cetera).
- In addition to the spatial signature of a single mine, minefield patterns and distances between mines may also effect local anomaly detection.
- Testing on one data set is not sufficient !
- Testing should be done on many data sets in areas with different soil types.

Current Algorithm Limitations

- Algorithms are limited by a priori assumptions:
  - Spectral Matchers typically assume a known spectral target signature or a set of training data
  - Anomaly detectors typically assume a known spatial profile
- While a priori assumptions will likely always be the limiting factor, we can make vast improvements by changing the kinds of assumptions made (e.g. instead of assuming a specific target signature one might assume that the signature satisfies a model)

Model Based Methods

- Model based methods may potentially alleviate the some of the requirements for training data.
- If we can predict target signatures with a few local soil measurements, then spectral matching methods become practical across soil/geomorphic regions.
- Thermal models and there associated inverse problems have been developed for the identification of landmines but require the estimation of soil thermal diffusivity, soil thermal emissivity, soil absorption, sky absorption, and convection heat transfer coefficients [5].
- Soil models are key to model based detection.

[5] T. Nguyen, D. Hao, P. Lopez, F. Cremer, and H. Sahli. Thermal Infrared Identification of Buried Landmines, Detection and Remediation Technologies for Mines and Minelike Targets X, *Proc. Of SPIE,* vol. 5794, pp 198-208, 2005.

Model Based Methods

- A model based hypothesis test might take on a form similar to:

  H_0: Signature = SoilModel + noise(σ_1,Σ_1)

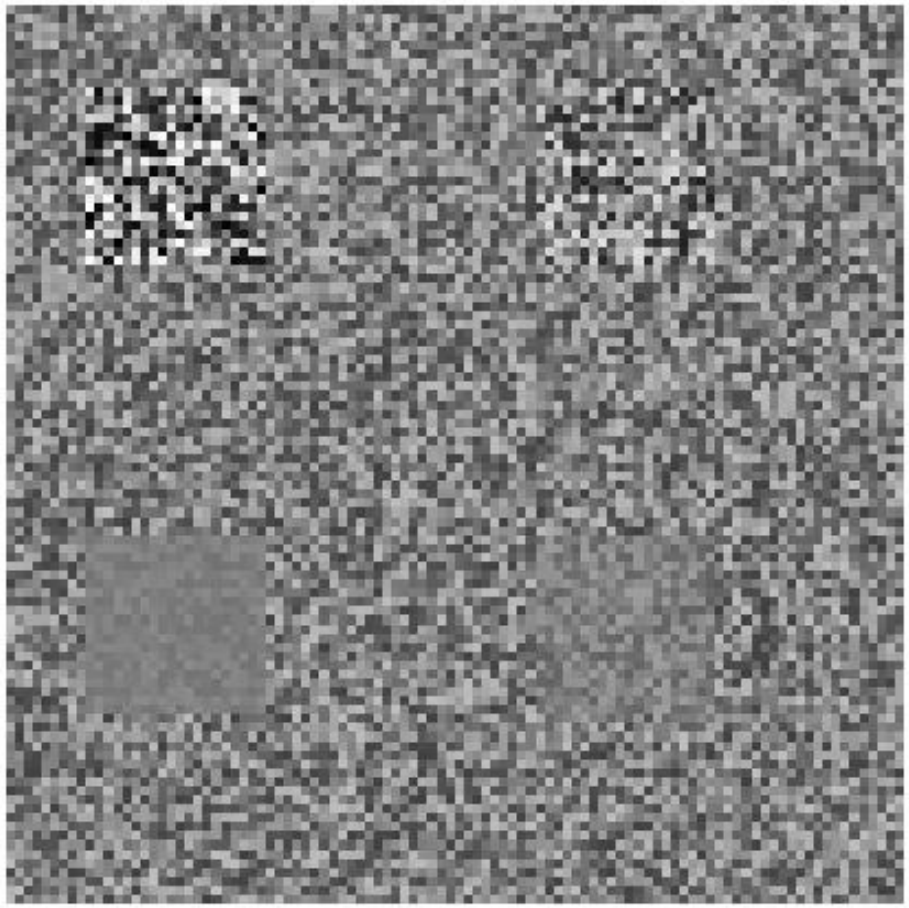
  H_1: Signature = DisturbedSoilModel + noise(σ_2,Σ_2)

- Ideally a model would predict not only the target signature but also the background signature and the expect variation in each.

- Further more, it would be spatially resolved.

Using Other Spatial Information

- With enough pixels on target, detection might be possible using solely higher order statistics.
- That is, even if a target and its background have the same mean signature, if the signature varies differently, then detection may still be possible.
- This motivates the use of higher resolution sensors and the study of signature variation phenomenology.



Spectral Analysis of Terrain Infrared Signatures

PI: Bryce Remesch & Michael Cathcart

Objective

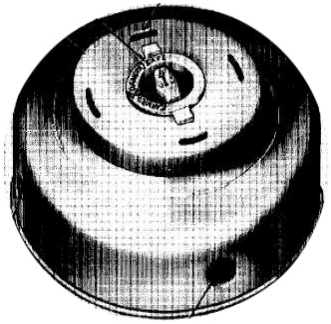
- **Landmine Detection via Hyperspectral Data Analysis**
  - False Target Reduction
  - Background Noise Compensation

- **Considerations**
  - Atmospheric Effects
  - Terrain Variation
  - Signature Masking
  - Natural Anomalies

Approach

- **Local Analysis**
  - Manual Background Selection
  - Statistical Spatial Analysis
  - Natural Variation Compensation
- **Global Analysis**
  - End-Member Analysis
  - Abundance Mapping
  - Classification Scheme Comparison

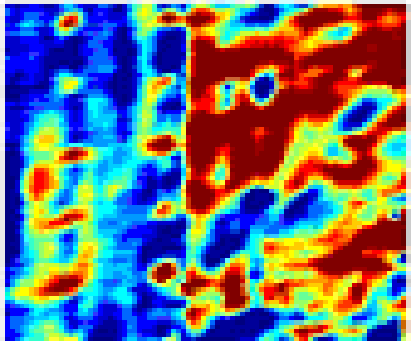
Scene Components



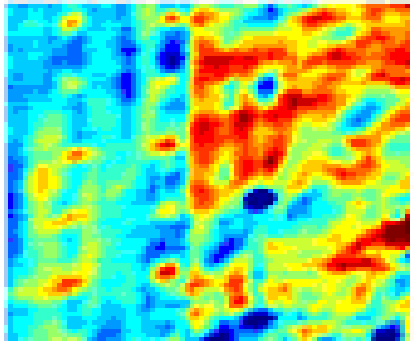
Group	Sub Category
Mines	M19, M20, RAM
Fiducials	IR Panels, Top Hats
Soil	Sandy Soil, Rocky Soil
Vegetation	Tree, Grass, Bush
Anomalies	Rocks, Pavement, Water, Holes

Atmospheric Corrections

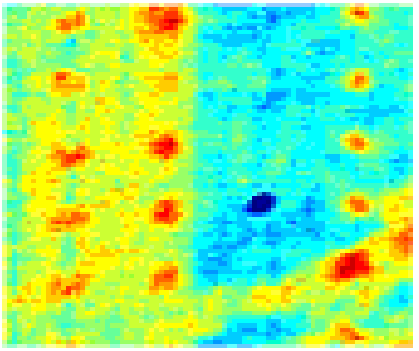
Apparent Radiance



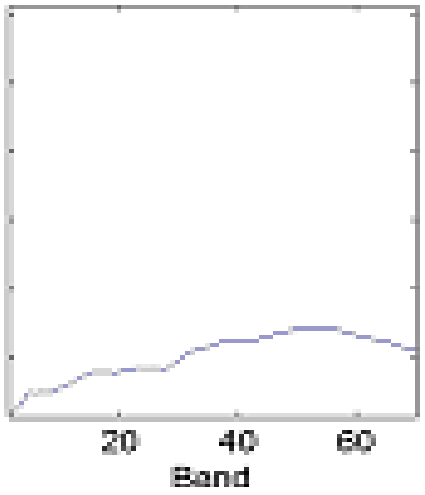
Ground Radiance



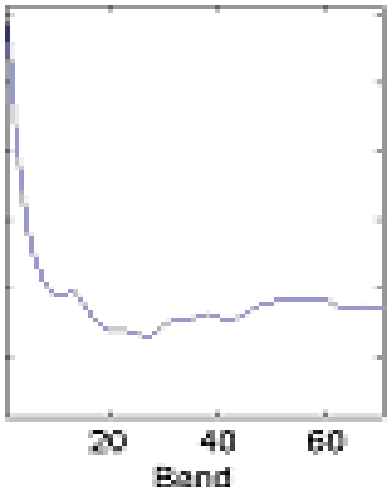
Ground Emissivity



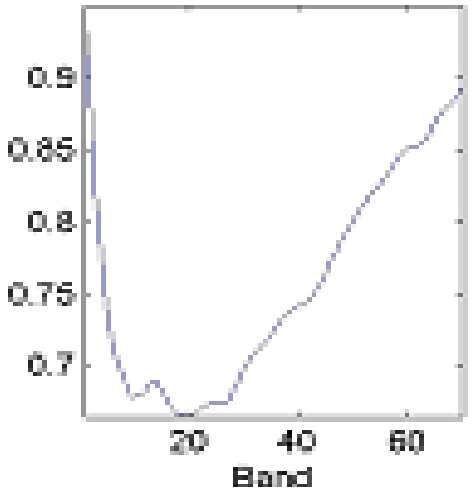
Average Radiance



Average Ground Radiance

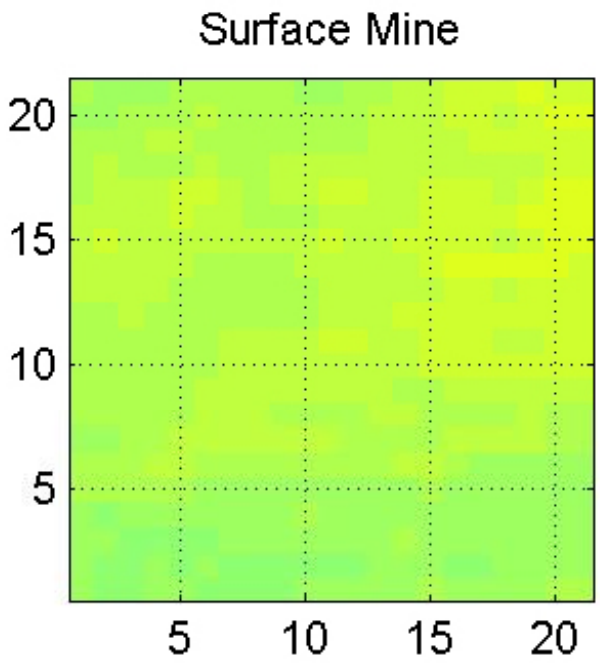


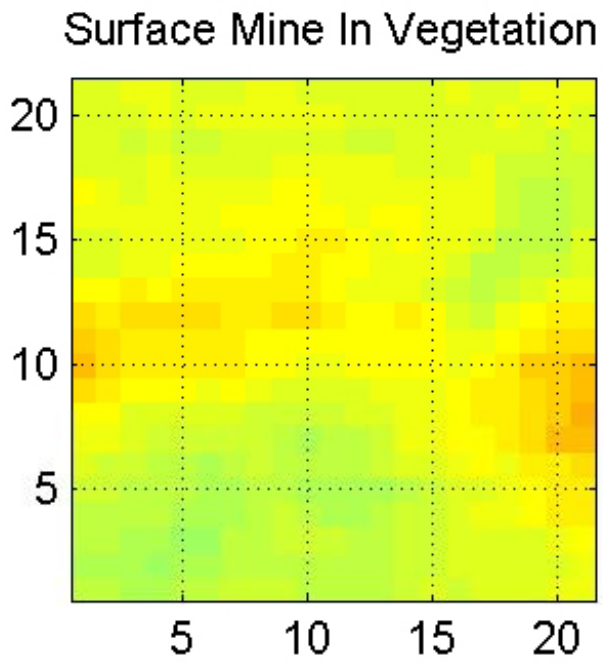
Average Ground Emissivity

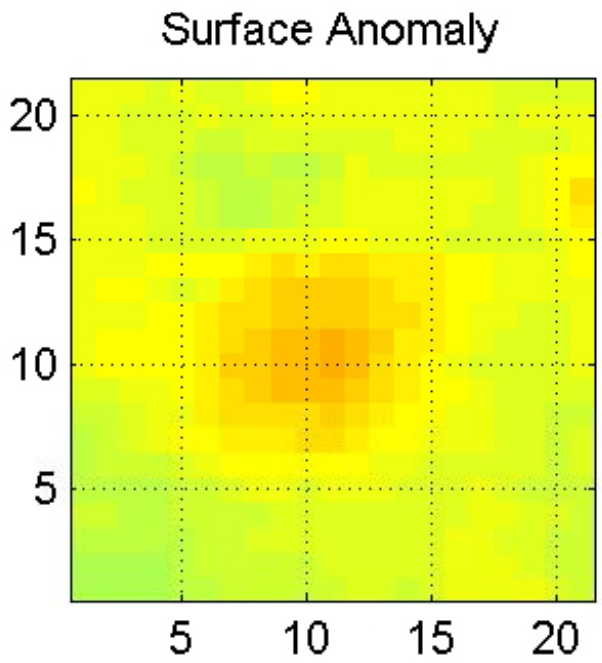


Surface Concerns

- Surface objects difficult to detect
- Vegetation almost completely masks signature
- Surface anomalies may have similar spatial symmetry

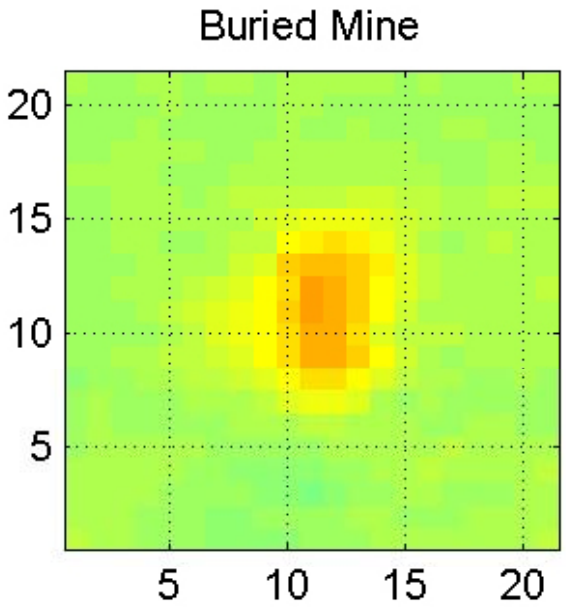


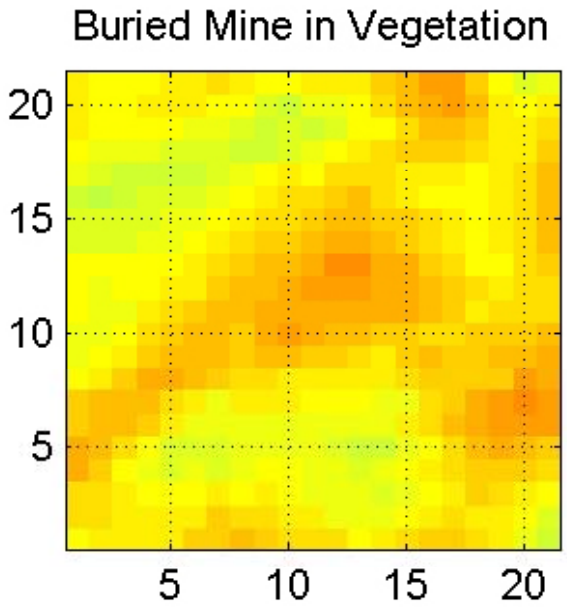


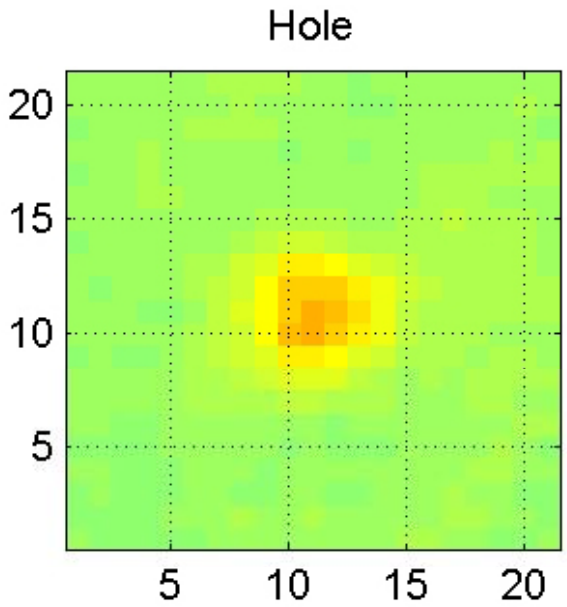


Subsurface Concerns

- Disturbed soil regions not mines themselves are detected
- Holes and regions of disturbed soil reduce target/false target discrimination
- Vegetation almost completely masks disturbed soil signature

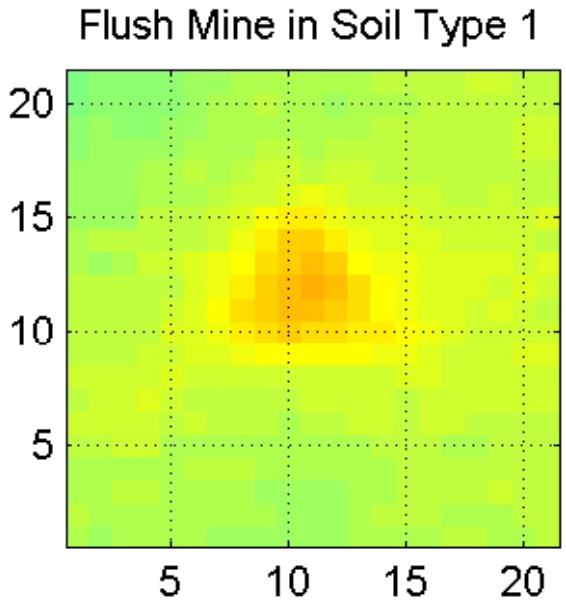


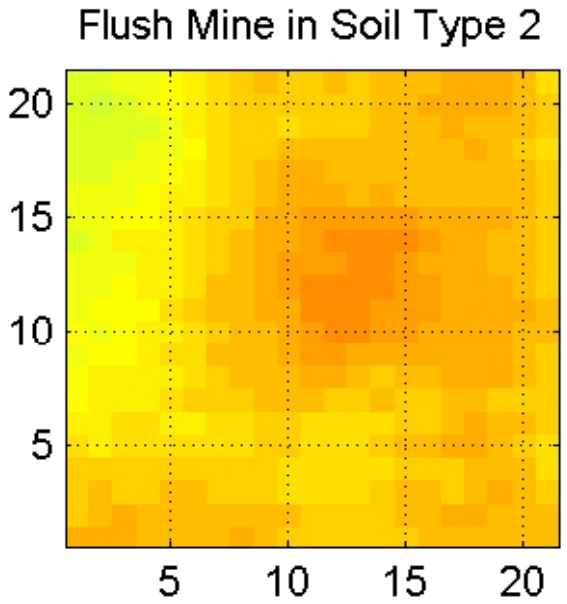


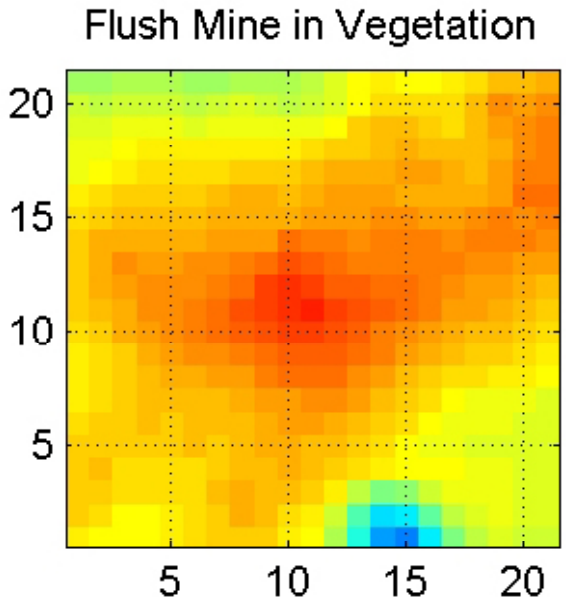


Background Variation & Masking

- Identical objects in different soil types have significantly different spectra
- Natural soil variation increases spectral noise
- Vegetation variation compounds confusion caused by natural soil variation







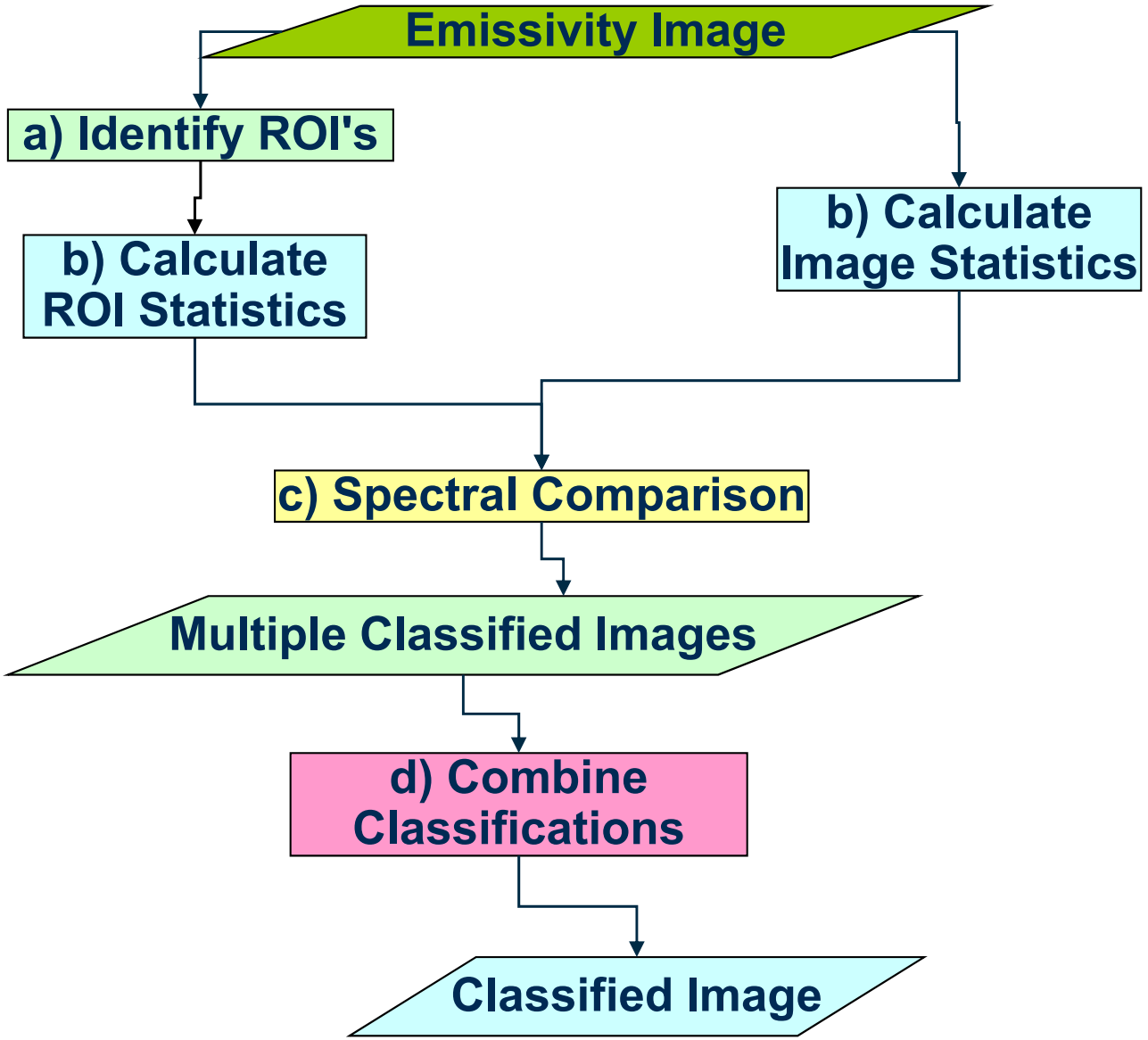
Analytical Techniques

- **Statistical Spatial Analysis**
  - Considers a small spatial range in each band (3x3 pixels)
  - Compares each pixel's statistical spectrum to the spectrum of every manually chosen region of interest
- **End Member Analysis**
  - Considers global scene
  - Employs multiple distance metrics to compare each pixel's spectrum to the spectrum of every end member
  - Uses a voting scheme to classify every pixel and determine an appropriate confidence interval

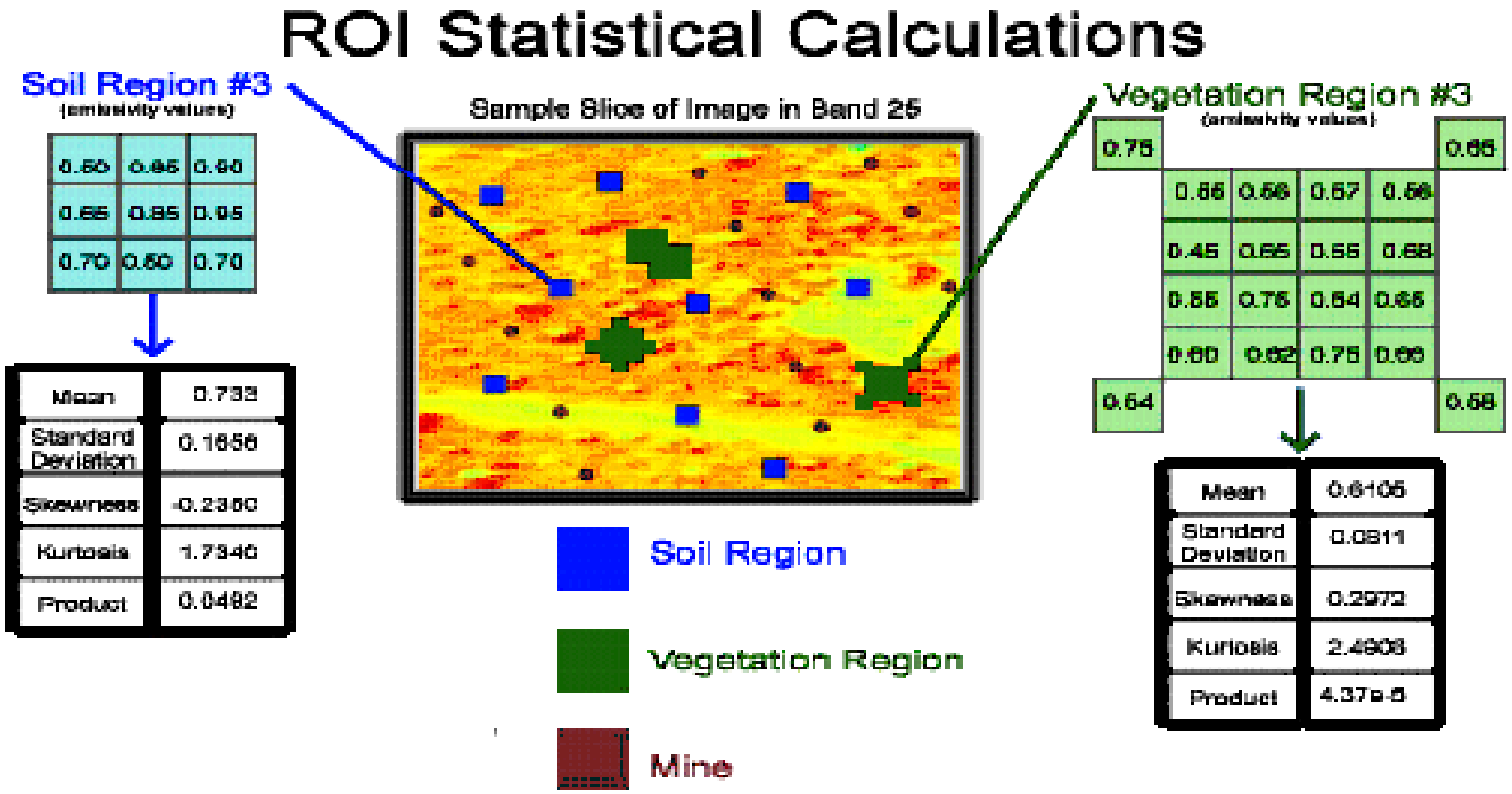
Statistical Spatial Analysis

- Identify regions of interest
  - Select multiple soil and vegetation regions spanning the image
- Calculate spatial statistics
  - Build library of statistical spectra for every region of interest
  - Calculate statistical spectra for every pixel in image
- Compare every library spectrum to every pixel spectrum
- Classify every pixel as soil or vegetation
- Combine classifications to determine background

Spatial Analysis



Statistical Spatial Analysis



Statistical Spatial Analysis

ROI Statistical Calculations

Soil Region

Band

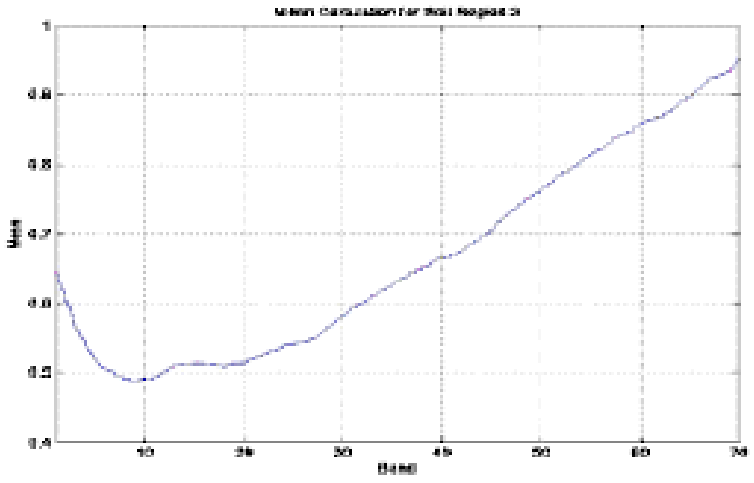
	1	...	24	25	...	70
Mean				0.73		
Standard Deviation				0.17		
Skewness				-0.24		
Kurtosis				1.73		
Product				0.05		

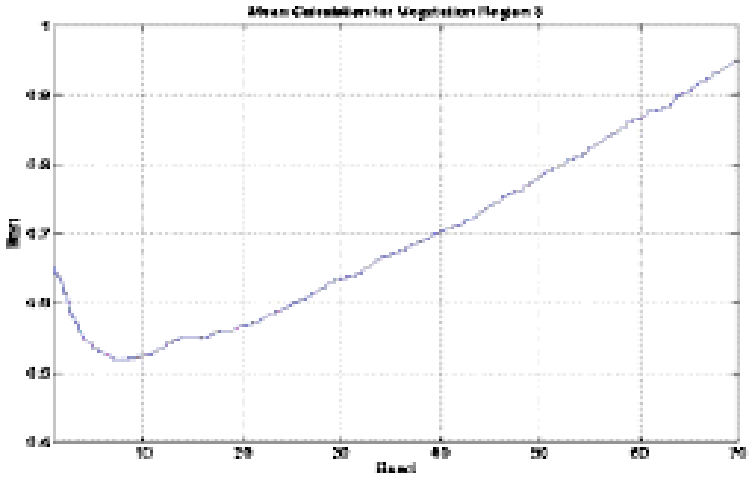
6 statistics calculated
for every band
for every soil region
and every vegetation region

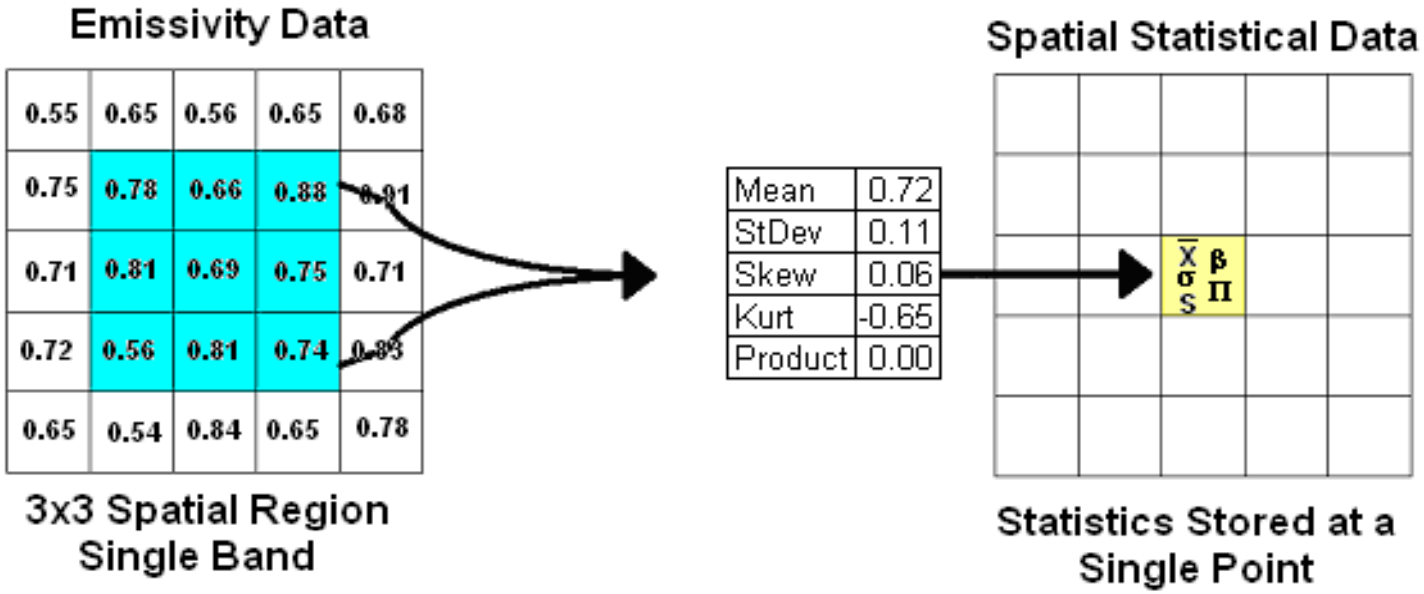
Vegetation Region

Band

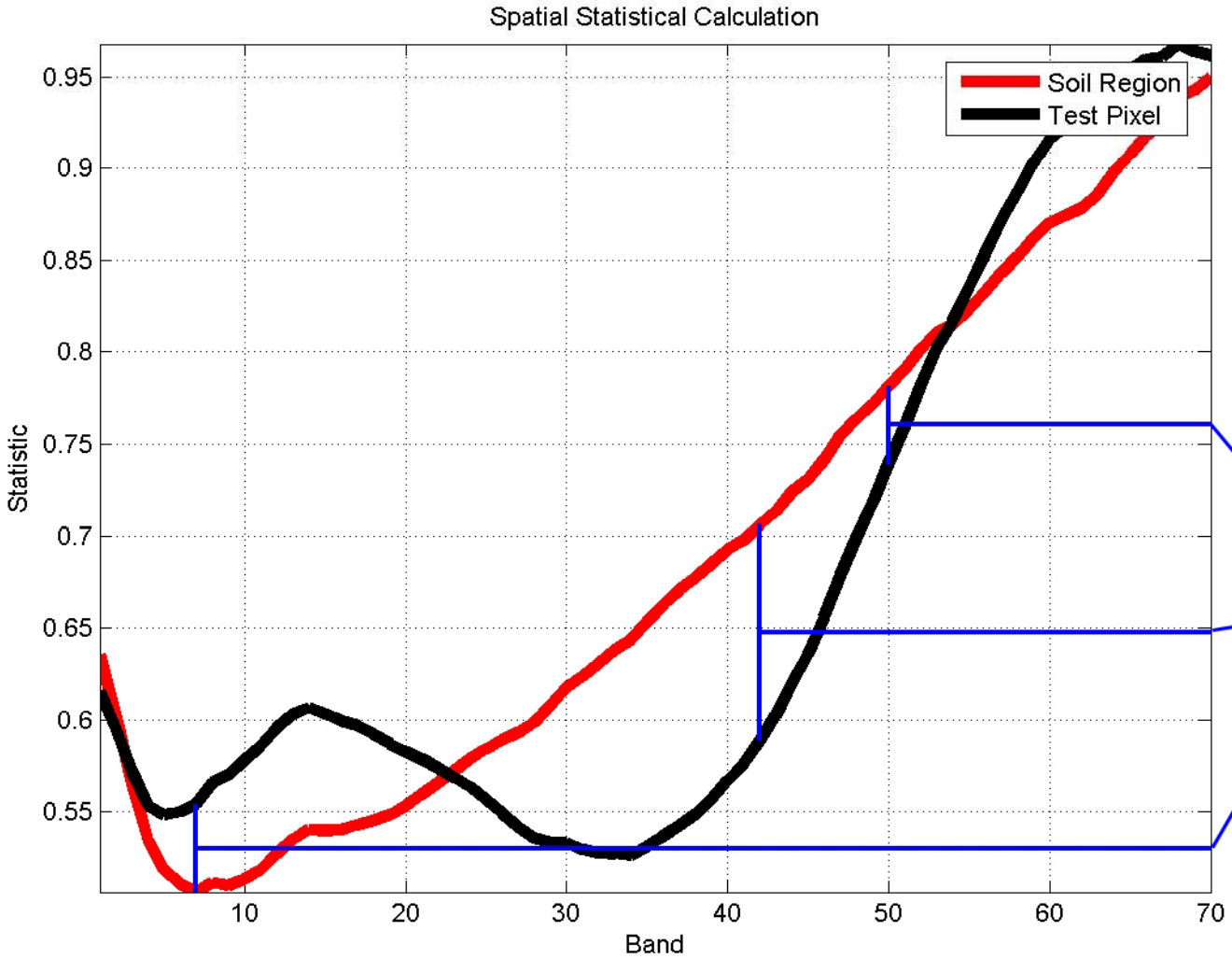
	1	...	24	25	...	70
Mean				0.61		
Standard Deviation				0.09		
Skewness				-0.30		
Kurtosis				2.49		
Product				0.00		





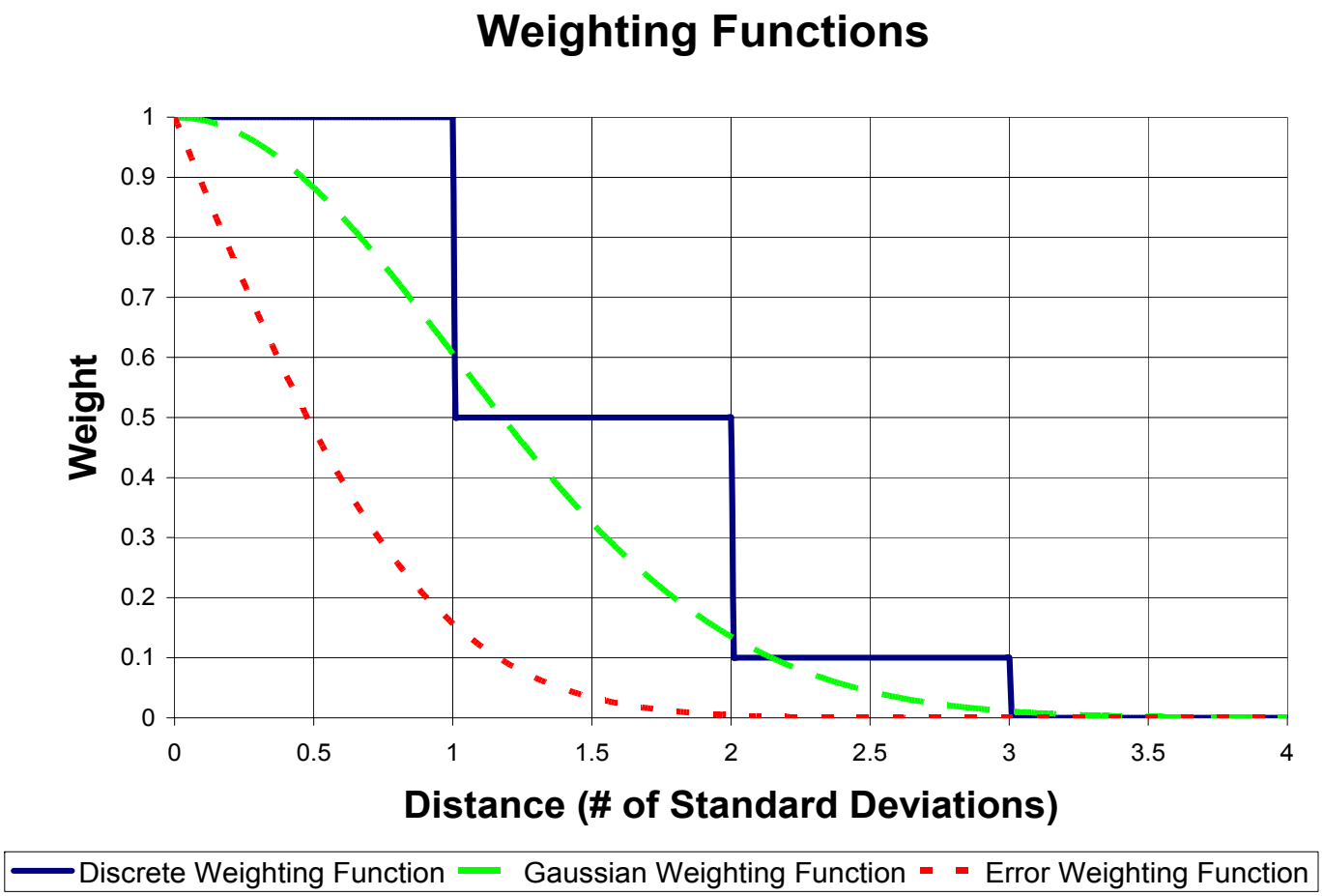


5 statistics are calculated on the values including and surrounding the center point. This calculation is carried out centered at every point in the image for every band.



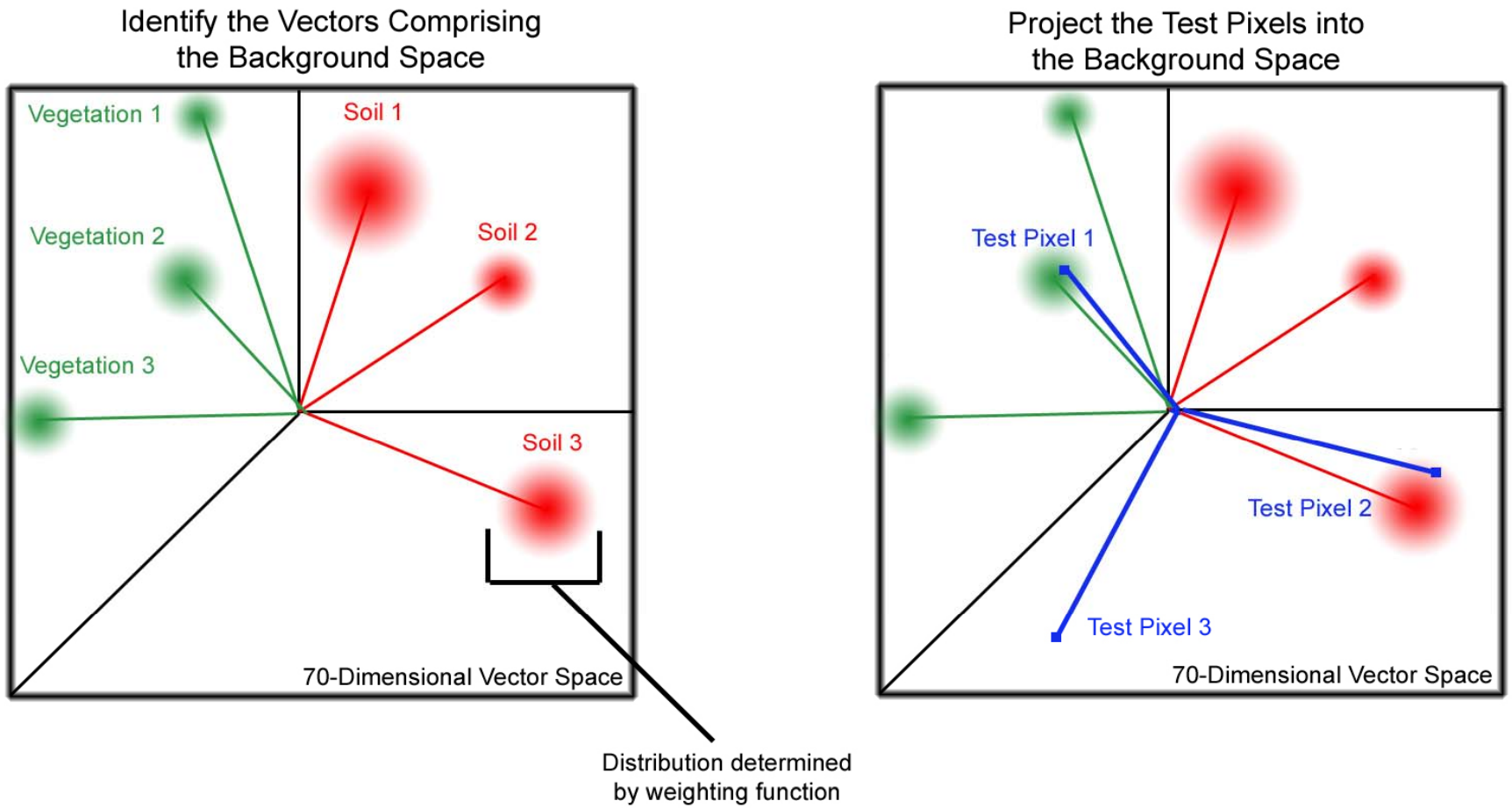
Band #	Weighted Values
…	…
50	0.95
…	…
42	0.70
…	…
7	0.94
…	…
Total Projection	0.90

Weighting Function



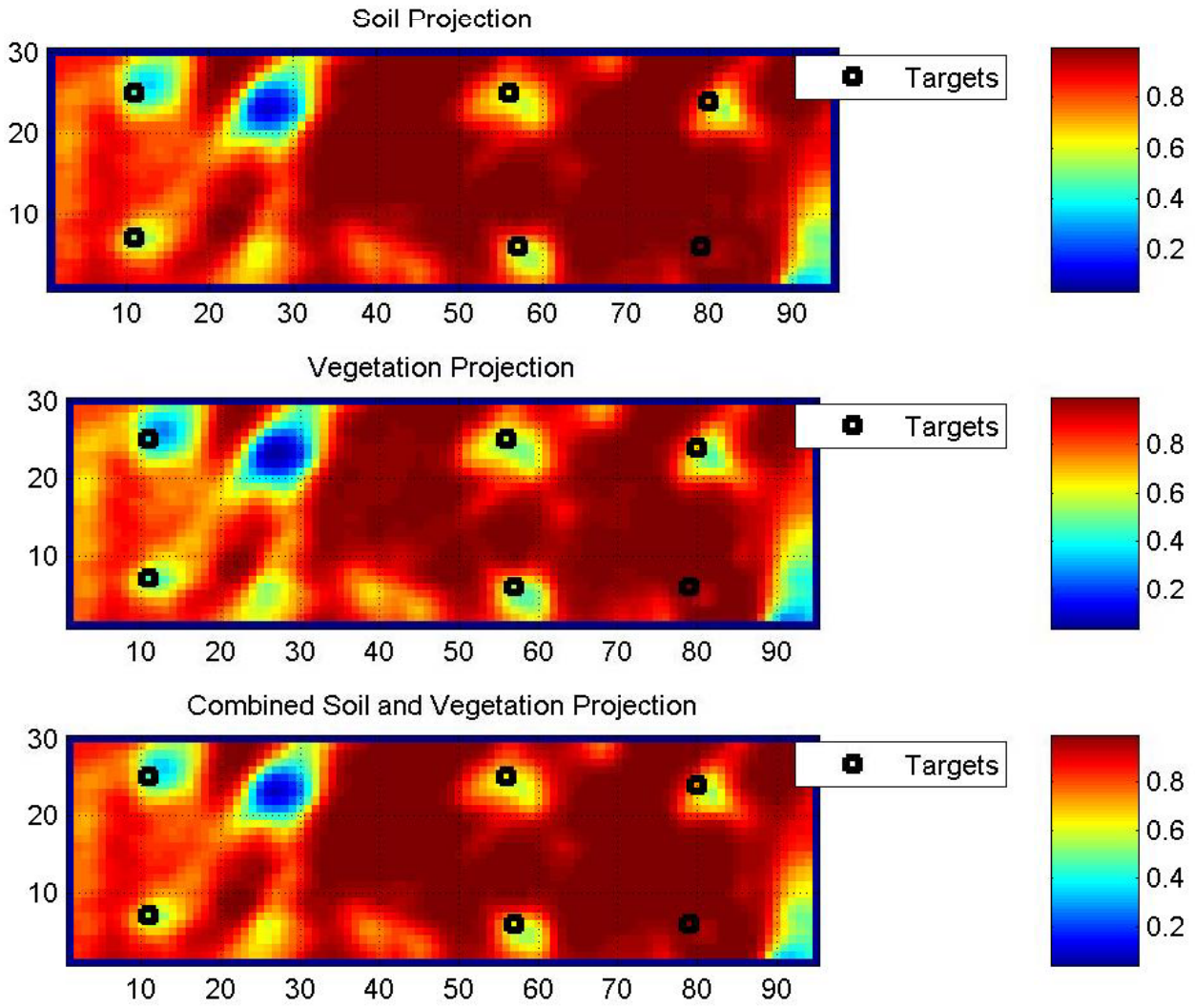
$$P = \sqrt{\frac{\sum_{i=1}^{70} W_i^2}{70}}$$

Vector Space

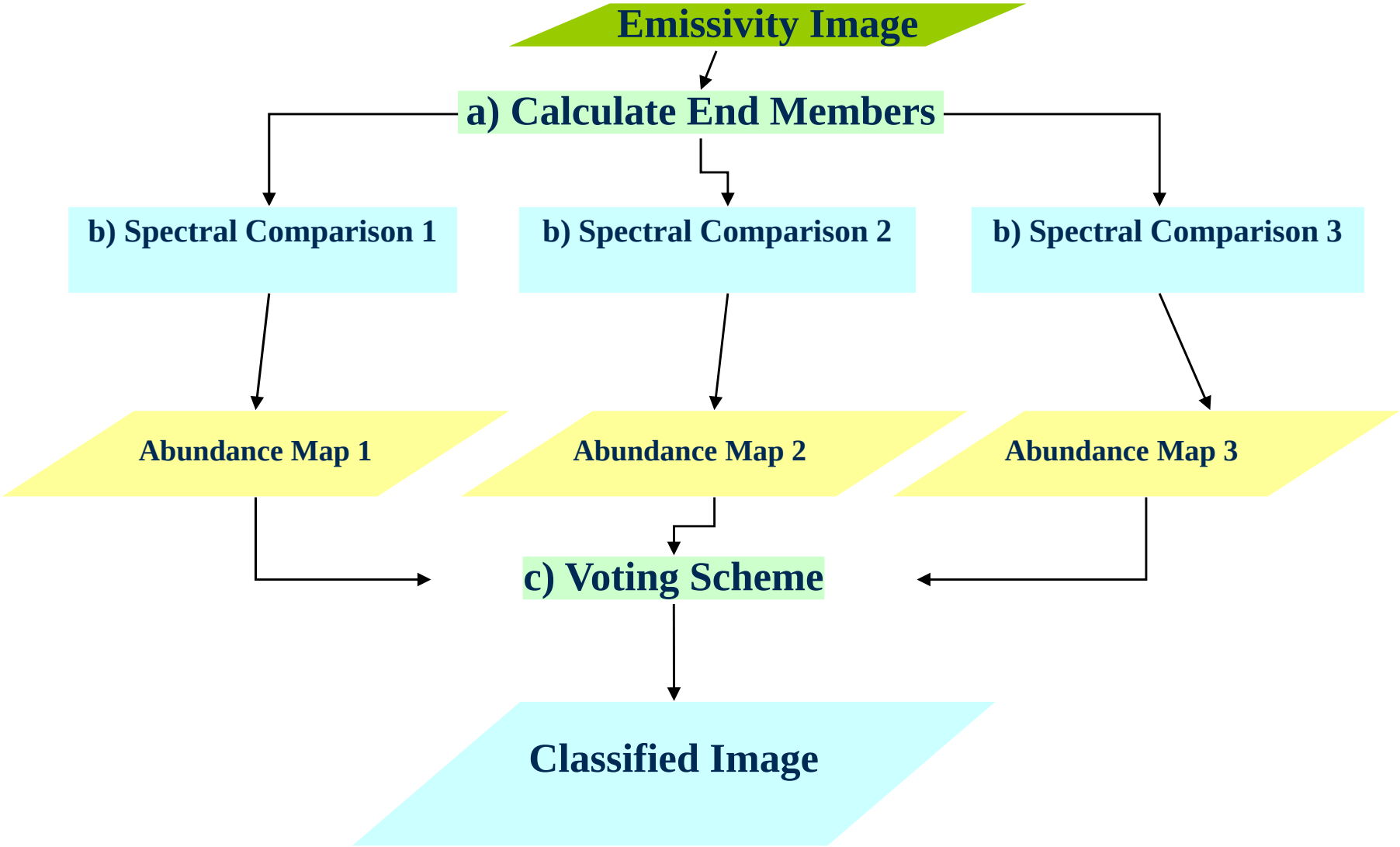


Test Pixel 1 determined to be 90% vegetation
Test Pixel 2 determined to be 70% soil
Test Pixel 3 determined to be neither vegetation nor soil

Results


Soil Projection
Targets
0.8
0.6
0.4
0.2
30
20
10
10 20 30 40 50 60 70 80 90
Vegetation Projection
Targets
0.8
0.6
0.4
0.2
30
20
10
10 20 30 40 50 60 70 80 90
Combined Soil and Vegetation Projection
Targets
0.8
0.6
0.4
0.2
30
20
10
10 20 30 40 50 60 70 80 90

End Member Analysis



End Member Analysis

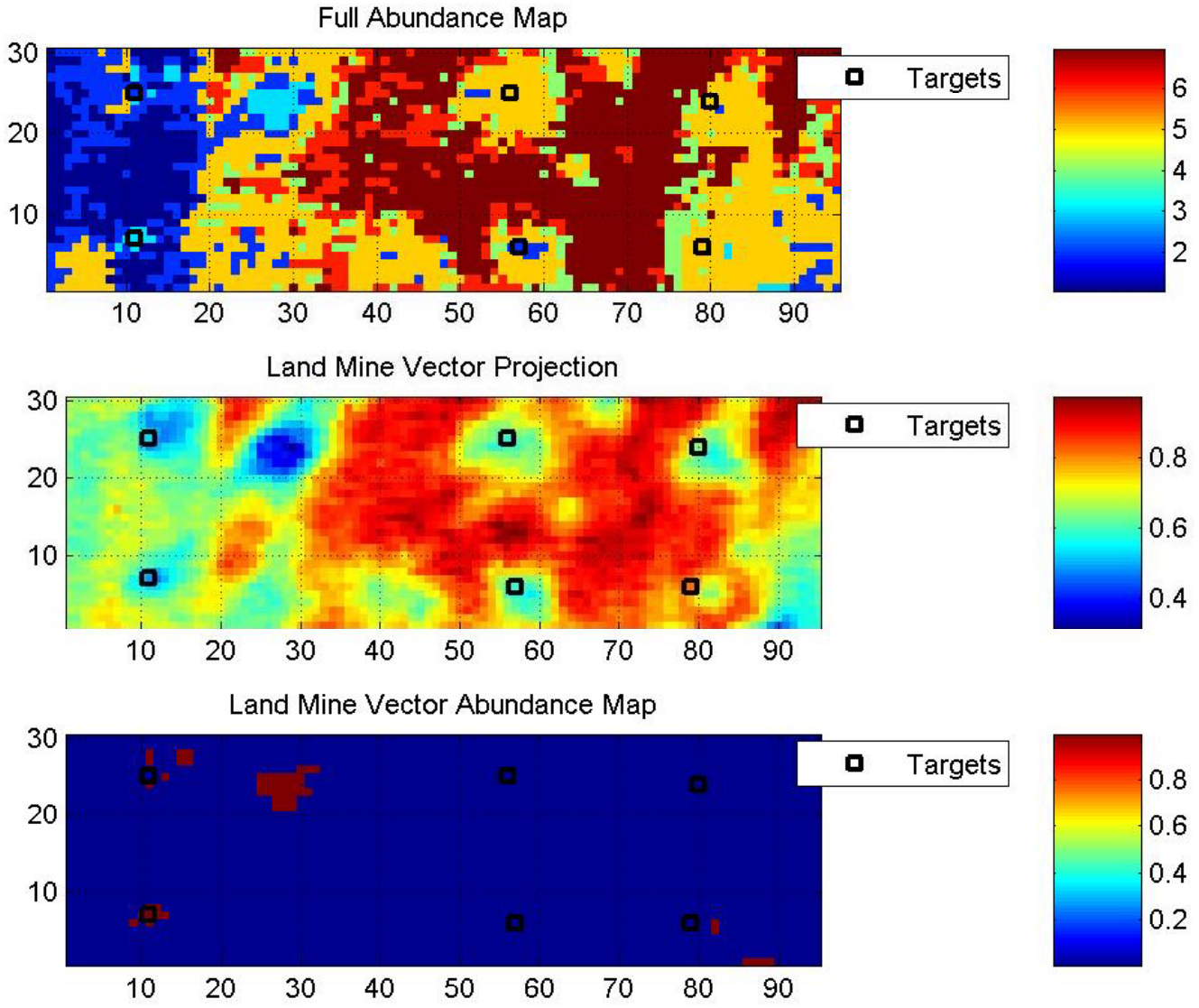
- Calculate end members for entire image
  - N-Finder Algorithm
- Compare every end member to every pixel
  - Use multiple distance calculation metrics
    - Euclidean distance, angular separation, dot product
- Determine most predominant end member in every pixel for each metric
- Create end member abundance map according to each metric
- Employ voting scheme to determine confidence value of pixel classification
- Generate overall abundance map

Winter, M. E., "N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data", Proceedings of SPIE Volume 3753: Imaging Spectrometry V, 10/1999

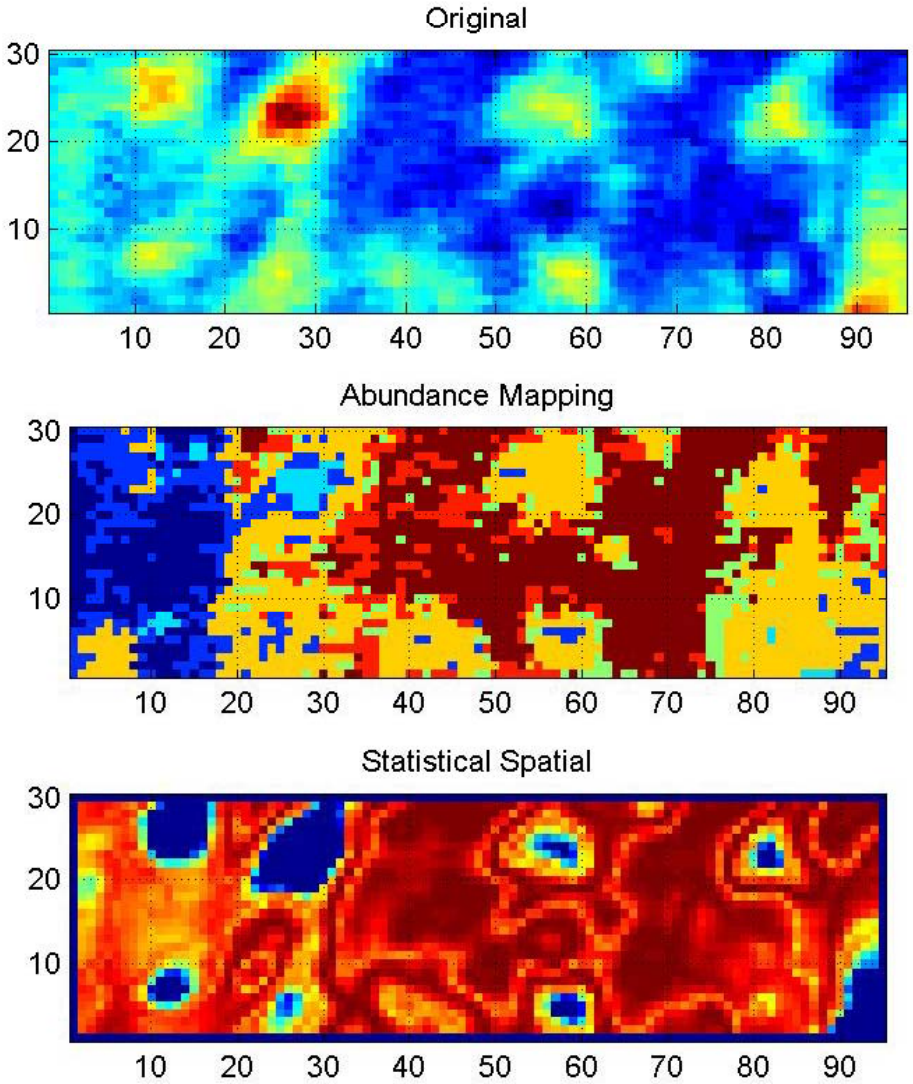
Voting Scheme

All 3 metrics confirm the same top end member	Confidence Level 1
Any 2 metrics confirm the same top end member	Confidence Level 2
All 3 metrics confirm a match between any of the 2 top end members	Confidence Level 3
Any 2 metrics confirm a match between any of the top 2 end members	Confidence Level 4
All 3 metrics confirm a match between any of the top 3 end members	Confidence Level 5
Any 2 metrics confirm a match between any of the top 2 end members	Confidence Level 6

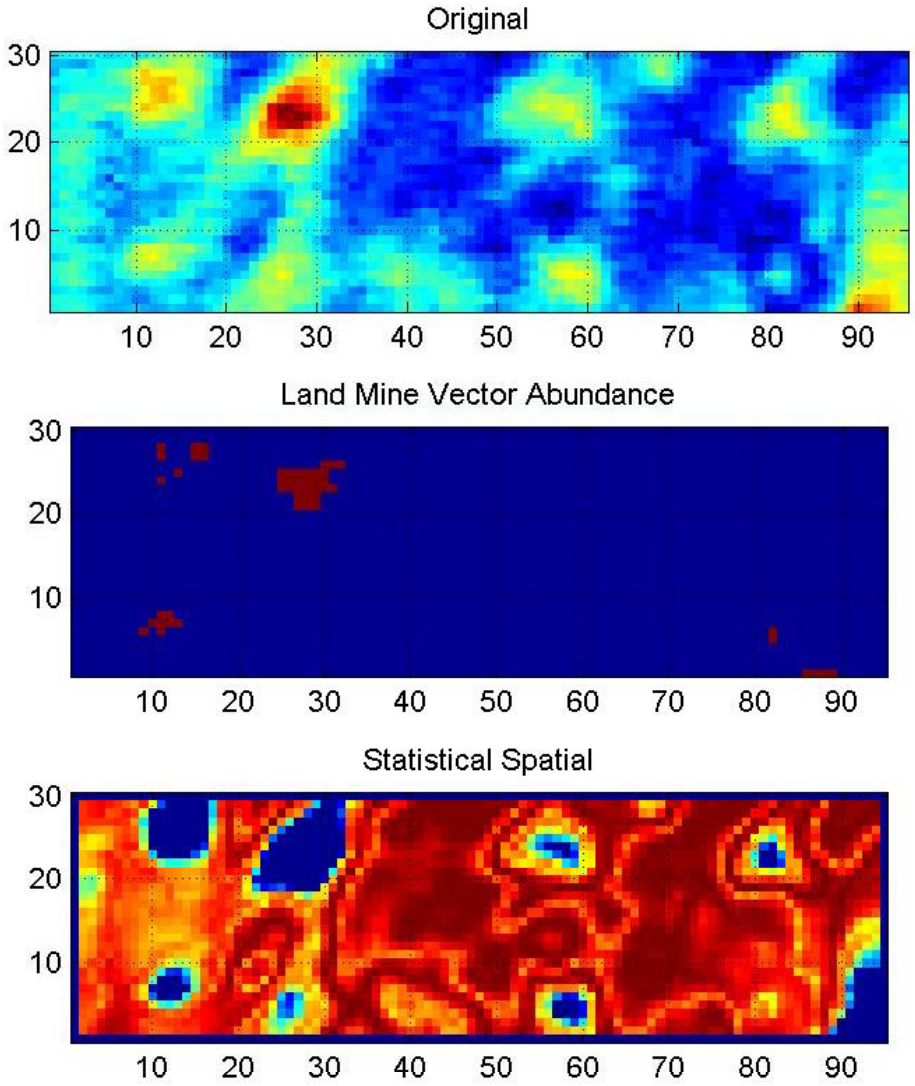
Abundance Mapping Results



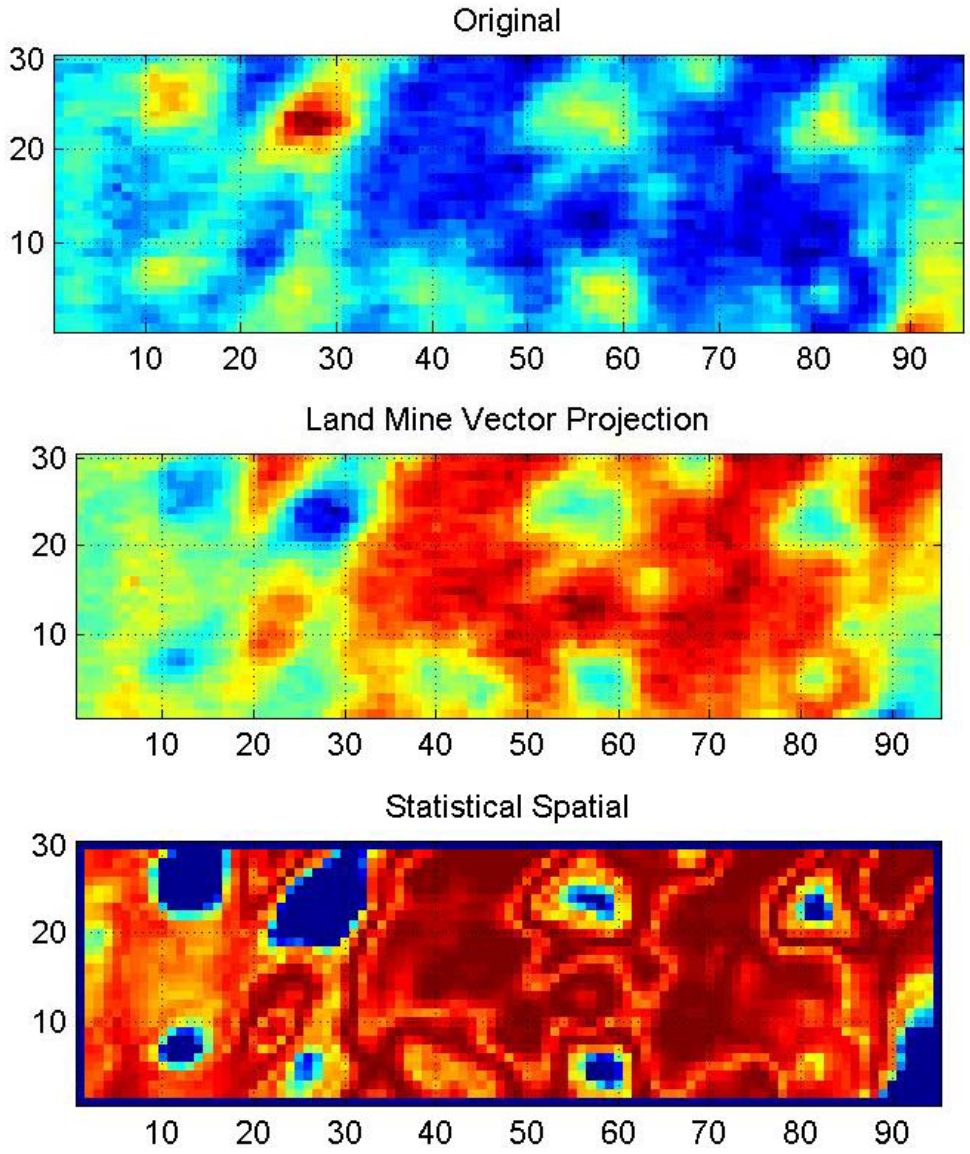
Comparison of EMA and SSA



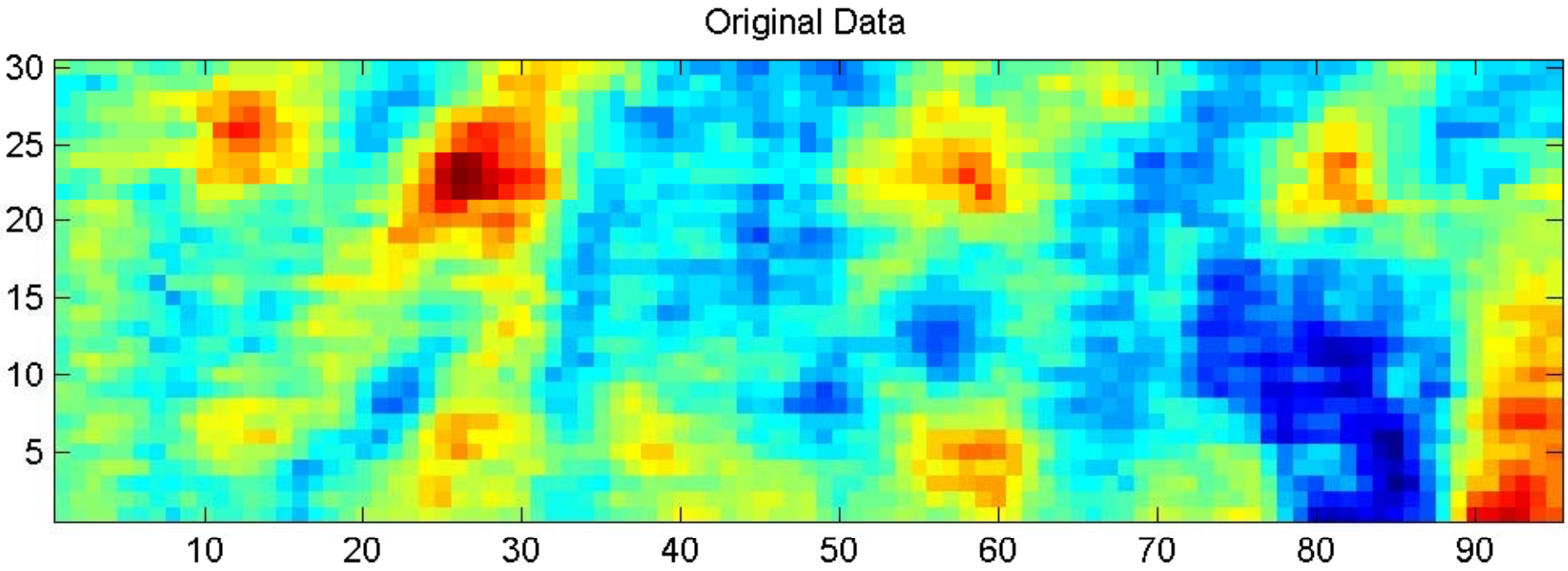
Comparison of EMA and SSA



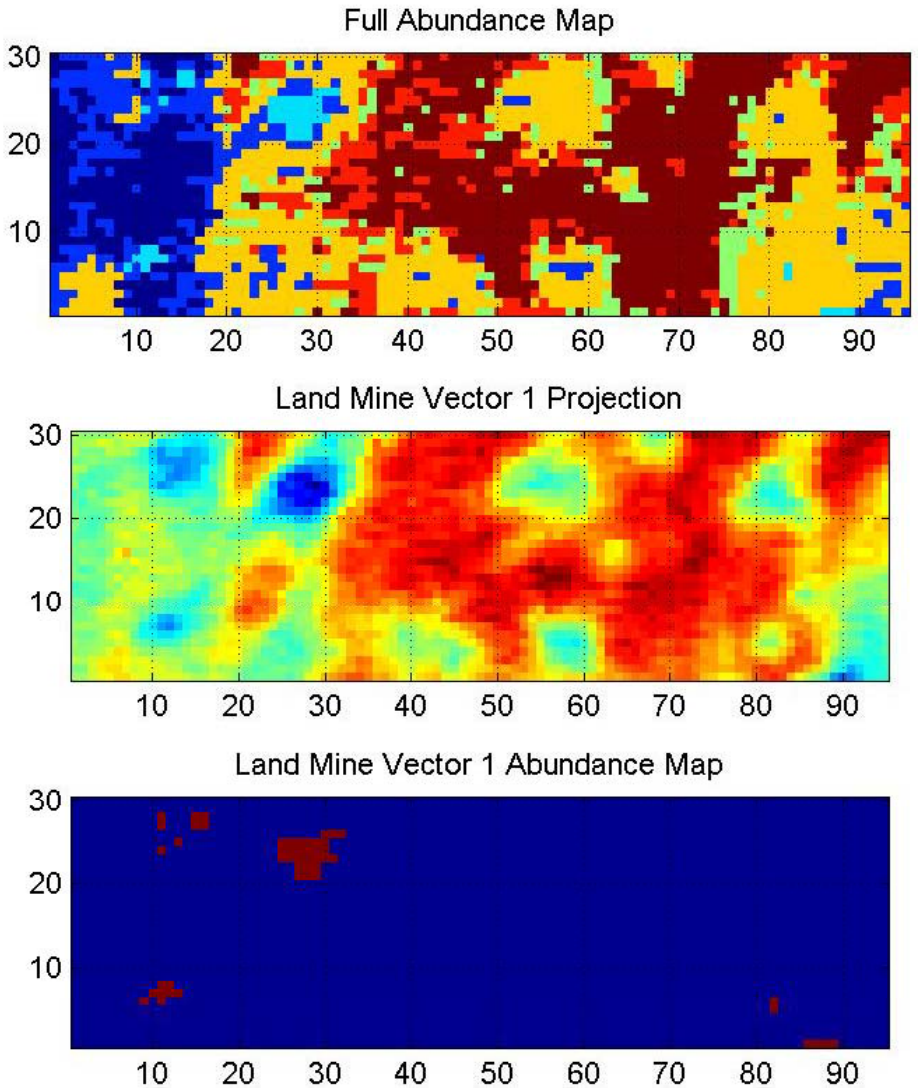
Comparison of EMA and SSA



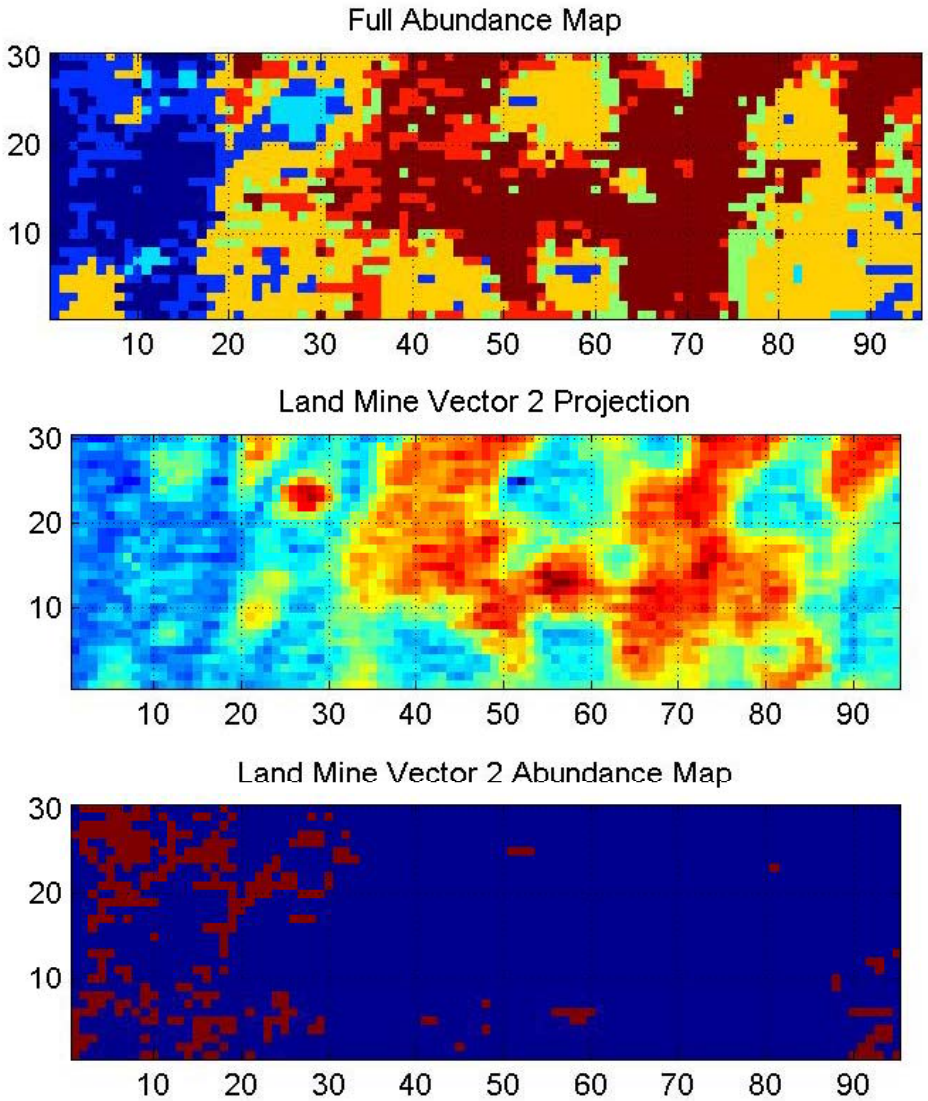
Vegetation Rich Region



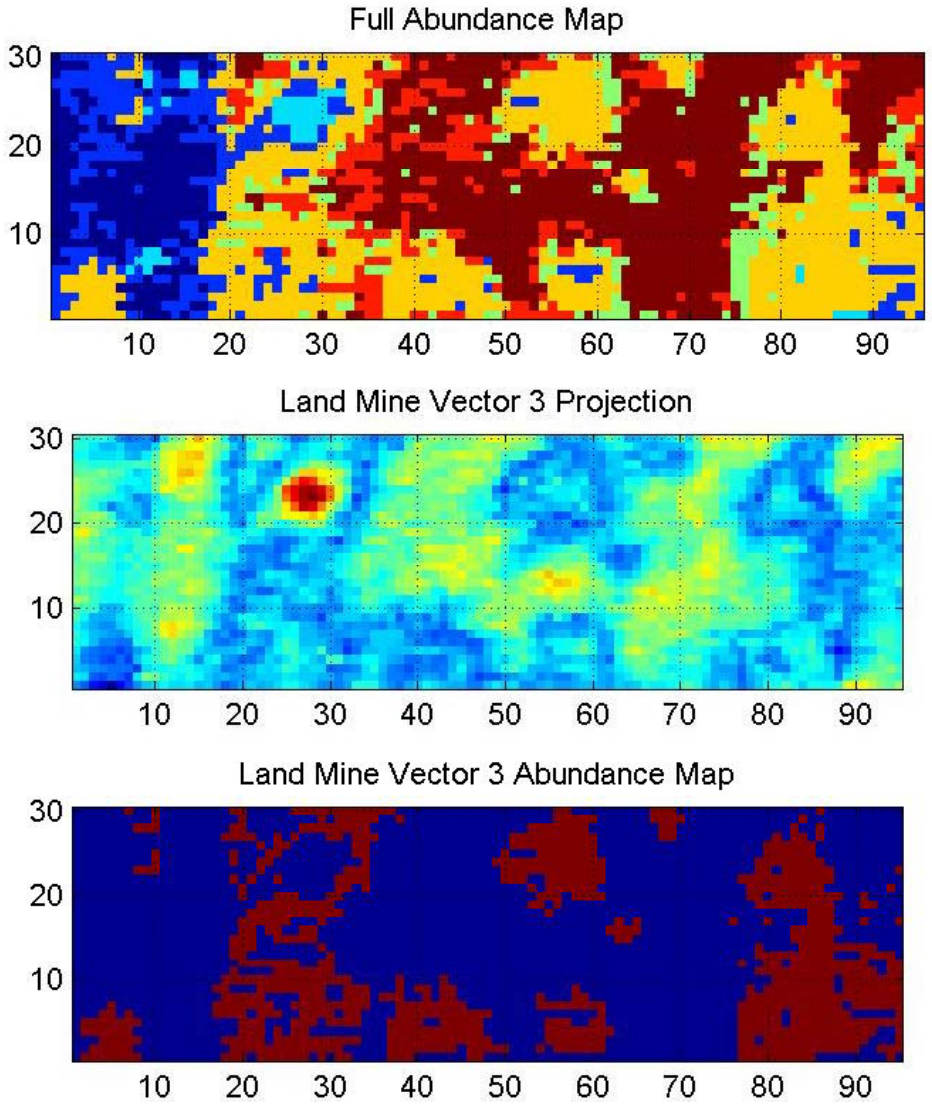
Multiple Land Mine Vector Consideration



Multiple Land Mine Vector Consideration



Multiple Land Mine Vector Consideration



Result Analysis

- **SSA Advantages**
  - Greater (improved) performance in vegetation rich environments
  - False target identification independent of false target abundance
  - Ability to identify the same targets in multiple soil types
  - Less vulnerability to masking and background variation effects
  - Background variation and band count are now coupled to accuracy independently
- **SSA Disadvantages**
  - Considerably longer (increased) processing time
  - Dependence on human interaction
  - No significant accuracy increase in open (relatively homogeneous) environments

Summary

- The SSA has similar performance to EMA in most situations but requiring greater processing time and resources
- SSA demonstrates greater performance in vegetation rich and anomaly rich scenes
- SSA is best used as a background suppression and false target reduction tool in most cases
- The fusion of SSA and EMA offers the potential for a more accurate but still computationally efficient algorithm (approach)

Additional Spectral Analysis of Terrain Infrared Signatures

PI: Bryce Remesch & Michael Cathcart

Problem Statement

- Develop an understanding of spectral variations observed in HS imagery
- Develop soil classification techniques
  - Spectral clustering
  - Vector space characterization

Technical Approach

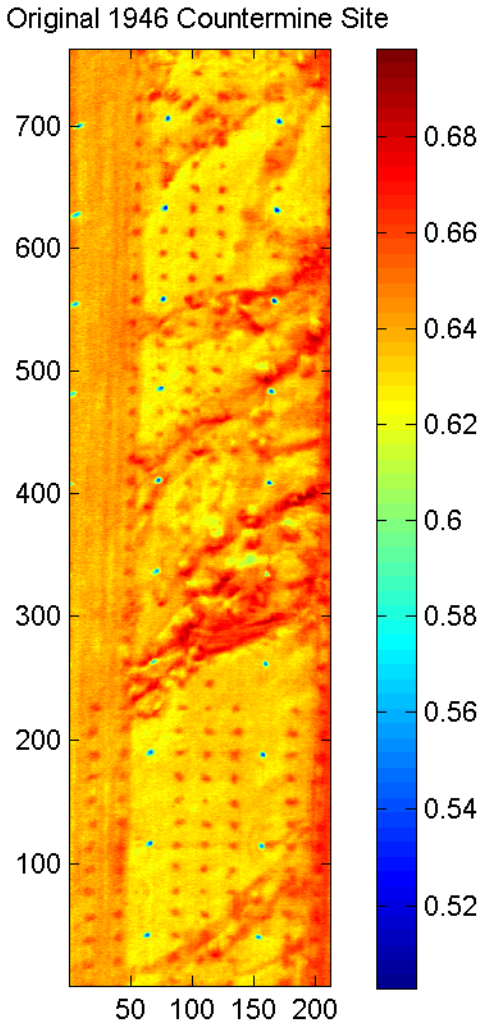
- Application of end-member analysis for identification of background regions
- Employ a combination of phenomenology and statistics to develop end-member derivation approaches
- Test each approach using spectral and spatial metrics (e.g., EM contribution mapping)

End Member Analysis

- Typical end member analysis approaches operate on image sizes of ~1000 pixels (20,000 in one case)
- WAAMD Scenes
  - Typical scene contains ~100K-200K pixels
  - Number of pixels leads to ~1E250 possible sets (choose sets of size 70 out of a population of 100K)
  - Even choosing sets of 7 out of the population leads to ~2E31 possible sets
  - Standard combinatoric approach to EM determination would require unrealistic computational time
- Thus, a smaller sample size is chosen for analysis
- Three selection methods used for choosing the image samples employed during the EM analyses
  - Random pixel sampling over entire scene
  - Selected sampling of known ground truth locations
  - Monte Carlo sampling to identify sub-regions within the scene

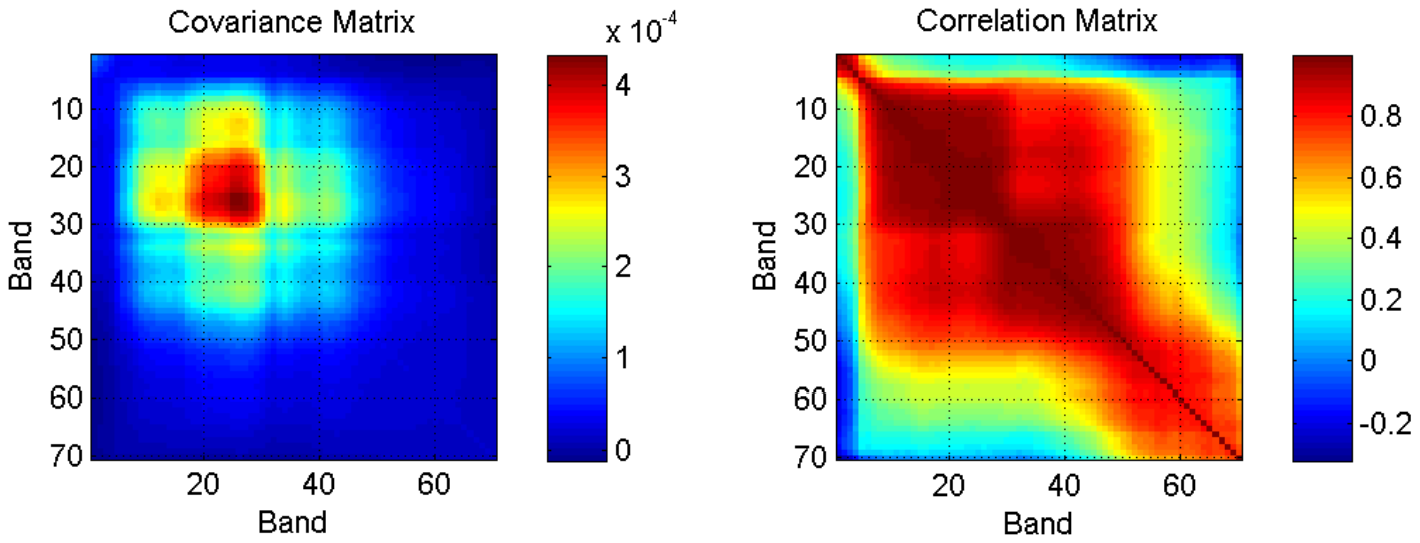
Image Data Characteristics

- End member analysis employed against WAAMD data
- Countermine site image number 1946



Distance Metric

- Inhomogeneous distribution of points in vectorspace
- Regions of high spectral correlation
- Appropriate distance metrics must be used to compensate for lack of orthonormality
  - Angular distance is independent relative dimensional variation
  - Mahalanobis distance takes into account relative dimensional variation

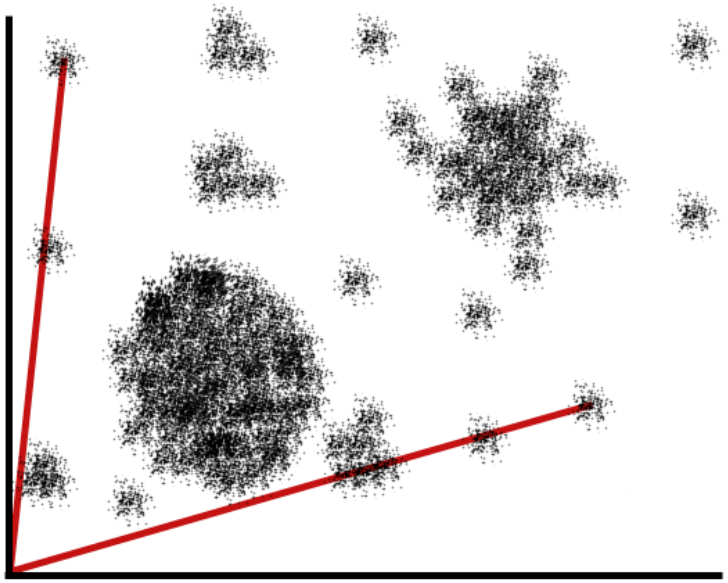


End Member Analysis

- **Multiple algorithms**
  - **N-FINDR**
  - **Greatest End-Member Separation (GEMS)**
  - **Least End-Member Distance (LEMS)**
  - **Discrete Space Characterization (DSC)**
- **Distinct set of end members from each method**

N-FINDR

- Developed by Michael Winter et al.
- Calculates the volume of the n-dimensional space
- Selects the set of end members enclosing the largest volume
- Focuses only on characteristics of EM set
- Ignores density and clustering of space



$$\mathbf{E} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \vec{e}_1 & \vec{e}_2 & \dots & \vec{e}_l \end{bmatrix}$$

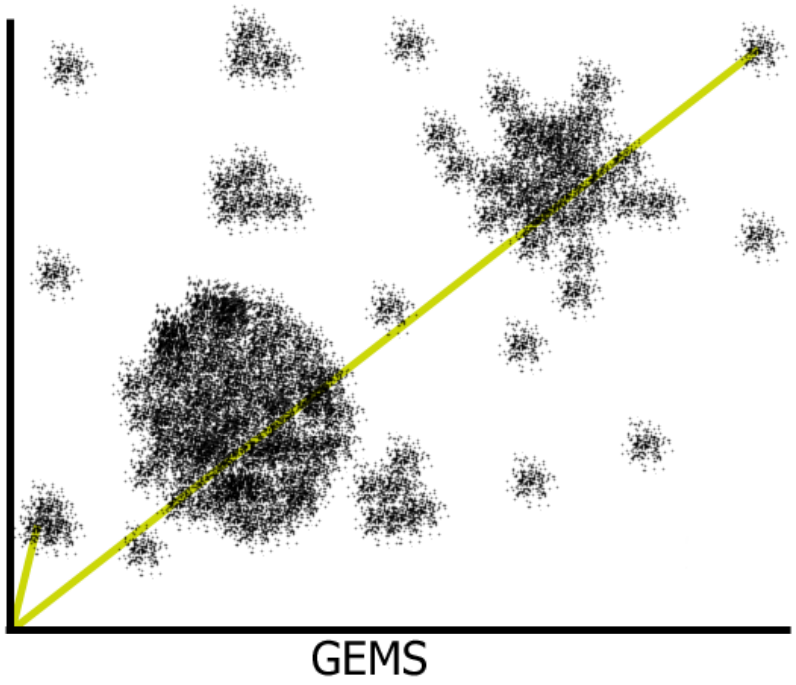
$$V(\mathbf{E}) = \frac{1}{(l-1)!} abs(|\mathbf{E}|)$$

$\vec{e}_l$ = column vector containing spectra of end member l

Dimensionality $= (l-1)$

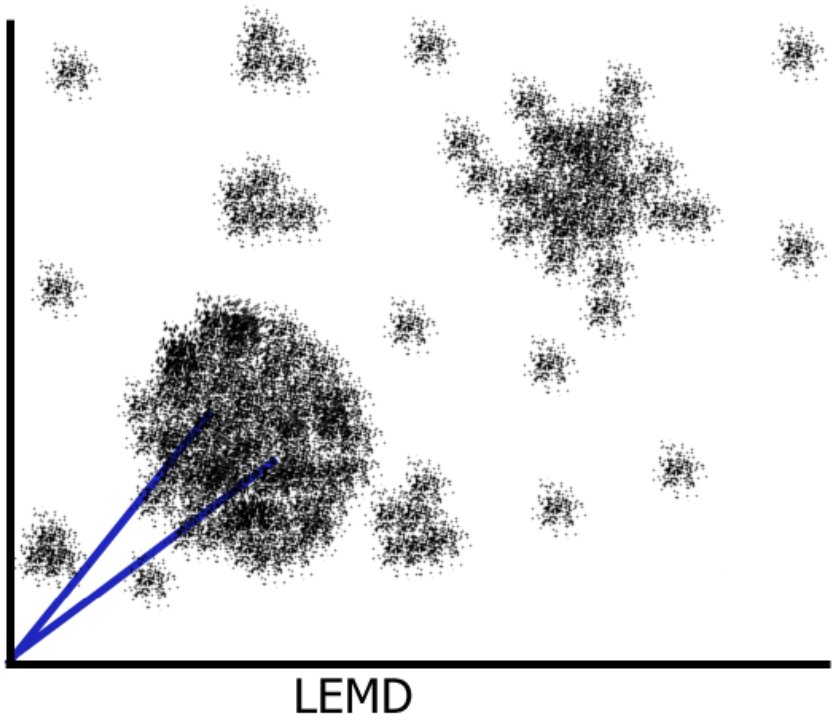
Greatest End-Member Separation (GEMS)

- Calculates the Mahalanobis distance between each end member and the rest of the rest of the end member population
- Selects the set with the largest mean distance
- Focuses only on characteristics of EM set
- Ignores density and clustering of space



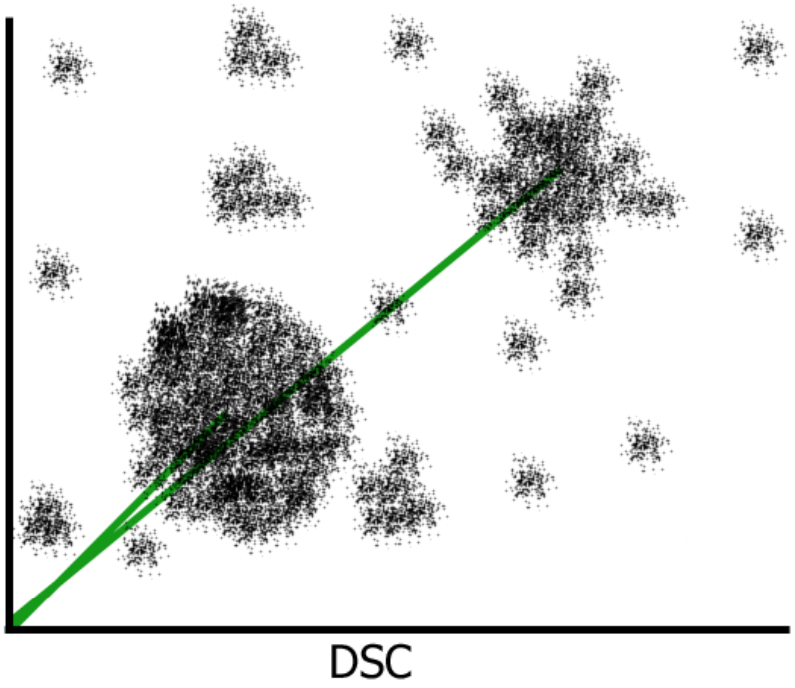
Least End-Member Distance (LEMD)

- Calculates the Mahalanobis distance between each end member and the rest of the scene
- Requires that at least 50% of the distances are smaller
- Selects the set with the smallest mean distance from the image
- High resolution for large or dense clusters,
- Low resolution for outlier



Discrete Space Characterization (DSC)

- Combines GEMS and LEMD
- Calculates the distance between every pair of end-members
- Calculates the distance between every end-member and every pixel
- Tests that at least half of the end members are further apart from each other and at least half of the end members are closer to the pixels
- At least half of the EM must be replaced in order to generate a new EM set



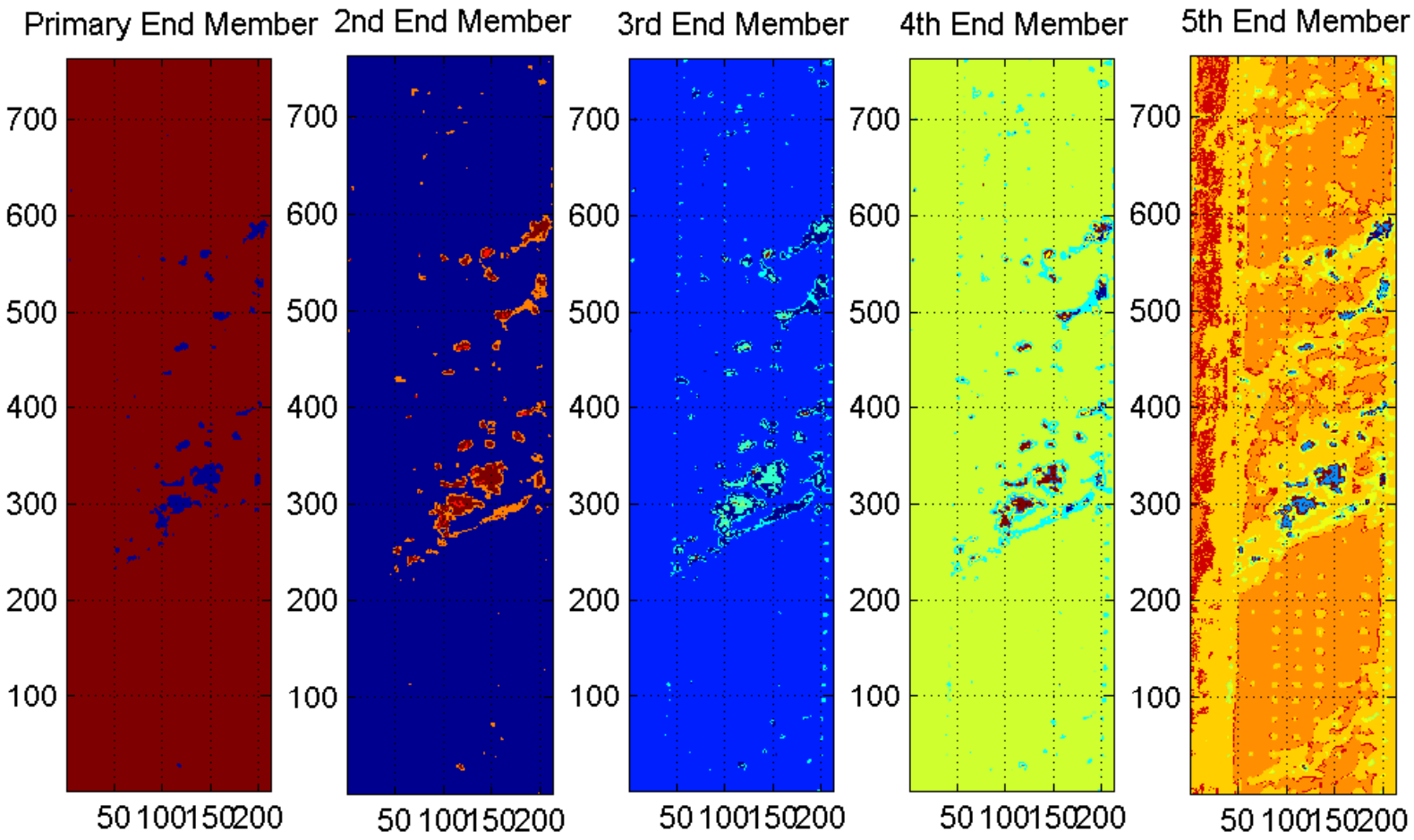
End Member Algorithm Comparison

- Each method generates 70 end members for a total of 280 EM
- De-population of this larger set by elimination of common EM
- Generates a 5^{th} set of EM
- Contains a total of 280 EM, ~100 of which are unique

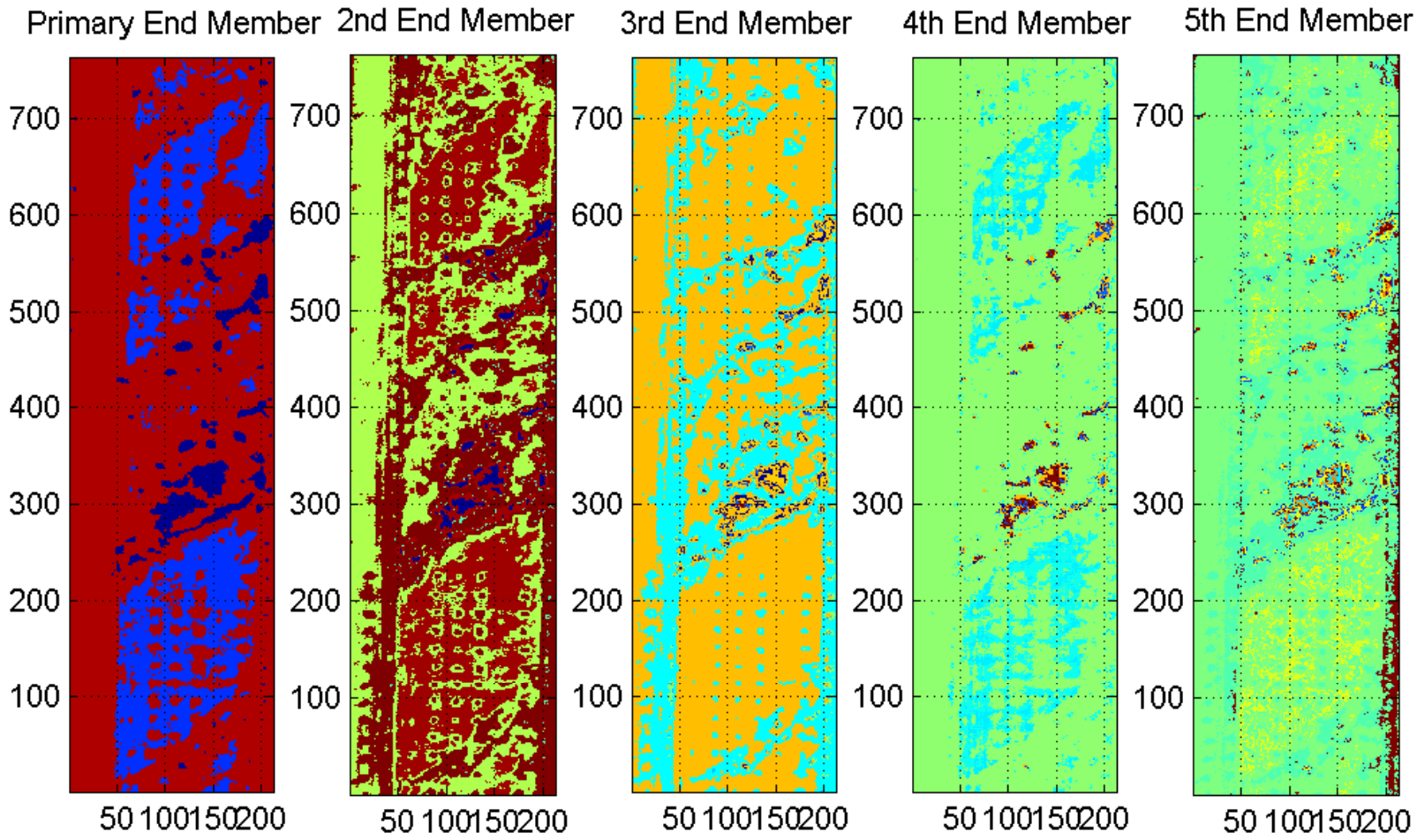
Contribution Mapping

- Angular distance between every end member and every pixel
- Pixel is classified based on its distance to each of the end members
- A pixel's primary contributor is it's closest end member, it's secondary contributor is its is second closest end member, etc.
- Pixel classification becomes more refined as more end members are considered
- Contribution order variation shows variability not only between classes but also within classes.

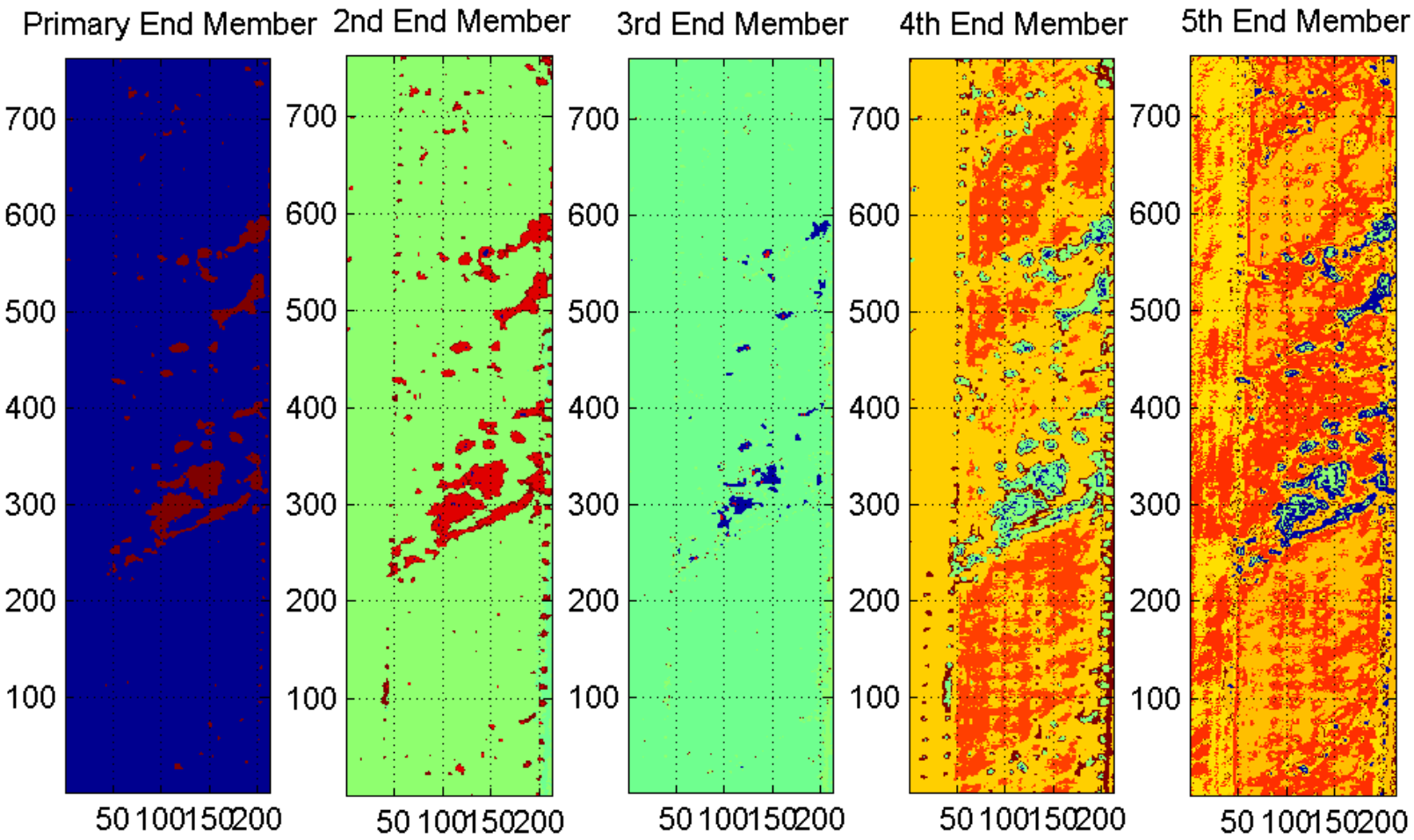
NFINDR


Primary End Member
2nd End Member
3rd End Member
4th End Member
5th End Member
700
600
500
400
300
200
100
50 100150200

LEMD


Primary End Member
2nd End Member
3rd End Member
4th End Member
5th End Member
700
600
500
400
300
200
100
50 100150200

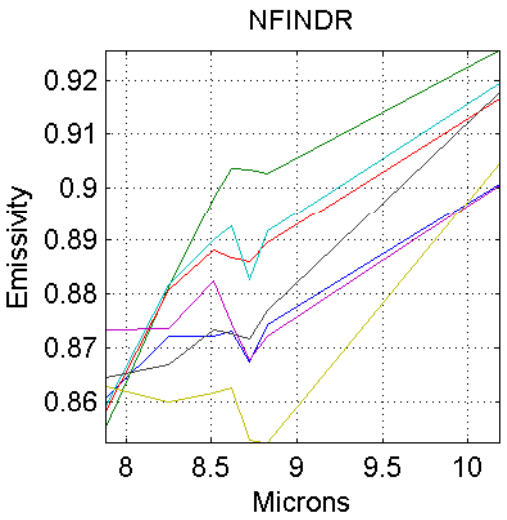
GEMS

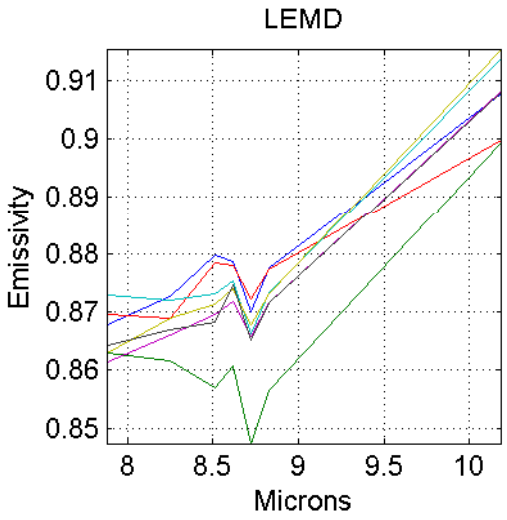

Primary End Member
2nd End Member
3rd End Member
4th End Member
5th End Member
700
600
500
400
300
200
100
50 100150200

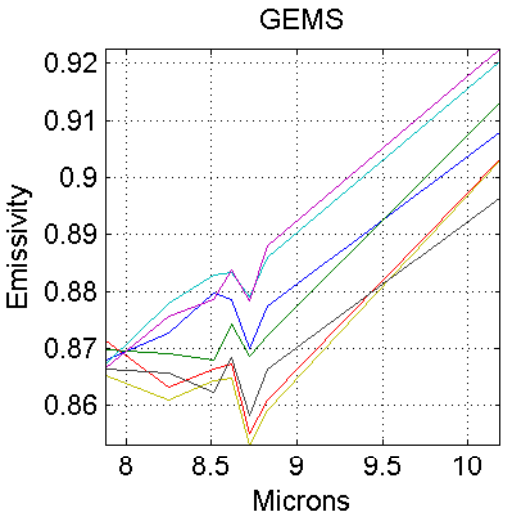
Dimension Reduction

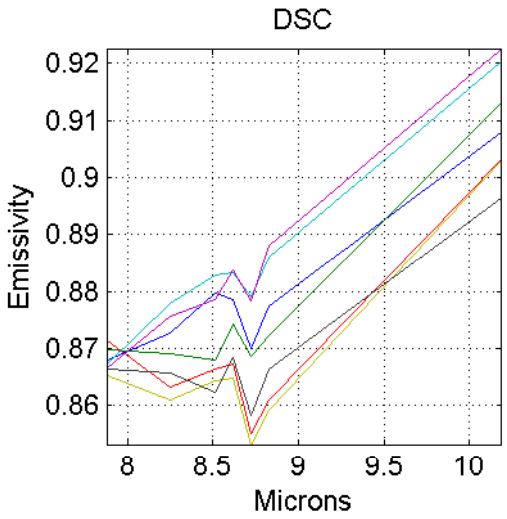
- 70 bands span ~7-11 microns
  - Not all bands contain distinguishable spectral features between the various materials
- Three approaches identified for examining dimensional reduction
- Phenomenological approach
  - Specific spectral band corresponding to undisturbed/disturbed soil
- Mathematical approach
  - Perform PCA on the entire scene
  - Select the 3 highest eigenvalue components
  - Extract spectral emissivity minima and maxima for each component
  - Reduce image dimensions by only using those bands corresponding to the extracted minima and maxima
    - Dimensions reduced to a range of 7 – 12 bands.
- Previous end-member algorithms applied to this reduced set

Disturbed Soil Feature Reduction

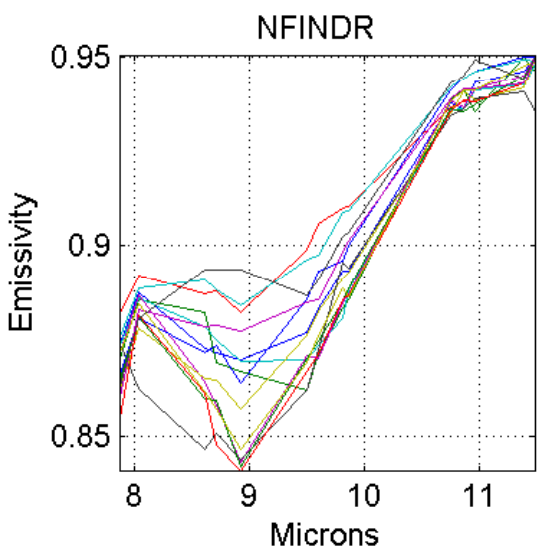


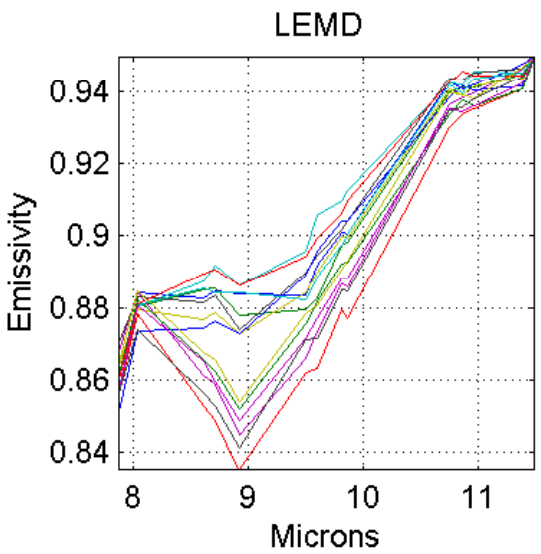


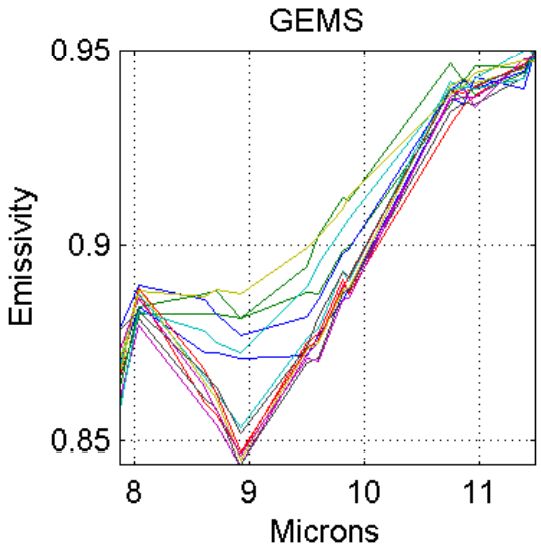


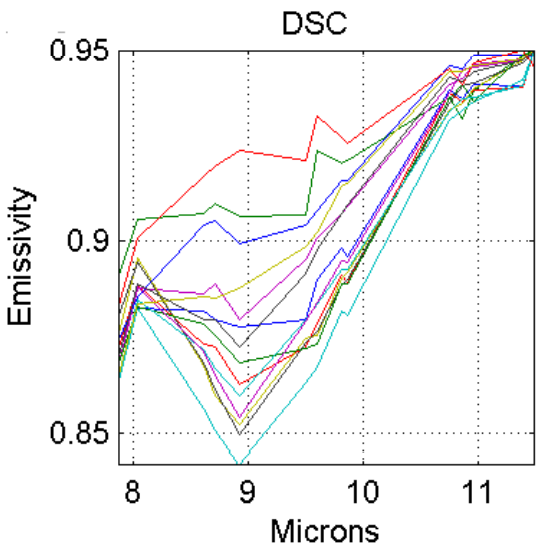


Minima Reduction

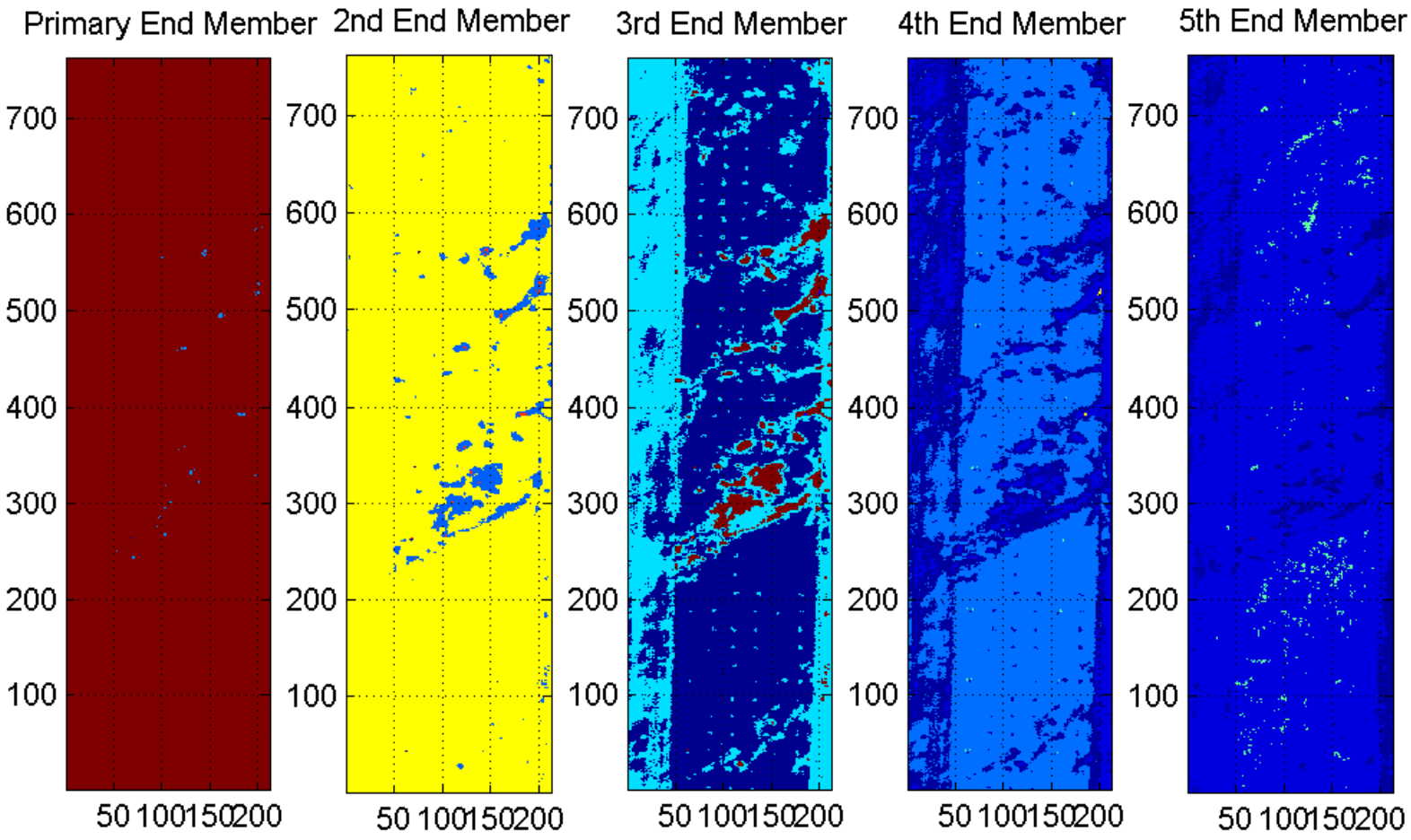




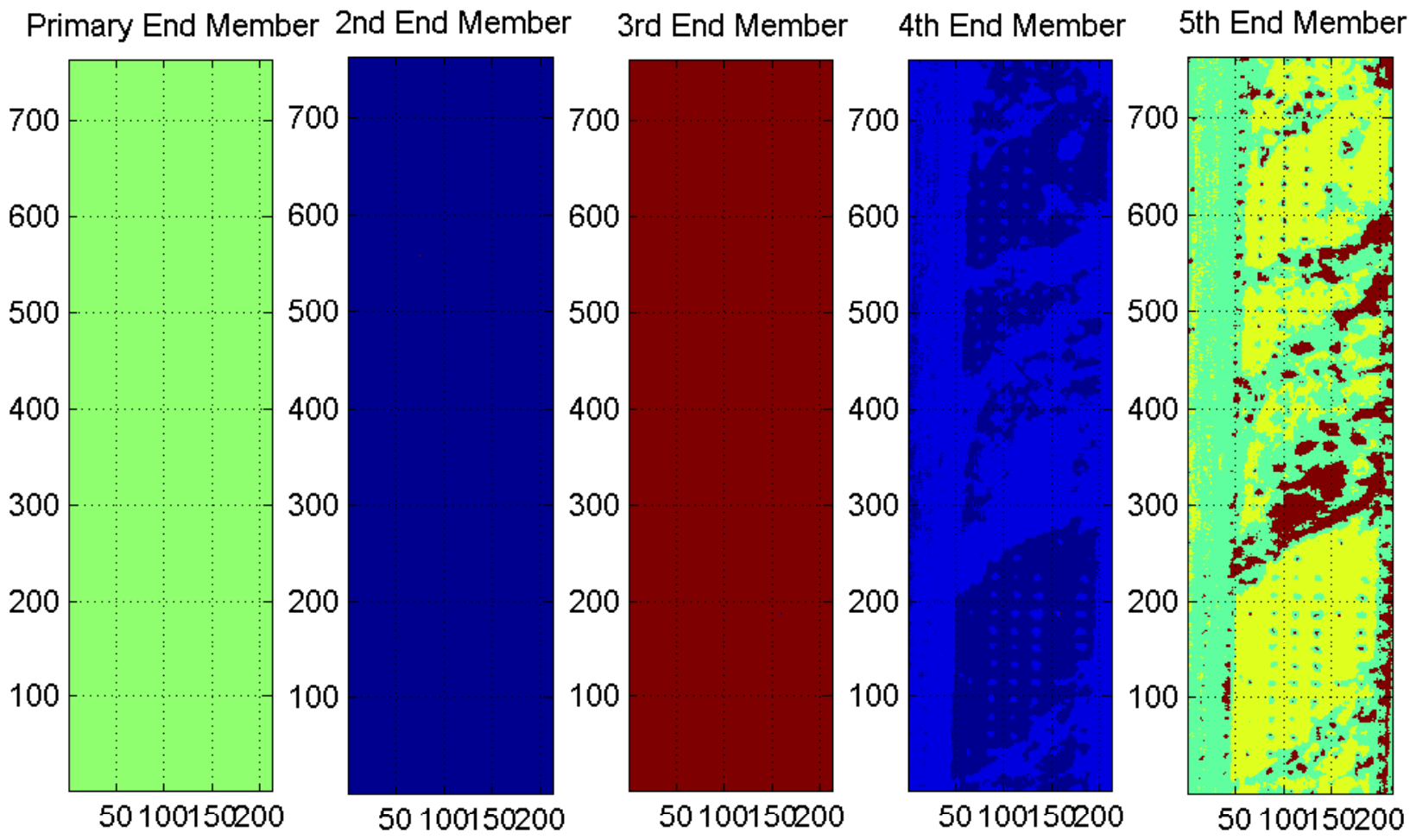




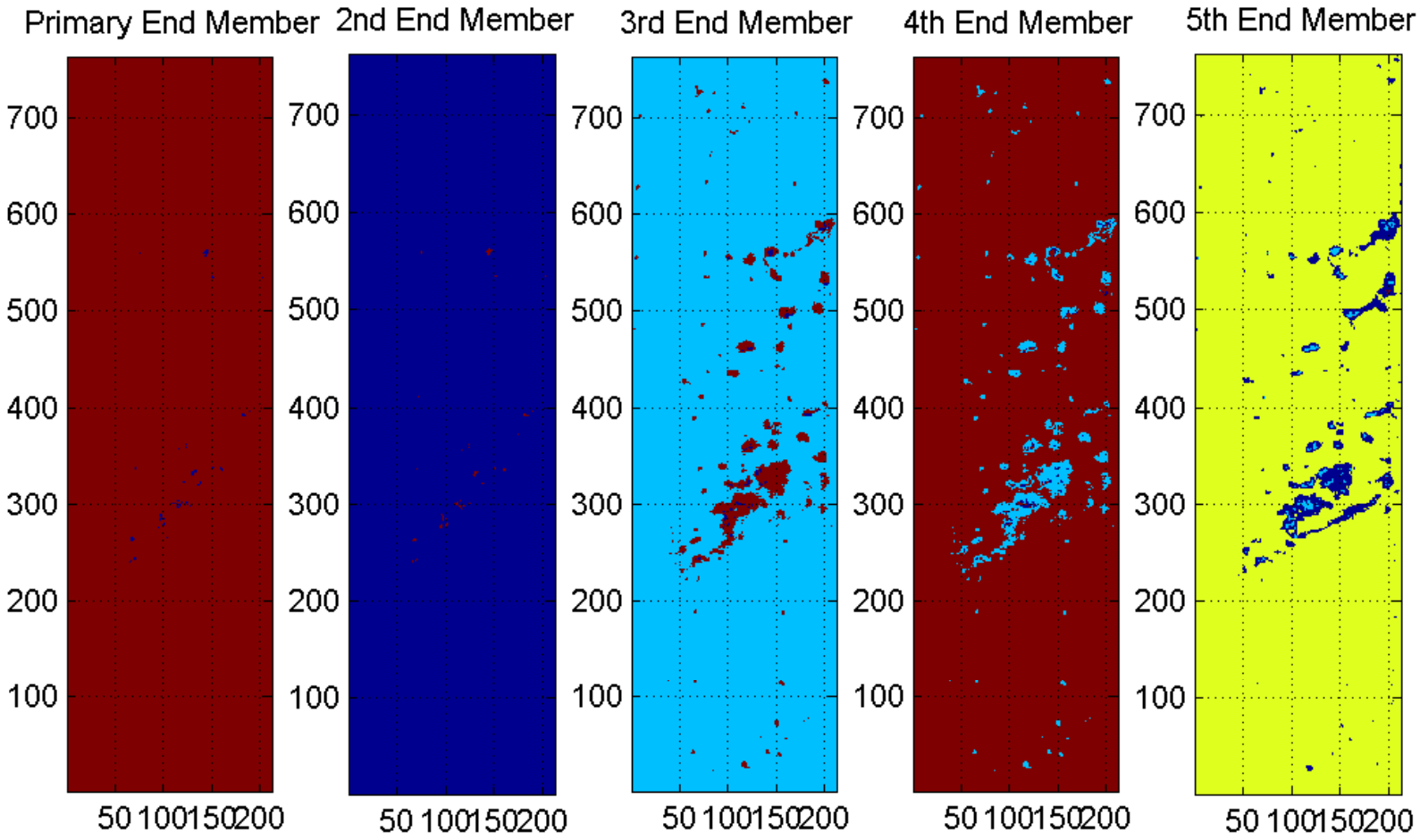
NFINDR
Minima Reduction



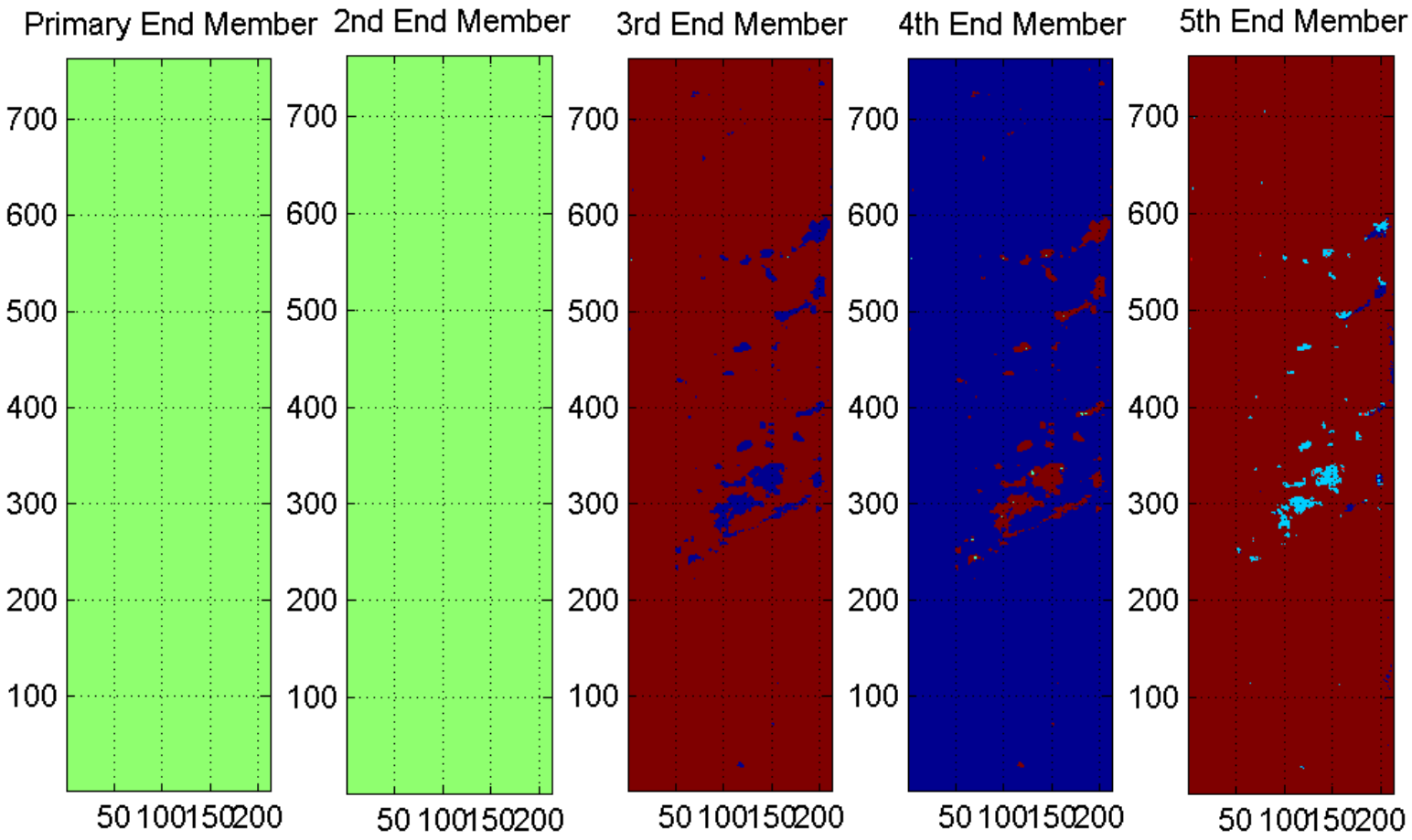
LEMD
Minima Reduction



GEMS
Minima Reduction



DSC
Minima Reduction



Summary

- Atmospheric Correction
- Random Image Sampling
- Multi-algorithm EMA
  - N-FINDR, LEMD, GEMS, DSC, Combined
- Dimensional Reduction
- Contribution Mapping

Hyperspectral LWIR model for studying landmine-soil interactions

PI: Ricardo Campbell, Sarah Greenwood, & Michael Cathcart

Digital Modeling – Overview

- **Objectives**
  - Analyze soil-mine interactions & dependencies
  - Develop detailed approaches for high spatial resolution M&S
  - Focus on M19 landmine
- **Model requirements**
  - High spatial resolution (cm level detail)
  - Detailed mine model (inclusion of internal components)
  - Differing soil properties next to mine (disturbed soil)
  - Spatial and time varying soil properties
  - Spectral signature generation
- **Modeling resources**
  - GTSIG; MATLAB; GTRENDER; MS EXCEL

Digital Modeling – Tasks

- Develop detailed thermal model for soil & mine
- Perform verification studies on digital model
- Generate optical signatures (initially LWIR)
- Develop spectral signature process
- Analyze impact of H_2O & porosity on soil properties & thermal signature
- Analyze disturbed soil properties & thermal signatures (ageing)
- Analyze effect of environment (soil composition, ageing, etc.)
- Develop requirements for high spatial resolution simulation (e.g., level of fidelity)

Landmine Phenomenology Study

- **Model classes**
  - Landmine (M19) – surface, flush, buried
  - False target (can/pipe) – surface, flush, buried
- **Initial studies focus on**
  - LWIR signature calculations
  - Spectral signature calculations
- **Verification & validation process**
  - 'Repeating' weather data
  - Comparison to field data
- **Correlation to other efforts**
  - Field data analysis
  - Polarization modeling
  - Analytical studies on soil properties

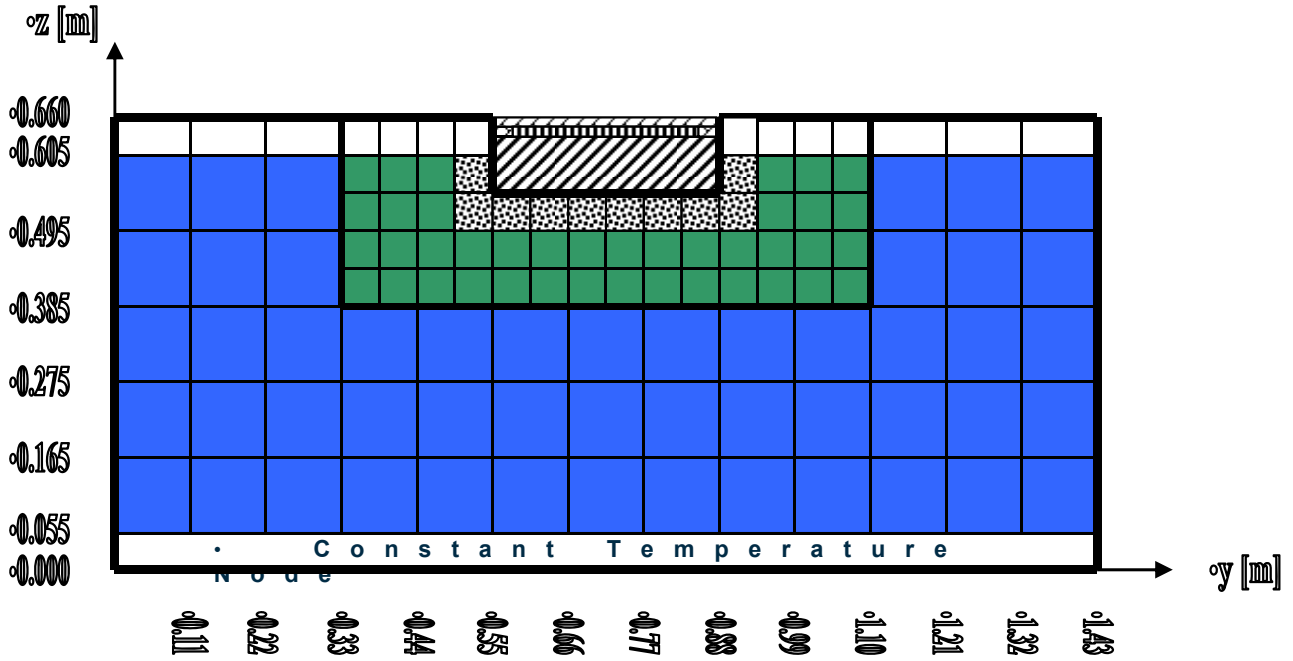
M19 Landmine

- Anti-tank mine
- 33 x 33 x 9.4 cm^3
- Low metal content
- Composition B Explosive

Thermal Modeling – Overview

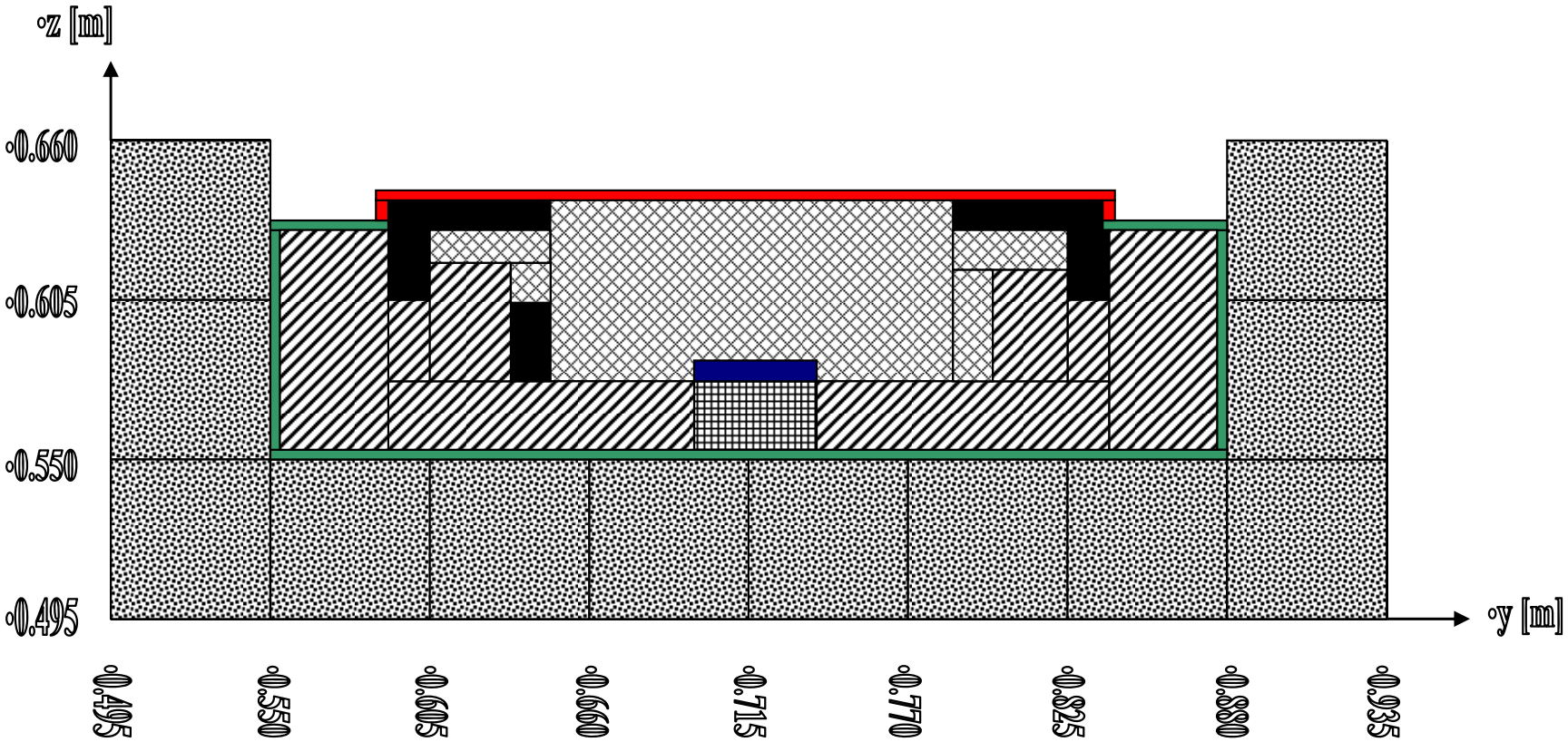
- Infrared model designed to study interaction between soil – atmosphere – landmine environment
- Spatially diverse model:
  - detailed landmine model; flush-buried
  - 'disturbed' soil area
  - surrounding undisturbed soil area
- Physical properties of each section can be adjusted to reflect specific soil conditions
- Accomplishments
  - Verification process performed
  - Signature study conducted (broadband & spectral)
  - Examined environmental dependencies of signatures

Soil – Mine Nodal Structure



- Large interior dry-sand nodes
- Small interior dry-sand nodes
- Small disturbed sand nodes surrounding the sides and bottom of the mine
- Small disturbed sand nodes covering the top of the mine

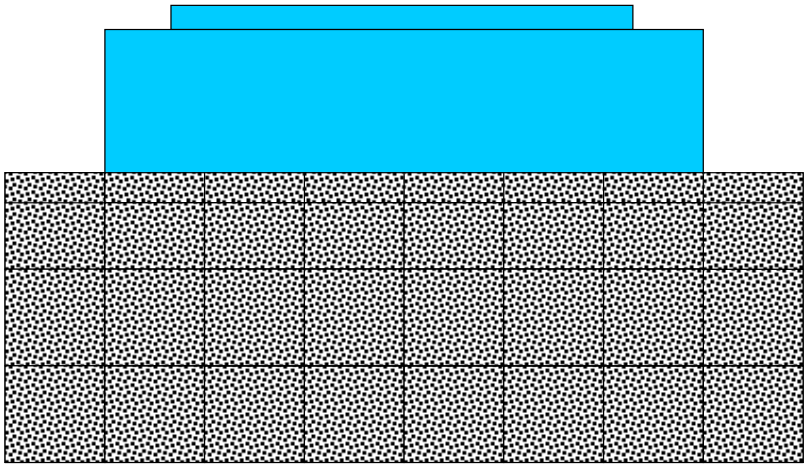
Mine - Interior Nodal Structure



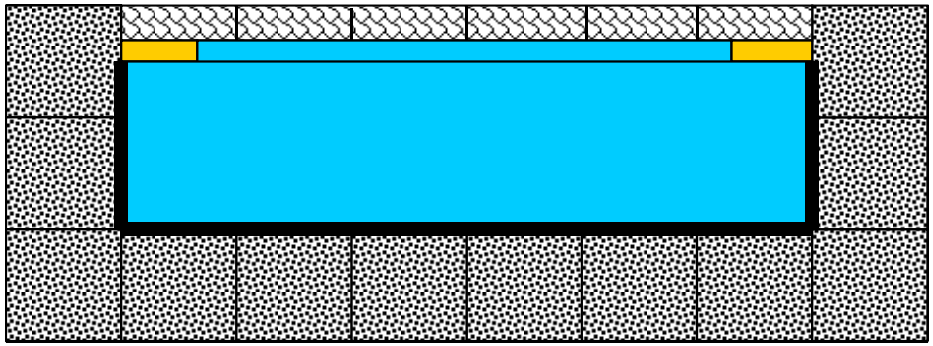
- Mine casing (glass reinforced plastic)
- Pressure Plate
- Composition B
- Interior plastic components
- Metal (copper & stainless steel)
- Air
- Tetryl booster pellet
- Disturbed soil

Landmine Target Modeling

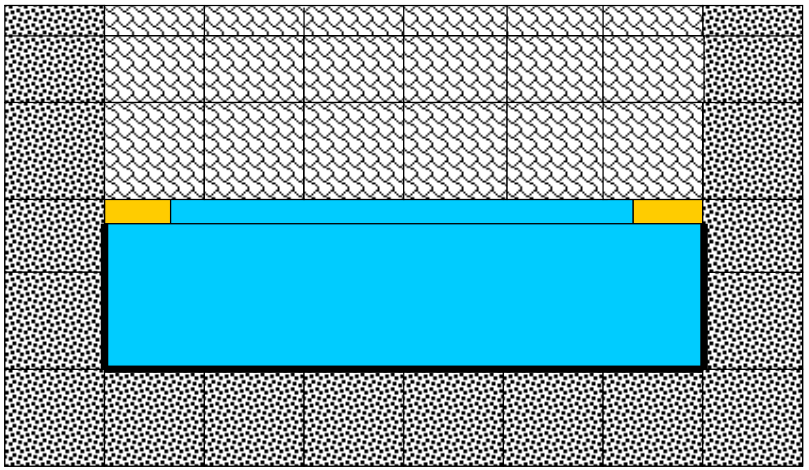
- Soil & target area (1.5 m^3)
- M19 structures
- Geometric resolution 5.5 cm x 5.5 cm
- Vertical resolution 1.6 cm to 5.5 cm
- > 14,000 thermal nodes
- Vertical and horizontal conduction paths



surface laid mine

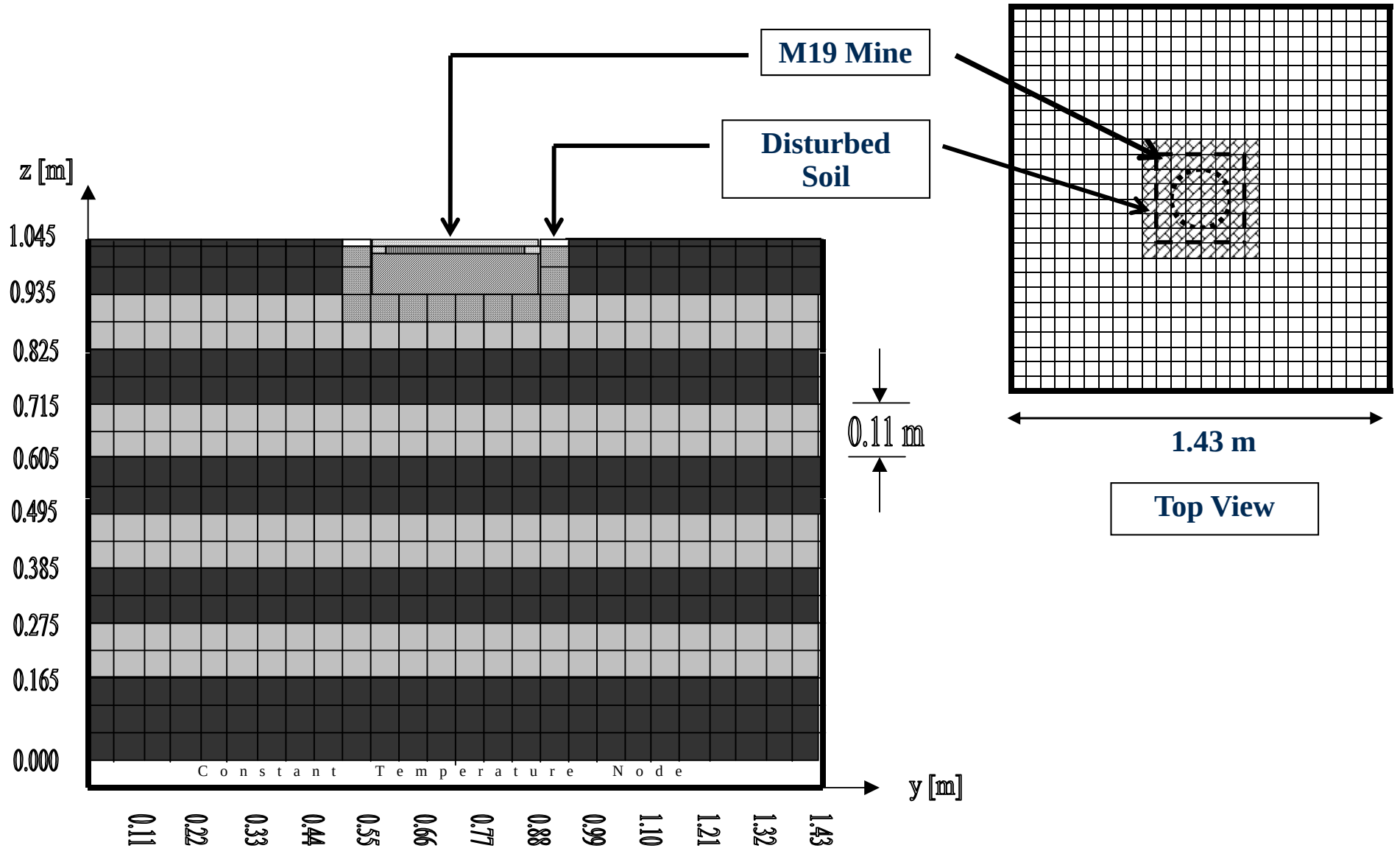


flush buried mine

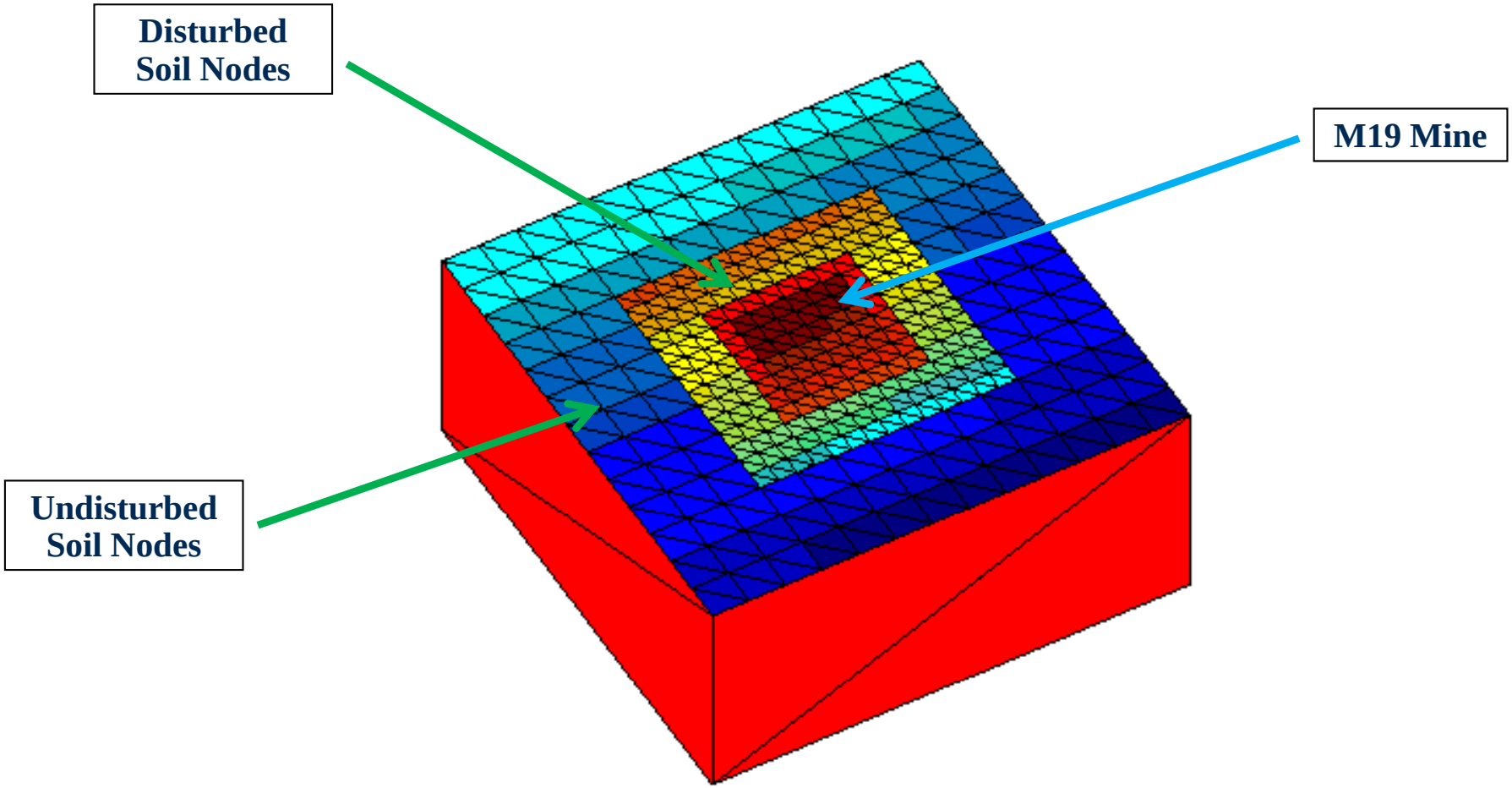


buried mine

Layered Soil Model – Flush buried mine



Soil Surface Structure



Landmine – Soil Model Parameters

Material	Density [kg/m^3]	Specific Heat [J/kg K]	Conductivity [W/m K]
Dry Sand	1520	800	0.33
Disturbed Sand	1140	600	0.165
Air	1.29	1000	0.026
Metal [a]	8397	423	227
Bakelite [b]	1400	1675	0.15
Polystyrene [c]	1040	1170	0.13
Polyethylene [d]	960	1850	0.48
Comp B [e]	1700	1108	0.268
Tetryl [f]	1700	1013	0.263

a – copper and steel	d – internal plastic components
b – pressure plate	e – explosive material
c – mine casing	f – booster

Soil Model Composition

Depth (cm)	Clay (%)	Silt (%)	Sand (%)
0 – 33	17.2	32.4	50.4
33 – 56	18.3	34.5	47.2
56 – 81	29.2	32.9	37.9
81 – 107	38.1	36.2	25.7

Soil Model – Physical Parameters

Layer	Depth	Density	Specific Heat	Thermal Conductivity
	(cm)	(kg/m3)	(J/kg K)	(W/m K)
1 – 7	0 – 33	1774	860	0.273
8 – 11	33 – 55	1790	863	0.269
12 – 15	55 – 77	1838	875	0.292
16 – 20	77 – 105	1900	889	0.299

Environmental Data

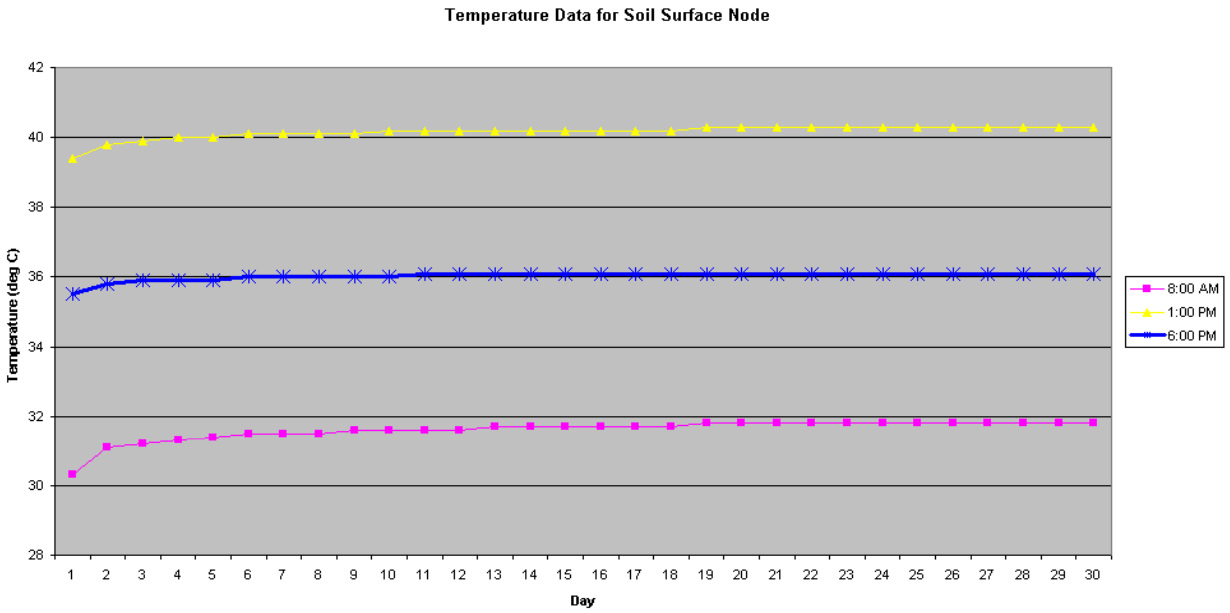
- Signature model requires meteorological data for specific location (solar insolation, air temperature, humidity. etc.)
- Weather data generated for two scenarios
  - Walnut Gulch, AZ; 30 day weather file
  - 'Repeating' weather; 30 day weather
- WAAMD data weather
  - Yuma met. data available
  - Requires conversion
- Soil moisture profiles

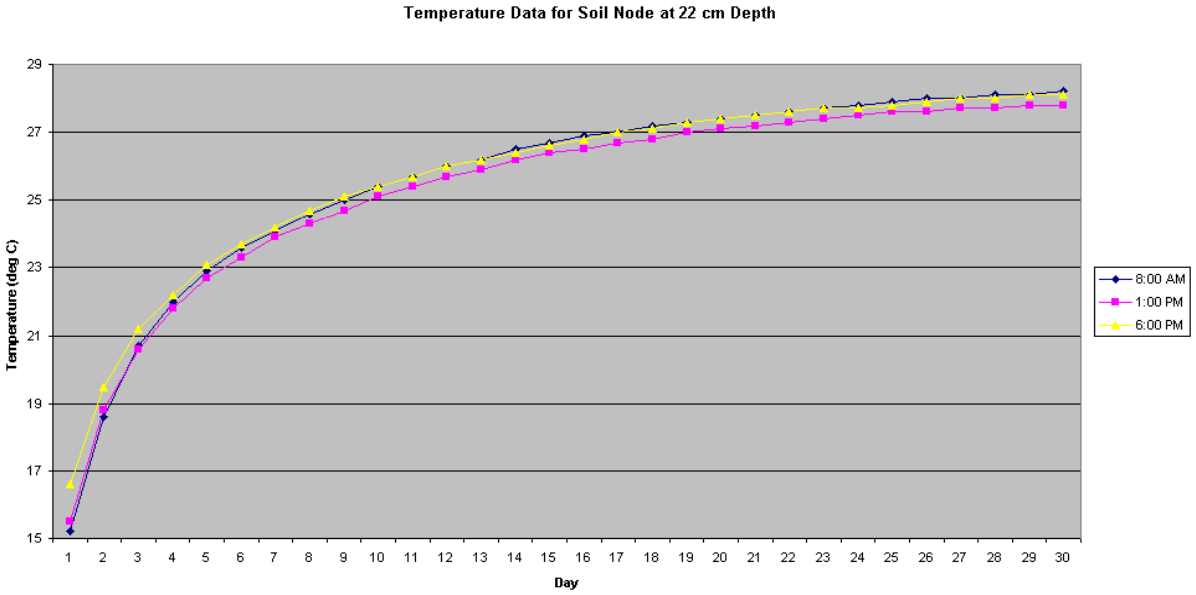
LWIR Signature Generation

- Create extended time weather files
  - Walnut Gulch weather data
- Select calculation time
- Generate input files
- Execute GTSIG
- Extract temperature & radiance data
- Perform qualitative & quantitative comparisons

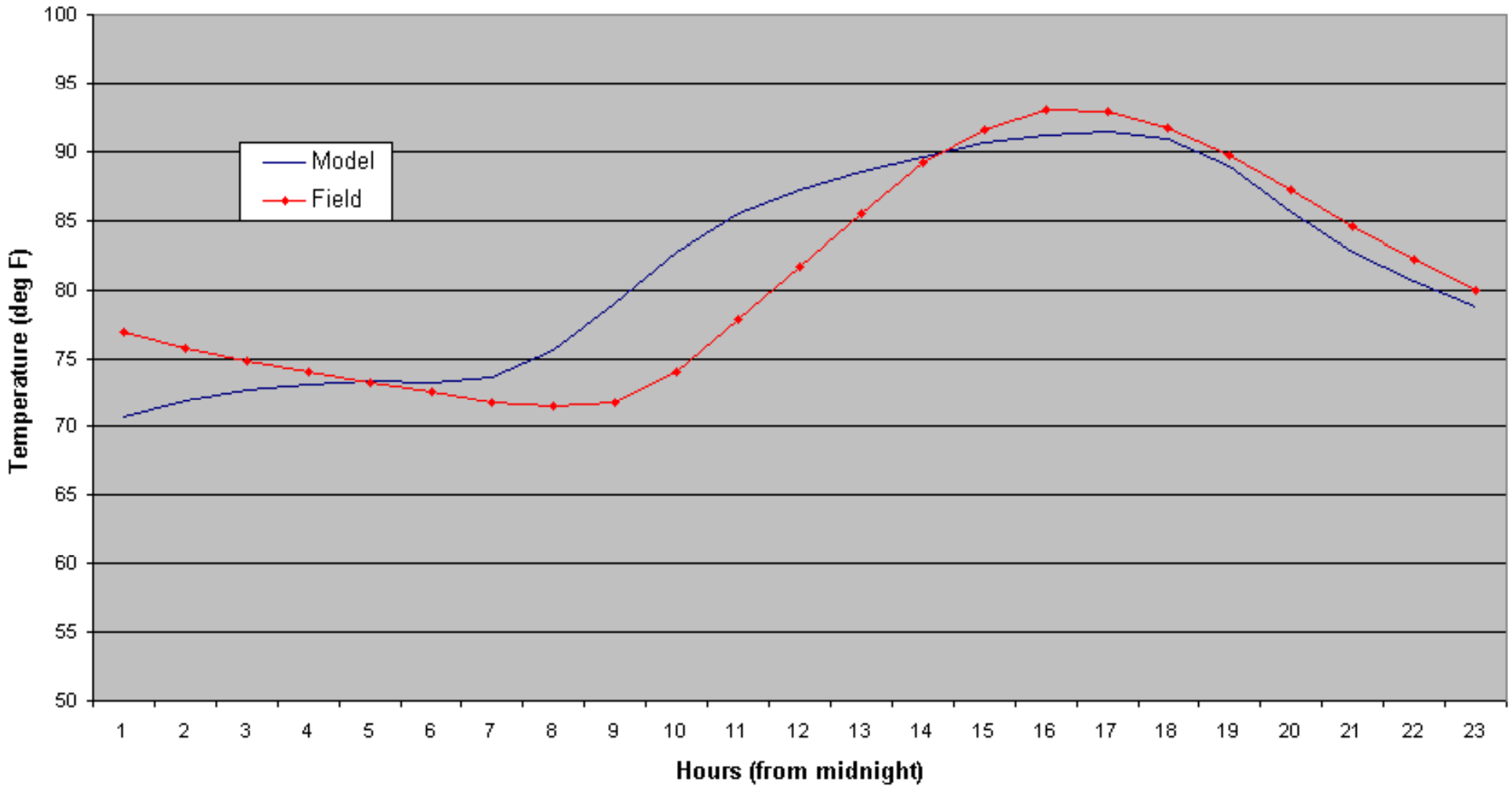
Constant Weather Results

- Test of thermal model
- 30 day repeating weather data set used
- Determine thermal relaxation constants

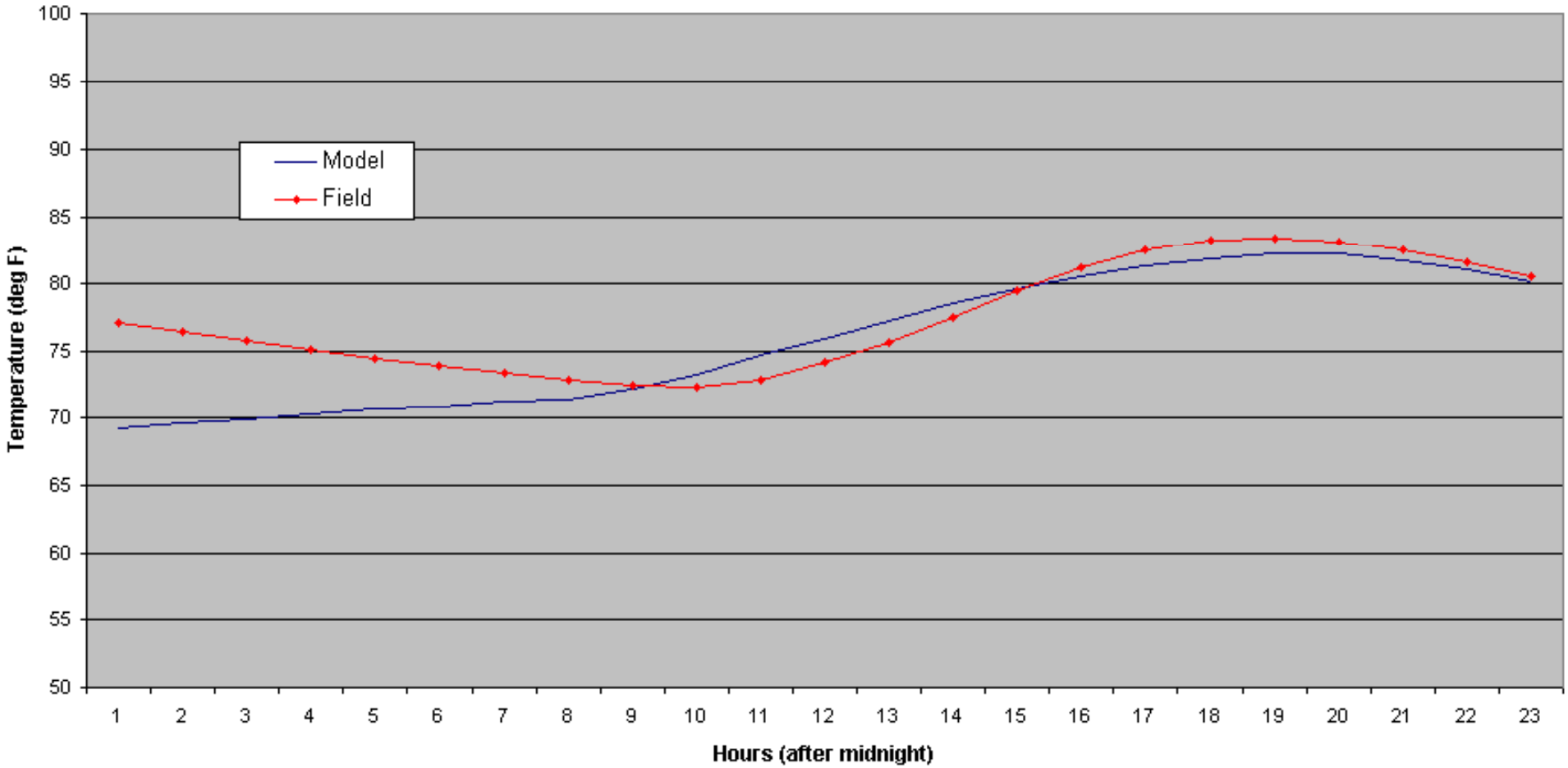




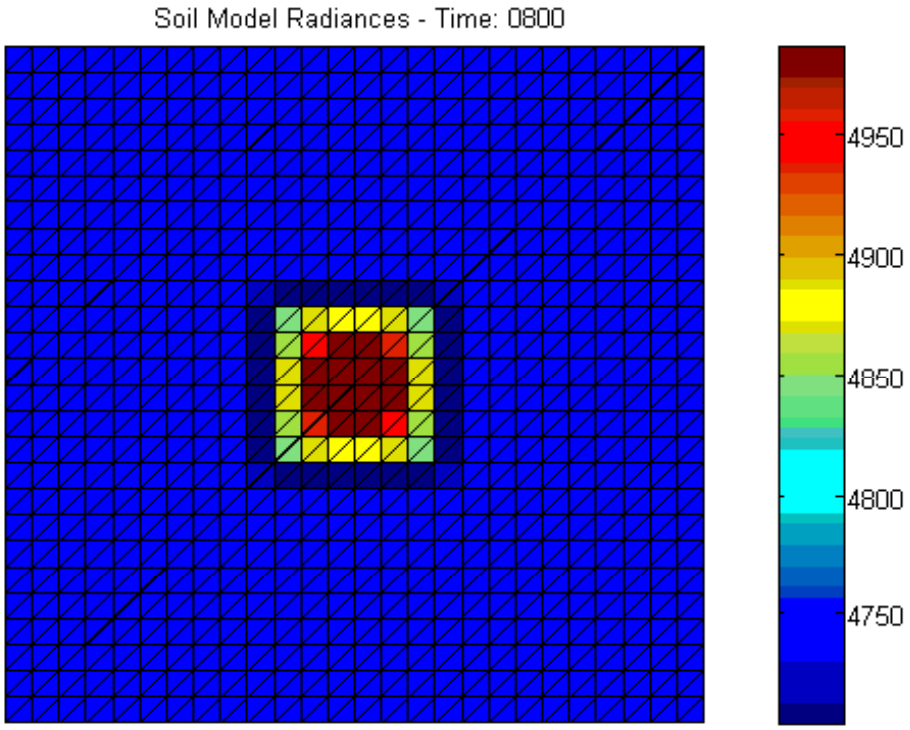
Walnut Gulch Data Comparison – 2” depth –



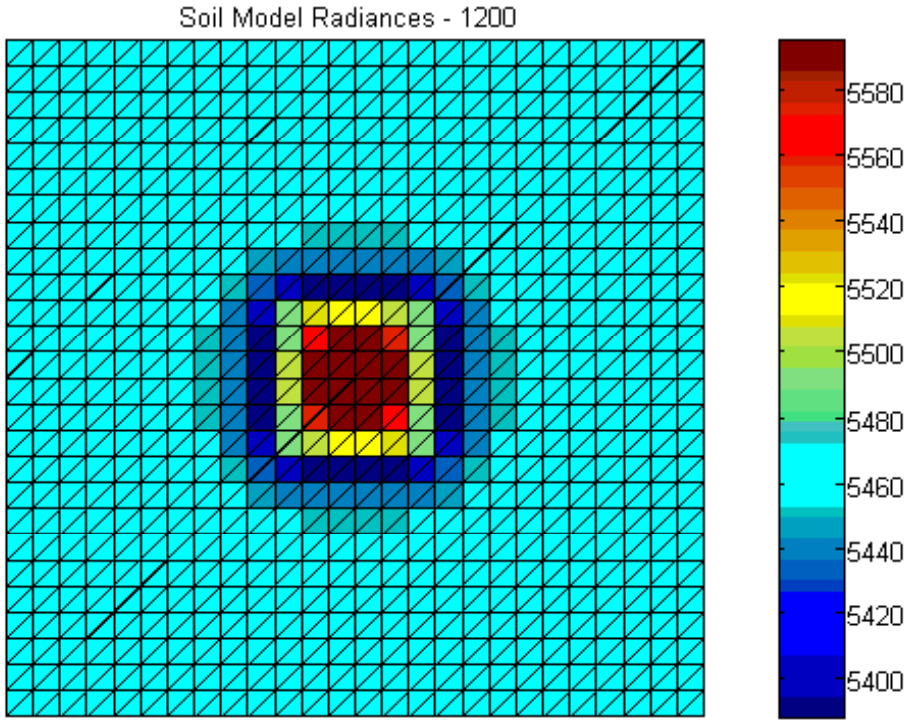
Walnut Gulch Data Comparison – 4” Depth –



Thermal Signature Results

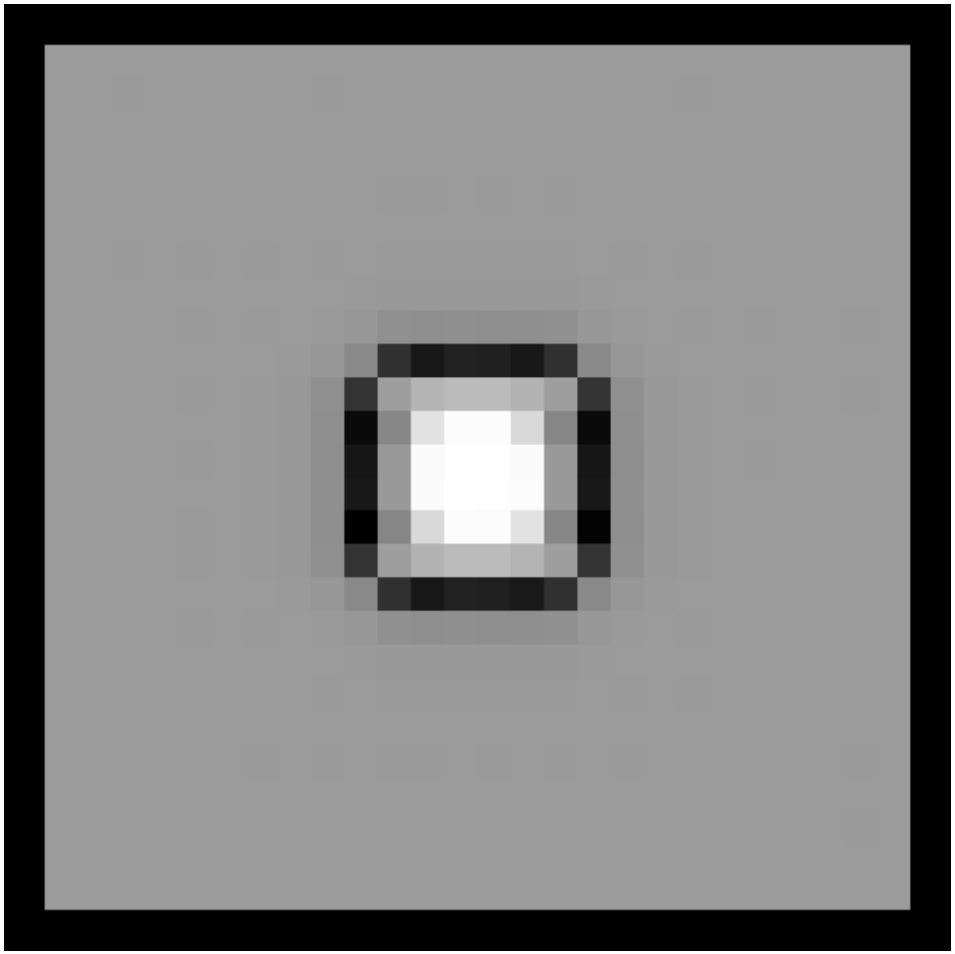


Flush-buried mine

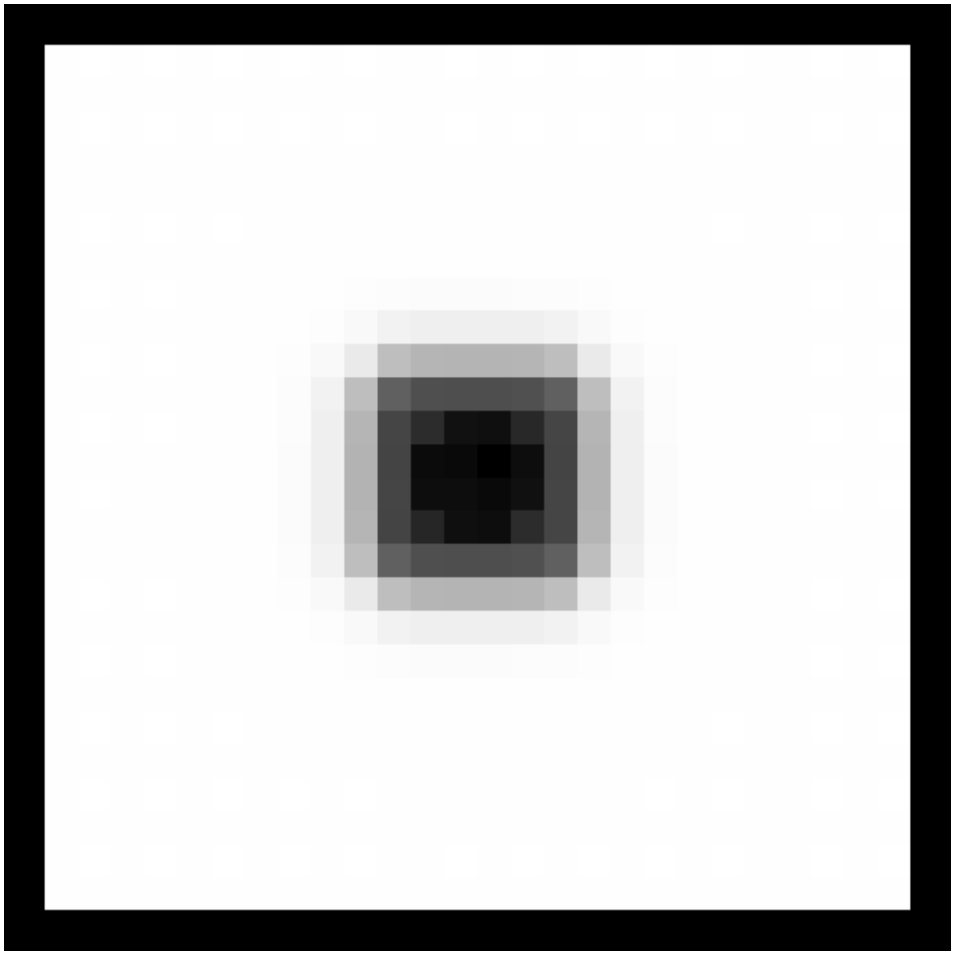


Thermal Signature Results – Undisturbed Soil

Undisturbed Soil at 1000 hrs

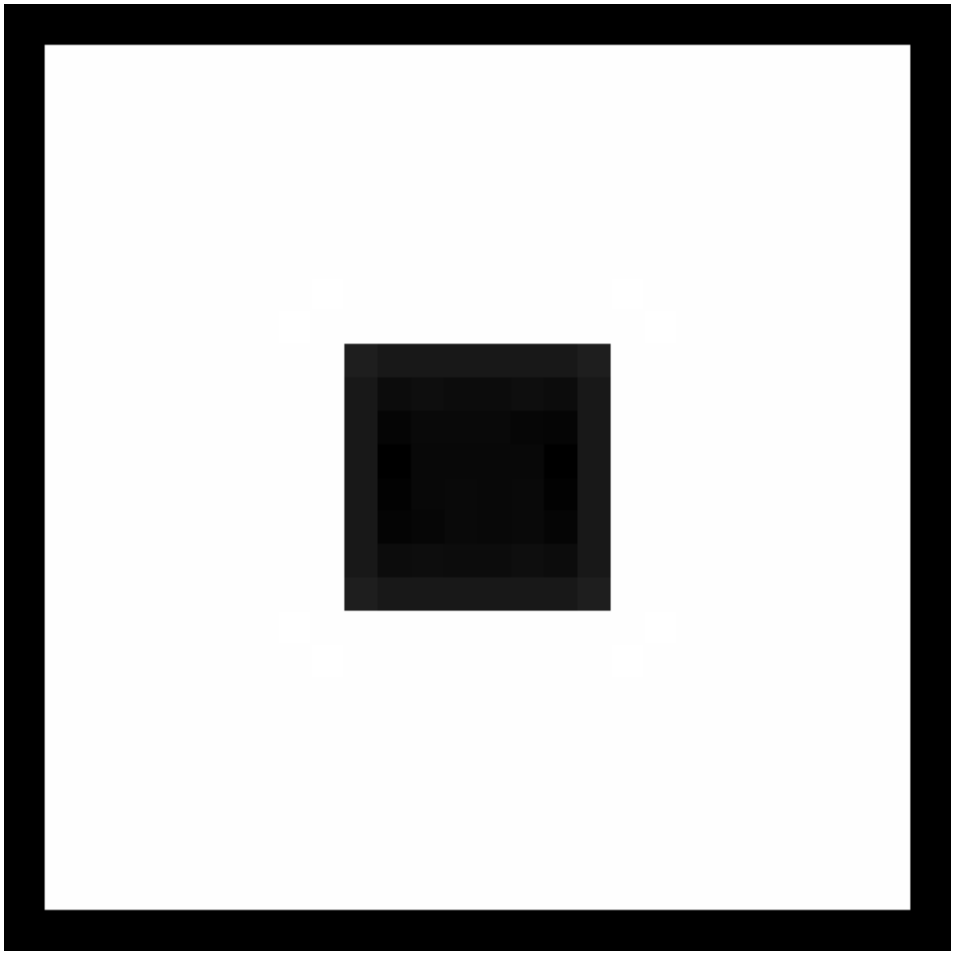


Undisturbed Soil at 2000 hrs

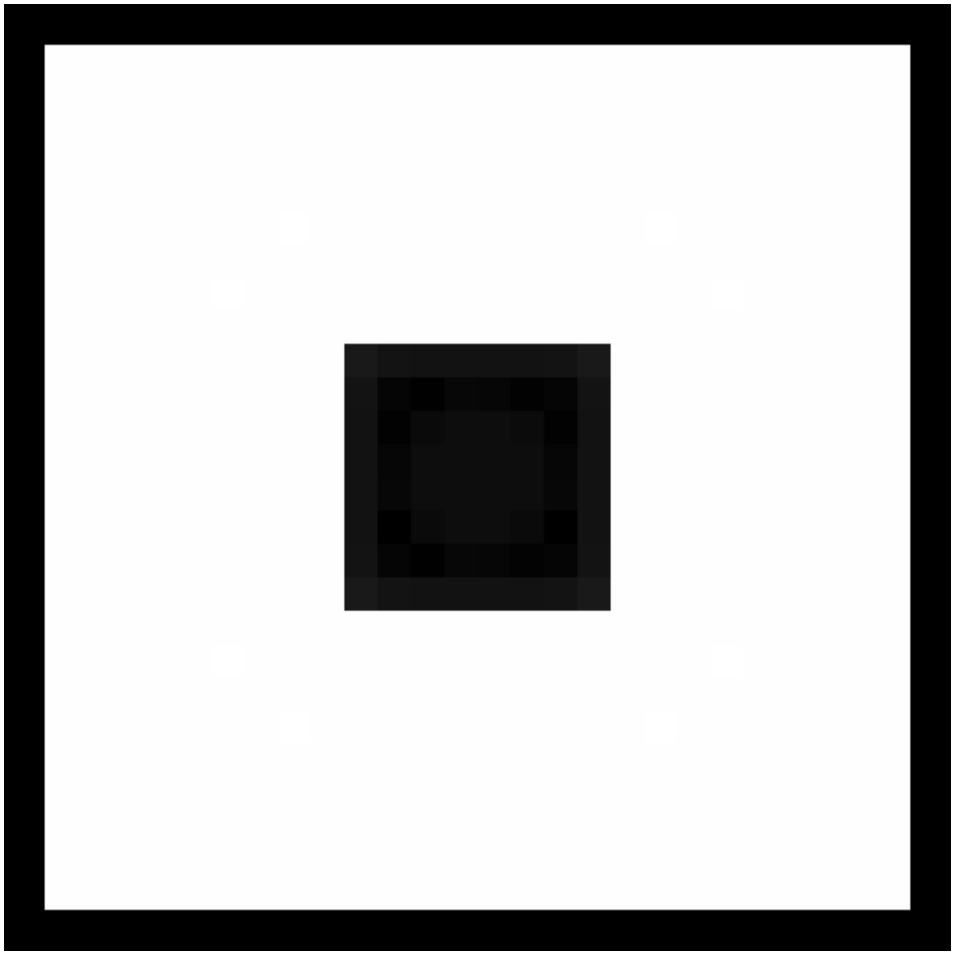


Thermal Signature Results – Disturbed Soil

Disturbed Soil at 1000 hrs

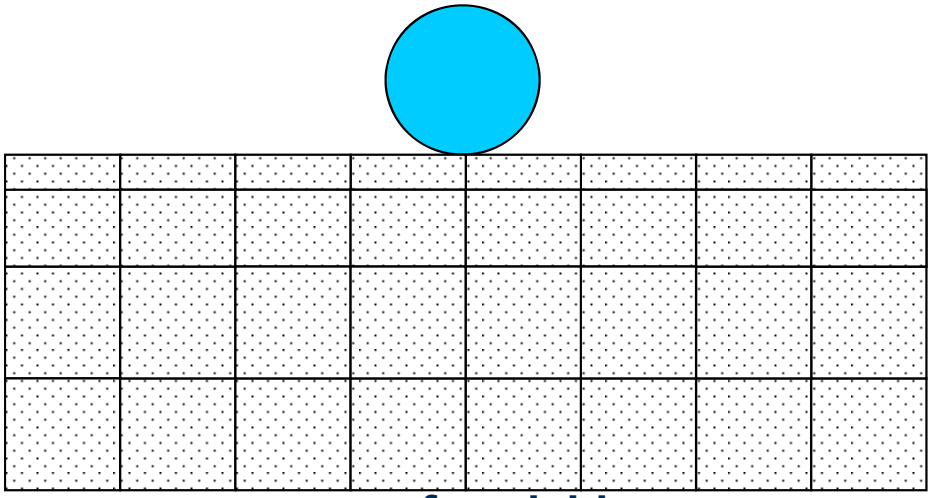


Disturbed Soil at 2000 hrs

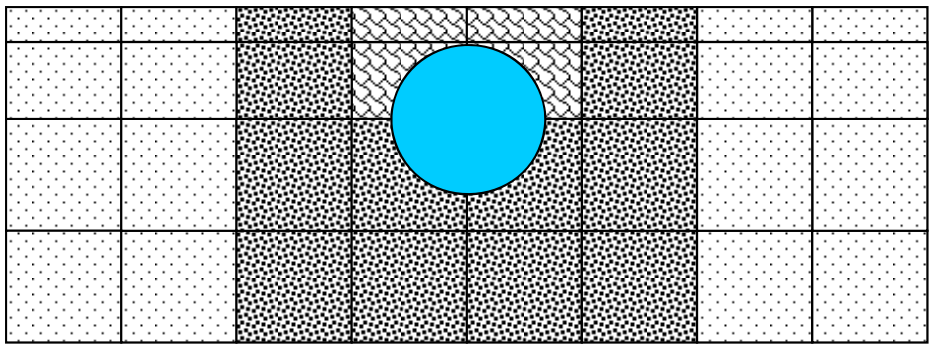


False Target Modeling

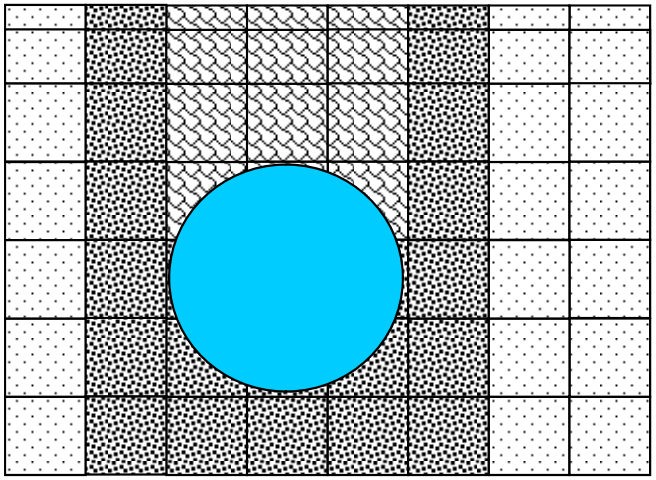
- Soil & target area (1.5 m^3)
- Cylindrical structures
- Geometric resolution 5.5 cm x 5.5 cm
- Vertical resolution 1.6 cm to 5.5 cm
- > 14,000 thermal nodes
- Vertical and horizontal conduction paths



surface laid

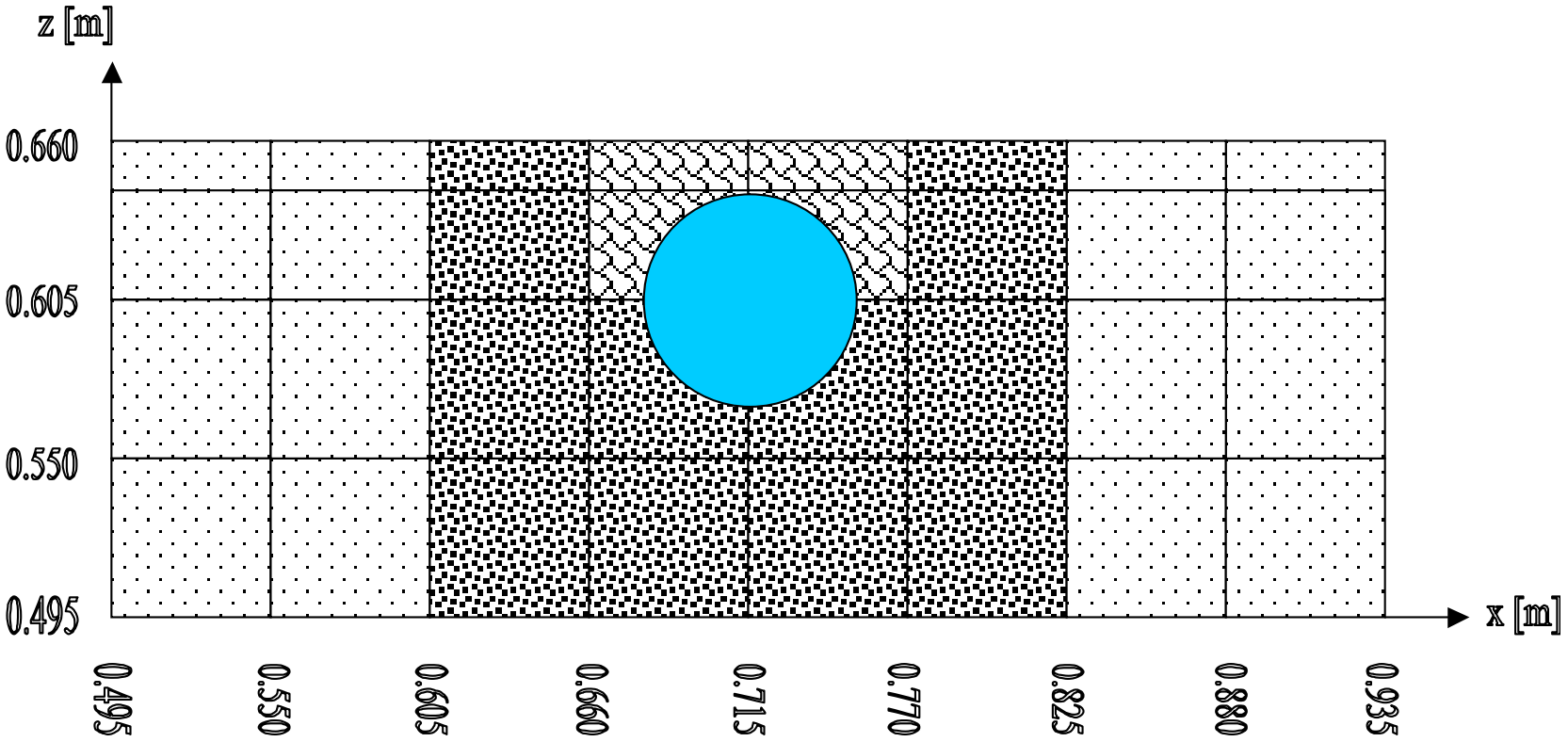


flush buried



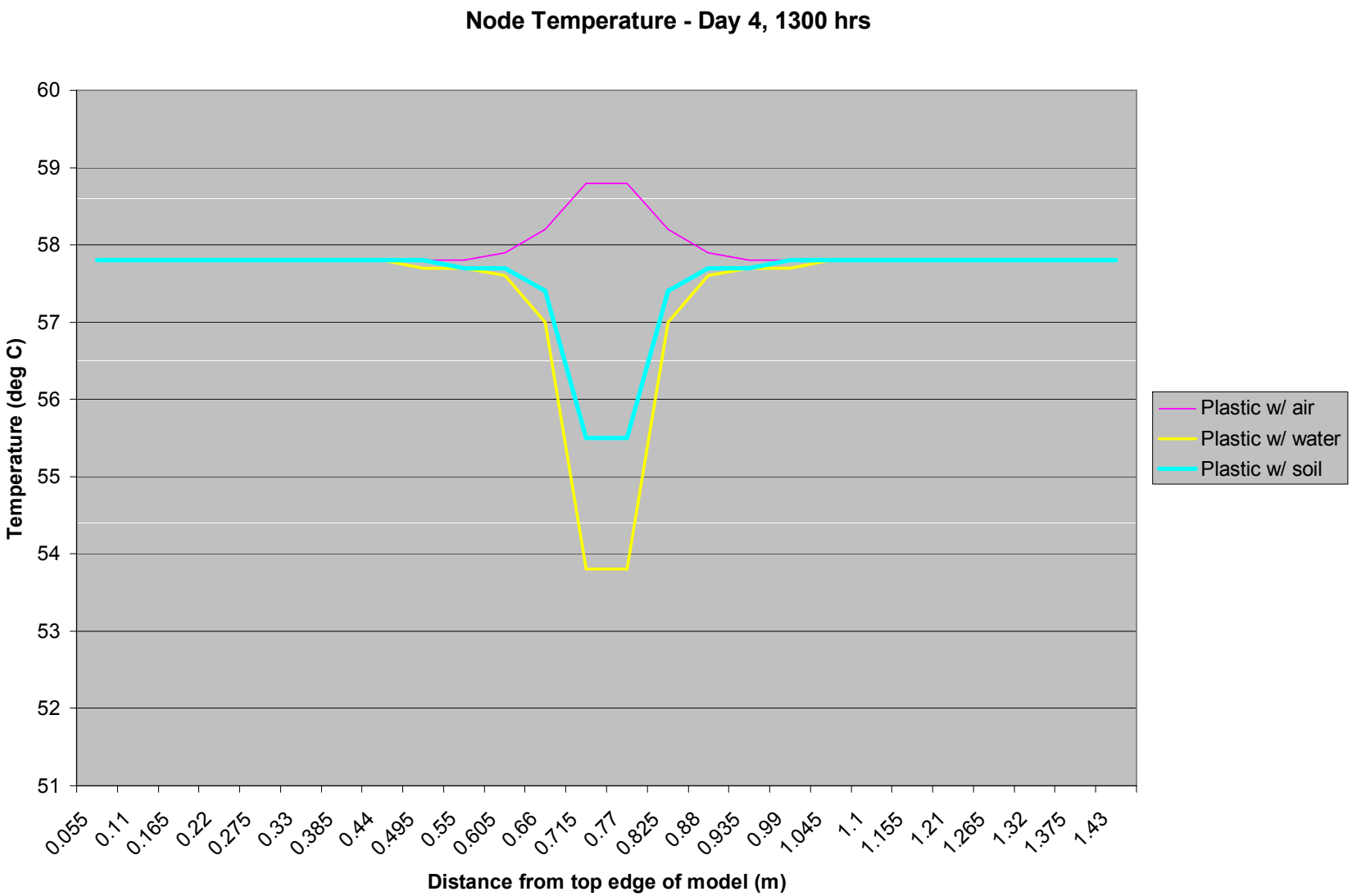
buried

Soil – False Target Nodal Structure



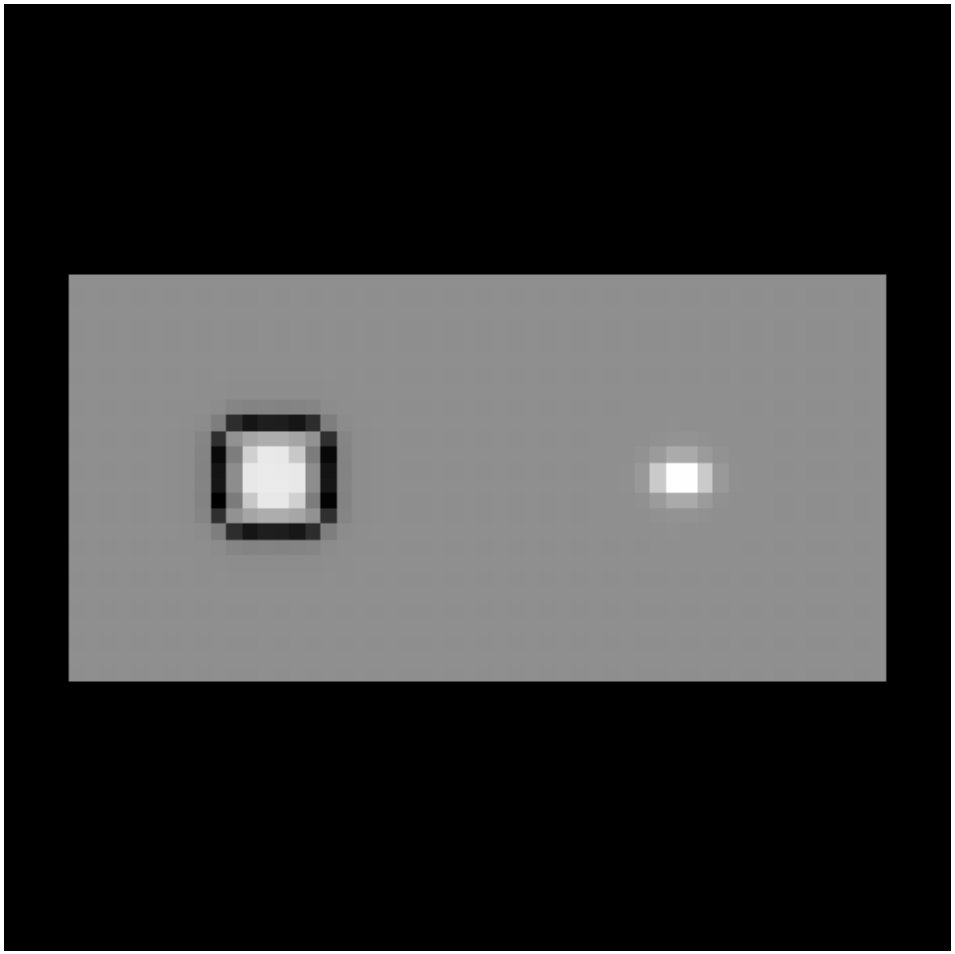
- Undisturbed soil
- Disturbed soil
- Disturbed soil over mine

False target temperatures

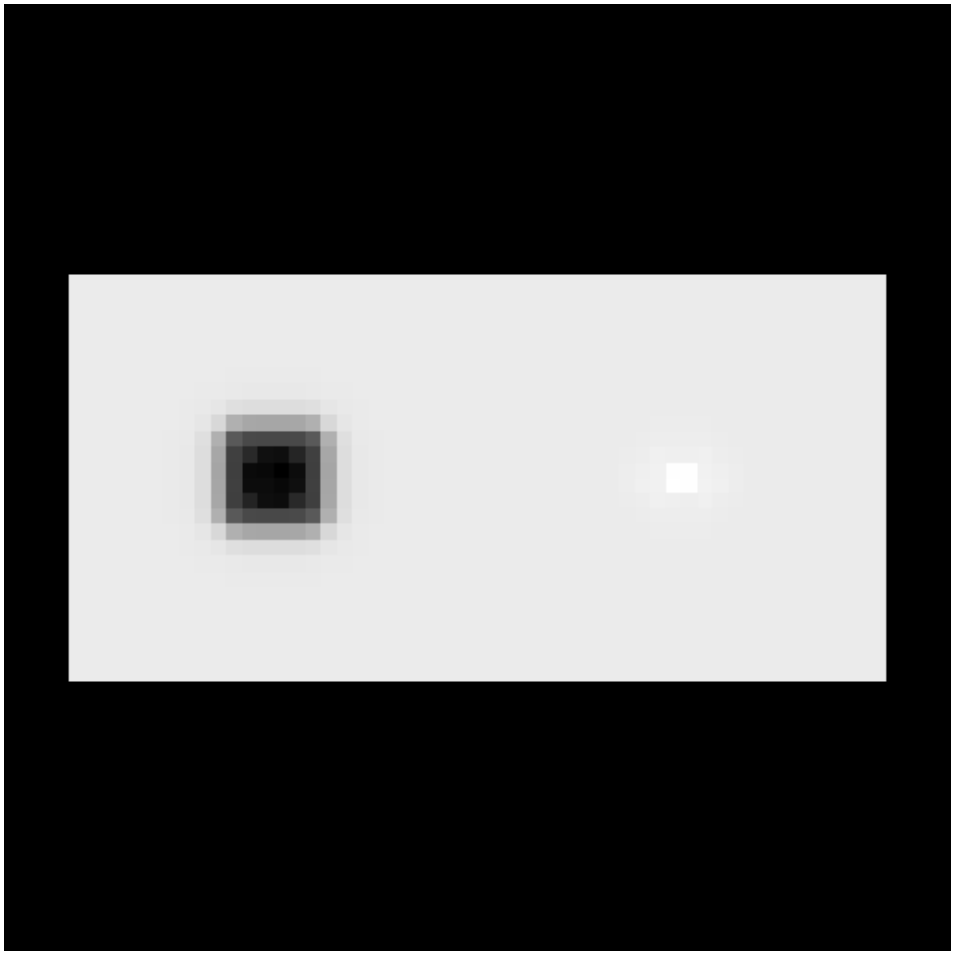


Thermal Signature Results – Flush Mine & Can

1000 hrs



2000 hrs



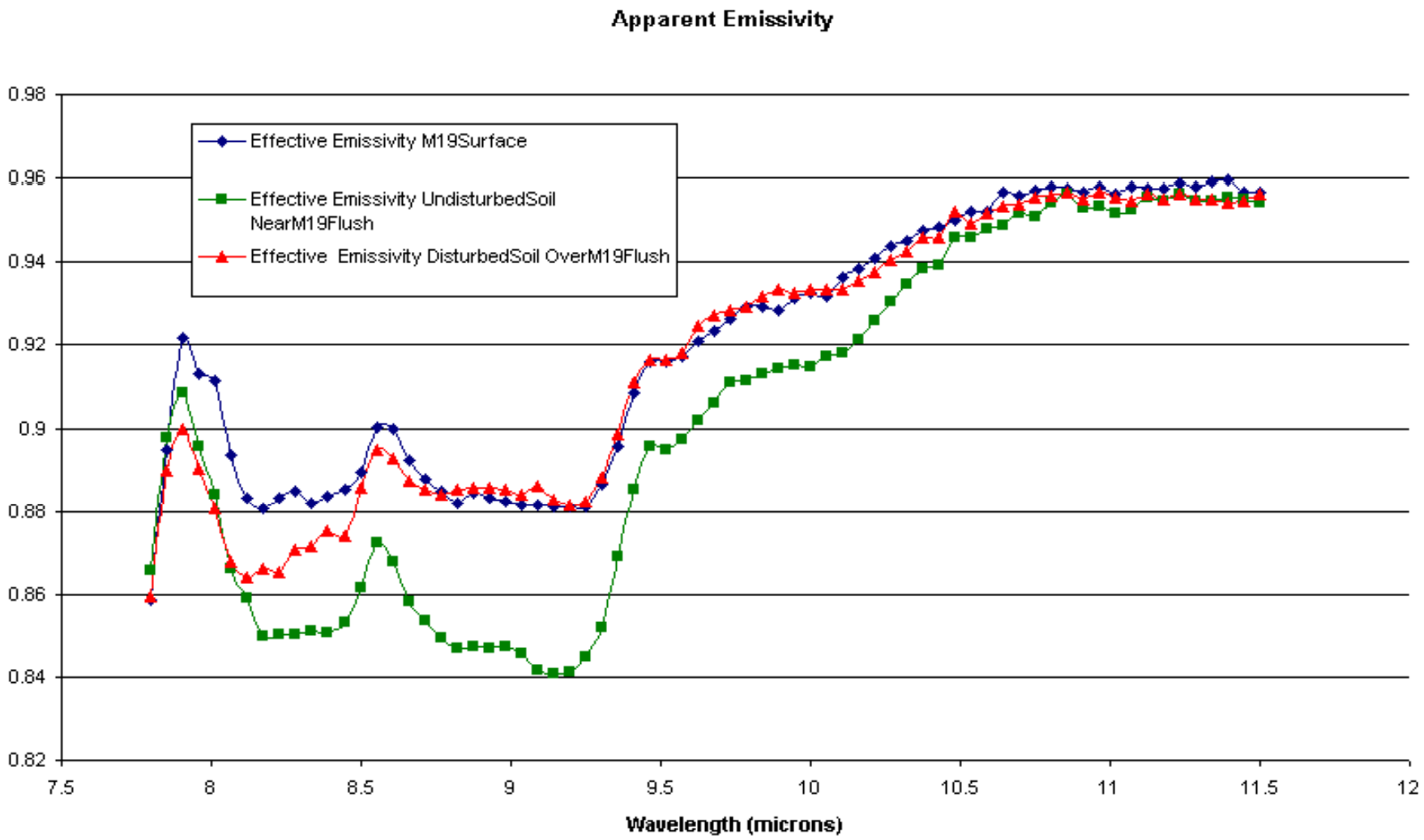
Spectral Signature Generation Process

- Determine AHI bands
- Derive apparent ε values for surface landmine, disturbed soil, undisturbed soil
- Generate input files for each band (incorporation of apparent ε data)
- Select calculation time (10th day)
- Execute GTSIG
- Extract radiance data
- Compare against AHI data

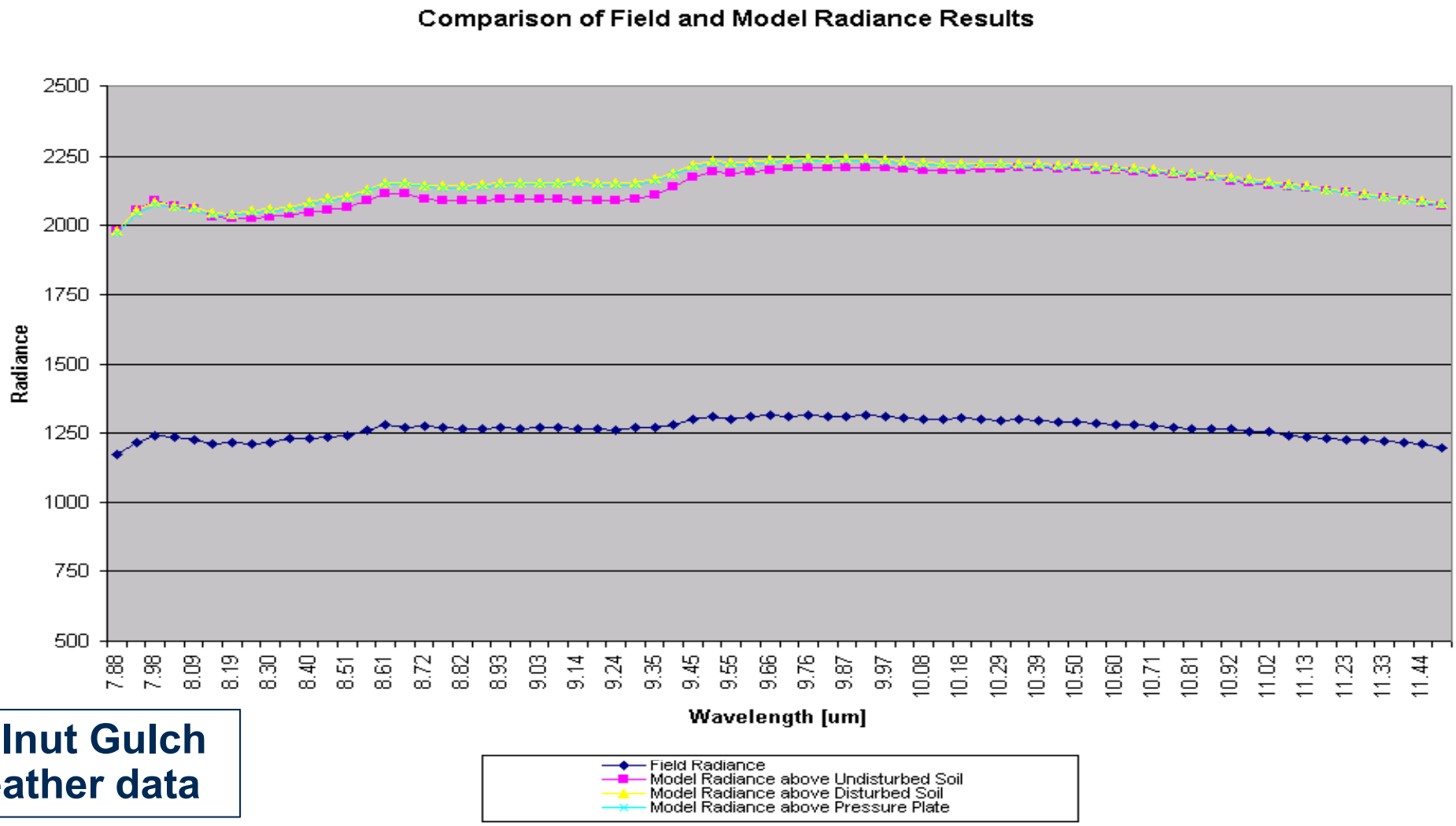
Spectral Signature Computation Steps

- Perform temperature / emissivity separation for arid data
- Determine emissivities at each wavelength:
  - Surface land mine
    - Undisturbed soil near land mine
    - Land mine itself
  - Flush & buried land mine
    - Disturbed soil over land mine
    - Undisturbed soil near land mine
- Incorporate spectral emissivity values with thermal data to generate spectral radiance curves
- Initial calculations used typical arid weather data from southwest (i.e., data from Walnut Gulch meteorological station)
- Updated calculations used NVESD weather data obtained from the southwestern test site during WAAMD data collection.

Derived Emissivity Data

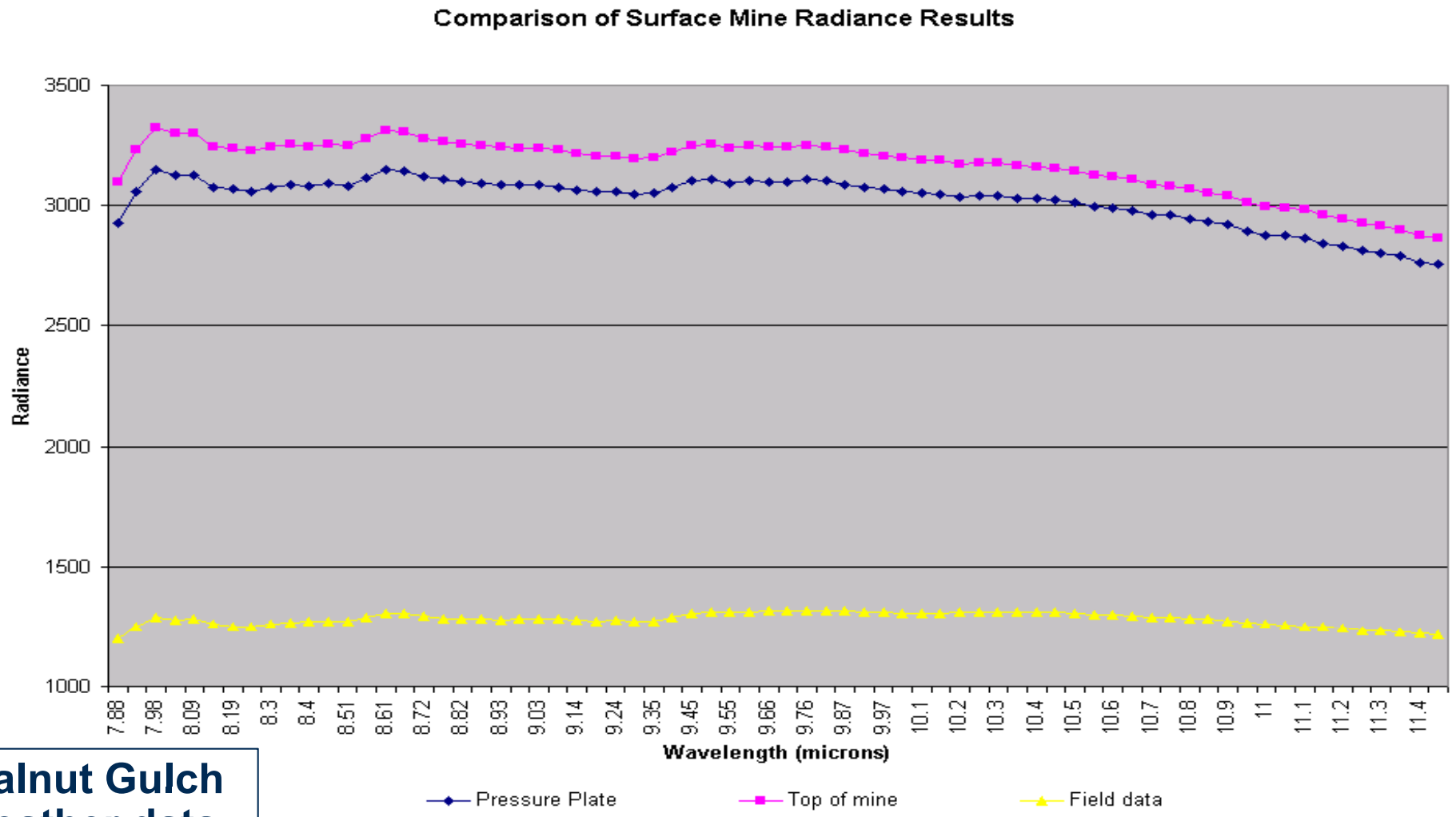


Comparison of Model & Field Data (flush buried case)



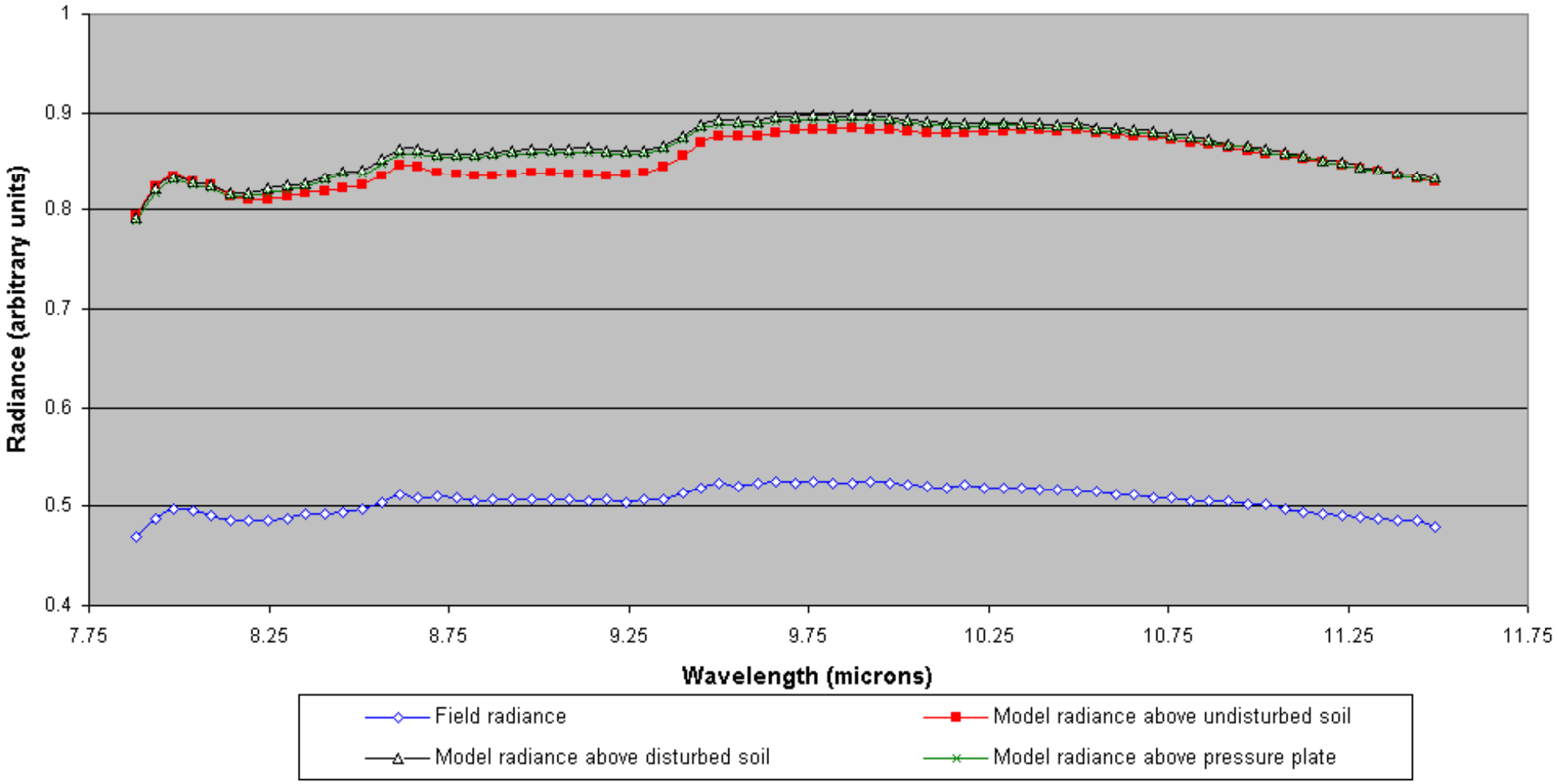
Walnut Gulch weather data

Comparison of Model & Field Data (surface case)



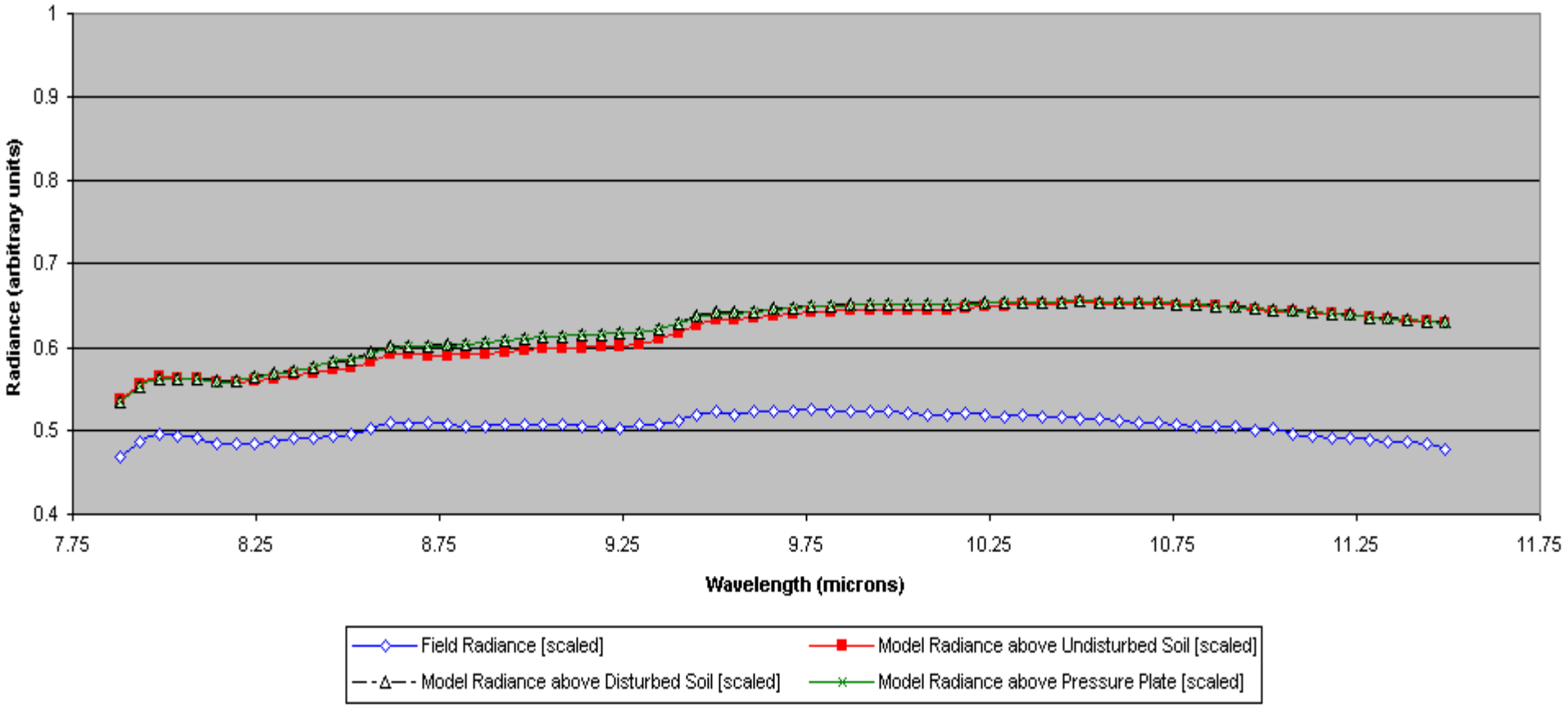
Walnut Gulch weather data

Comparison of Model & Field Data (WG weather data)



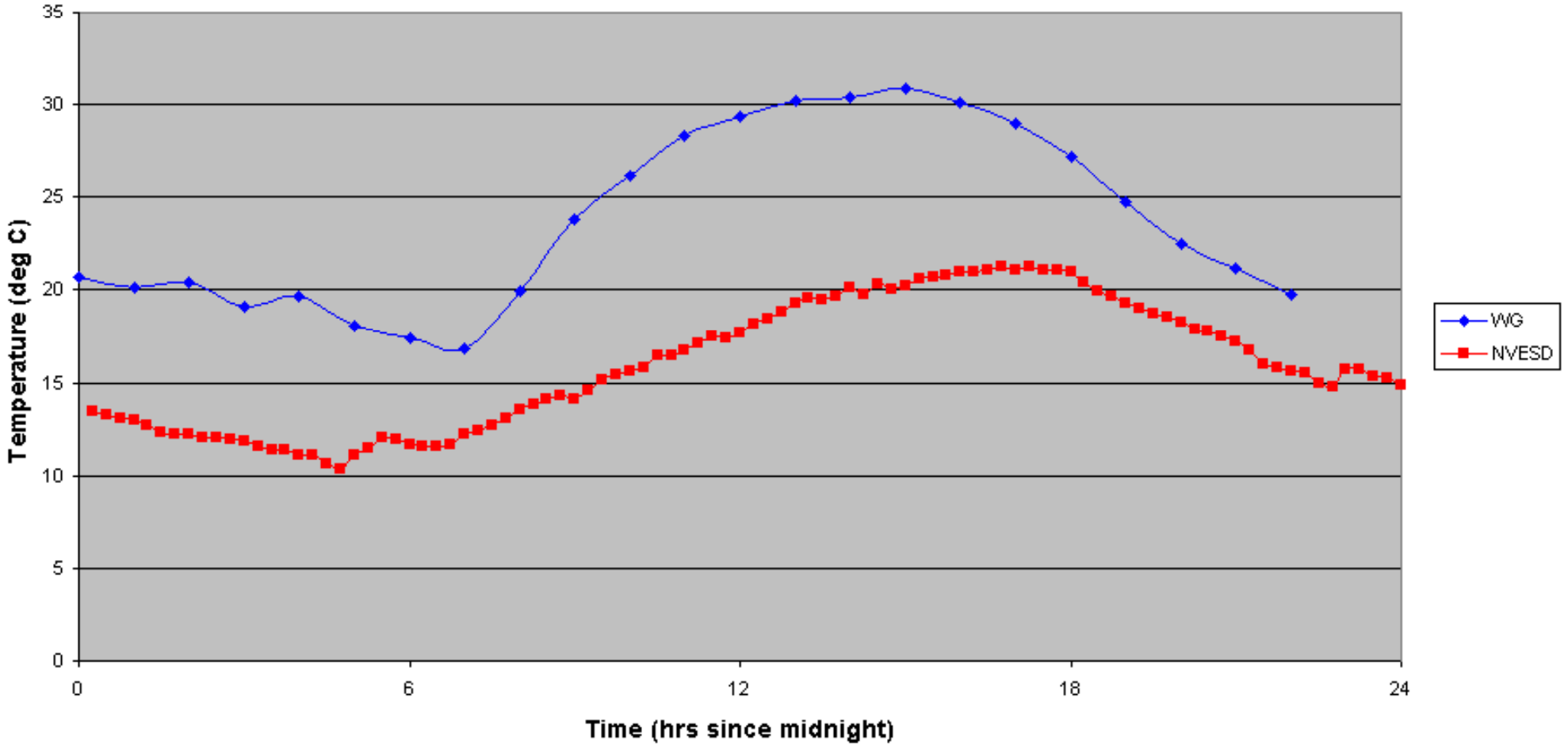
Scaled radiance results

Comparison of Model & Field Data (NVESD southwestern weather data)



Scaled radiance results

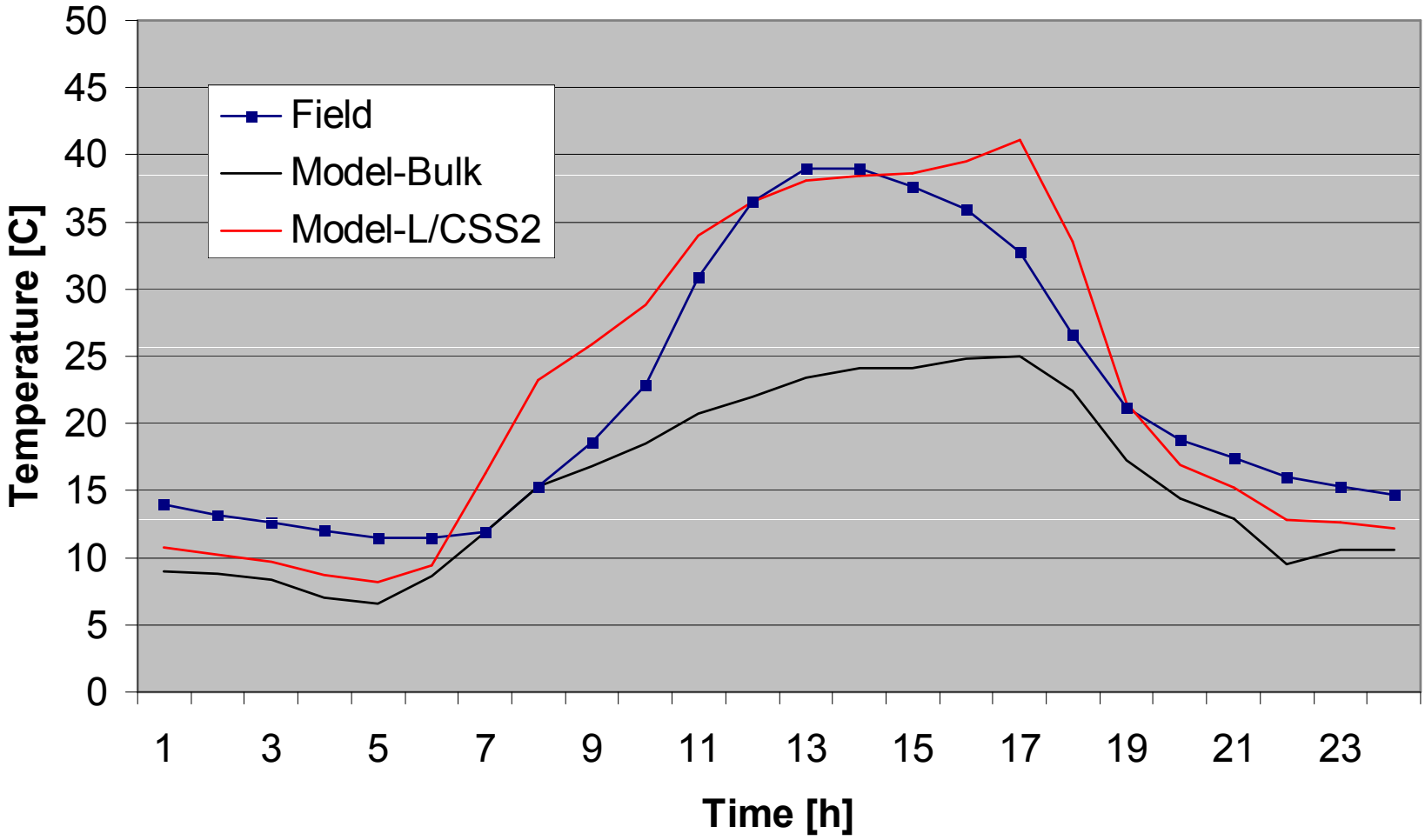
Comparison of the air temperature data for the two weather data sets



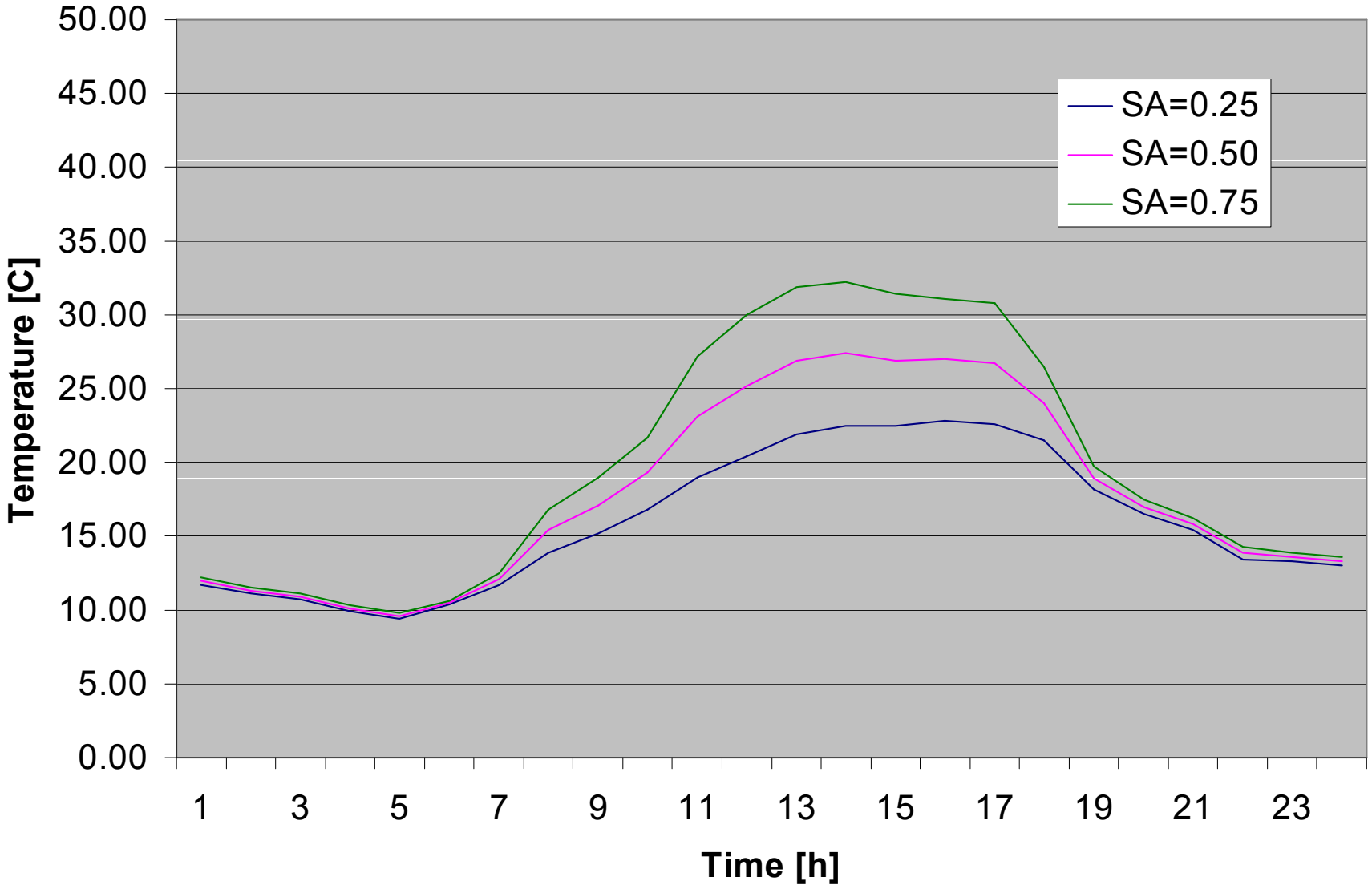
Comparison illustrates one reason why the radiance data for the two weather cases varies so much.

Soil and Environmental Parameters - Sensitivity Analysis

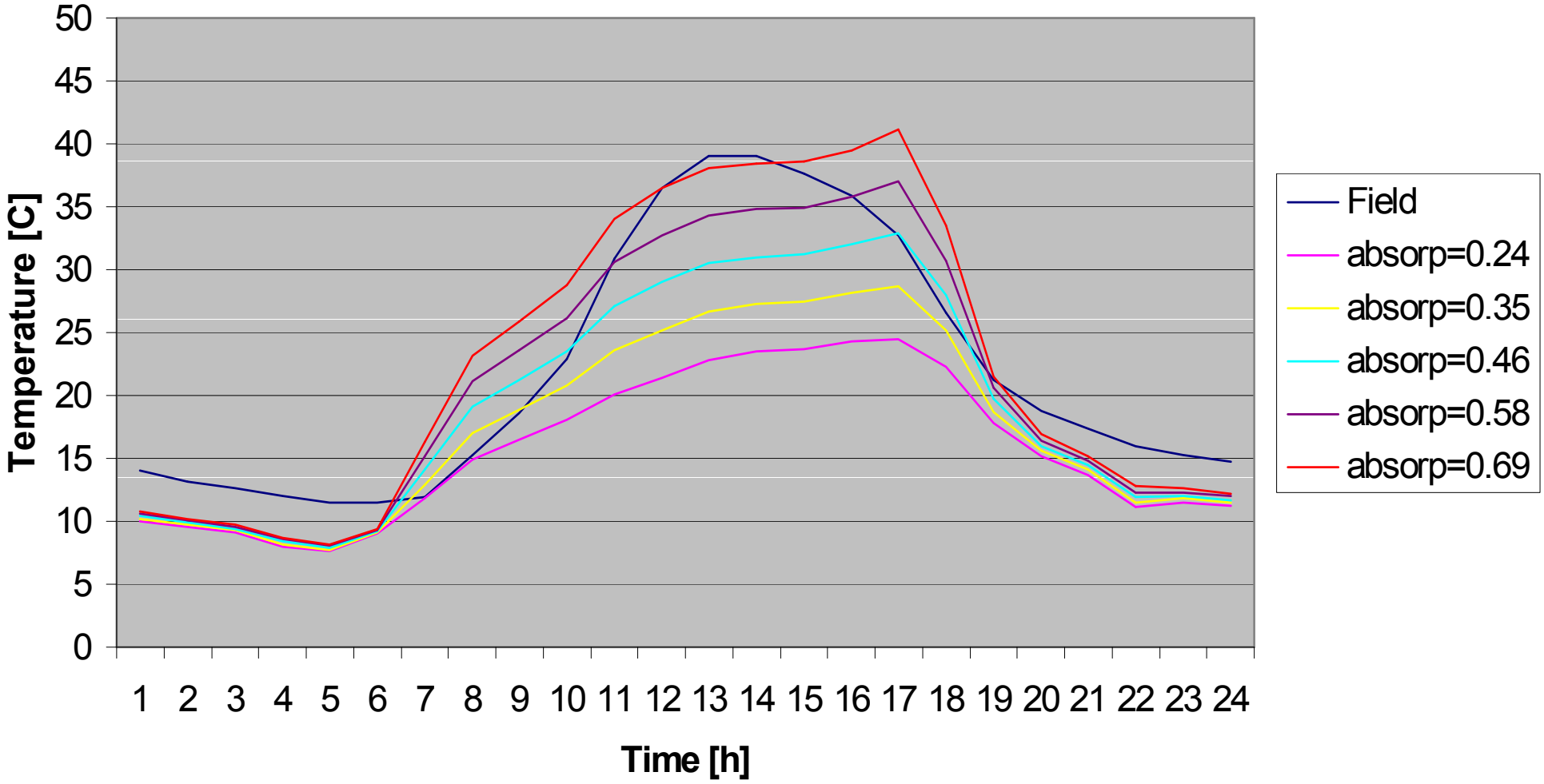
Diurnal Temperature Comparison (surface)



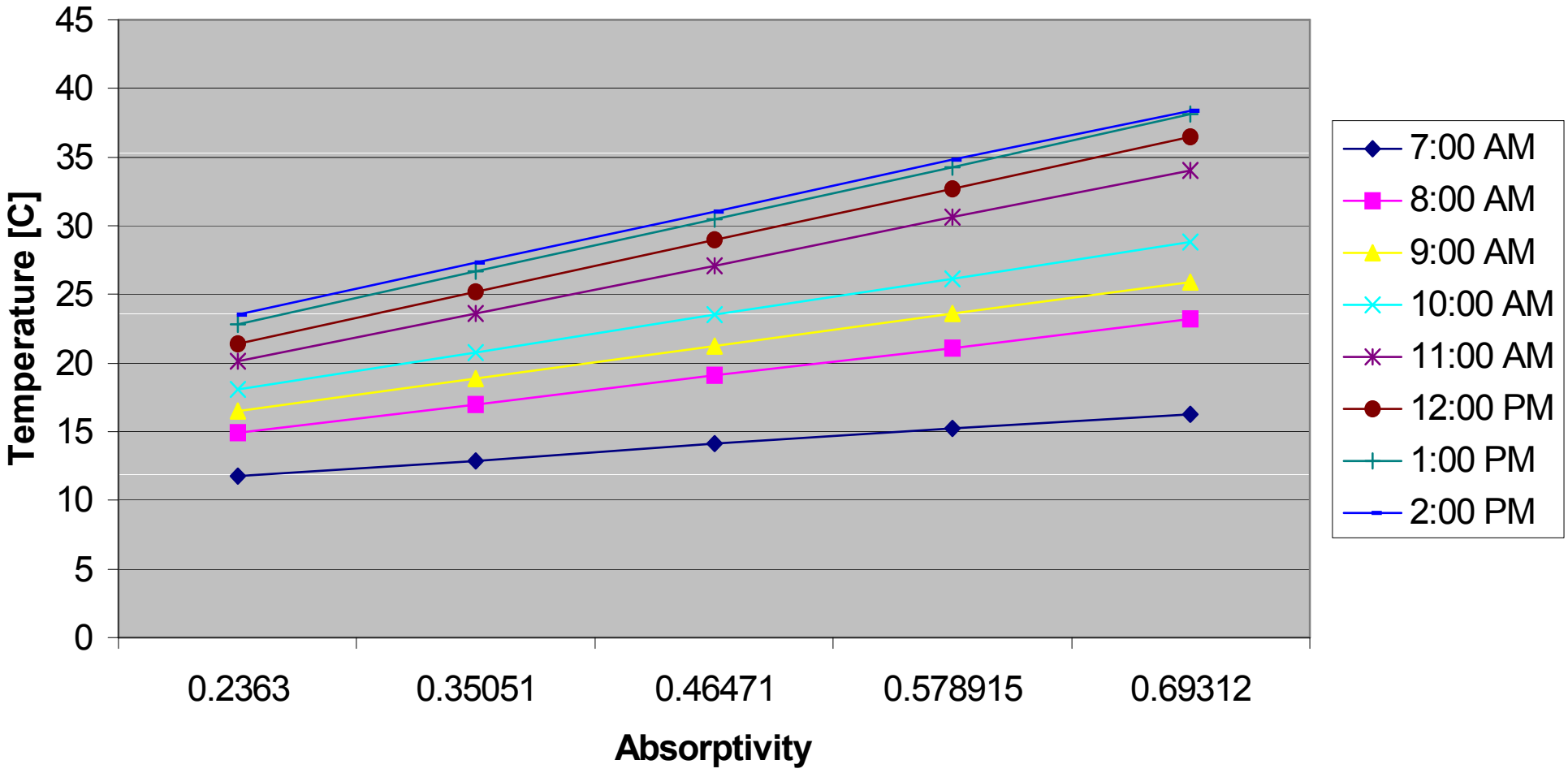
Surface Diurnal Temperature Profile with Solar Absorptivity Variation



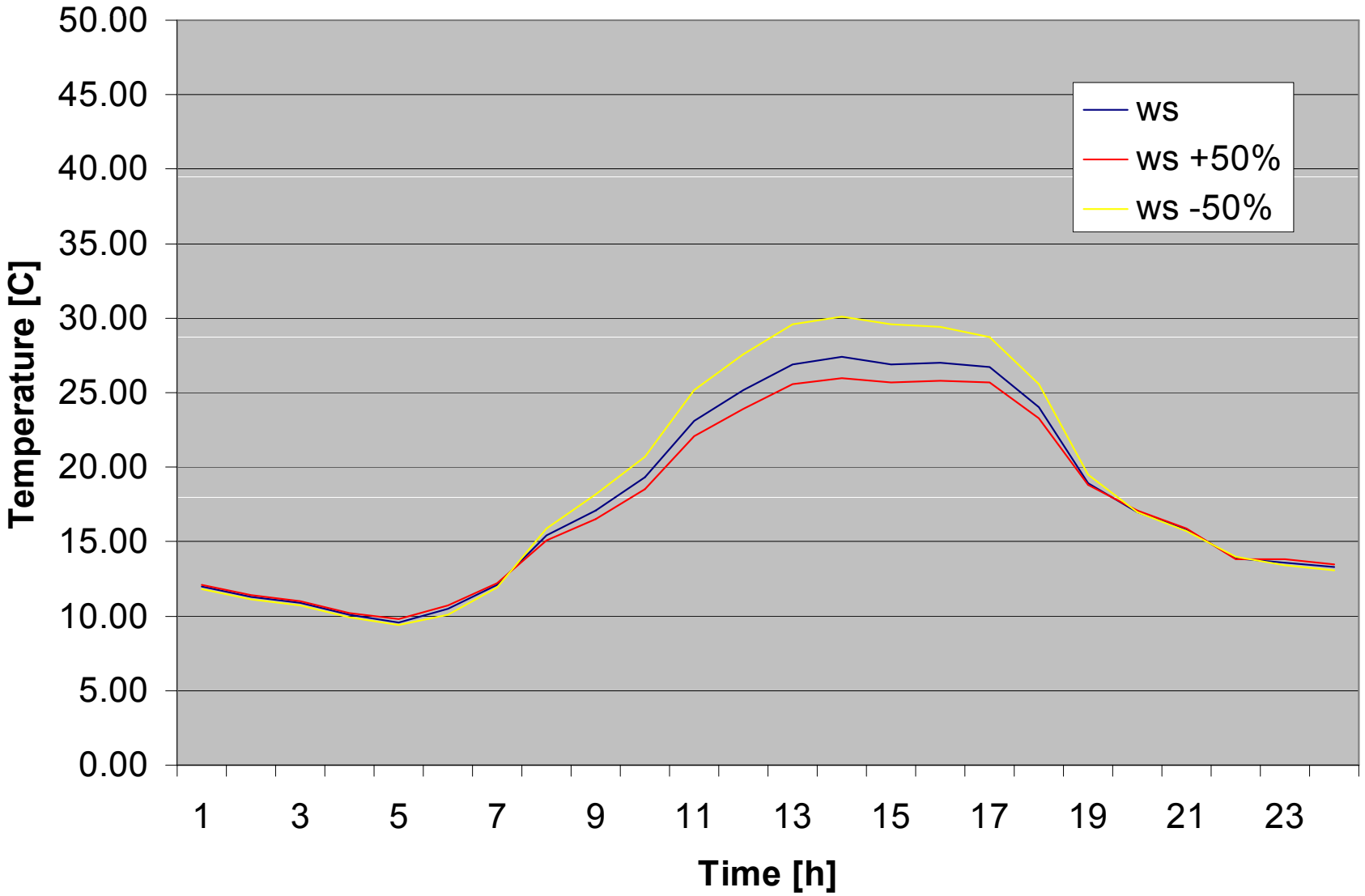
Diurnal Surface Temperatures (weighted solar absorptivity)



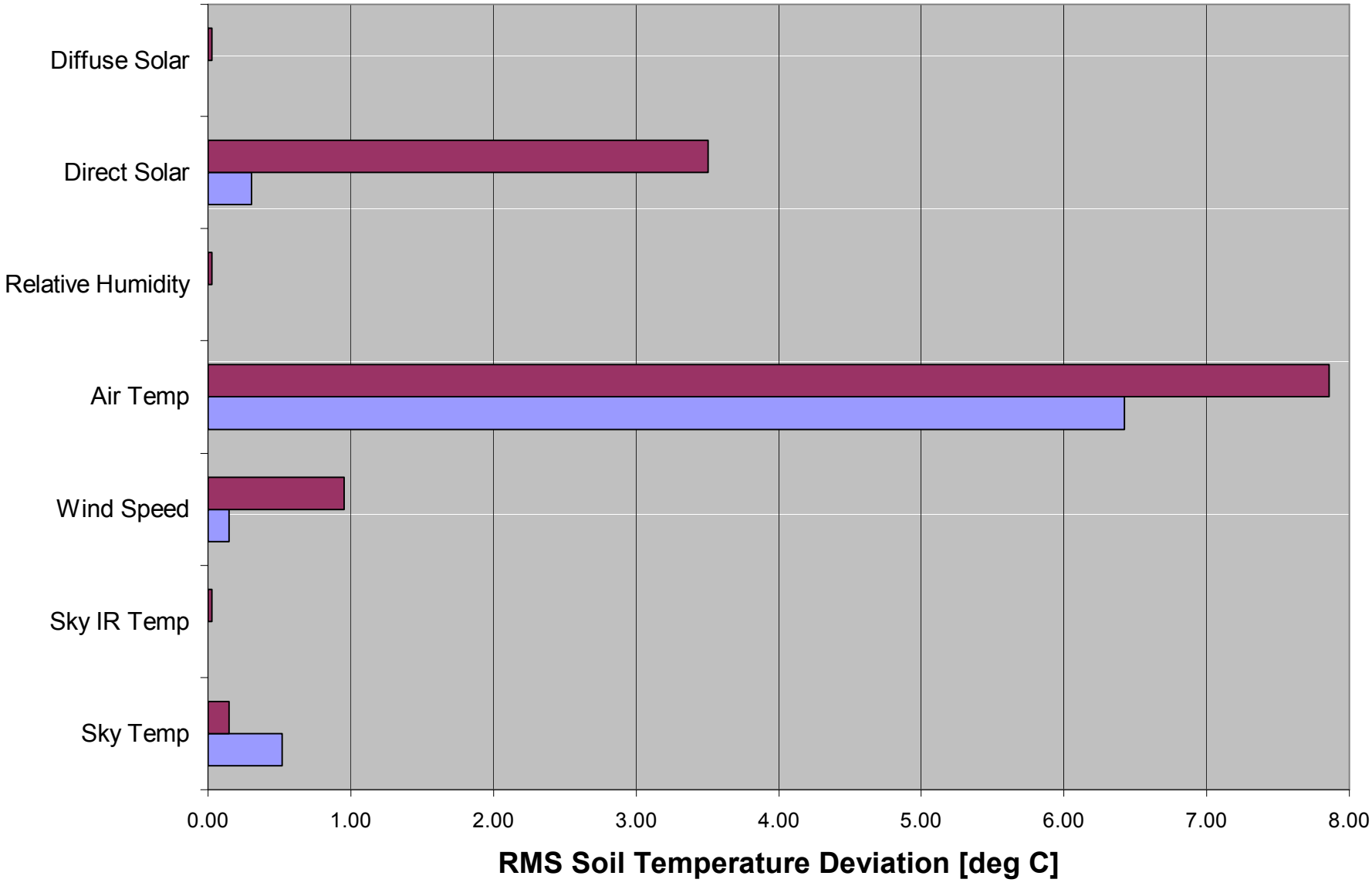
Solar Absorptivity Impact on Diurnal Surface Temperatures



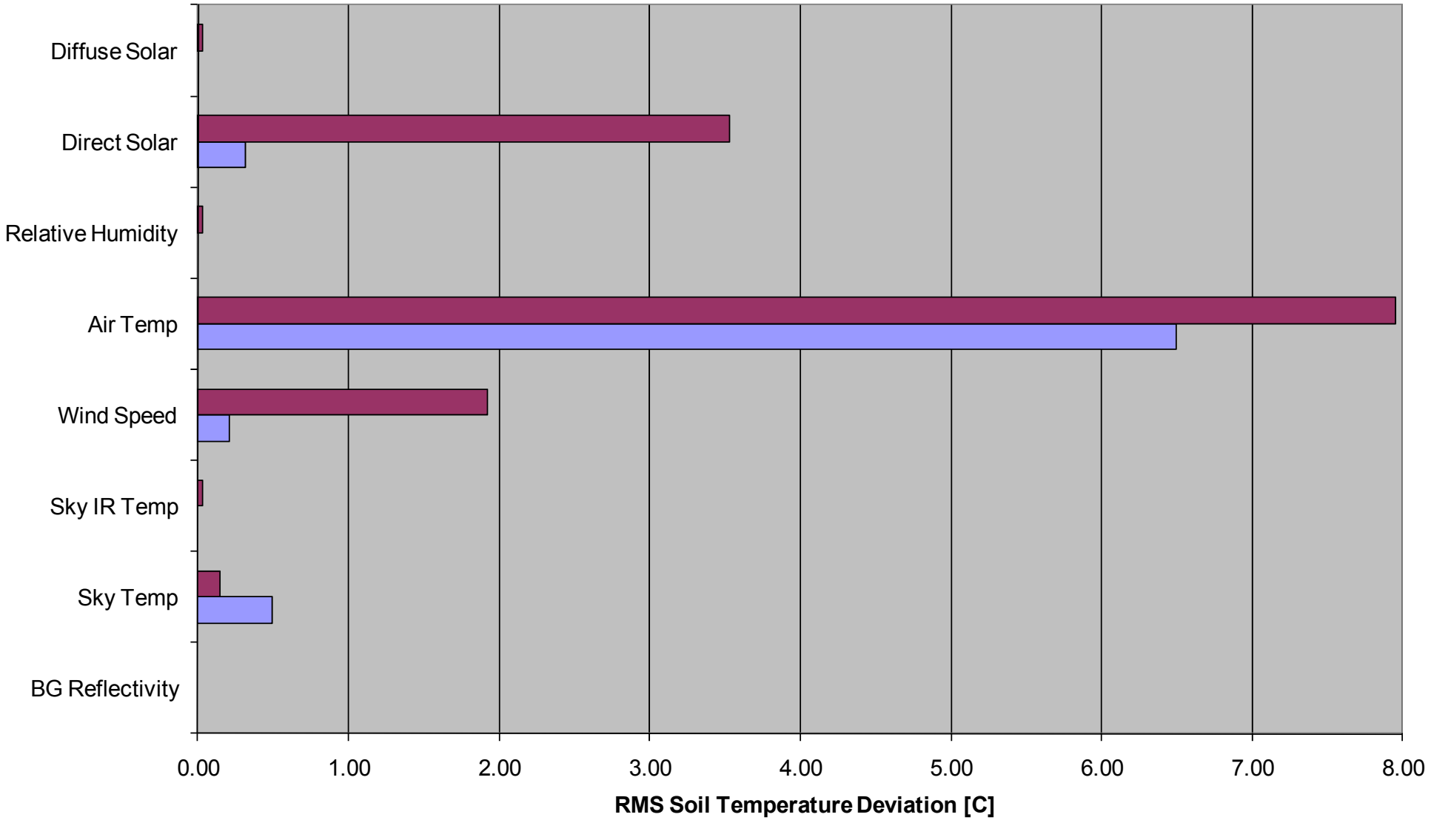
Surface Temperature Profile with Wind Speed Variation



Environmental Parameter Variation (+50%)

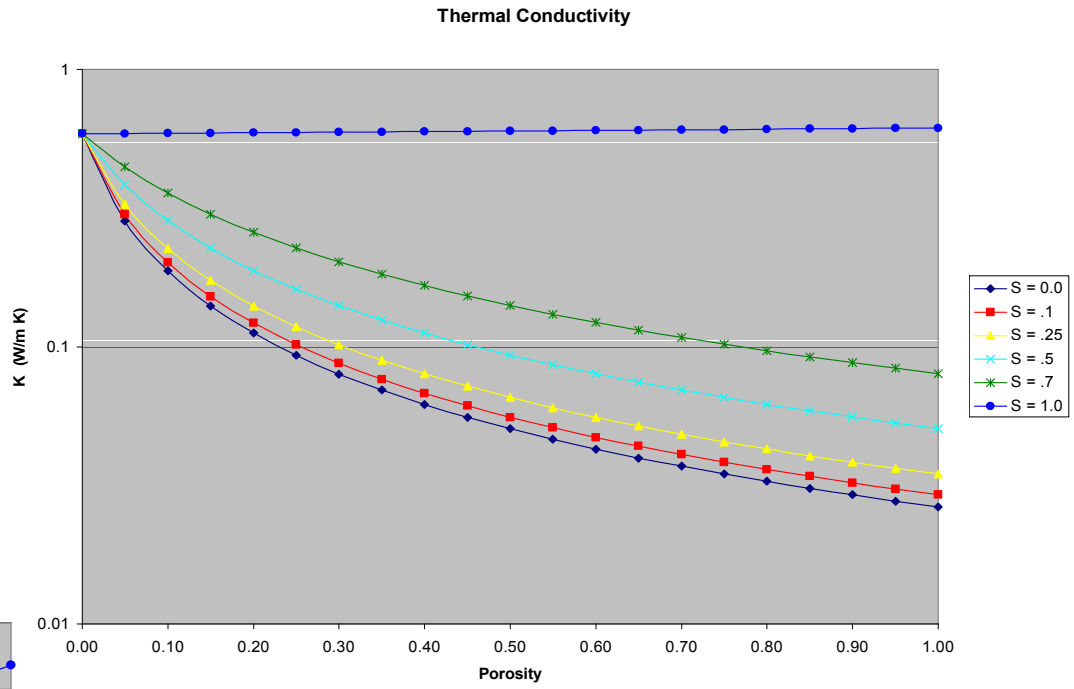


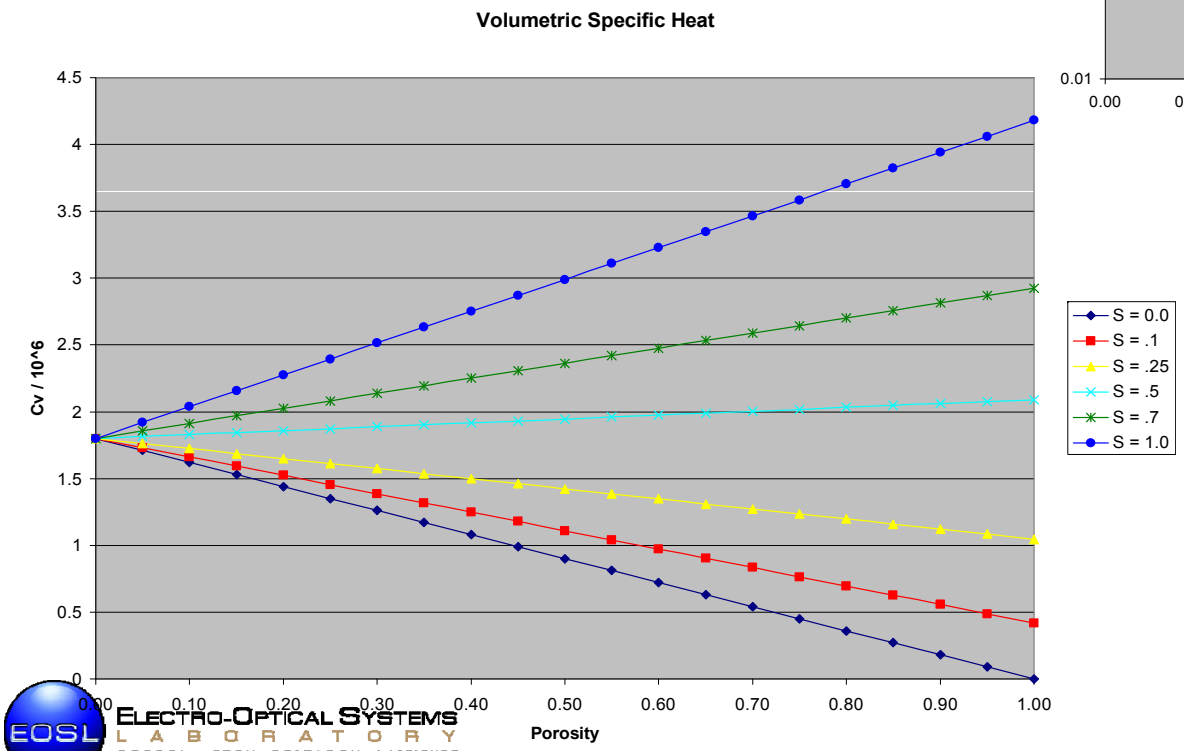
Environmental Parameter Variation (-50%)



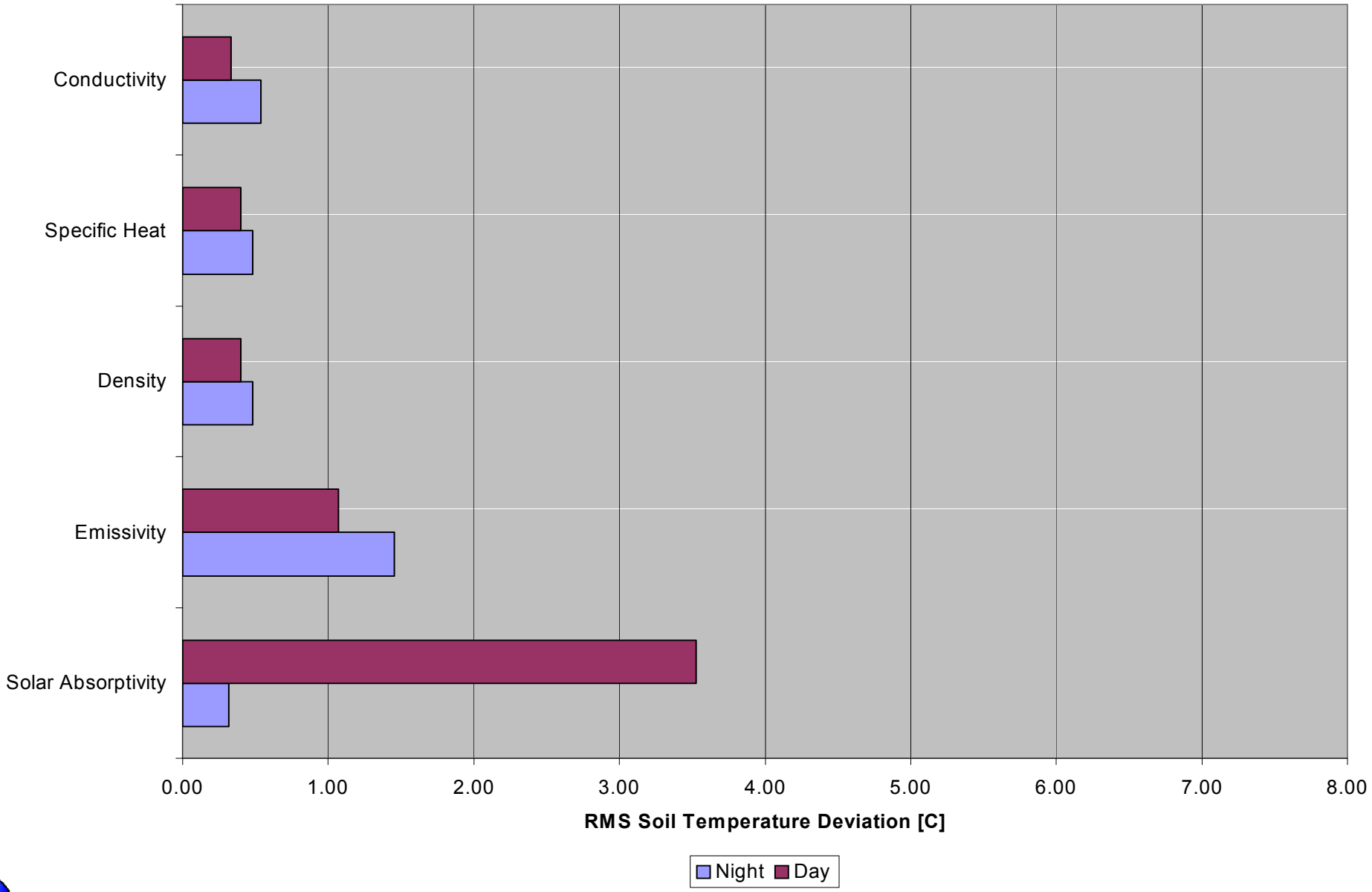
Soil Parameter Effects

- Thermal properties vary with
  - soil constituents
  - porosity
  - H_2O
- Linear mixing model

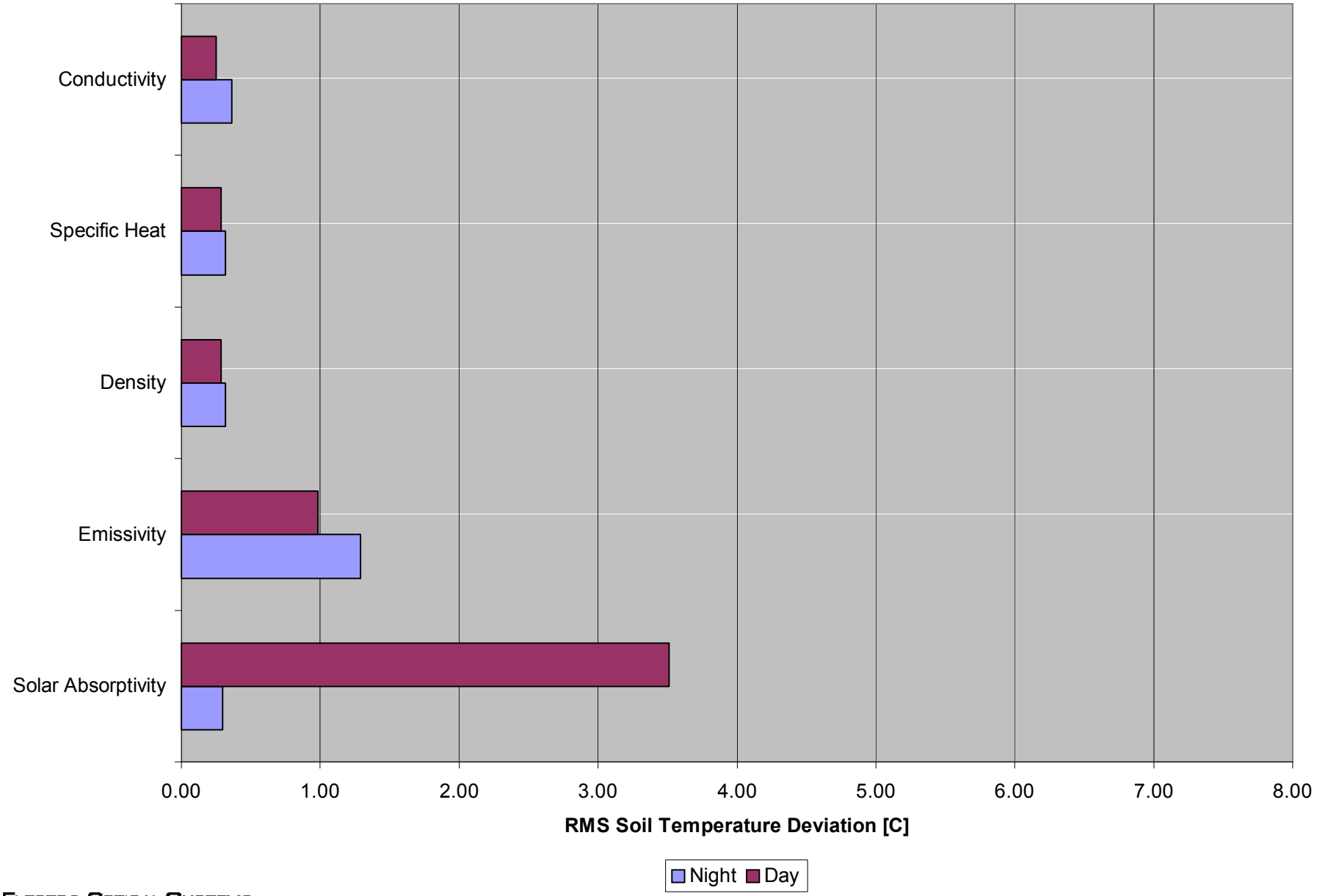




Soil Parameter Variation (-50%)



Soil Parameter Variation (+50%)



Digital Modeling – Summary

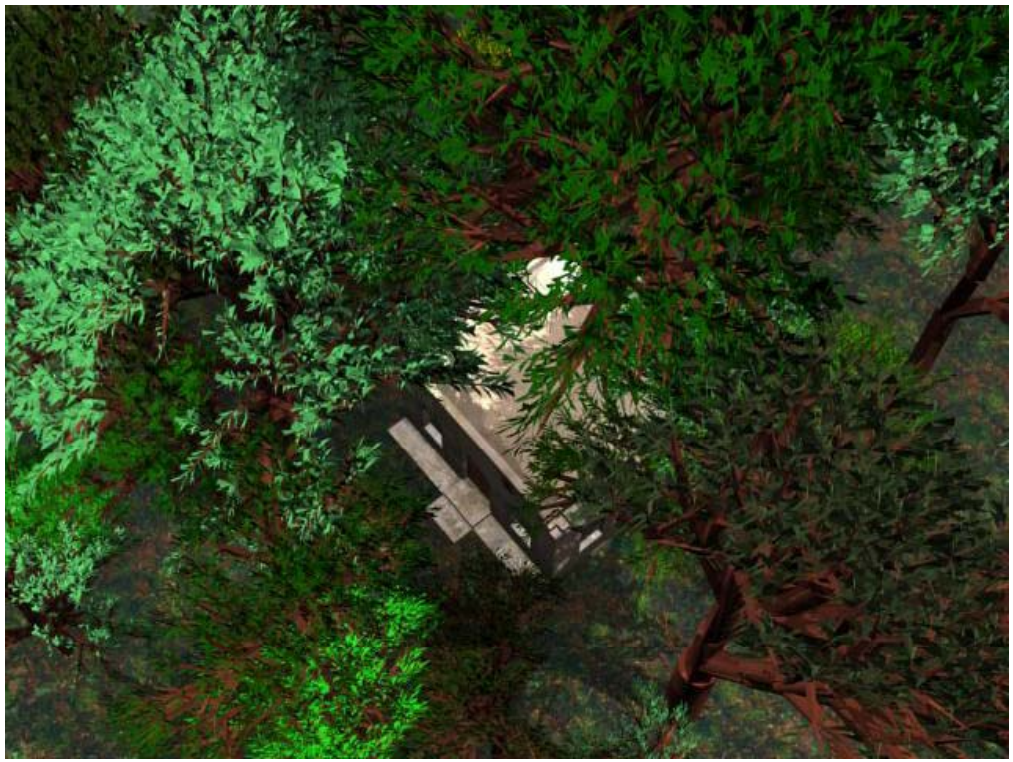
- **Accomplishments**
  - Surface, flush, & buried landmine models created
  - Surface, flush & buried false target models created
  - Preliminary signature study conducted (total & spectral)
  - Verification process performed
  - Comparison made to measured AHI data
- **Issues**
  - Yuma weather data corresponding to time of measurements unavailable
  - Additional validation data, tests, and measurements needed

Exposure effects on optical properties of building materials

PI: Sarah Lane, Tim Harrell, & Michael Cathcart

Background

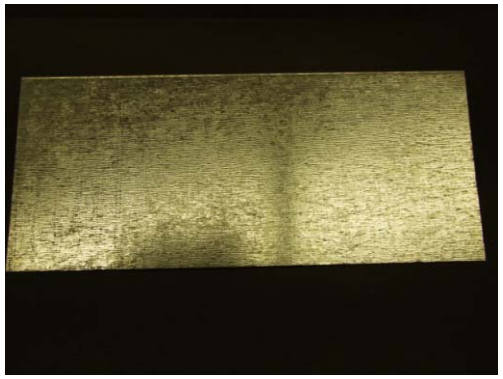
- Optical property data needed to support
  - Urban simulation
  - Obscured object detection
  - Signature modeling & analysis
- One of the major sources of error in urban simulations and signature evaluation is inaccurate reflectivity and emissivity data on building materials.
- Data on temporal changes potentially useful for age determination
- No unclassified data source available that spans Vis – LWIR



Approach

- Obtained samples of several typical building & outdoor materials; divided into:
  - One exposure set
  - One control set
- Constructed a rack to hold exposure samples
  - Samples clamped to rack to maintain orientation, etc
- Placed samples on roof of laboratory building
- Conducted reflectance and emittance measurements on a periodic basis for exposure sample set
  - 3 times per week for ~3 weeks
  - Approximately once per week thereafter (weather dependent)
  - Reflective: 0.3 – 2.5 μ
  - Emittance: 2.5 – 16.0 μ
- Recorded meteorological data using a weather station; continuous recording

Materials List



- Acrylic
- Aluminum Flashing
  - Bare
  - Black Painted
  - White Painted
  - Clear Coat Enamel
  - Galvanized

- Gray Paver
- Black Electrical Tape
- Gray Duct Tape
- Polycarbonate

- Tarps
  - Blue, all purpose
  - Brown, heavy duty
  - Canvas
  - PVC/Rubber
- Asphalt Shingle
- PVC Gutter Cover

Measurement Instruments

- **Novalynx Weather Station with Sensors and Data Logger**
  - **Pyranometer**
  - **Temperature**
  - **Relative Humidity**
  - **Rain gauge**
  - **Anemometer**
  - **Barometric pressure**

Measurement Instruments

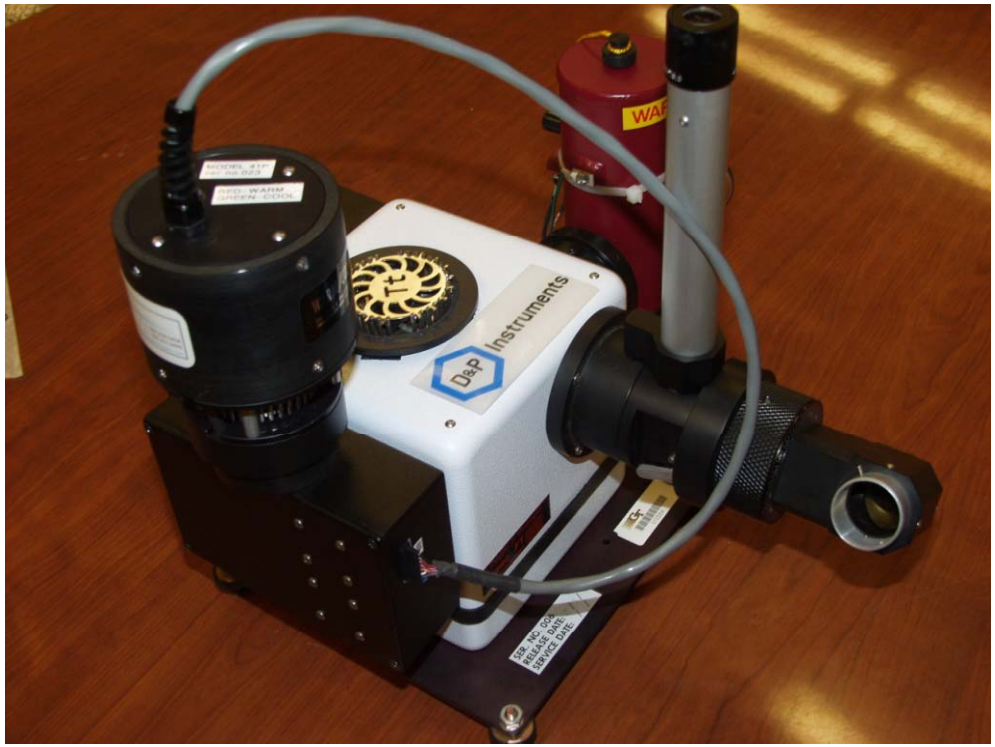
- **Cary UV-VIS-NIR Spectrometer**
  - **Internal Diffuse Reflectance Accessory used for reflectance measurements from 300 to 2500 nm.**
  - **Flat and protruding PTFE plates used for reflectance reference.**

Figure of DRA from Varian, Inc.

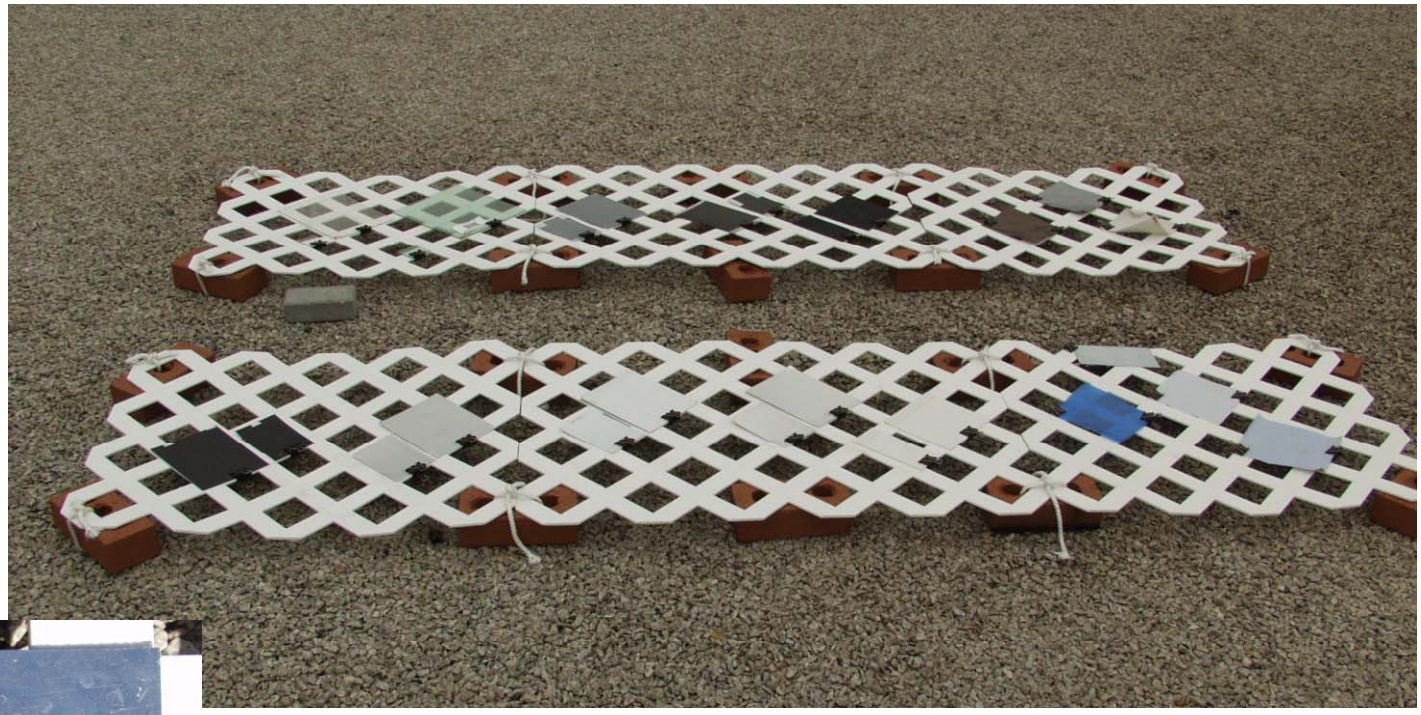
(http://www.varianinc.com)

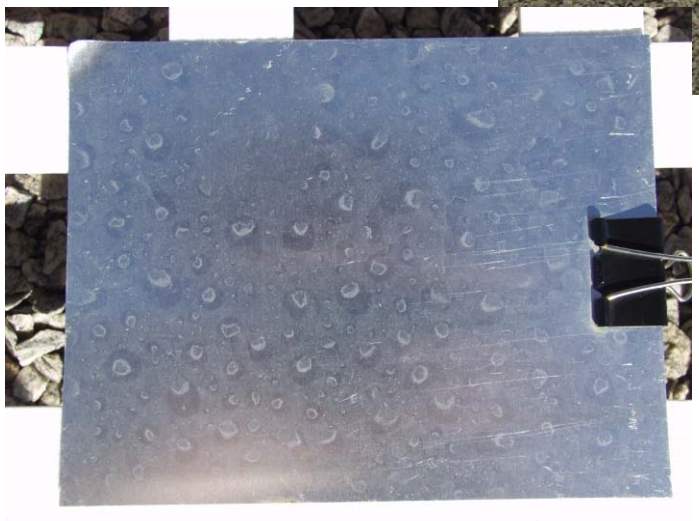
Measurement Instruments

- **D&P TurboFT FTIR Spectrometer**
  - Dual detector; InSb (~2-5 µm) & MCT (~5-16 µm)
  - Blackbody accessory for onsite calibration of radiance measurements
  - Spectral resolution 4 cm-1
  - Field portable



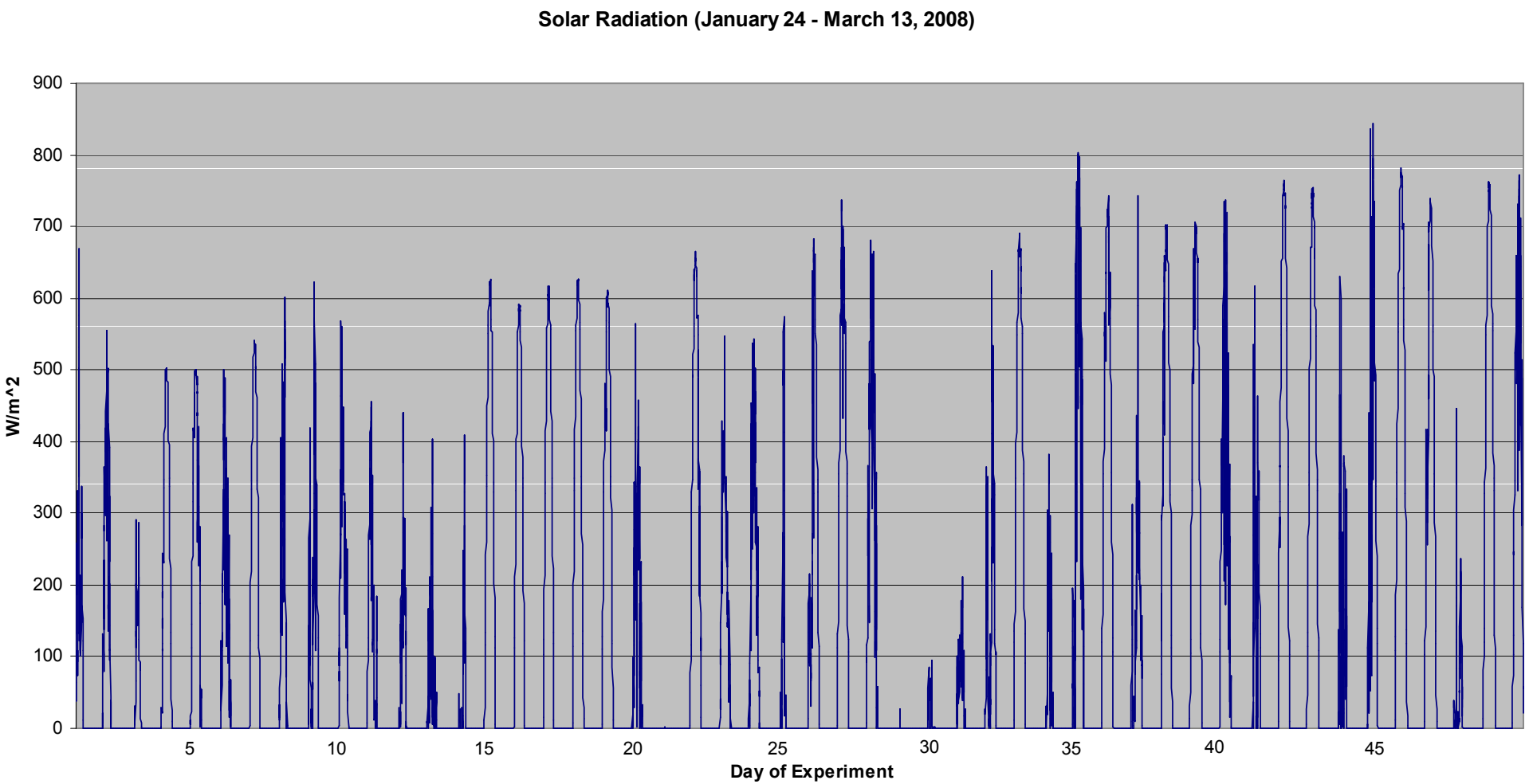
Measurement Setup



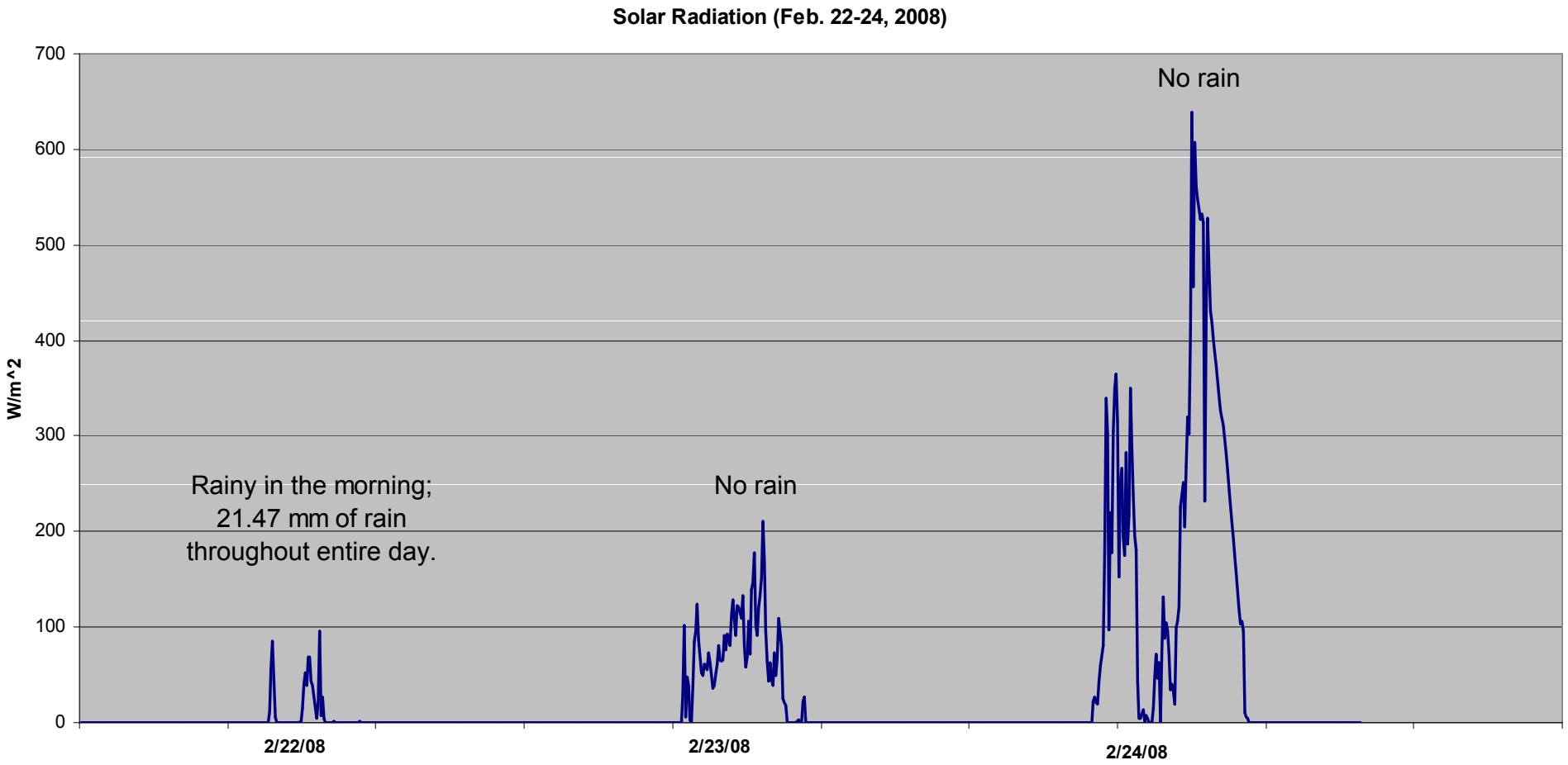


Exposure Rack and Sample Detail (Aluminum at Day 15)

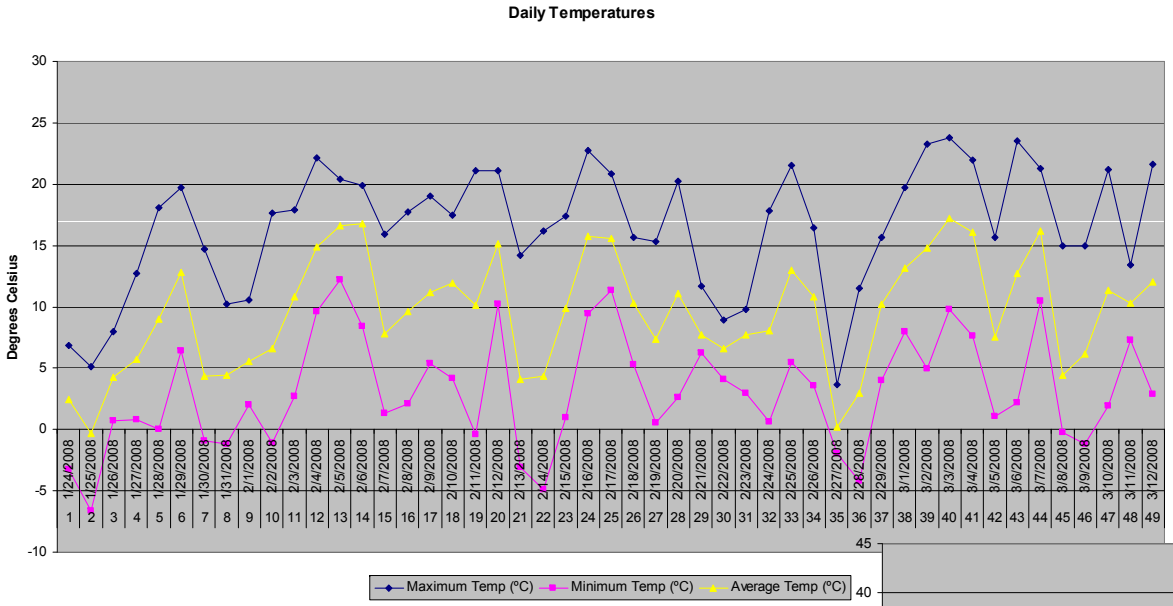
Solar Exposure – 50 days

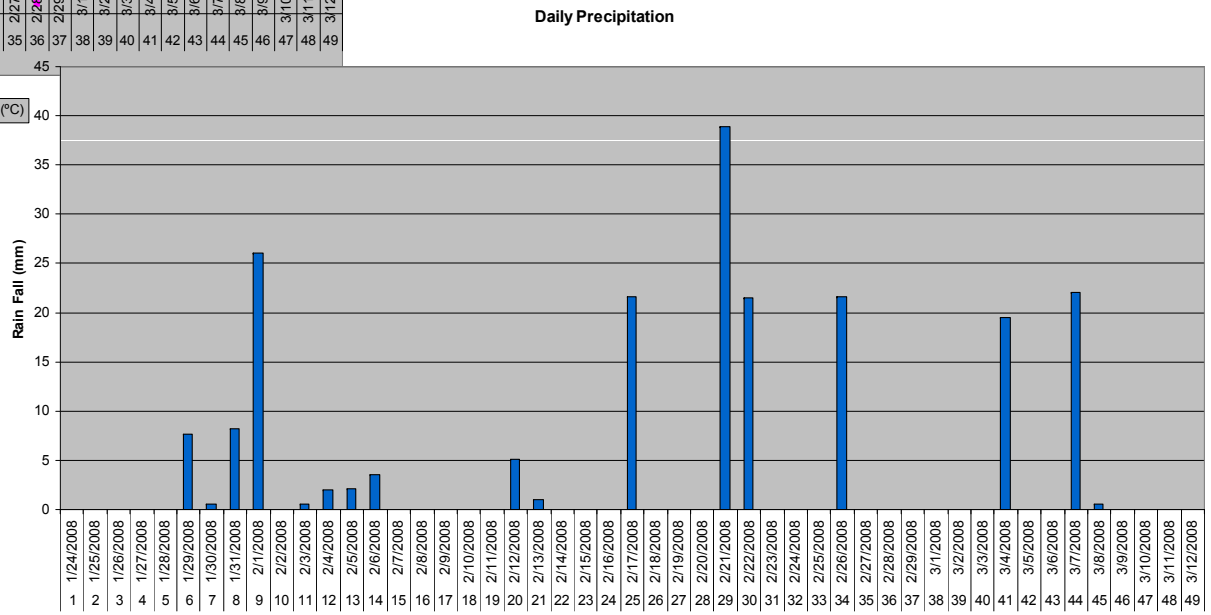


Solar Exposure – 3 days



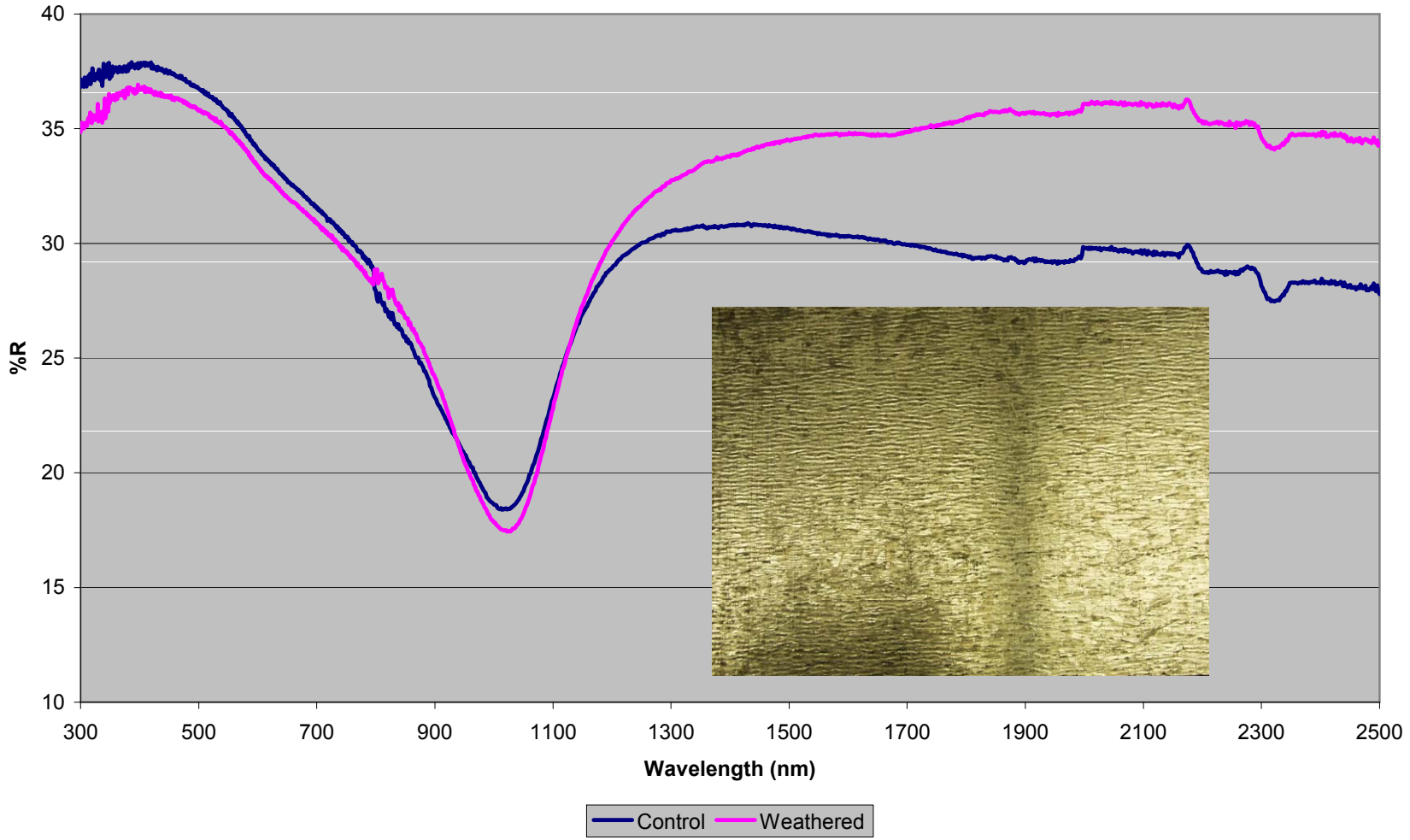
Temperature and Precipitation – 50 day Averages



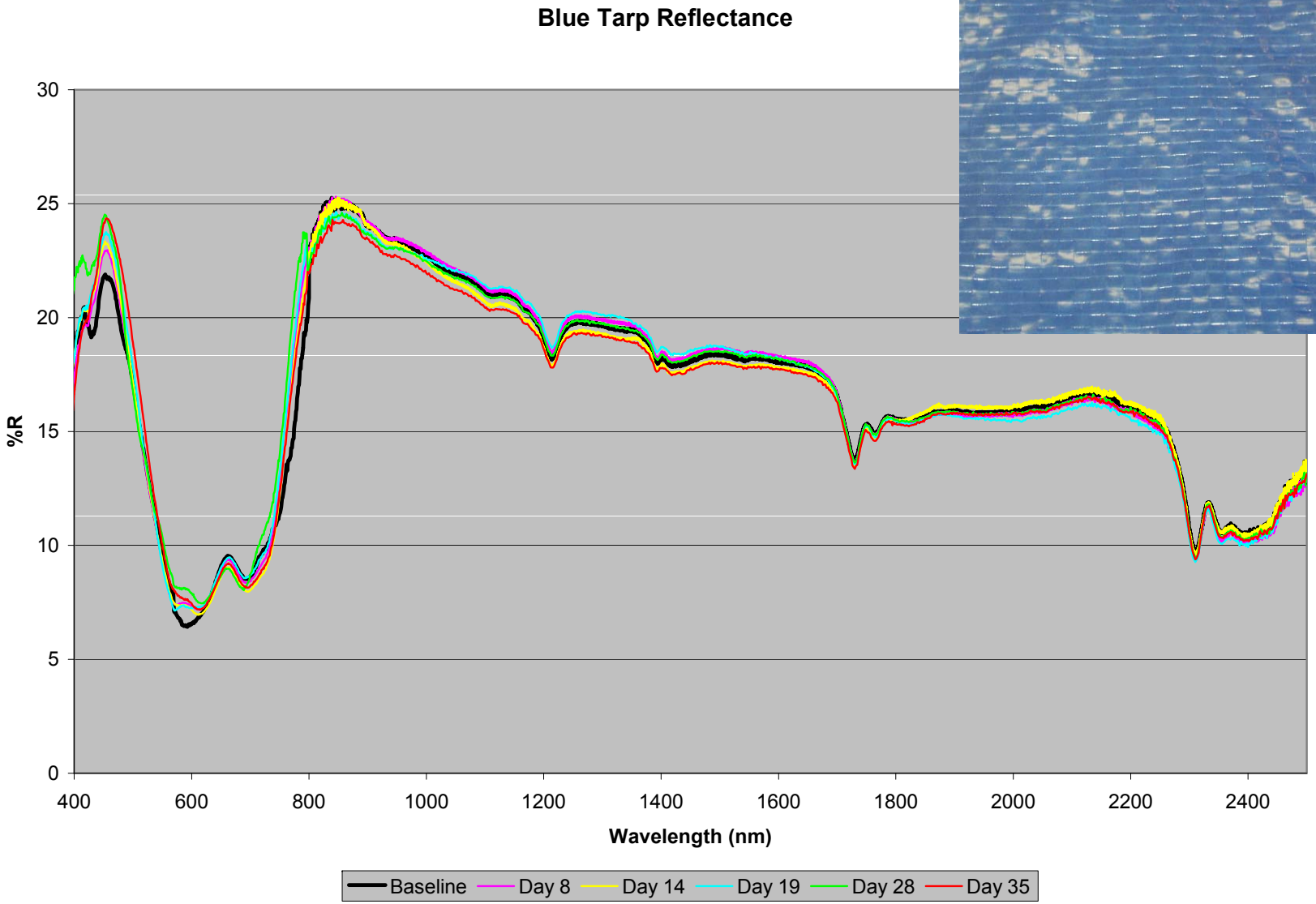


Galvanized Aluminum – Reflectance Measurements

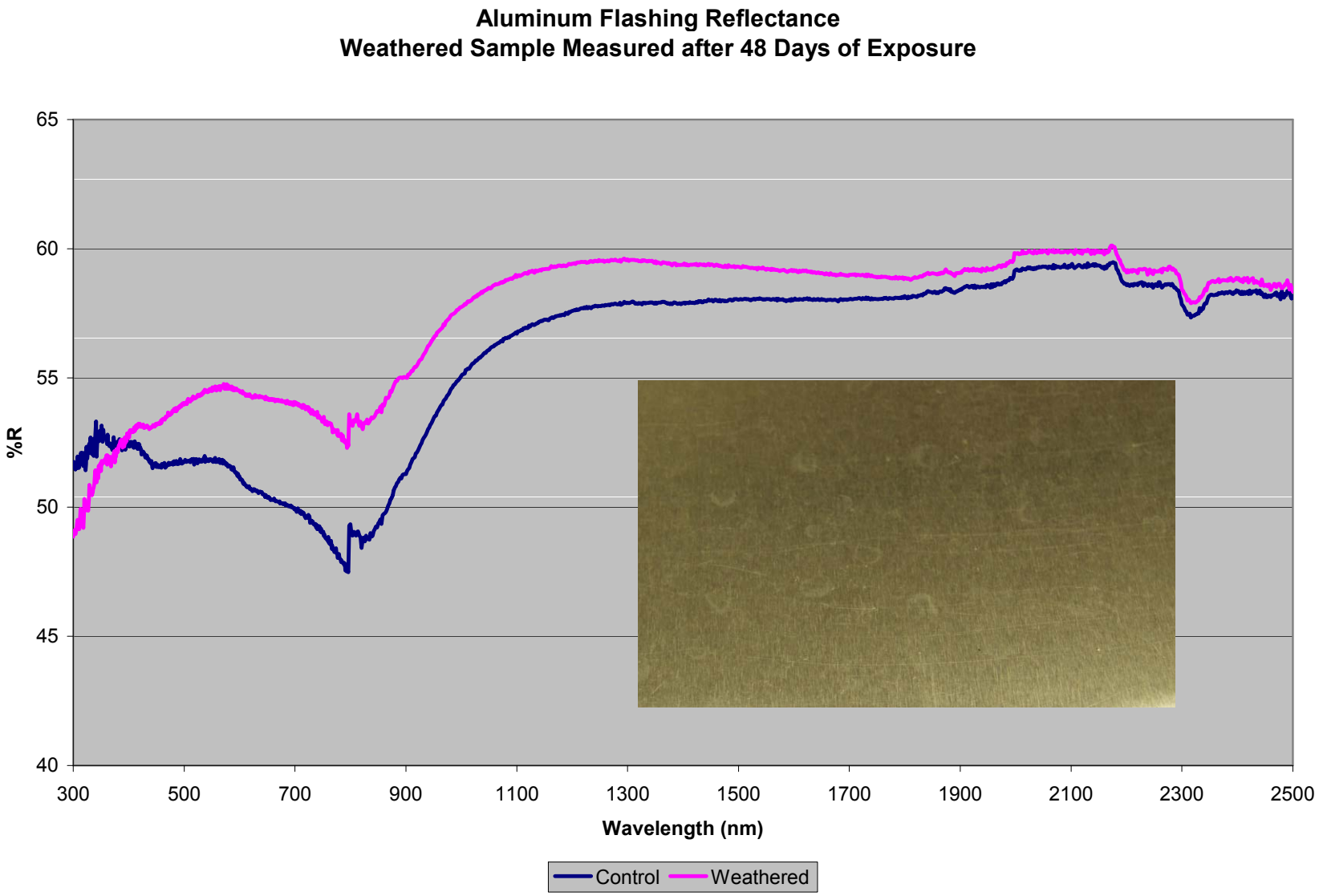
Galvanized Aluminum Reflectance
Weathered Sample Measured after 48 Days of Exposure



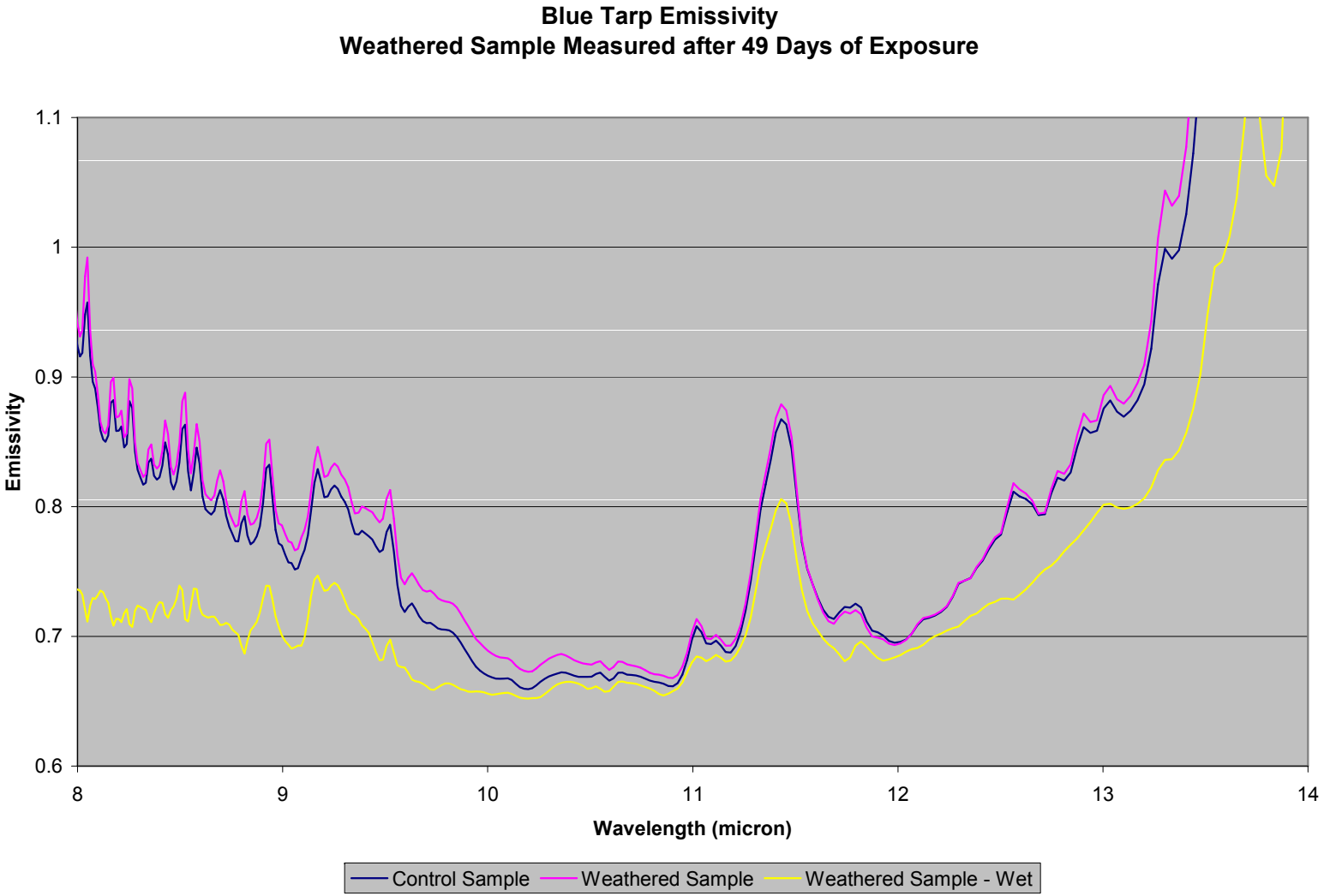
Tarp – Reflectance Measurements



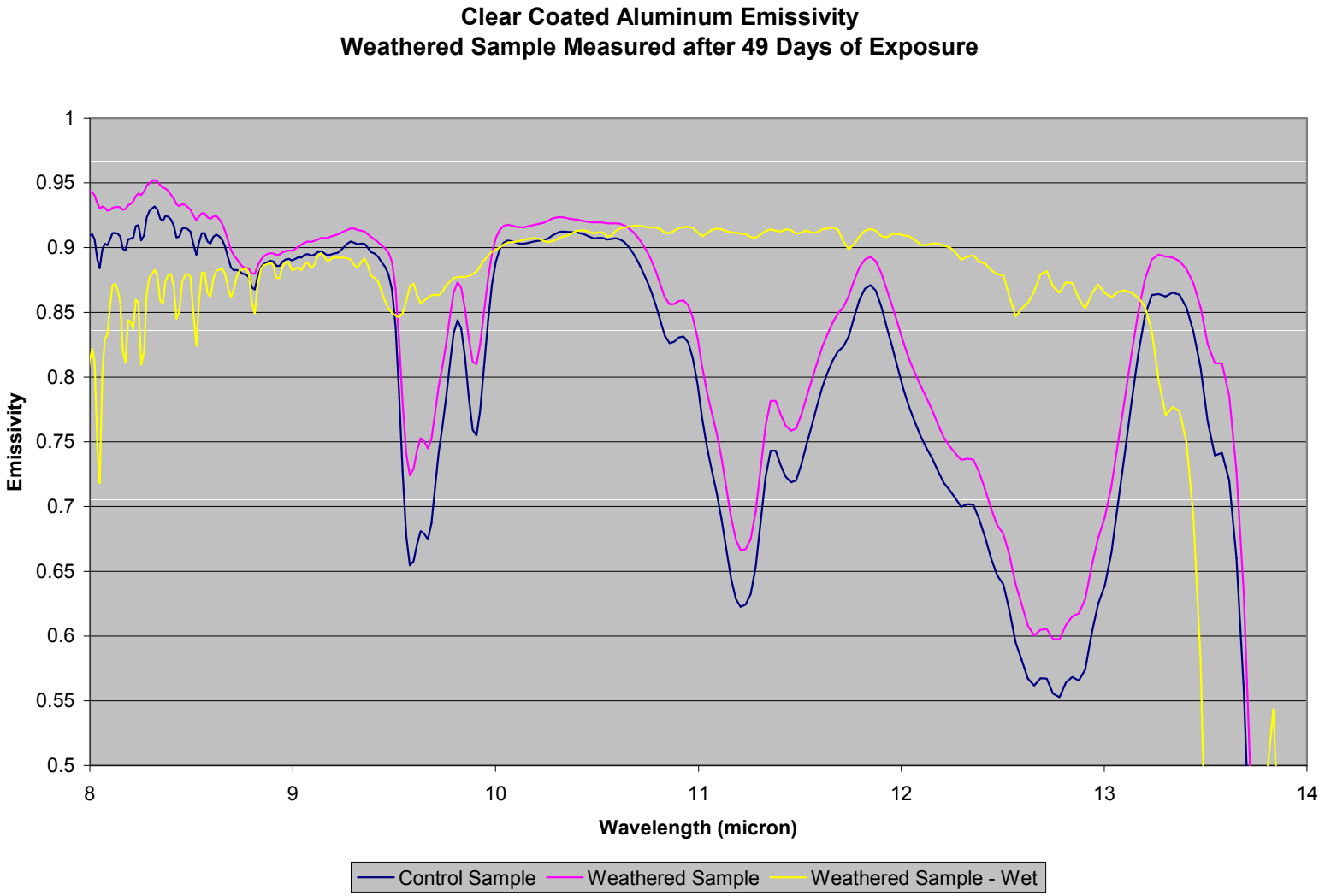
Aluminum – Reflectance Measurements



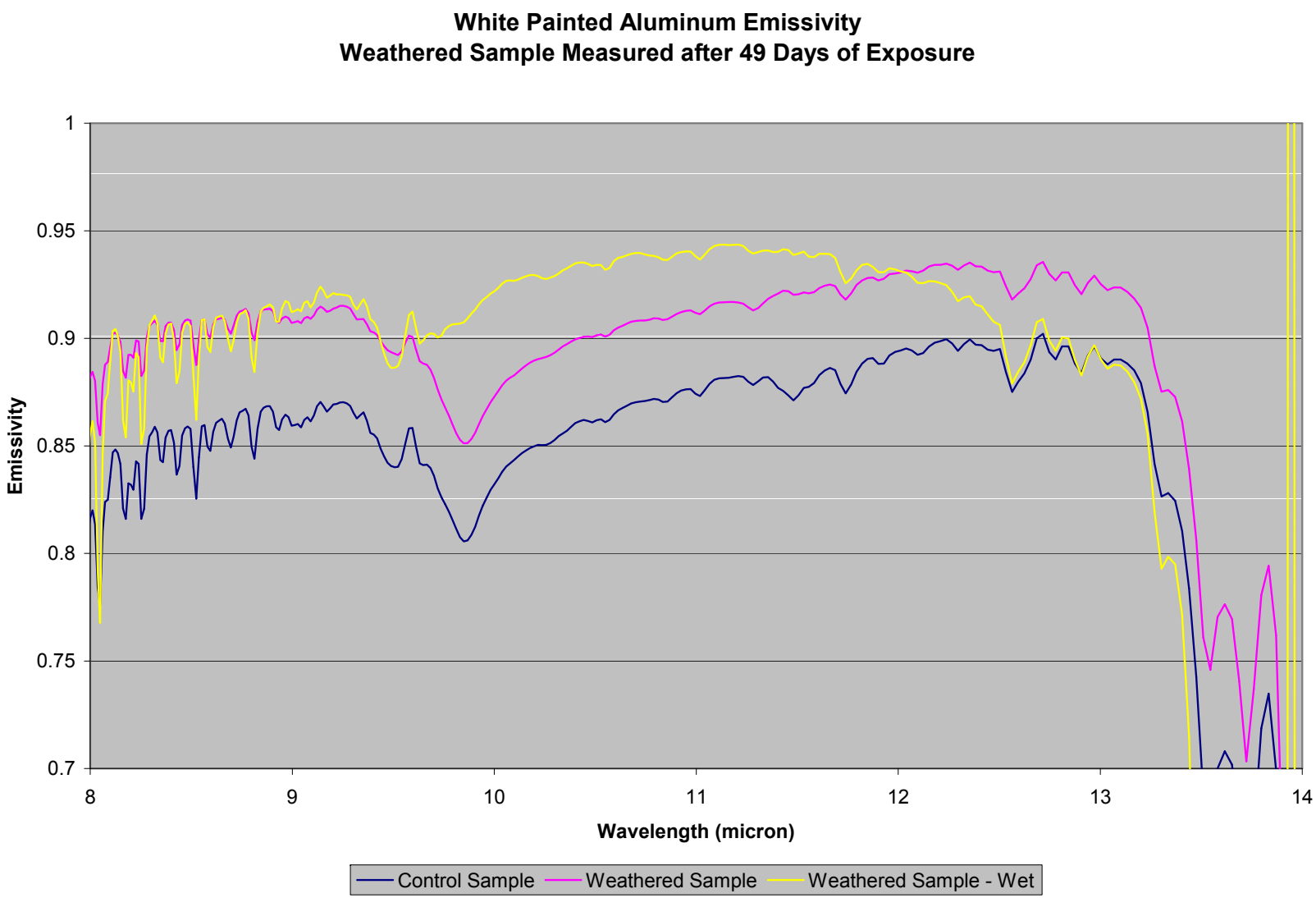
Blue tarp – Emittance Measurements



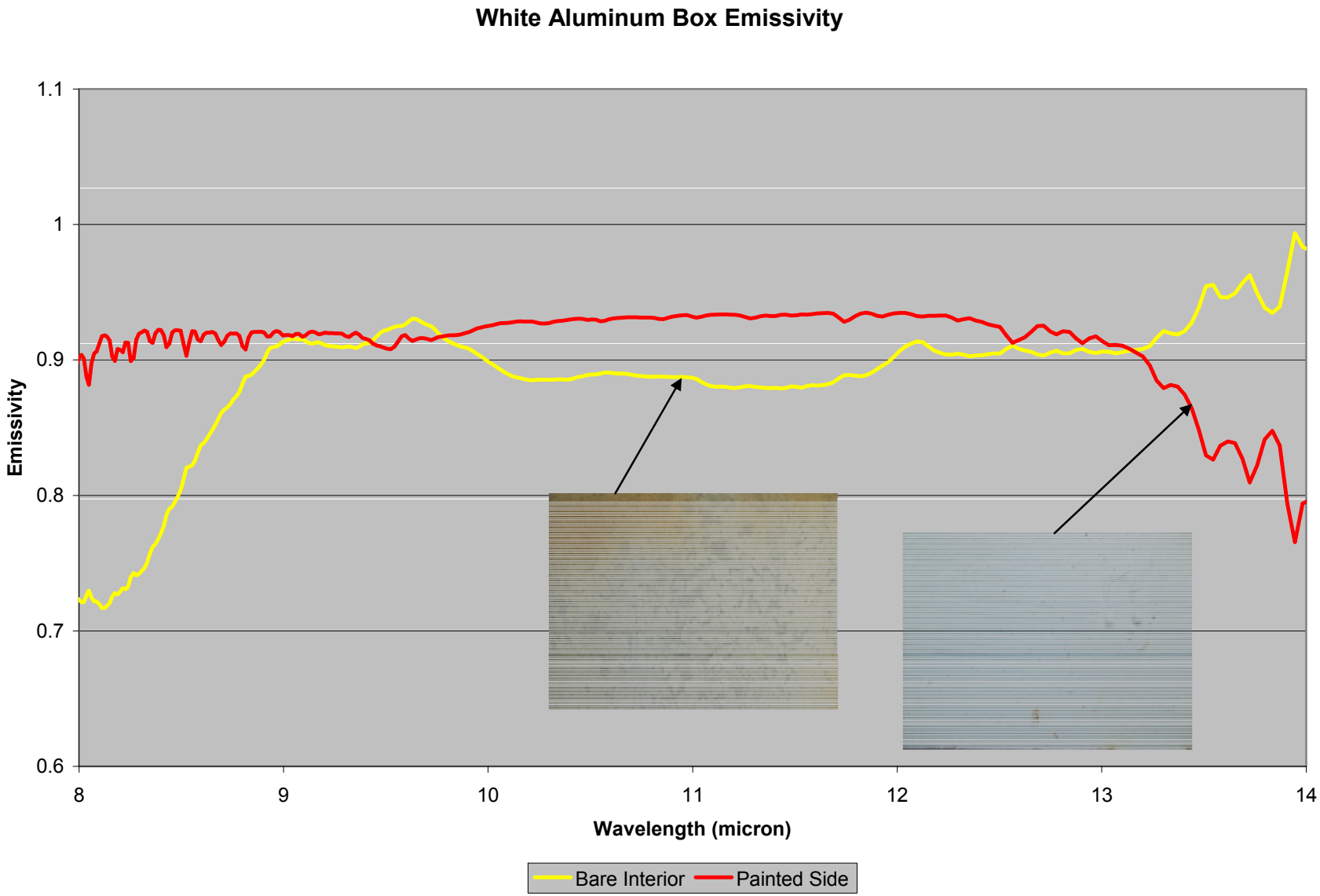
Aluminum – Emittance Measurements



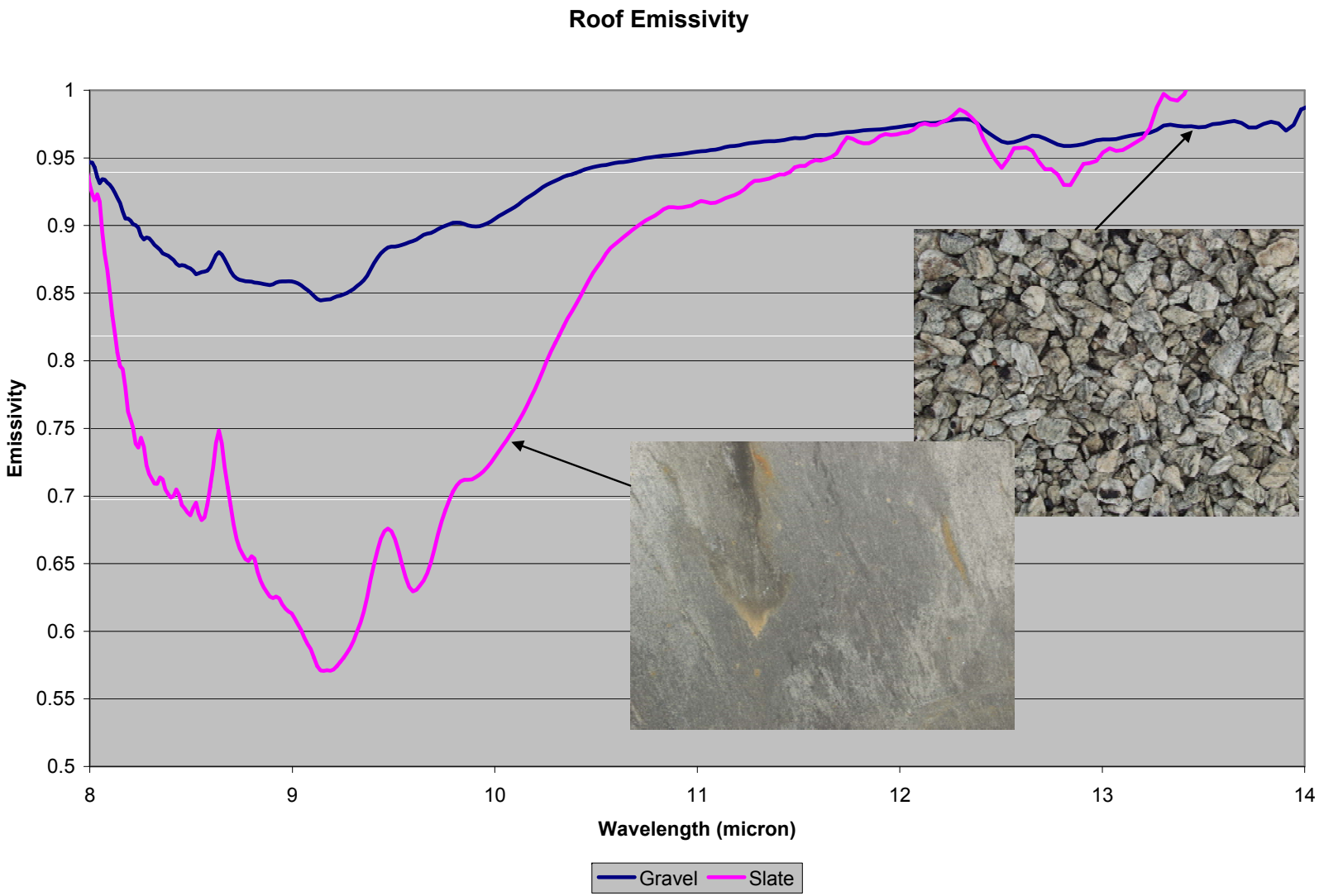
Aluminum – Emittance Measurements



Aluminum – Emittance Measurements



Resident roof materials – Emittance Measurements



Summary

- Described an exposure experiment
  - Typical building & outdoor materials selected
  - Materials placed on rack on roof
  - Reflectance and emittance measurements periodically conducted
- Observed changes in several materials; typically small changes in reflectance, emittance or both (after ~49 days)
- Issues
  - Emissivity computation sensitive to measurement conditions; errors can result without careful attention to temperature and sky measurement
- Continue experiment; reduce frequency of measurements
  - Material exposed for long period shows measurable change in emissivity
  - Widen selection of building and outdoor materials
  - Add more replicates of the samples we have

Disturbed Soil Characterization Workshop

January 15 – 17, 2008

Host: Georgia Institute of Technology

Disturbed Soil Characterization Workshop: Post-Meeting Summary

Michael Cathcart

Georgia Institute of Technology

Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XV

SPIE Defense & Security Symposium 2010

April 6, 2010 Orlando, FL

michael.cathcart@gtri.gatech.edu 404-407-6028

Purpose

- Re-examine the disturbed soil issue
  - Increasing use of buried/concealed explosives
  - Changes in military operational environments
- Address issues related to
  - Phenomenology
  - Detection
  - Exploitation
- Identify research problems and approaches
  - Support to the warfighter
  - Environmental measurements

Background

- Current targets of interest are difficult to detect
  - Frequently they are buried or obscured in some fashion.
  - Subtle changes in the local environment may be the only clues available for detection
- After several years of intensive research by various DoD organizations it was time to see what this research had yielded and where future research programs should focus.
- These objectives were particularly relevant given the current operational state of affairs in Iraq and Afghanistan (i.e., roadside bombs, IED's, UXO, etc).
- In addition, it is believed that future conflicts will need to deal with similar asymmetric warfare issues.

Programmatics

- **Workshop co-sponsored by**
  - Army Research Office
  - Night Vision and Electronic Sensors Directorate
  - US Army Corp of Engineers Engineering Research and Development Center
- **Objective:**
  - Define the basic science questions that need to be addressed across the full spectrum of military applications to fully exploit this phenomenon.
- **2-1/2 Day event hosted at Georgia Tech in 2008**
- **Participants attended from both US and foreign organizations**

Workshop Format

- Technical presentations and discussion groups
- Presentations
  - Cover multiple topics of interest (operational to exploratory)
  - Provide background on current efforts to exploit this phenomena
- Discussion groups
  - Near Surface Soil Phenomenology
  - Sensor Technology
  - Algorithm Development
- Three areas to address
  - Assess the current state of the art
  - Determine the problems confronting research in each area
  - Define approaches to overcome these problems and advance the state of the art

“Near-Surface Phenomenology” Overview

- A need exists to understanding the contrast in soil properties and processes between the disturbed and undisturbed soil.
  - Properties include mineralogy and grain size, bulk density, water content, etc.
  - They drive the electromagnetic, heat and mass transfer, and mechanical properties of the soil.
- Traditional studies concentrate in areas of agriculture, natural resource discovery, environmental impacts, structural applications, etc
  - Geologist, geophysicists, civil & construction engineers
- “Disturbed” soil is recent phenomena of interest (some efforts related to tracking, intelligence, etc have been pursued)

Direct "Disturbed Soil Phenomenology" – Summary

- ***Reststrahlen, LWIR signature changes in soil***
  - *Change in the size distribution of particles in disturbed soil*
- ***Thermal IR imaging indicates temperature distribution signatures***
  - *Changes in effective thermal conduction and heat capacity leading to dynamic changes in temperature with diurnal cycle*
- ***Differential reflectance with respect to surrounding soils as observed by imaging radar***
  - *Surface roughness changes due to soil disturbance; possible mineral content changes in the surface soil*
- ***Acoustic / Seismic signal changes***
  - *Changes in speed of sound in disturbed soil, foreign objects reflectance, and acoustical resonances*
- ***Long term temporal changes in EO-IR and RF soil signatures***
  - *Precipitation, bio-changes, temperature cycling, radiation*
- ***Frost patterns over buried articles***
  - *Thermal property changes; water pooling over buried article; moisture content difference of disturbed soil; particle size changes*
- ***Changes in blue band signatures; possible chemical signatures; moisture content changes***
  - *differential soil moisture movement, TBD*

"Near-Surface Phenomenology" – Additional Discussions

- Indirect phenomenologies to be exploited
  - Variation of size envelopes (landmines, tire tracks, fields)
  - Manmade vs. natural
- Temporal behavior
- Debate on definition of disturbed soil, disturbances, etc
  - Recommended name change: "Disturbed surface"
  - Accounts for soil, vegetated, & rocky surfaces

“Near Surface Phenomenology” Research Barriers

- **Barriers**
  - Defining exactly what is meant by “disturbed soil”
  - Identifying the processes that create disturbed soil
  - Determining which soil properties to measure
  - Determining anticipated signature changes
  - Field verification of laboratory measurements & models
- **Two primary research directions**
  - Fundamental studies into the general nature of disturbed soils
  - Applied investigations that address military operational needs
- **General recommendations**
  - Move “back to the basics”
  - Employ new methods/approaches

“Sensor Technology” Overview

- **Various sensors modalities were reviewed.**
- **Limited field data for most sensing modalities.**
- **No operational sensor developed specifically for disturbed soil detection.**

Ground Penetrating Radar	3rd Gen Image Intensifier
Synthetic Aperture Radar	Acoustic
V/NIR Imagers	Seismic
SWIR Imagers	Chemical
MWIR Imagers	Animal (odor detection)
LWIR Imagers	Human vision
LIDAR	Computer Vision
Hyperspectral Imagers	Multi-spectral Imagers

“Sensor Technology” Summary

- **Hyperspectral**
  - LWIR reststrahlen feature exploitation
  - Large data volume
- **Multispectral**
  - Band selection
- **Human vision & visual**
  - Color differences, texture changes, size
  - Human detection demonstrated
  - Automated detection difficult
- **Lidar**
  - Limited utility demonstrated
- **Thermal**
  - Temperature differential
  - Diurnal crossover impact
- **Radar detection**
  - GPR shows differential reflection effect
  - Higher frequency for size changes
- **Acoustic/seismic**
  - limited to a few controlled field studies

“Sensor Technology” – Additional Notes

- Limited field studies in most cases to properly assess sensing modality
- False alarm understanding is needed.
- Optical techniques sample near surface only
  - Disturbance can be disguised (move soil, cover with other soil)
- Diurnal thermal crossover is a function of several factors
  - Soil moisture, time of day, amount of energy received, degree of soil disturbance, soil hydraulic properties, soil thermal properties, etc
  - Linkage between these not well understood
- Temporal changes in soil signatures

"Sensor Technology" – Research Barriers

- **Three areas of research barriers: technology, operational, programmatic**
- **Technology**

Barrier	Sensor Type	Comments
Pixels on target	Imaging	More pixels on target yields improved detection/identification
Bandwidth	RF	Deeper penetration with lower frequencies
Spectral Resolution	Spectral	Better separation of spectral features
Acquisition Range	All	Disturbed soil signals may be faint
Data logging	All	Recording of relevant ancillary data to improve sensor performance
Data transmission	All	Current communication bandwidths limit transmitted data volume
Post-processing	All	Computational resources and algorithm complexity limit speed of calculations.

“Sensor Technology” – Research Barriers (cont.)

- **Operational**
  - Relate to environmental factors that limit sensor performance
  - Propagation, illumination conditions (differences, shadowing), diurnal changes (heating, moisture, etc)
- **Programmatic**
  - High sensor development costs and deployment timeline
  - Existing sensors employed for data collections
  - New sensors specifically for “disturbed soil” seems impractical

“Algorithm” – Summary

- **Significant work on exploitation of optical and LWIR data (NVESD & DARPA)**
  - **Thermal, spatial, and spectral discriminants**
- **Landmine detection has been the primary objective of research/technology efforts**
  - **Size, shape, and texture features provide potential discrimination features**
- **WAAMD program**
  - **Investigated optical, infrared, and radar sensors (singly and data fusion)**
  - **Established performance levels; graded performance of various algorithms**
  - **Signature-based and anomaly-based methods evaluated**

"Algorithm" – Additional Notes

- False alarms
  - Phenomenology establishes a feature
  - Marginal separation exists between the disturbed soil and background spectral features
  - Natural environmental variations lead to significant false alarms
- Results focused on landmine detection
  - Spatial characteristics useful
- Few studies on detecting "disturbed soil" from other activities: digging, vehicles, walking, etc
- Operational experience indicates a need to re-examine use of thermal sensing

"Algorithm" – Research Barriers

- Well-documented field data in operationally-relevant environments
  - Algorithms frequently optimized at test sites that do not replicate the anticipated operational locations
- Well-documented data sets from operational environments
- High spatial resolution data (spatial differences)
- High dynamic range data (differentiate small signals)
- Multi-modal data sets to enable more extensive studies of sensor and data fusion

Summary

- **Phenomenology research**
  - Investigate basic physics of the disturbance process
  - Identify the impact of disturbance on all soil properties
  - Identify the observables in all sensor bands
- **Sensor research**
  - Collect multi-modal sensor data
  - Improve sensor characteristics (i.e., sensitivity, ground sampling distance)
  - Collect additional environmental data
- **Algorithm research**
  - Exploit multi-sensor data
  - Develop approaches that use scene contextual information
  - Investigate methods to tune/select algorithms based on operational environment

Spectral Automatic Target Detection/Recognition (ATD/R)

University of Maryland

Research Summary

Personnel

- Prof. Rama Chellappa:
  - Principal Investigator
  - Graduate Students: Joshua Broadwater, Hirsh Goldberg, and Dalton Rosario
- Dr. Reuven Meth:
  - Faculty Research Assistant
  - Worked on early change detection and subpixel detection algorithms
- Dr. Joshua Broadwater:
  - Graduated May 2007 with Ph.D. in Electrical and Computer Engineering
  - Developed subpixel detectors and adaptive threshold estimates
- Mr. Hirsh Goldberg:
  - Graduated May 2007 with Master's in Electrical and Computer Engineering
  - Worked with Dr. Nasrabadi at Army Research Laboratory on Kernel Anomaly Detectors
- Mr. Dalton Rosario:
  - Ph.D. Candidate and ARL staff member
  - Developed semi-parametric method for anomaly detection

Collaborations

- **Army Research Laboratory**
  - Worked with Dr. Nasrabadi investigating kernel anomaly detection algorithms
  - Published joint papers in SPIE and IEEE Trans. Geo Sci.
- **Night Vision and Electronic Sensors Directorate**
  - Received VIS/NIR/SWIR HSI data from Ms. Miranda Schatten and the Wide Area Airborne Mine Detection (WAAMD) program
  - Produced “blue ribbon” detection results
- **Clark Atlanta University**
  - Worked with Dr. Lance Kaplan and Dr. Peter Molnar
  - Provided detection results for use in their algorithm development
- **The Johns Hopkins University Applied Physics Laboratory**
  - Worked with Dr. Amit Banerjee to develop kernel based methods for endmember and abundance estimates.
- **Rochester Institute of Technology**
  - Provided feedback to Dr. David Messinger on DIRSIG results for both VIS/NIR/SWIR and LWIR imagery
- **University of Florida**
  - Provided detection results to Dr. Paul Gader for use in his choquet fusion algorithms.
- **University of Hawaii**
  - Received LWIR HSI imagery from Dr. Tim Williams for buried mine and IED detection

Anomaly Detection

- Semi-parametric anomaly detection
  - Uses a logistic regression model for the background class
  - Uses an exponentially twisted density function to specify a threshold for anomaly detection
  - Work presented at SPIE and IEEE Aerospace conferences.
- Kernel-based anomaly detection
  - Compare linear based methods to their kernel counterparts
  - Research found kernel methods can outperform their linear counterparts
  - Kernel methods have the added difficulty of identifying the correct kernel and kernel parameters (i.e., not all kernels and corresponding parameters are better than linear methods).
  - Research continues on identifying the “best” kernel and parameter settings
  - Work presented at SPIE and in IEEE Trans. on Geosci.

Subpixel Detection

- In-scene fully automated atmospheric correction
  - Based on Piech and Walker (1974)
  - Automatically finds signatures in the scene that are impulse functions of the atmospheric profile
  - Estimates atmospheric parameters from these signatures
  - Work shows the method provides signatures better matched to true signatures found in imagery than MODTRAN in a number of cases
  - Early work presented at IGARSS 2005 and to be submitted to JOSA 2008.
- Background Effects
  - Show the number of endmembers can greatly influence the ability of a structured detector
  - Developed a method to identify the number of endmembers for a detector based on a synthetic mixed pixel.
  - Developed a kernel method to extract endmembers and abundances for future kernel detectors
  - Work presented at IGARSS 2007 and submitted to IGARSS 2008.

Subpixel Detection (cont.)

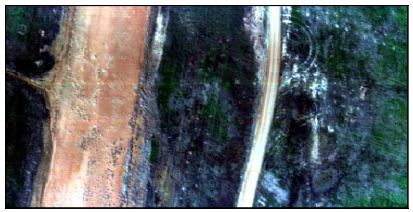
- Hybrid Detectors
  - Combine structured detectors with physically meaningful estimates
    - Endmembers represent true spectral signatures
    - Abundances are non-negative
    - Abundances sum to one
  - Detectors have three advantages over other methods
    - Partially insensitive to the number of endmembers used
    - Provide improved performance for “weak” targets (e.g. low reflectance, similar to background)
    - Have CFAR-like properties suppressing the background into similar ranges of values better than standard CFAR methods.
  - Initial work presented at IGARSS 2004 and MSS 2005.
  - Final work published in IEEE Trans. Pattern Analysis and Machine Intelligence.

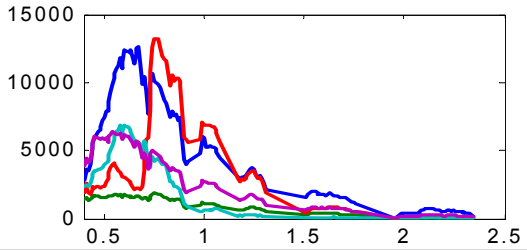
Adaptive Thresholds

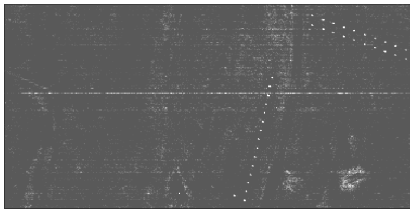
- Problem with “CFAR” detectors
  - Classical CFAR detectors are based on Gaussianity assumptions which are not found in HSI data
  - Theoretical thresholds found using such distributions do not match those seen in practice
  - Thresholds are only applicable to CFAR detectors ignoring non-parametric methods
- Extreme Value Theory
  - Fisher-Tippett theorem is the “Central Limit Theorem” for tail distributions.
  - Can model both CFAR detector and non-parametric detector results.
  - Can set a false alarm rate even in the presence of outliers (e.g. targets)
  - Initial work presented at ICASSP 2006.
  - Final work submitted to NVESD for approval to IEEE Trans. on Signal Processing 2008.

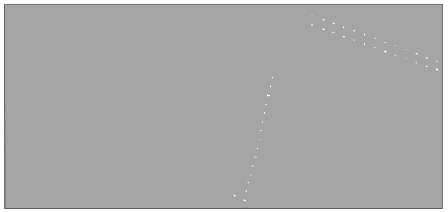
University of Maryland MURI Summary

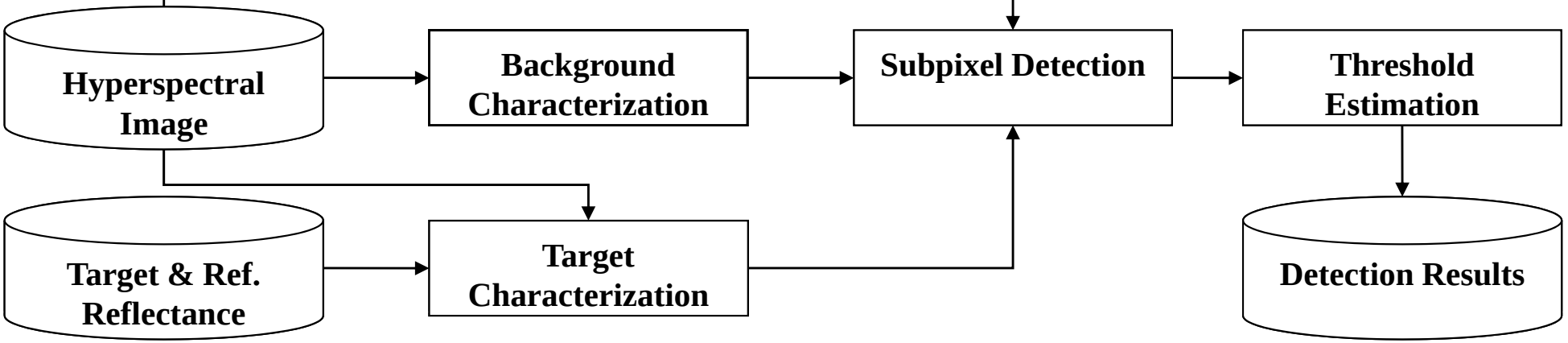
The University of Maryland developed an end-to-end subpixel detection system that uses only the hyperspectral image, a reference signature, and the desired target signature. Results show near optimum performance.

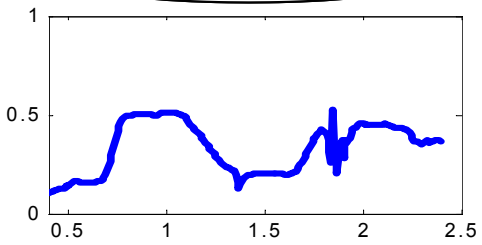


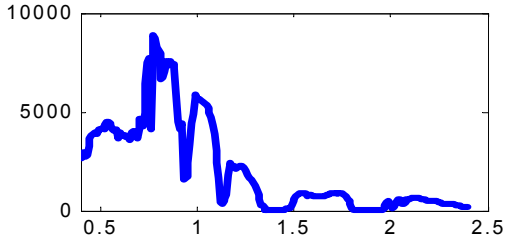












Improvements were made in each one of these areas leading to cumulative results that were beyond WAAMD expectations (e.g. P_d of 0.9 at 10^{-5} fa/m^2 across multiple targets and images).

MURI Related Publications

- J. Broadwater and R. Chellappa, "An Adaptive Threshold Method via Extreme Value Theory," under review by NVESD for *IEEE Trans. on Signal Processing*, 2008.
- J. Broadwater, "Effects of Endmember Dimensionality on Subpixel Detection Performance," submitted to *2008 IEEE International Geoscience and Remote Sensing Symposium*, Boston, MA, July 2008.
- J. Broadwater and R. Chellappa, "Hybrid Detectors for Subpixel Targets," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 29, 2007, pp. 1891-1903.
- J. Broadwater, A. Banerjee, P. Burlina, and R. Chellappa, "Kernel fully constrained least squares abundance estimates," *Proc. of the 2007 IEEE Geoscience and Remote Sensing Symposium*, Barcelona, Spain, 2007, pp. 4041-4044.
- A. Banerjee, P. Burlina, and J. Broadwater, "A machine learning approach for finding hyperspectral endmembers," *Proc. of the 2007 IEEE Geoscience and Remote Sensing Symposium*, Barcelona, Spain, 2007, pp. 3817-3820.
- J. Broadwater, *Physics-Based Detection of Subpixel Targets in Hyperspectral Imagery*, Ph.D. Dissertation, University of Maryland, College Park, MD, April 2007.
- H. Goldberg, H. Kwon, and N. Nasrabadi, "Kernel Eigenspace Separation Transform for Subspace Anomaly Detection in Hyperspectral Imagery," *IEEE Geoscience and Remote Sensing Letters,* vol. 4, 2007, pp. 581-585.
- H. Goldberg and N. Nasrabadi, "A comparative study of linear and nonlinear anomaly detectors for hyperspectral imagery," *Proceedings of SPIE*, vol. 6565, 2007.
- H. Goldberg, *A Performance Characterization of Kernel-Based Algorithms for Anomaly Detection in Hyperspectral Imagery*, Master's Dissertation, University of Maryland, College Park, MD, April 2007.

MURI Related Publications (cont.)

- J. Broadwater and R. Chellappa, "Physics-based detectors applied to long-wave infrared hyperspectral data," *Proceedings of the 2006 Army Science Conference*, Orlando, FL, November 2006.
- J. Broadwater and R. Chellappa, "An Adaptive Threshold Method for Hyperspectral Target Detection," *Proc. of the 2006 Acoustics, Speech and Signal Processing (ICASSP) Conference*, Toulouse, France, 2006, pp. 1201-1204.
- D. Rosario, "A Nonparametric F-Distribution Anomaly Detector for Hyperspectral Imagery," *Proc. of the 2005 IEEE Aerospace Conference*, 2005, pp. 2022-2029.
- J. Broadwater, R. Meth, and R. Chellappa, "Average relative radiance transform for subpixel detection," *Proc. of the 2005 IEEE International Geoscience and Remote Sensing Symposium,* Seoul, South Korea, 2005, pp. 3565-3568.
- J. Broadwater and A. Banerjee, "A Hybrid Method for Automatic Detection of Sub-pixel Targets," *Proc. of the MSS CC&D Conference*, SPAWAR Charleston, SC, 17 February 2005, NOFORN/ITAR.
- D. Rosario, "A semi-parametric approach using the discriminant metric SAM (spectral angle mapper)," *Proceedings of SPIE*, vol. 5426, 2004, p. 58.
- J.B. Broadwater, R. Meth, and R. Chellappa, "Dimensionality Estimation in Hyperspectral Imagery Using Minimum Description Length," *Proceedings of the 2004 Army Science Conference*, Orlando, FL, November 2004.
- J. Broadwater, R. Meth, and R. Chellappa, "A hybrid algorithm for subpixel detection in hyperspectral imagery," *Proc. of the 2004 IEEE Geoscience and Remote Sensing Symposium,* Anchorage, AK, vol. 3, 2004, pp. 1601-1604.
- D. Rosario, "Highly effective logistic regression model for signal (anomaly) detection," *Proc. of the 2004 Acoustics, Speech, and Signal Processing (ICASSP) Conference*, vol. 5, 2004, pp. 817-820.

Spectral Automatic Target Detection/Recognition (ATD/R)

2003 Research Efforts

Spectral ATD/R Goals

Goal 1:
Development of enhanced HSI target detection methodologies

Goal 2:
Utilization of phenomenology to enhance HSI ATD/R

2003 Research Topic Outline

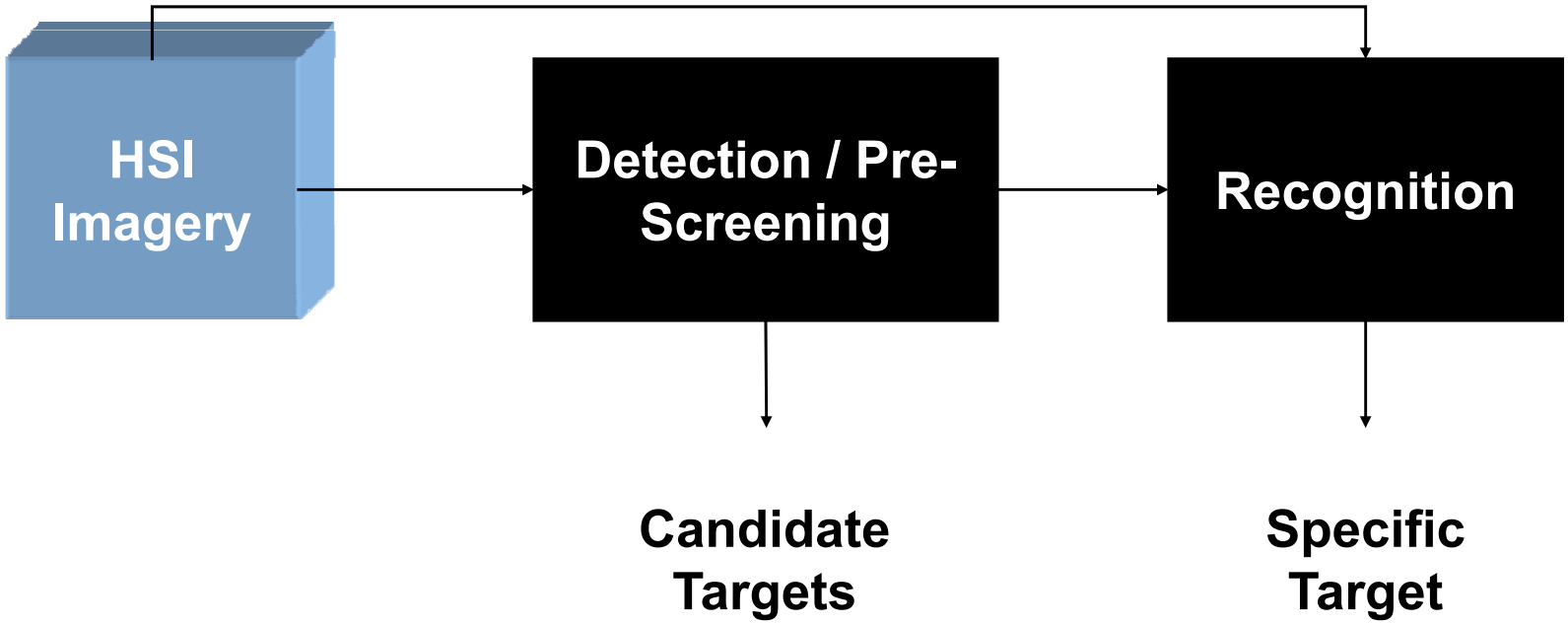
- Detection/recognition methodology
- Anomaly detection
  - RX, GMM, Endmember Stochastic
  - Fusion
- Endmember extraction
- Recognition / classification

RX – Reed & Xiaoli
GMM – Gaussian Mixture Model

Impact of Research

- Focus of attention / data reduction
- Object discrimination (general class)
- Object recognition (specific type)
- Mine detection

Terminology / Methodology



HSI Target Detection/Recognition

- Anomaly based (detection)
  - RX
  - GMM
  - Endmember Stochastic
- Spectral signature based (recognition)

RX Anomaly Detection

- Generalized Likelihood Ratio Test (GLRT)
  - Kelly, MIT-LL, ’89
  - Reed & (Xiaoli) Yu (RX), IEEE Trans. ASSP, ’90
- Gaussian background statistics
- Additive deterministic target
- Background covariance same for target and non-target hypotheses

RX Algorithm

- Local Normal Model:

$$H_0 : x \sim N(\mu, \Sigma_x),\ H_1 : x \sim N(s, \Sigma_x)$$

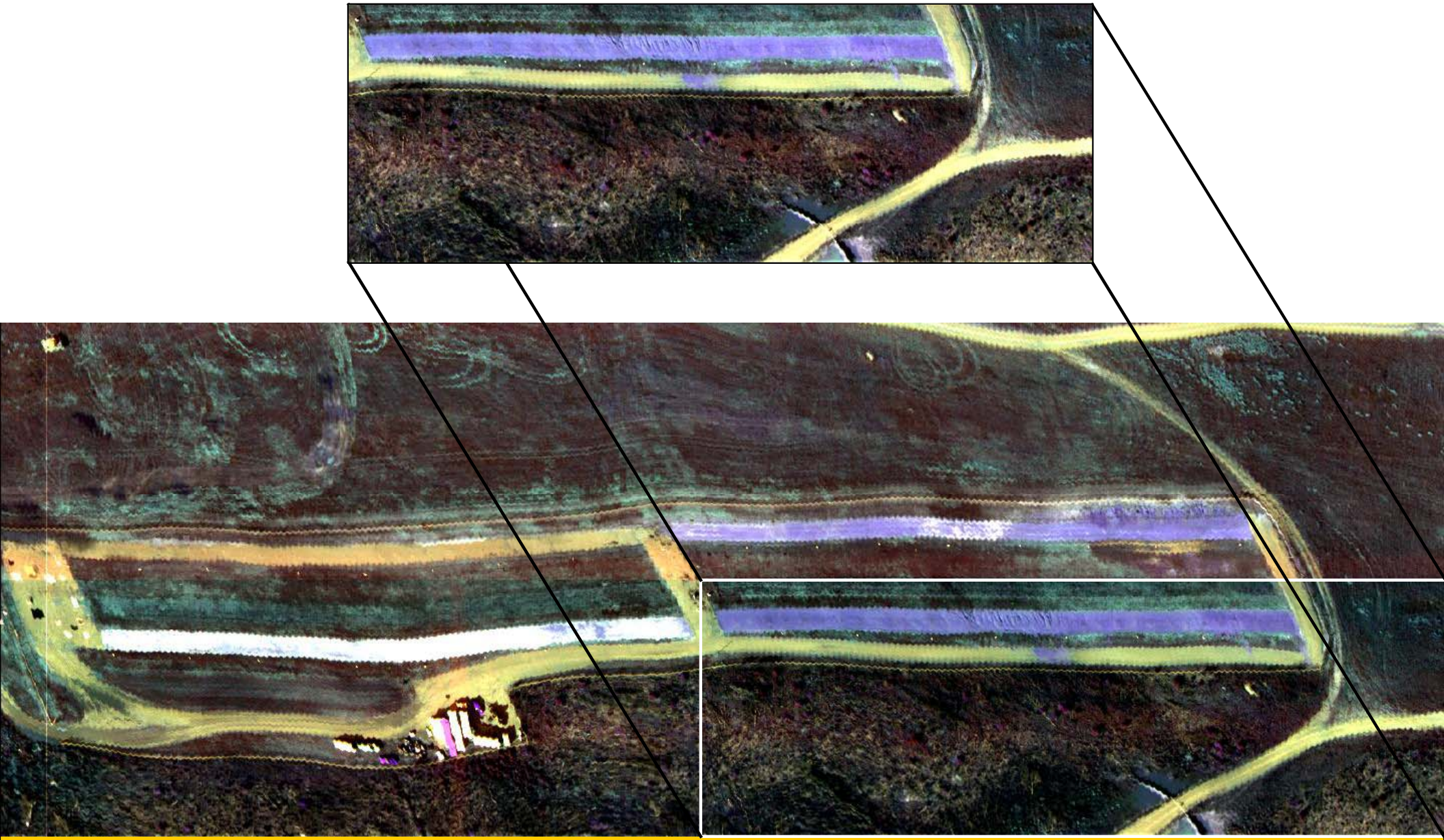
- Using GLRT formulation results in

$$RX(x) = (x-\mu)^T \left(\frac{N}{N+1}\Sigma_x + \frac{1}{N+1}(x-\mu)(x-\mu)^T \right)^{-1} (x-\mu)$$

- For large N simplifies to the more common formulation:

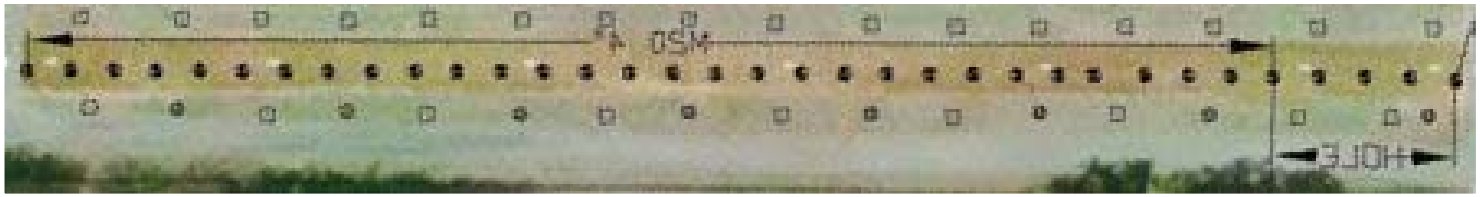
$$RX_L(x) = (x-\mu)^T (\Sigma_x)^{-1} (x-\mu) \underset{H_0}{\overset{H_1}{\gtrless}} \tau$$

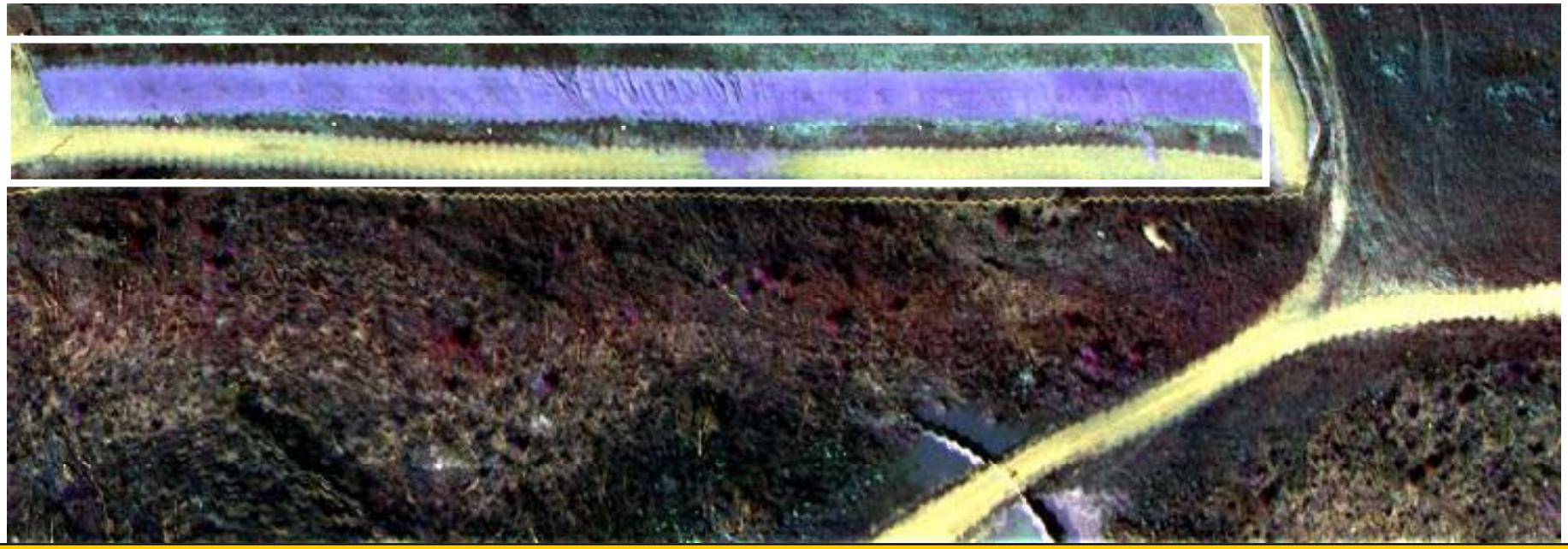
STI False Color Image



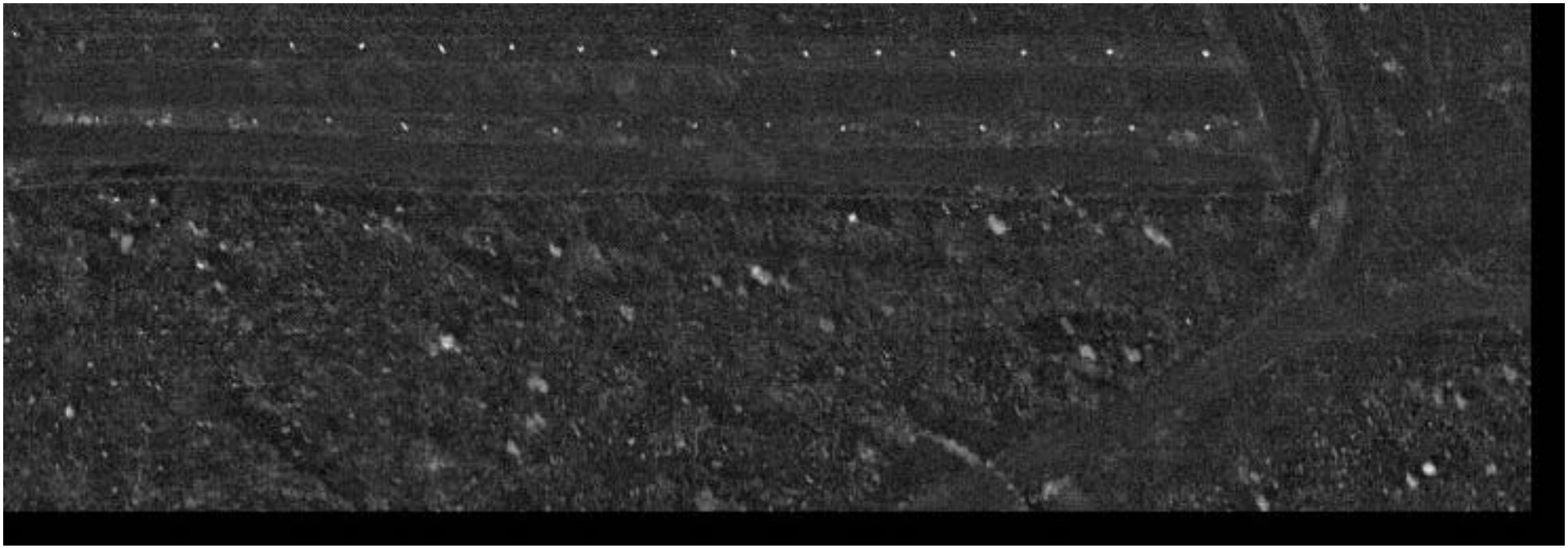
Ground Truth

Ground Truth





RX Detection Statistic



RX

Gaussian Clutter Model

Is single Guassian sufficient to properly model the background?

Not Globally!

GMM-based Anomaly Detection

- Enhanced background statistical model
  - Stein et al., IEEE Signal Processing, '02
- Mixture model (since clutter is a mixture)
  - Background is often multi-modal
  - May apply globally and/or locally
- Global estimate of clutter statistics
  - Adaptive updating of background model
- Anomaly detection based on clutter model
- Gaussian mixture model (GMM)
  - Automated/unsupervised estimation via EM
  - May generalize to kernel-based estimation

Automated Background Modeling/Segmentation

- Gaussian mixture model (GMM)

$$f(x \mid \theta) = \sum_{j=1}^{k} \alpha_j \frac{1}{\sqrt{(2\pi)^d \left|\Sigma_j\right|}} \exp\left\{-\frac{1}{2}\left(x-\mu_j\right)^T \Sigma_j^{-1}\left(x-\mu_j\right)\right\}$$

- Parameters $\theta=\{\alpha_j,\mu_j,\Sigma_j\}$, $j=1,\ldots,k$
- Initialization via K-means, search via Expectation Maximization (EM)

EM Estimation

- Expectation

$$w_{tj} = \frac{\alpha_j f\left(x_t \mid \mu_j \Sigma_j\right)}{\sum_{i=1}^{k} \alpha_i f\left(x_t \mid \mu_i \Sigma_i\right)}, \quad j = 1,\ldots,k, \quad t = 1,\ldots,n$$

- Maximization

$$\widehat{\alpha}_j \leftarrow \frac{1}{n}\sum_{t=1}^{n} w_{tj}$$

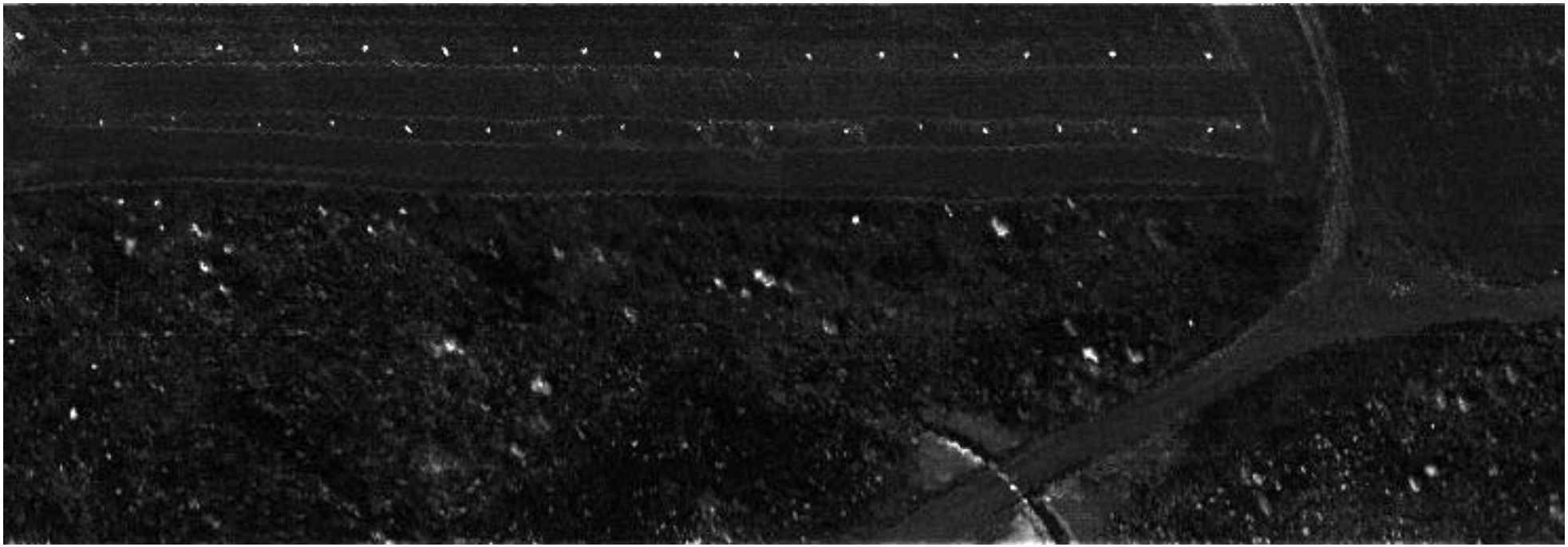
$$\widehat{\mu}_j \leftarrow \frac{\sum_{t=1}^{n} w_{tj} x_t}{\sum_{t=1}^{n} w_{tj}}$$

$$\widehat{\Sigma}_j \leftarrow \frac{\sum_{t=1}^{n} w_{tj} (x_t - \widehat{\mu}_j)(x_t - \widehat{\mu}_j)^T}{\sum_{t=1}^{n} w_{tj}}$$

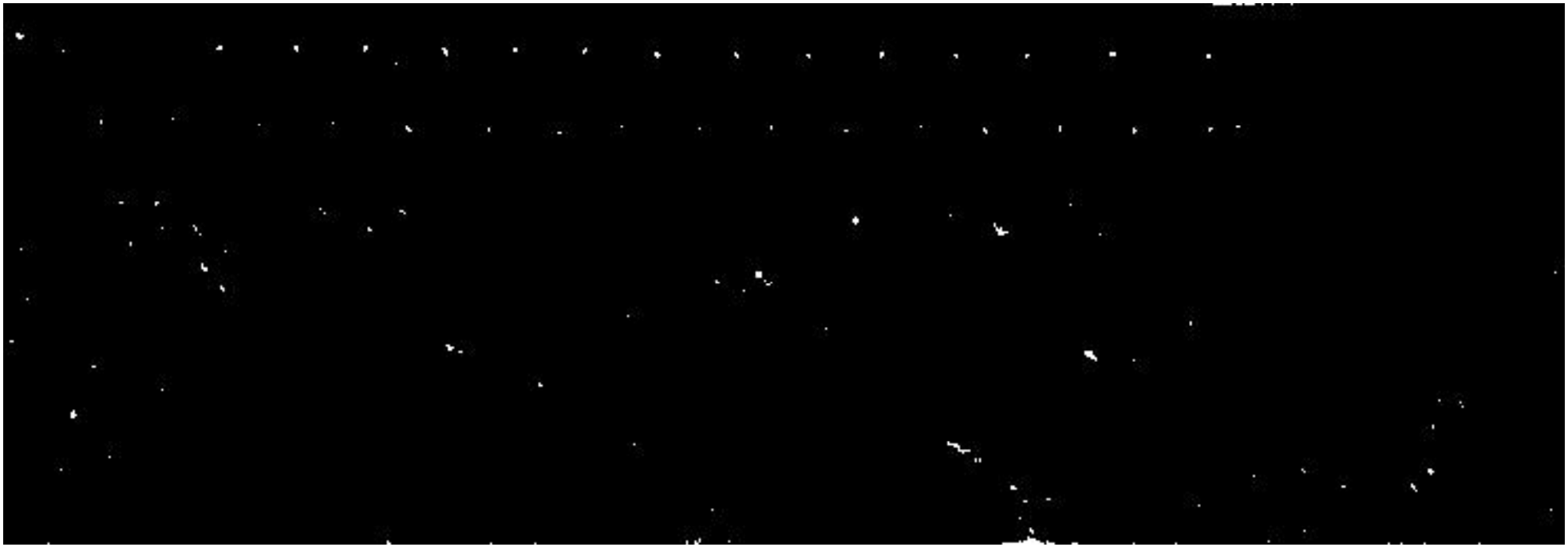
Detection

- MAP
  - Threshold MAP estimate to detect anomaly
  - Priors (estimated via GMM) may not be indicative of anomaly
- ML
  - Use Mahalanobis distance to each individual Gaussian to detect anomaly
- LRT based on GMM density

GMM-Based Anomaly Statistic



GMM Anomaly



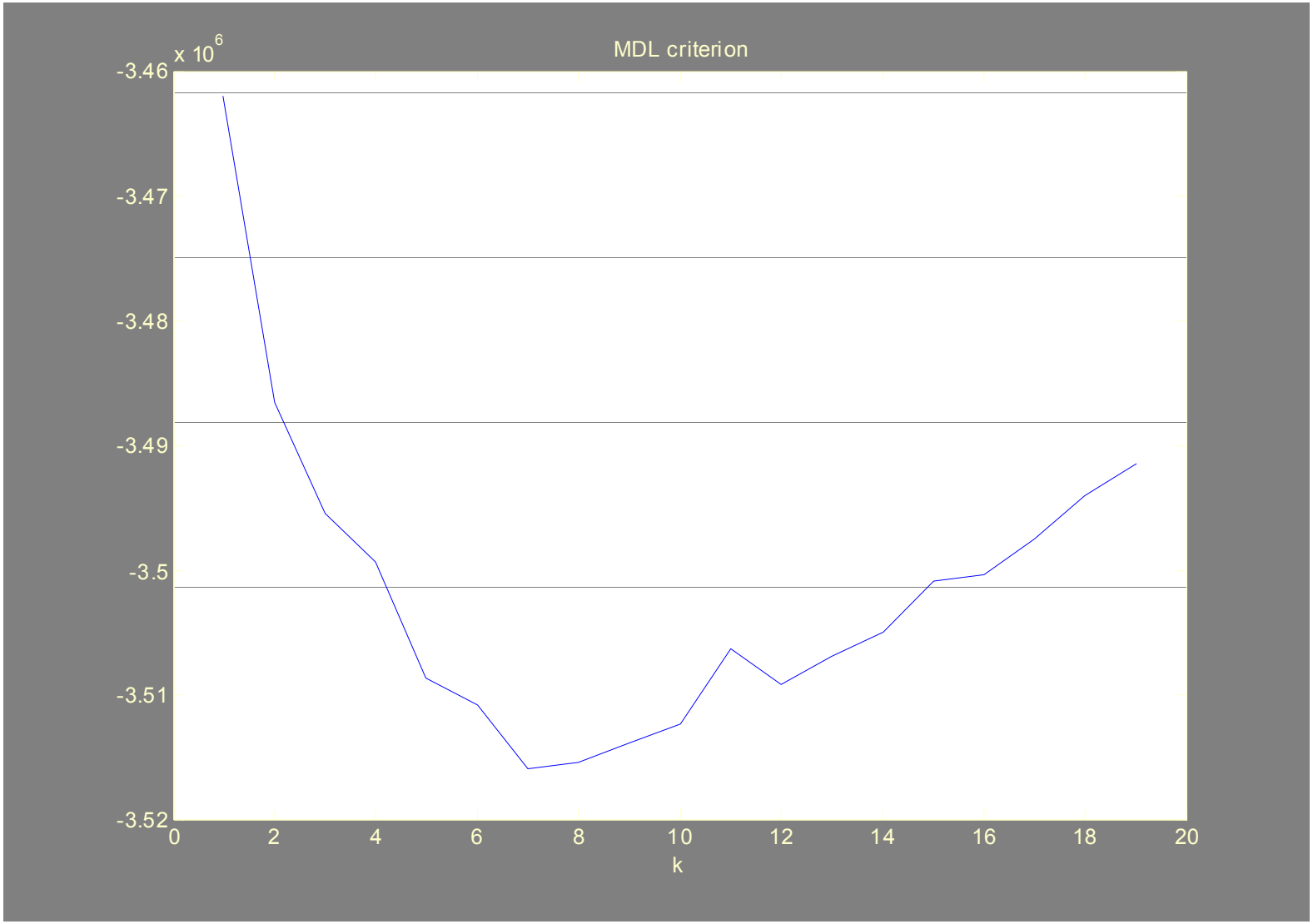
Number of Mixture Components (k)

- Minimum descriptor length (MDL)
- Maximize information and compression

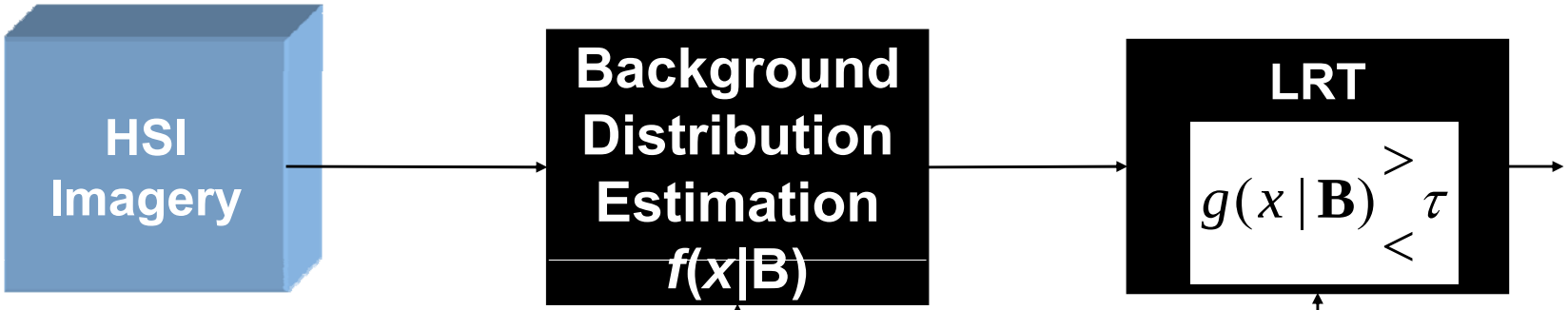
$$\min_{\theta,k}\left\{-\log f(x\,|\,\theta)+\frac{l_k}{2}\log N\right\}$$

$$l_k=(k-1)+kd+k\left(\frac{d(d+1)}{2}\right)$$

K via MDL


MDL criterion
x 10^6
-3.46
-3.47
-3.48
-3.49
-3.5
-3.51
-3.52
0
2
4
6
8
10
12
14
16
18
20
k

LRT Detection



$$f(x|\mathbf{B}) \sim N(\mu_B, \Sigma_B) \quad \text{Gassian (RX)}$$

$$f(x|\mathbf{B}) \sim \sum_i \alpha_i N(\mu_{B_i}, \Sigma_{B_i}) \quad \text{GMM (GMRX, etc.)}$$

$g(x,\mathbf{B}) = (x-\mu)^T \Sigma^{-1} (x-\mu)$	RX
$g(x,\mathbf{B}) = (x-\mu_i)^T \Sigma_i^{-1} (x-\mu_i)$	GMRX ML class conditional
$g(x,\mathbf{B}) = -\log \alpha_i + (x-\mu_i)^T \Sigma_i^{-1} (x-\mu_i)$	GMRX MAP class conditional
$g(x,\mathbf{B}) = f(x\|\mathbf{B})$	GMM LRT decision

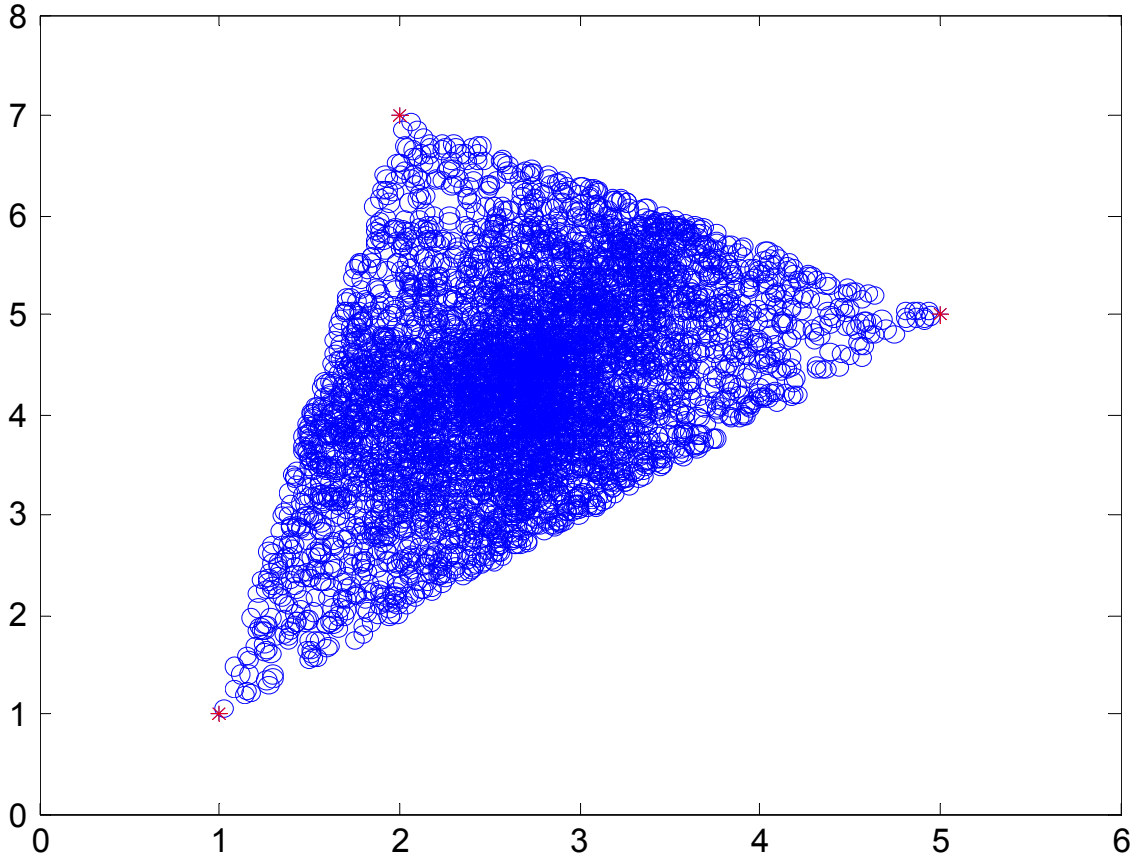
N-FINDR Algorithm

- "Shrinkwrapping" method to determine the endmembers in a spectral unmixing problem
  - Winter '99, 13th Intl. Conf. on Applied Geologic Remote Sensing
- Idea: Maximize volume of simplex over the space where E is a matrix formed from the N candidate endmember pixels:

$$V = \frac{abs(|E|)}{(N-1)!}$$

- Iteratively search through all pixels and all combinations until maximum is achieved
- The pixels that maximize the volume are the endmembers

Simplex in 2D



N-FINDR (Continued)

- To calculate N endmembers, the data must be reduced to the N-1 dimension using:
  - PCA: Select N-1 eigenvectors of the sample covariance matrix corresponding to the N-1 largest eigenvalues via:

$$\hat{\Sigma}_x = \mathbf{WDW}^T$$

  - MNF: A variant of PCA based on both the noise and signal covariances where left eigenvectors are calculated from:

$$\Sigma_n \hat{\Sigma}_x^{-1}$$

  - MNF can be viewed as a projection that maximizes the SNR that does not necessarily map onto an orthogonal basis

N-FINDR (Continued)

- Inversion used to map each pixel to a combination of endmembers (abundances)
  - Unconstrained Least Squares (S matrix of endmembers):

$$\hat{a}^U = (S^T S)^{-1} S^T x$$

  - Constrained Least Squares with Additivity (Za=b):

$$\hat{a}^F = \hat{a}^U - (S^T S)^{-1} Z^T \left[Z (S^T S)^{-1} Z^T \right]^{-1} (Z\hat{a}^U - b)$$

  - Fully constrained solution (additivity and nonnegativity) solved via quadratic programming

N-FINDR STD Algorithm

- Stochastic Target Detection (STD) identifies anomalies as those endmembers that have few pure pixels in the image
  - Yu et al., IEEE Trans. Image Proc., '97
- Histogram each abundance image to find those containing few pure endmembers. Denote this set Y (target-like) and the remaining endmembers in set X (background-like).
- Generate a bank of matched filters to find the anomalies in the image

N-FINDR STD (Continued)

- Define a replacement model :

$$H_0 : x = m_x + c_x, y = m_y + c_y$$

$$H_1 : x = m_x^\tau + c_x, y = m_y^\tau + c_y$$

$$m = \text{class mean}, c = \text{class variability}$$

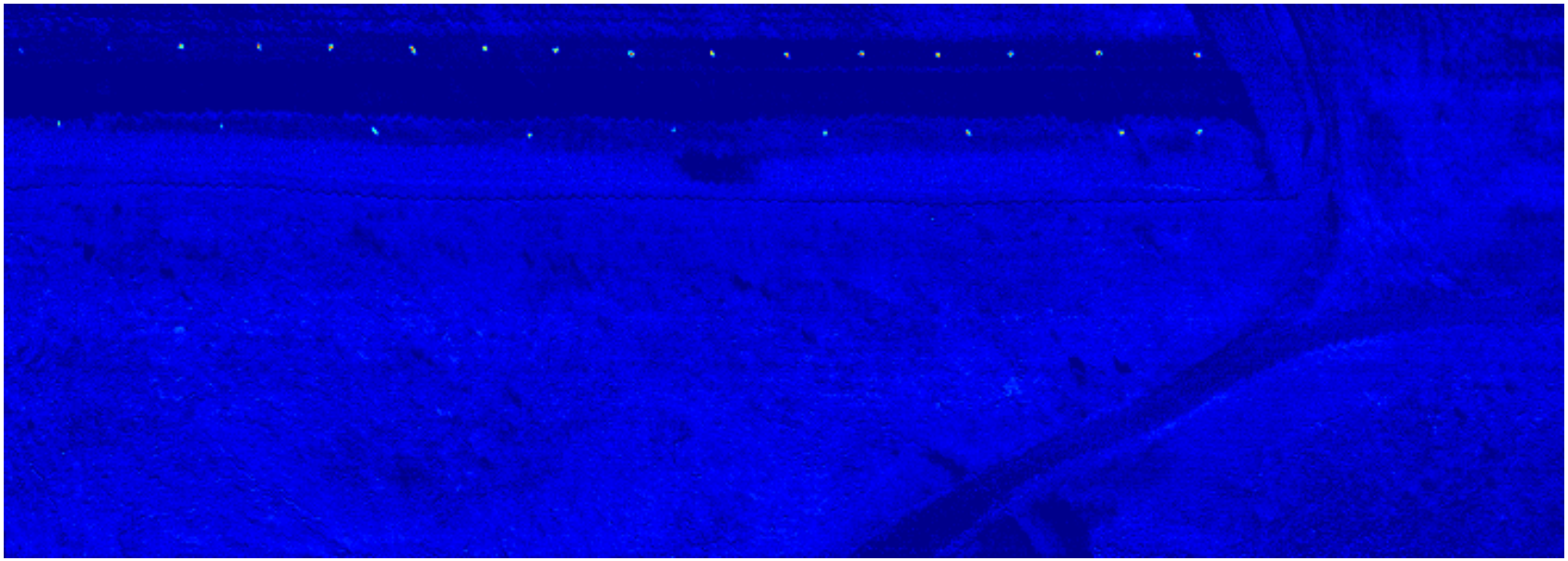
- Under H_1 define the means as:

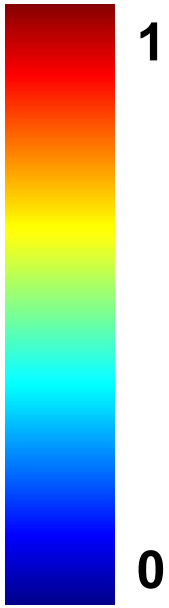
$$m_x^\tau = (0,\cdots,0)^T, m_y^\tau = (0,\cdots,0,1,0,\cdots,0)^T$$

- Calculate the GLRT as:

$$\lambda_\tau(x,y) = (m_y^\tau - m_y)^T(\Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy})^{-1}(y - \Sigma_{yx}\Sigma_{xx}^{-1}(x - m_x))$$

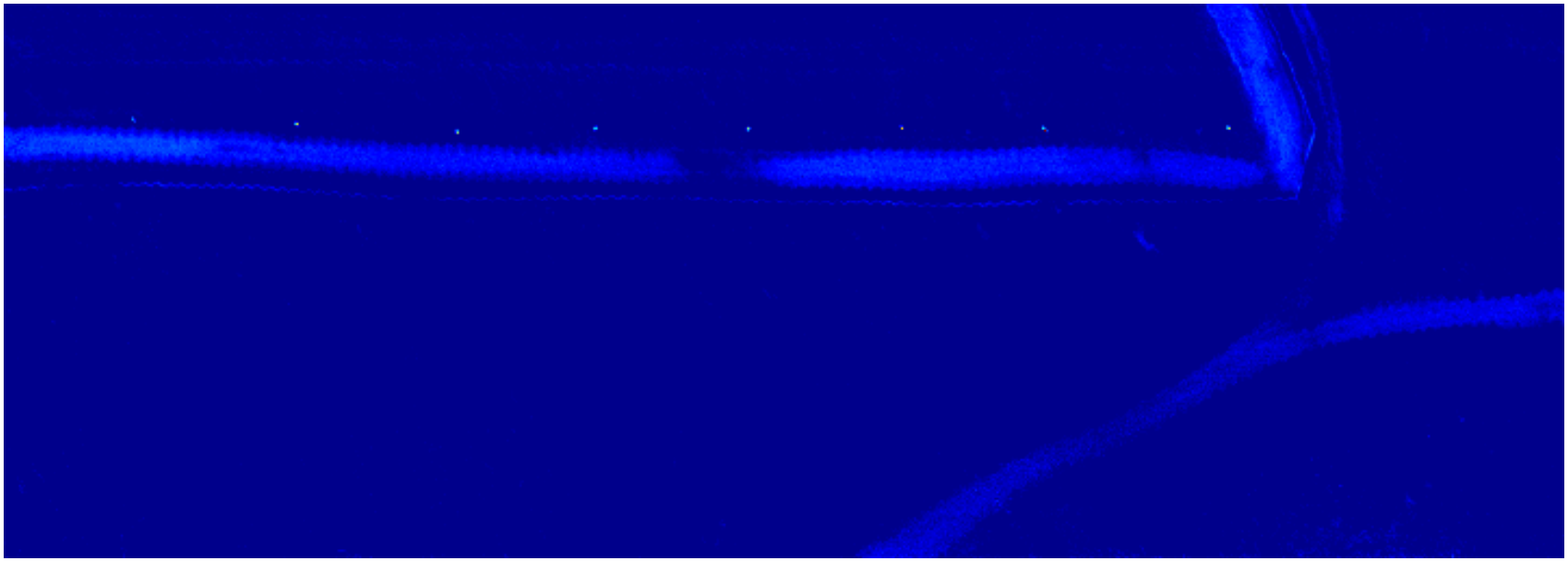
Abundance Maps

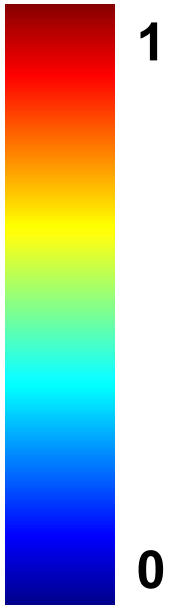




IR Panels

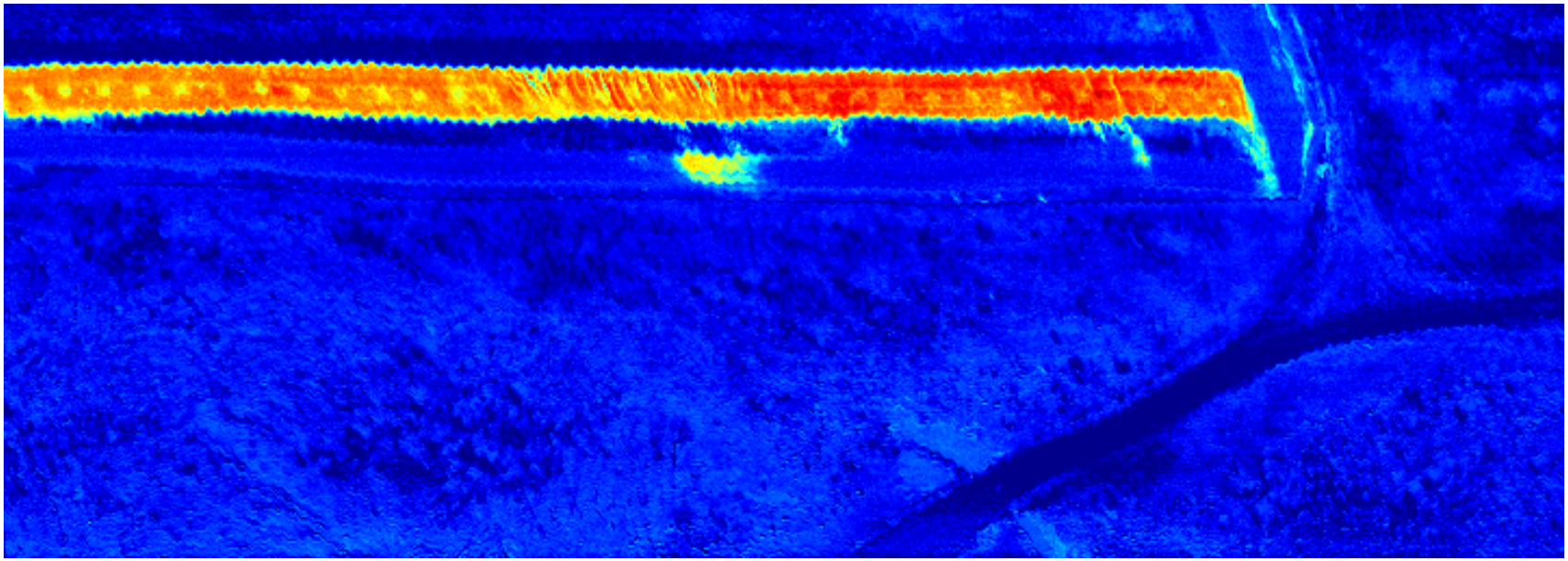
Abundance Maps





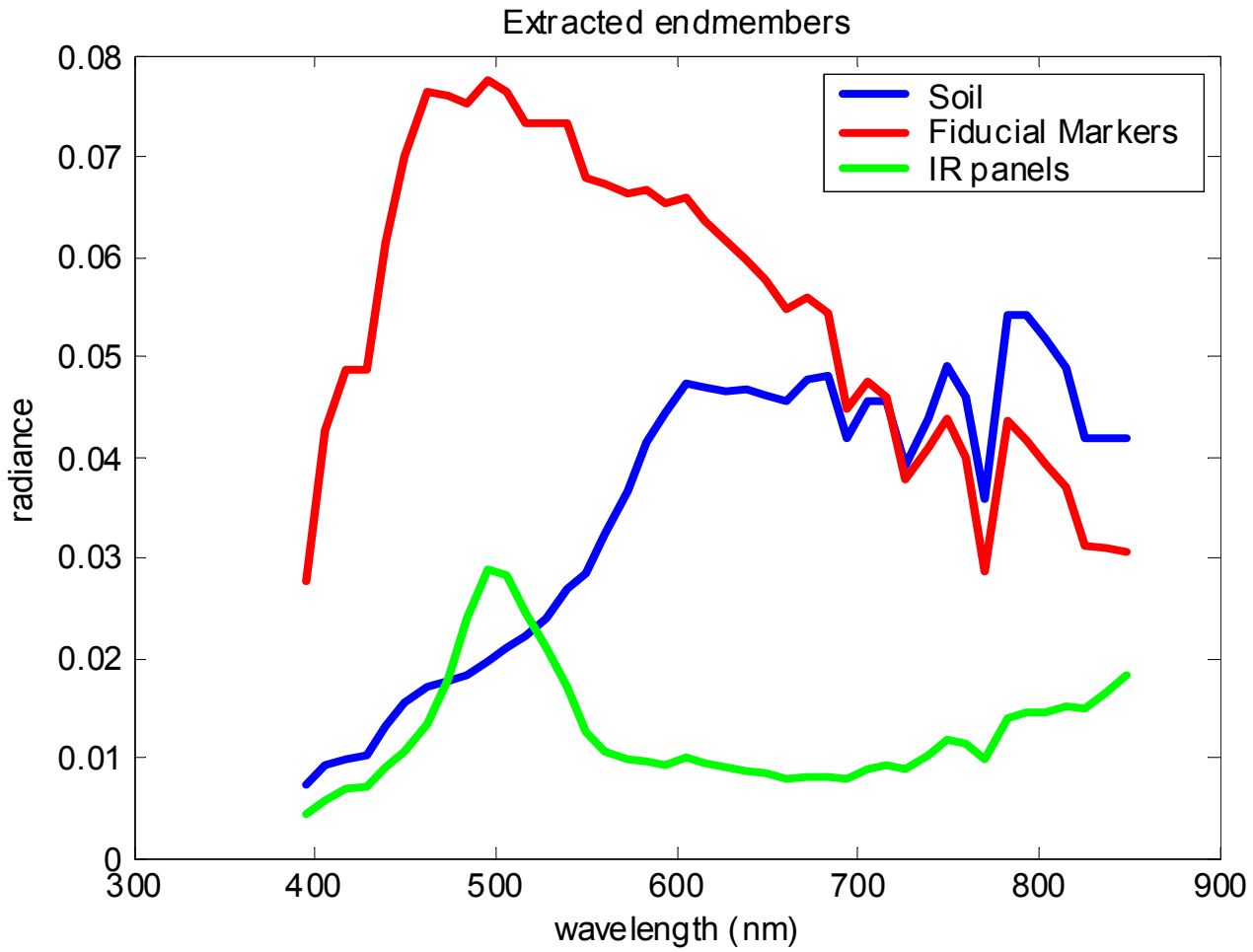
Fiducial Markers

Abundance Maps

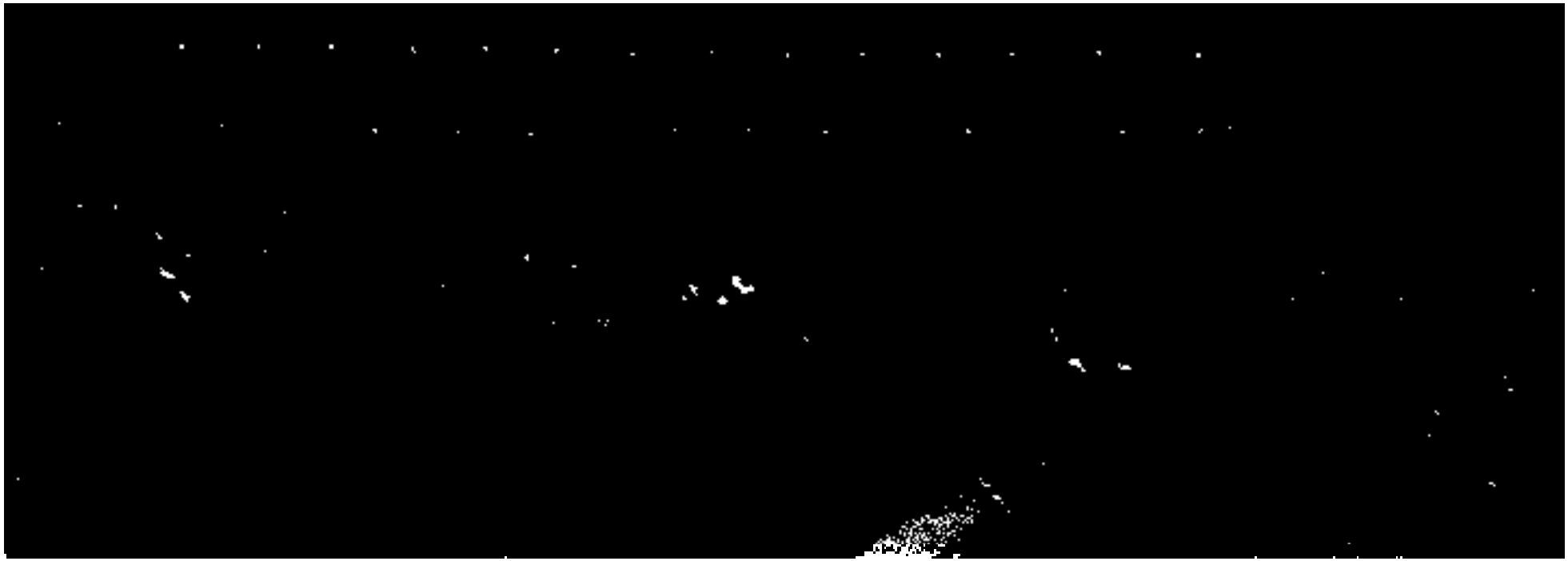


Soil

N-FINDER Extracted Spectra



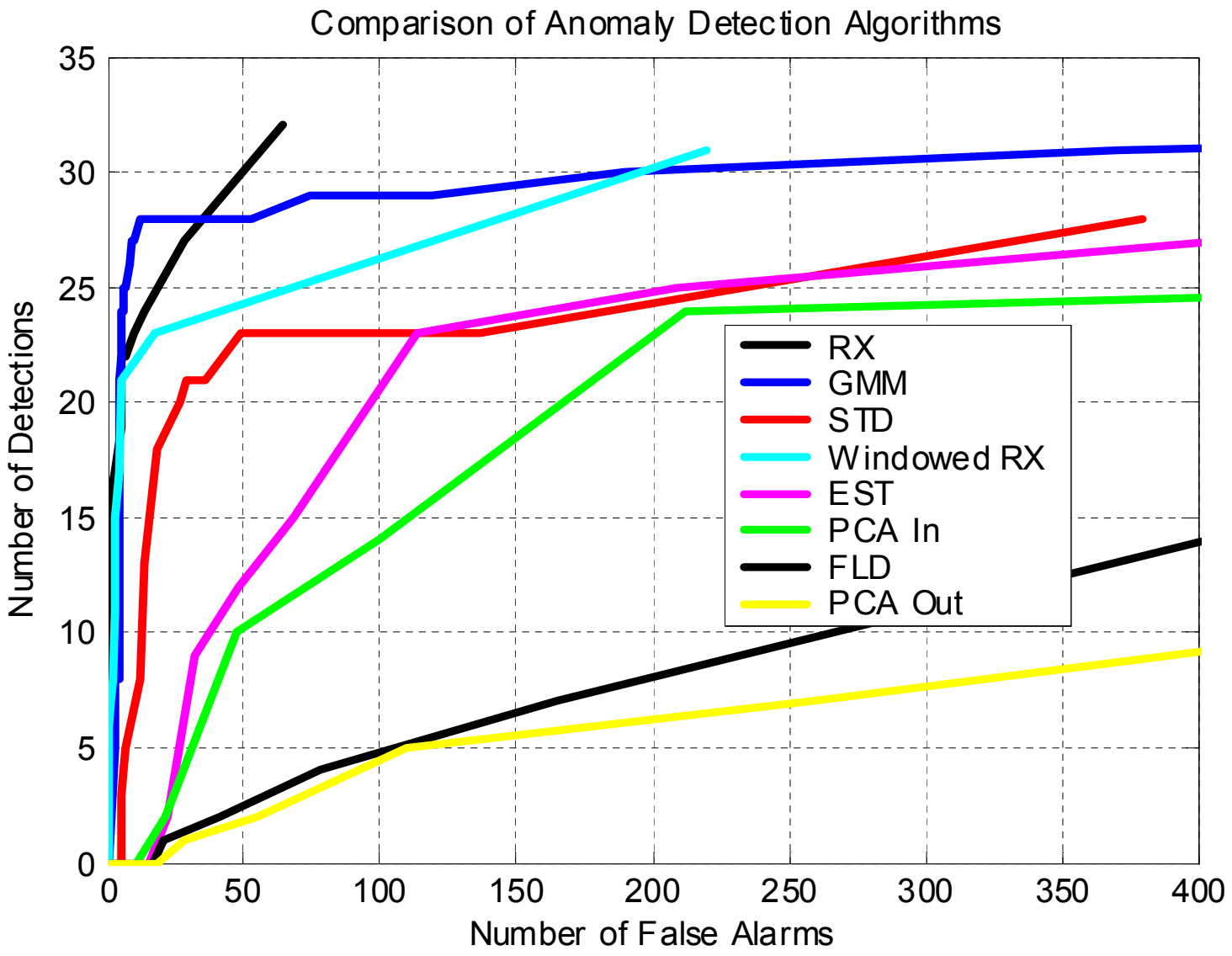
Stochastic Target Detection



Performance

- Initial results, computed for single image only
- Results shown for multipixel “targets” (IR panels, fiducial markers)
- Algorithms include those developed at ARL

Performance



Multiple Algorithm Fusion (MAF)

- Each detector has different underlying assumptions in the model
- Fuse results to exploit strengths
- Simple fusion methods like ANDing don't account for confidence of detector
- Utilize joint PDF to provide enhanced fusion

Recognition

- Utilize target information within detection (presently, LRT)
- Phenomenology will provide characteristics of target (& background)
- Incorporate phenomenology into LRT
- Utilize phenomenology to investigate subspace algorithms
- Small target sample sizes present poor conditioning

LDA

- Class separability characterized by within-class and between-class scatter matrices (S_W, S_B)

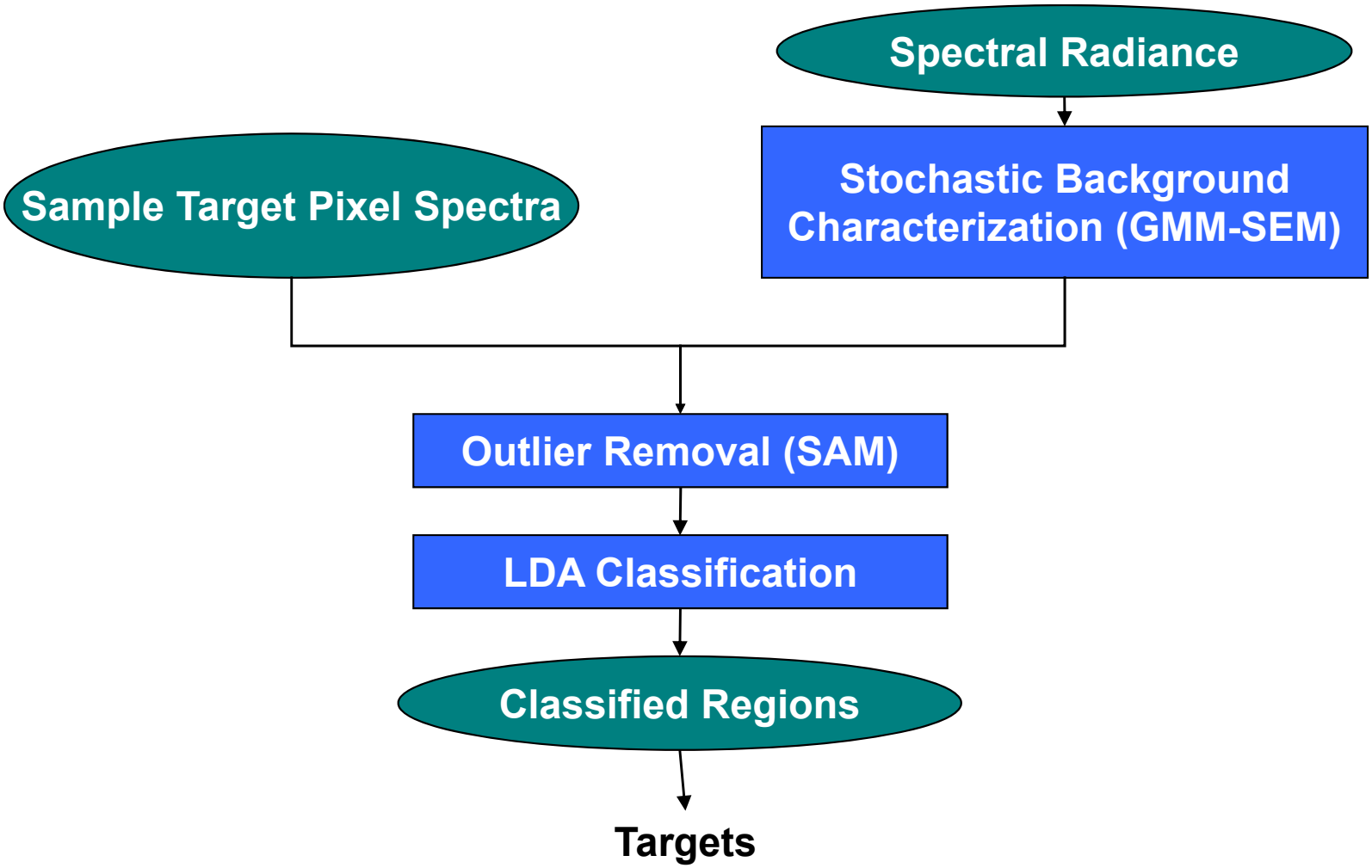
$$S_W = \sum_j P_j E\{(X - \bar{x}_j)(X - \bar{x}_j)^T \mid X \in \omega_j\})$$

$$S_B = \sum_j P_j E\{(\bar{x}_j - E(X))(\bar{x}_j - E(X))^T\}$$

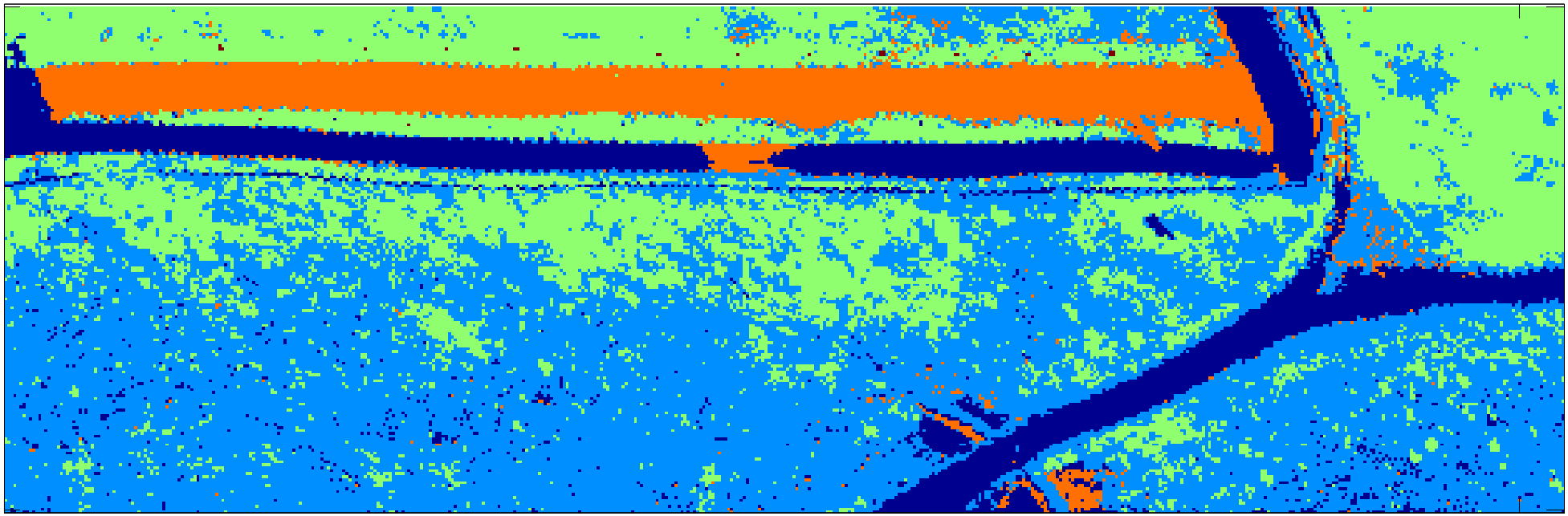
$$\sum_j (e_j(x - \bar{x}_k))^2 \leq \sum_j (e_j(x - \bar{x}_i))^2, \quad \forall i \neq k$$

- Transforms data to maximizes the "ratio" of between-class and within-class scatter $(S_W)^{-1}S_B$
- Transformed space basis from eigenvectors of $(S_W)^{-1}S_B$

LDA with Target Cueing



LDA Example

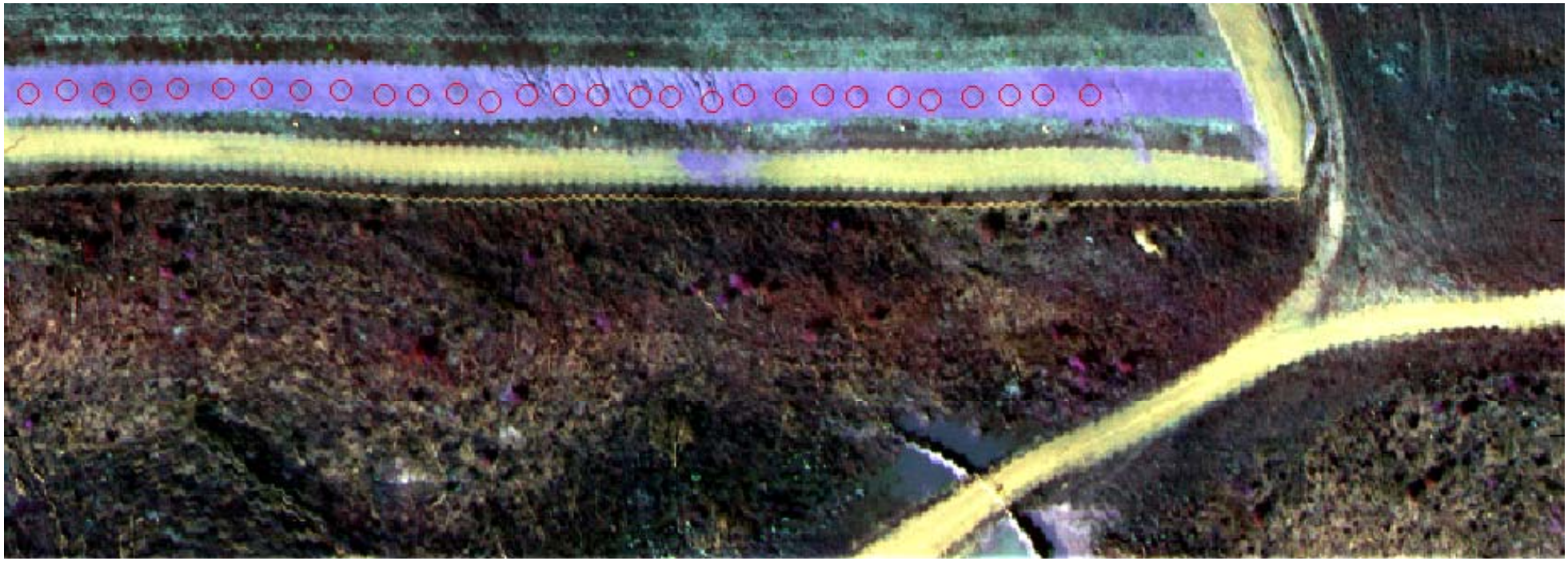


Background GMM, Sample target spectra utilized

GMM LDA – IR panels

Buried Mines

3-Band Image, Mine "Truth" Circled



Buried Mines

Mine Ground “Truth”

Buried Mines

GMM LDA

RX Characteristics

- Effective on local anomalies
  - Local anomalies are not always of interest (Isolated tree, shadow, …)
  - Function of clutter window (distanced from pixel under test - PUT)
- In scene clutter training
- Computationally intensive at every pixel

GMM Characteristics

- Not dependant on window size
  - Multiple target sizes (e.g. mines & vehicles)
  - Effective on location dependant resolution
  - May detect large region (use to update model)
- Discards local anomalies of prevalent objects (e.g. isolated tree, shadow)
- In scene clutter training
- Efficient - Clutter statistics need not be recomputed for every pixel

NFINDR - STD Characteristics

- Assumes targets are pure endmembers
- Sensitive to outliers (spikes, registration artifacts, etc.)
- In scene training

LDA Characteristics

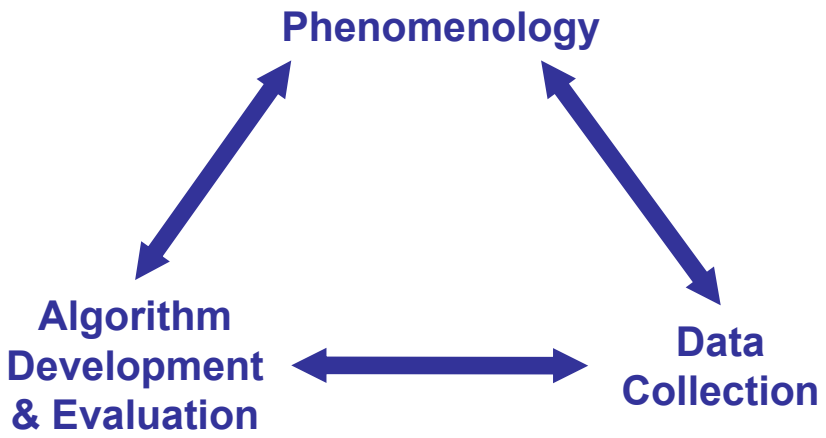
- Requires knowledge of scene constituents
- Specifically designed to enhance class separation
- Additional samples enhance modeling

Subspace Approaches

- Reduce dimensionality and enhance detection/recognition
- Applicable to detection and recognition methods
  - Subspace RX
  - Atmospheric invariance

Phenomenology

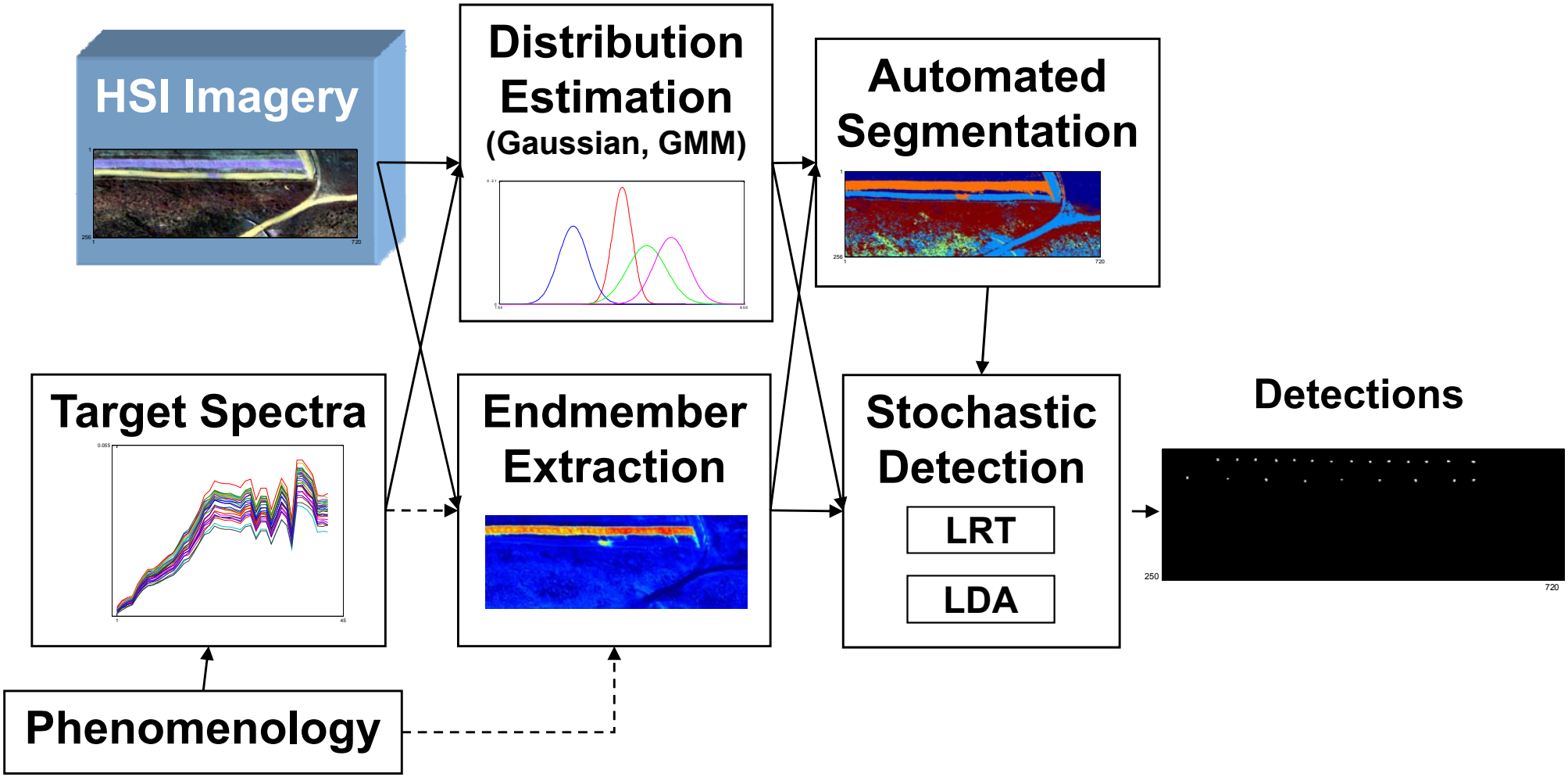
- Phenomenology - Scientific analysis/characterization of variations in data
- Statistics of target spectra may be used directly in
  - GMM formulation to provide distribution for target to be used in LRT
  - LDA to provide target characteristics
- Phenomenology for characterizing scene target variations
  - Characterization of temporal spectral signature (surface and buried mines)
- Use of phenomenology to define and localize the target (and possibly background) subspace will be investigated
- Data collections should reflect phenomenological variations that will be investigated/provided



2003 Research Accomplishments

- State of the art anomaly detection – (RX, GMM, Endmember)
- Endmember extraction module (full constraint variant)
- Recognition via target spectra
- Initial identification of areas where phenomenological studies will help

Spectral ATR - Summary



Spectral Automatic Target Detection/Recognition (ATD/R)

2004 Research Efforts

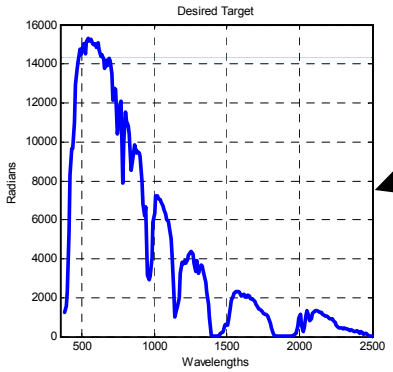
2004 Research Topic Outline

- Automated HSI ATR System
  - Atmospheric Compensation
  - Background Estimation
  - MDL
- Small target (pixel/subpixel) detection
- Semi- and non-parametric detection
- Background characterization
- Atmospheric compensation

Impact of Research

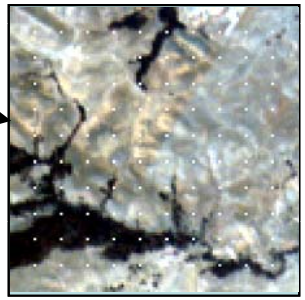
- Focus of attention / data reduction (anomaly detection)
- ATD/R (Automatic target detection / recognition)
  - Mine detection
- Reconnaissance
- Change detection

UMD HSI Automated System



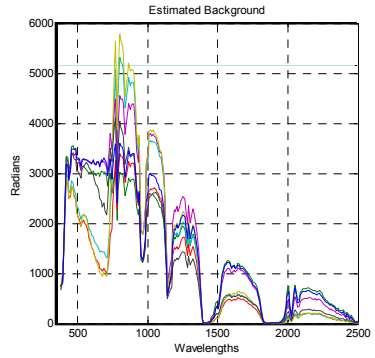
Target Signatures

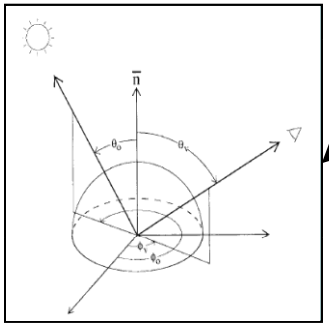
HSI



Forward Atmospheric Projection

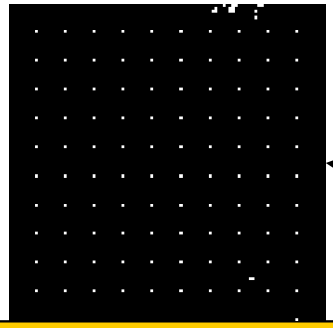
Background Estimation

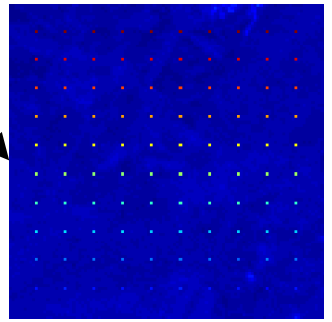




Detector

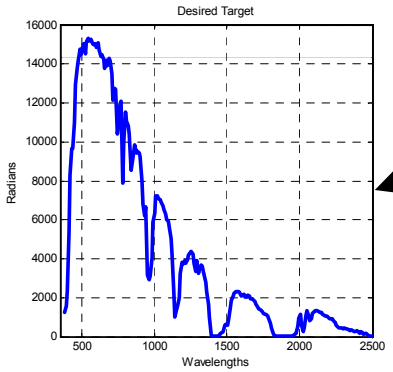
Post Processing





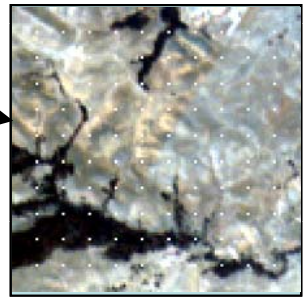
Output

Atmospheric Compensation



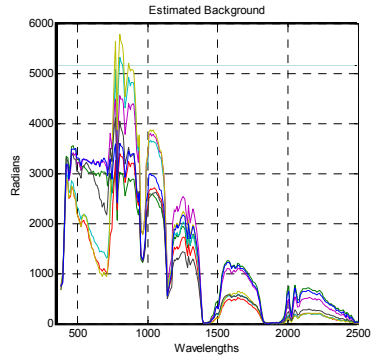
Target Signatures

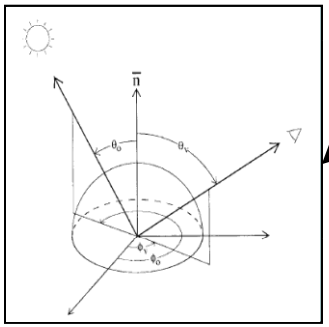
HSI



Forward Atmospheric Projection

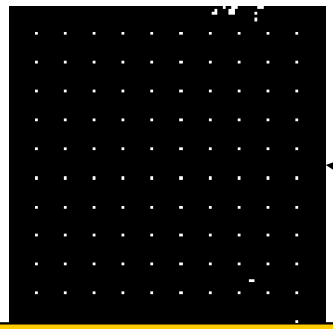
Background Estimation

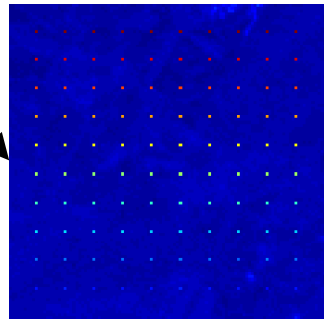




Detector

Post Processing





Output

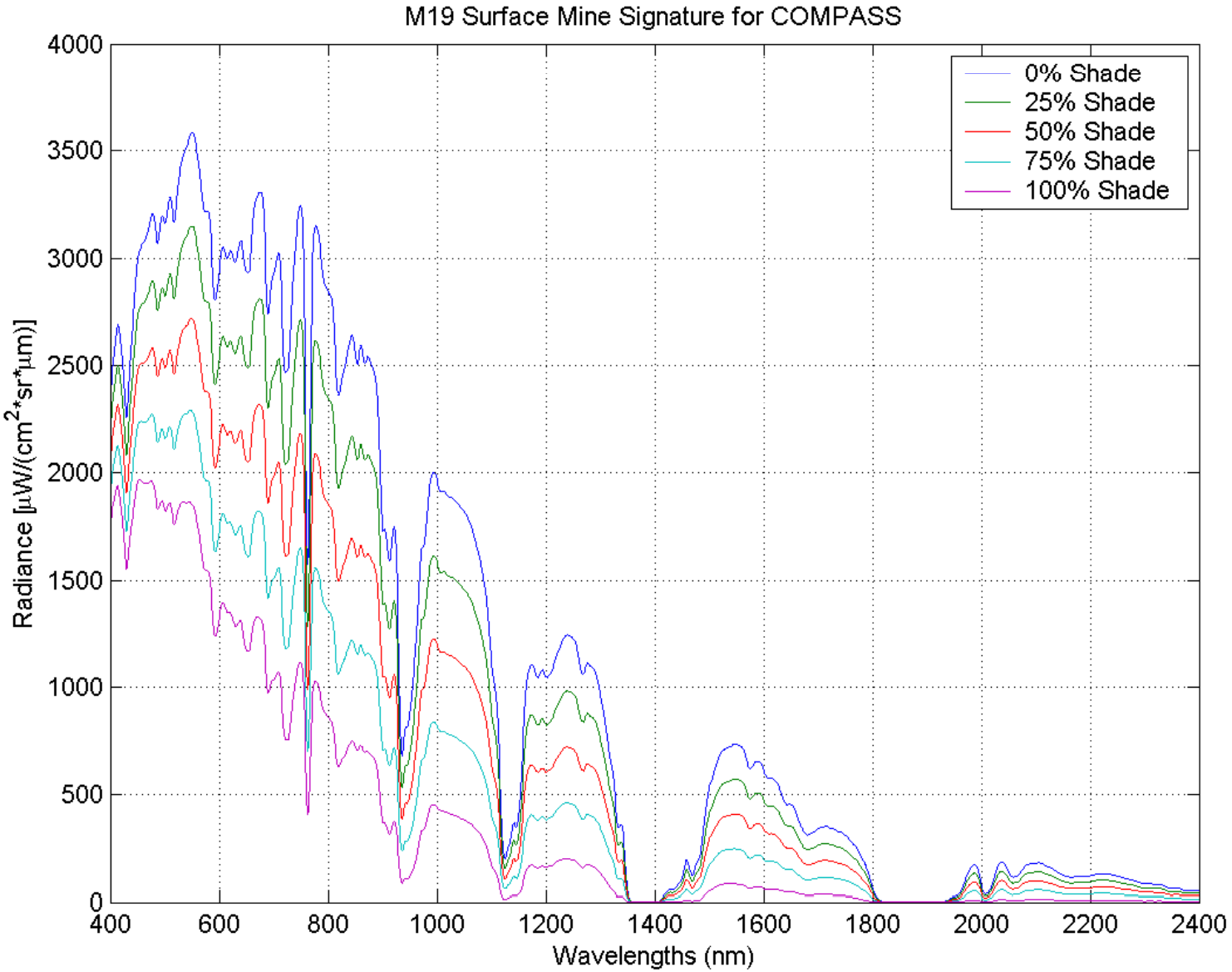
Atmospheric Compensation

- Atmospheric effects modeled to transform reflectivity/emissivity values into radiance at the sensor
- Forward projection that generates model unique for image acquisition conditions (Kolodner and Murphy, 2002)
- Forward projection for multitude of environmental conditions (Healey and Slater, 2001)
- Advantages / disadvantages

Forward Projection (Kolodner)

- **Forward projection uses MODTRAN 4.0 to map reflectivity/emissivity signatures into radiance signatures that would be seen at the sensor**
- **Pros:**
  - **Does not require an inversion as with FLAASH or ATREM which makes simplifying assumptions**
  - **Can be implemented in near-real time**
  - **Can account for different shading conditions**
  - **Can create signature before sensor collects imagery using military weather forecasts**
    - **Forecasts are given every three hours**
    - **Forecast stations are within 6 miles of every location on the Earth**
- **Cons:**
  - **Requires radiosonde or weather data which may be impossible to get for past image collections**

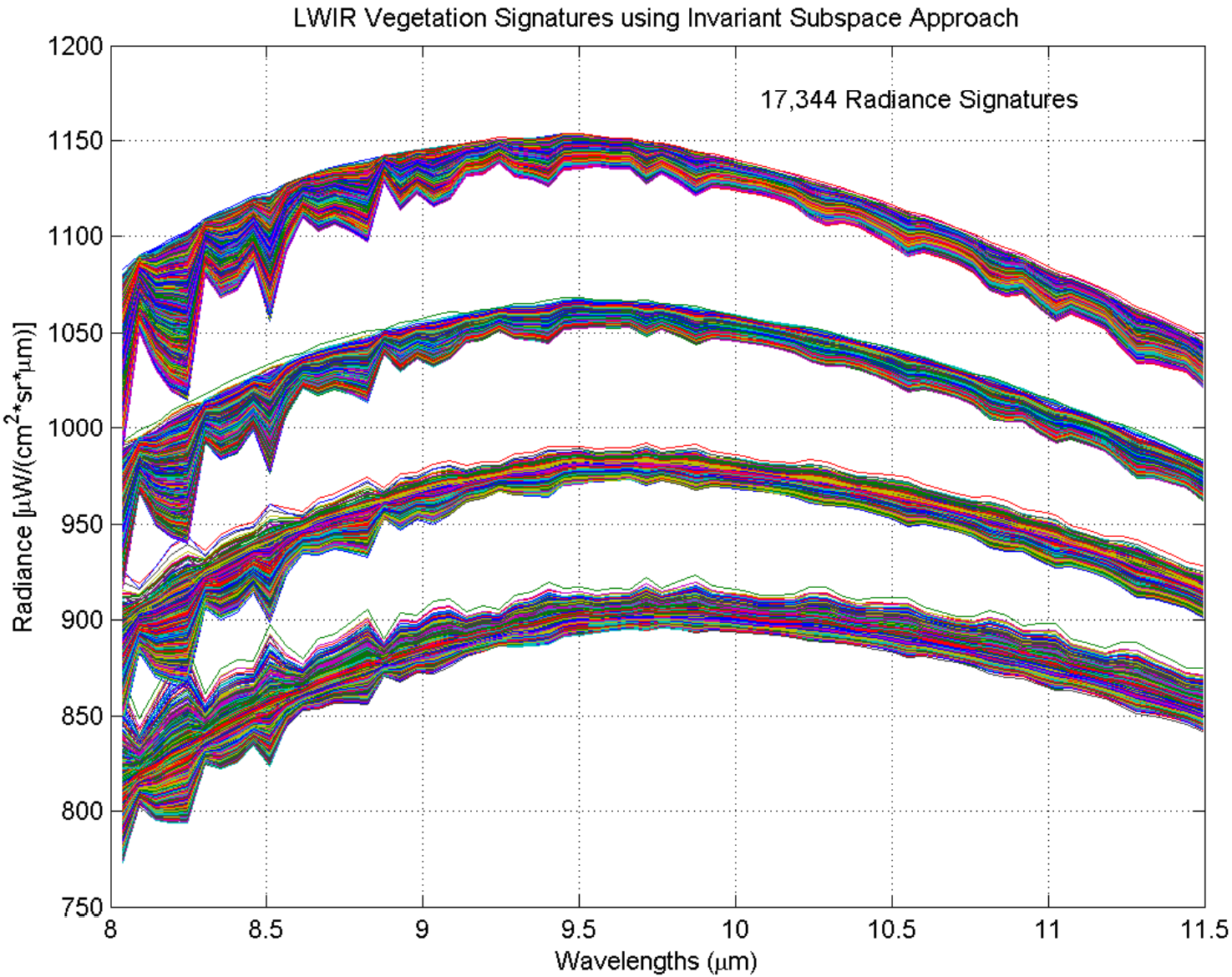
Forward Projection Example



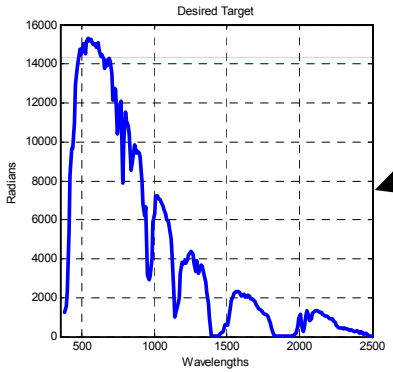
Forward Projection (Healey)

- Generates over 17,000 radiance signatures for each target spectra to account for all atmospheric possibilities
- Pros:
  - Does not require an inversion as with FLAASH or ATREM which makes simplifying assumptions
  - Method does not require weather information allowing it to be used on both future and past image collections
  - Once a target subspace has been created, it is not necessary to perform the calculations again
- Cons:
  - Generation of the target subspace is time consuming and does not allow for quick turnaround of new signatures
  - Method requires the common subspace within each class to be well distanced from that of other classes. Past result indicate that this does not adversely affect performance.

Invariant Subspace Example

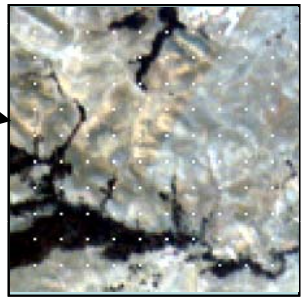


Background Estimation



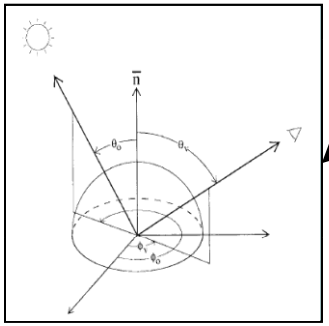
Target Signatures

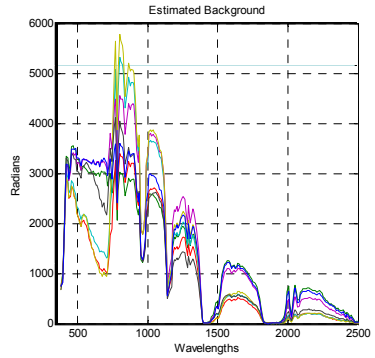
HSI



Forward Atmospheric Projection

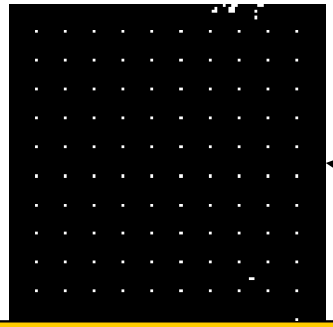
Background Estimation

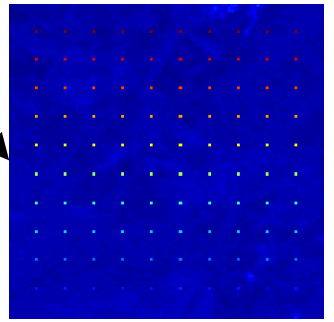




Detector

Post Processing





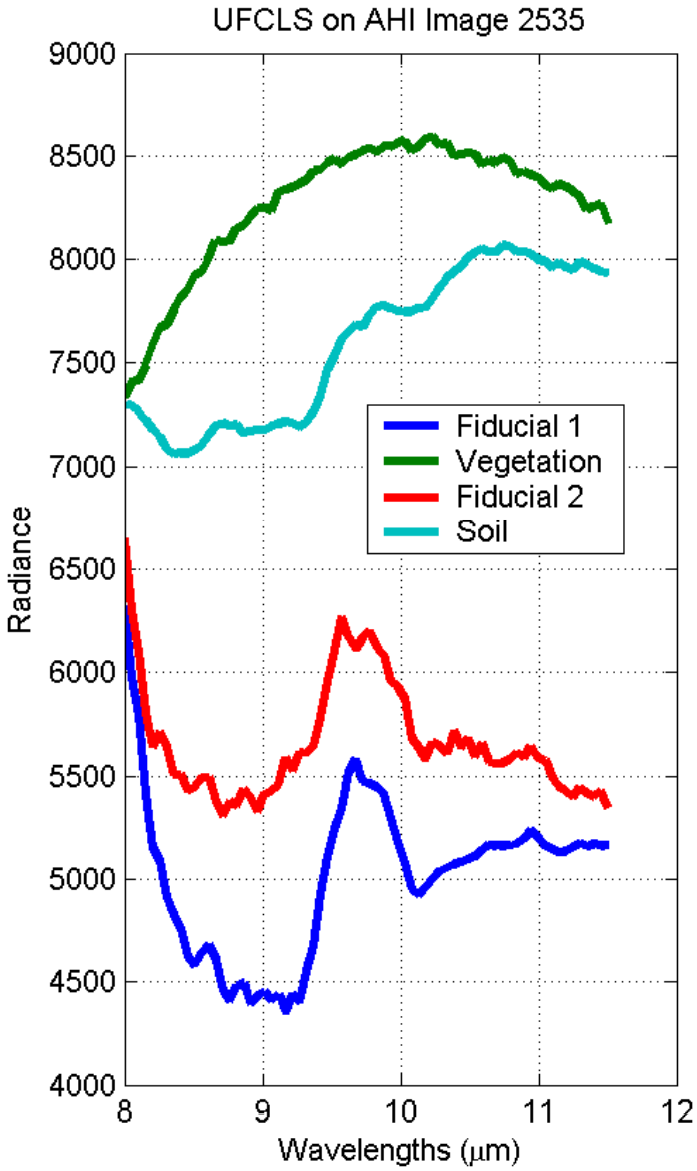
Output

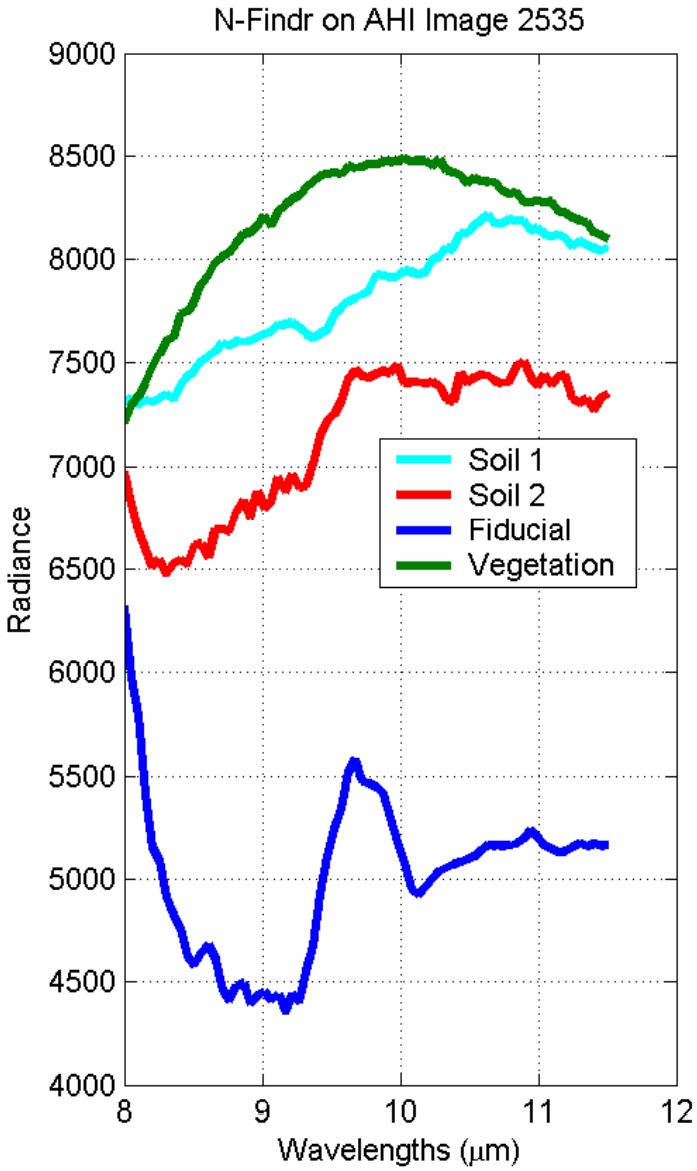
Background Estimation

- **Two methods utilized to find background endmembers:**
  - **N-FINDR (Winter 1999): Uses "shrinkwrapping" approach to identify the vertices of the simplex that encloses most pixels in the image**
  - **Unsupervised Least Squares (Chang 2001): Unmixes the image given a target signature and finds the endmembers as those pixels that produce the greatest mean squared error.**
- **Methods do not provide a rigorous way to identify how many endmembers should be used for ATD/R purposes**
  - **Solution: Applied a Minimum Description Length (MDL) criterion to the methods based on the difference between the actual image and their current solution with M endmembers. M is chosen such that the MDL attains a minimum value.**

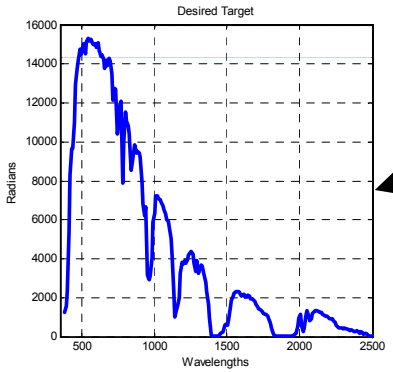
$$MDL = \sum_{i=1}^{N} (\mathbf{x}_i - \hat{\mathbf{E}}\hat{\mathbf{a}}_i)^T (\mathbf{x}_i - \hat{\mathbf{E}}\hat{\mathbf{a}}_i) / \hat{\sigma}^2 + dL \log N$$

Background Estimation Examples



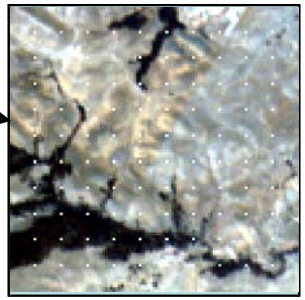


Detectors



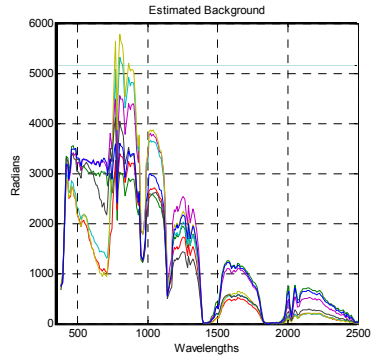
Target Signatures

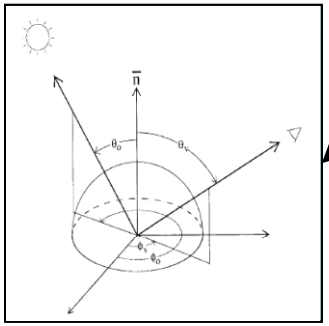
HSI



Forward Atmospheric Projection

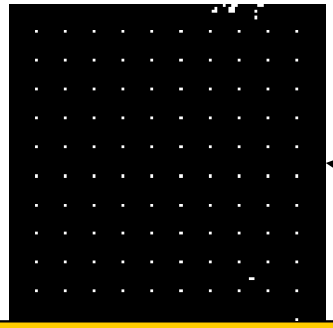
Background Estimation

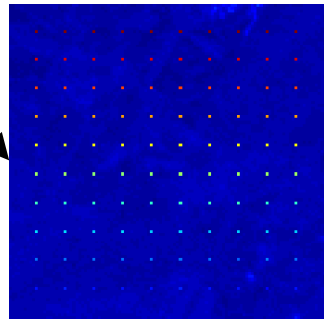




Detector

Post Processing





Output

Fully Constrained Least Squares Detector (FCLS)

- Idea: Create a detector based directly on accurate abundance estimates (Heinz, Chang 2002)

- Model:

$$H_0 : x = Ba_b + w, \quad w \sim N(0, \sigma_w^2 I)$$

$$H_1 : x = Sa_t + Ba_b + w = Za + w$$

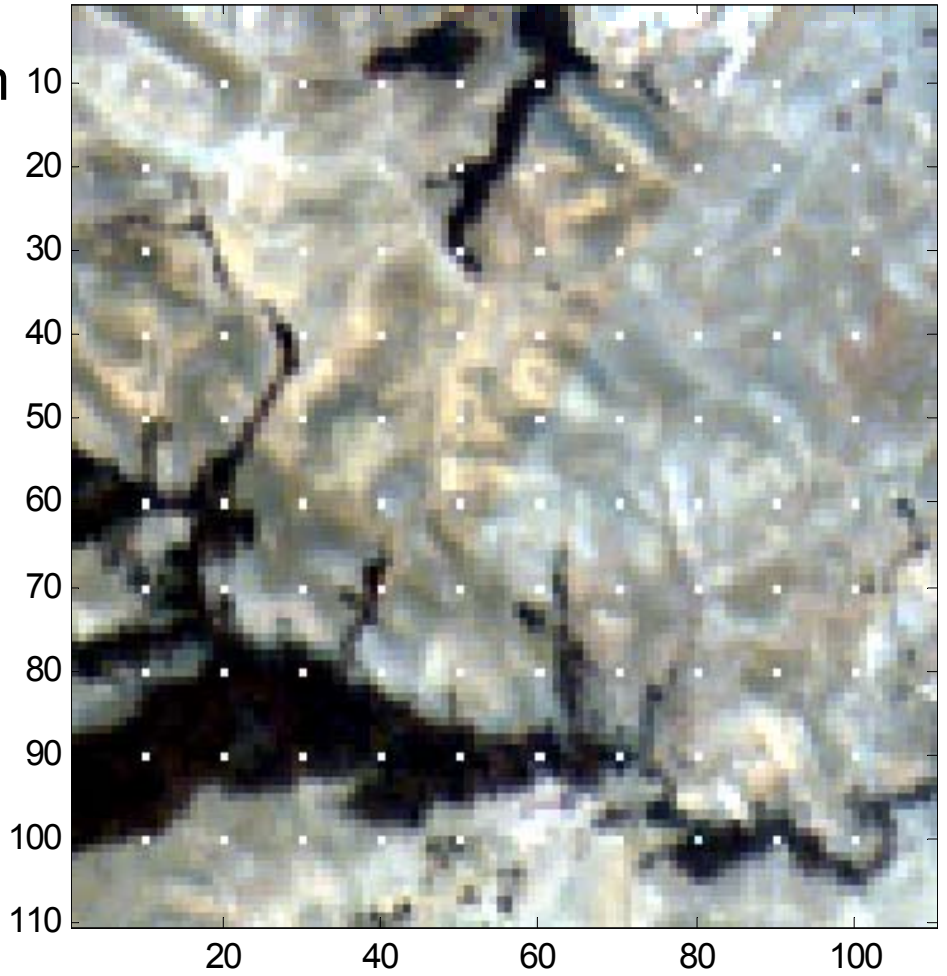
- Solution:
  - Estimate the abundances *a* given a target and background subspace such that
    - Abundances sum to one (Lagrange multipliers)
    - Abundances are non-negative (quadratic programming)
  - Directly threshold the target abundances a_t to create detection statistic

- Issues:
  - False Alarms: For targets that have signatures that are “close” to the background, can have a significant number of false alarms
  - Optimality: Does not provide a test to identify whether a target is *statistically* different from the background

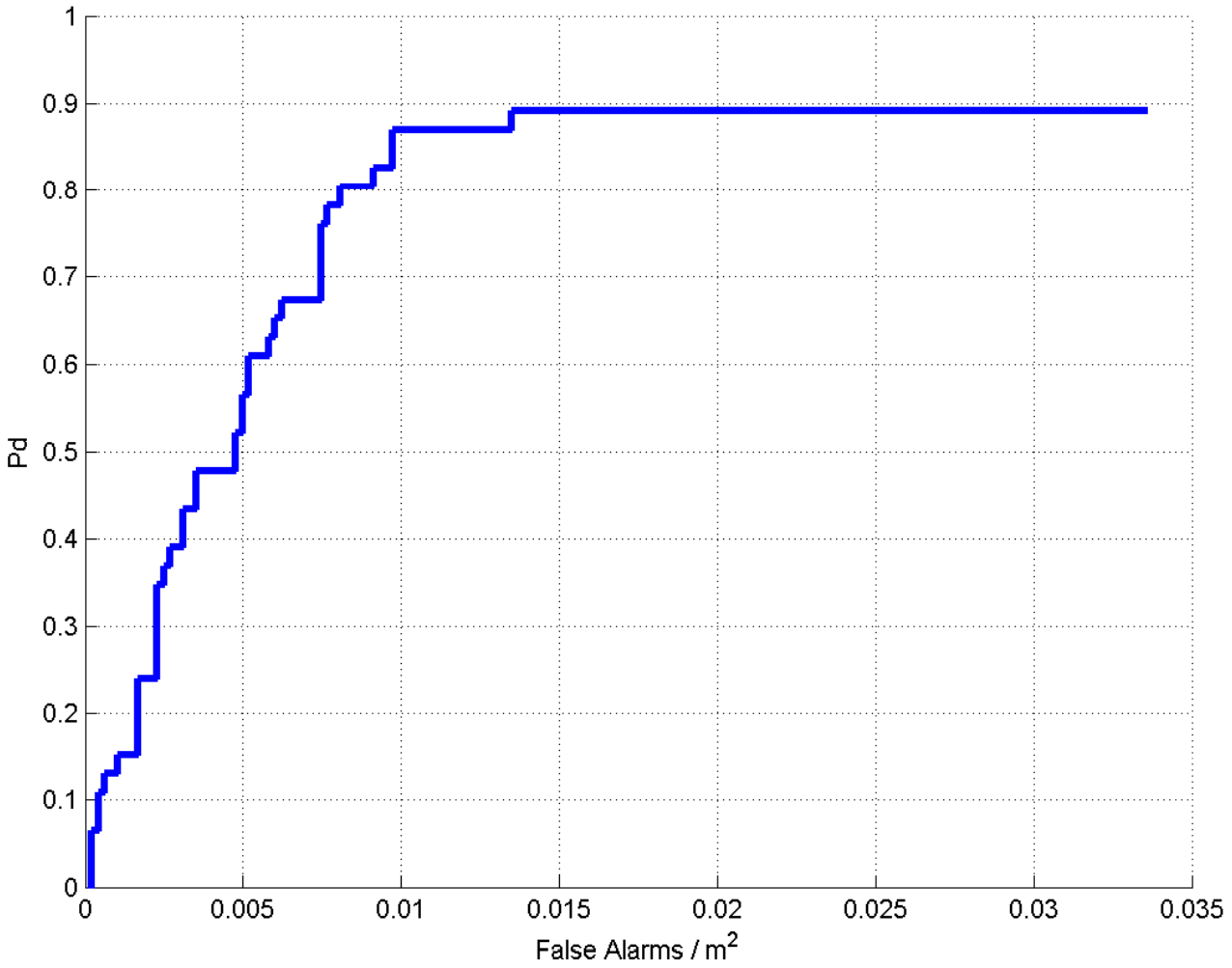
FCLS Example

- Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) data
- Wavelength range is from the 400 nm to the 2500 nm range
- Injected target signature (roof) into 100 pixels as varying ratios to simulate sub-pixel targets

Abundance (True)	Abundance (Mean)	Abundance (Std Dev)
1.0	1.0000	0.0000
0.9	0.9006	0.0015
0.8	0.8006	0.0022
0.7	0.6999	0.0051
0.6	0.6024	0.0062
0.5	0.5055	0.0087
0.4	0.4062	0.0079
0.3	0.3065	0.0075
0.2	0.2012	0.0077
0.1	0.1109	0.0087



FCLS Example: AHI Image 2535



Adaptive Matched Subspace Detector (AMSD)

- Idea: Create a detector based on Maximum Likelihood Estimates (MLE) and a Generalized Likelihood Ratio Test (GLRT) (Manolakis, Siracusa, Shaw 2002)
- Model:

$$H_0 : x = Ba_b + w, \quad w \sim N(0, \sigma_w^2 I)$$

$$H_1 : x = Sa_t + Ba_b + w = Za + w$$

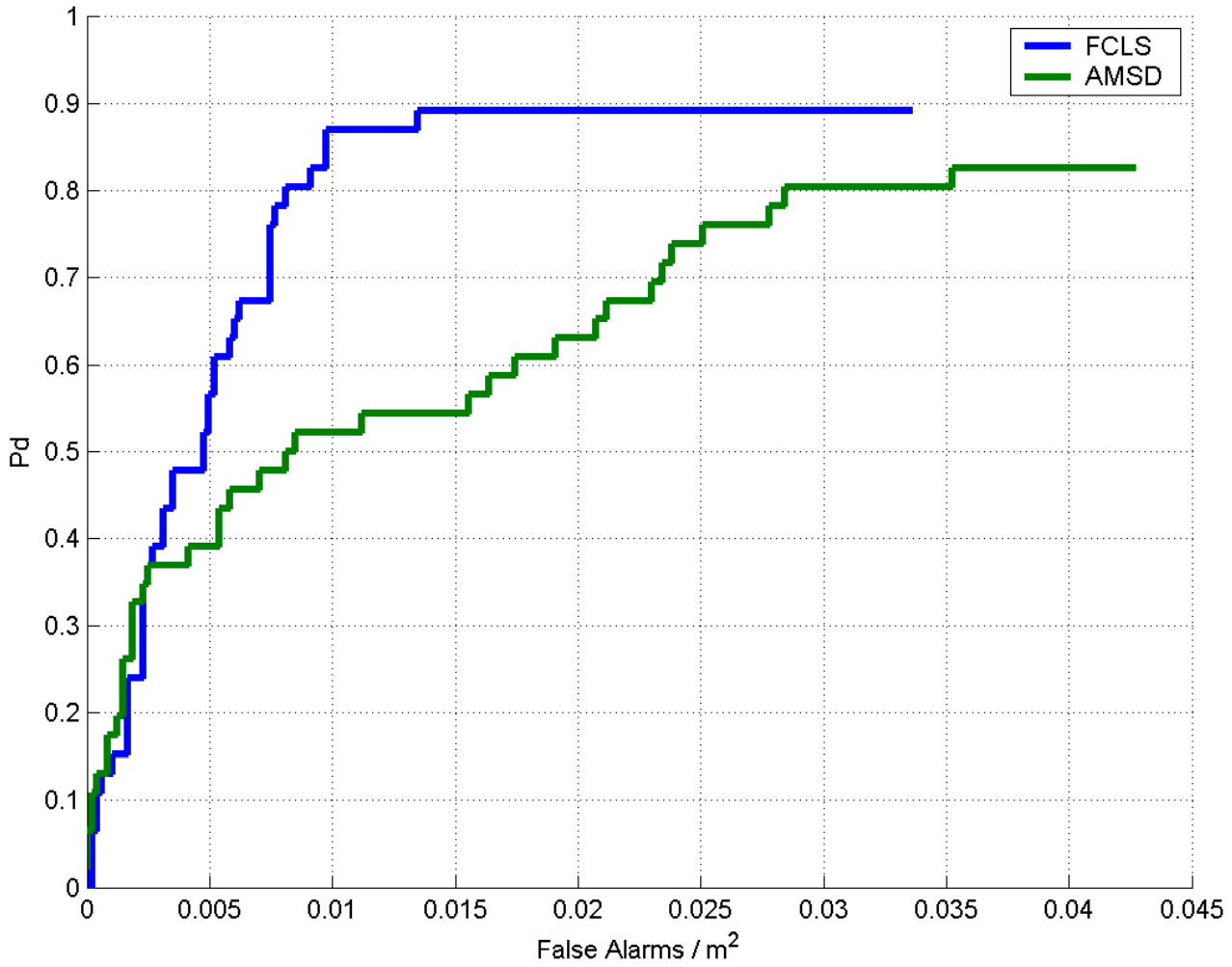
- Solution:
  - Estimate the unknown parameters using MLE
    - Abundances without non-negativity or sum-to-one constraint
    - Noise variance
  - Form a GLRT using the estimates to create the detection statistic

$$P_W^{\perp} = I - W(W^T W)^{-1} W^T, W \in \{B, Z\}$$

$$D(x) = \frac{x^T (P_B^{\perp} - P_Z^{\perp}) x}{x^T P_Z^{\perp} x}$$

- Issues:
  - Phenomenology: Violates the physical model that the abundances need to be non-negative and sum-to-one

AMSD Example: AHI Image 2535



Hybrid AMSD Algorithm

- Idea: Create a detector using both phenomenology and statistical inference (Broadwater, Meth, Chellappa 2004)
- Model:

$$H_0 : x = Ba_b + w, \quad w \sim N(0, \sigma_w^2 I)$$

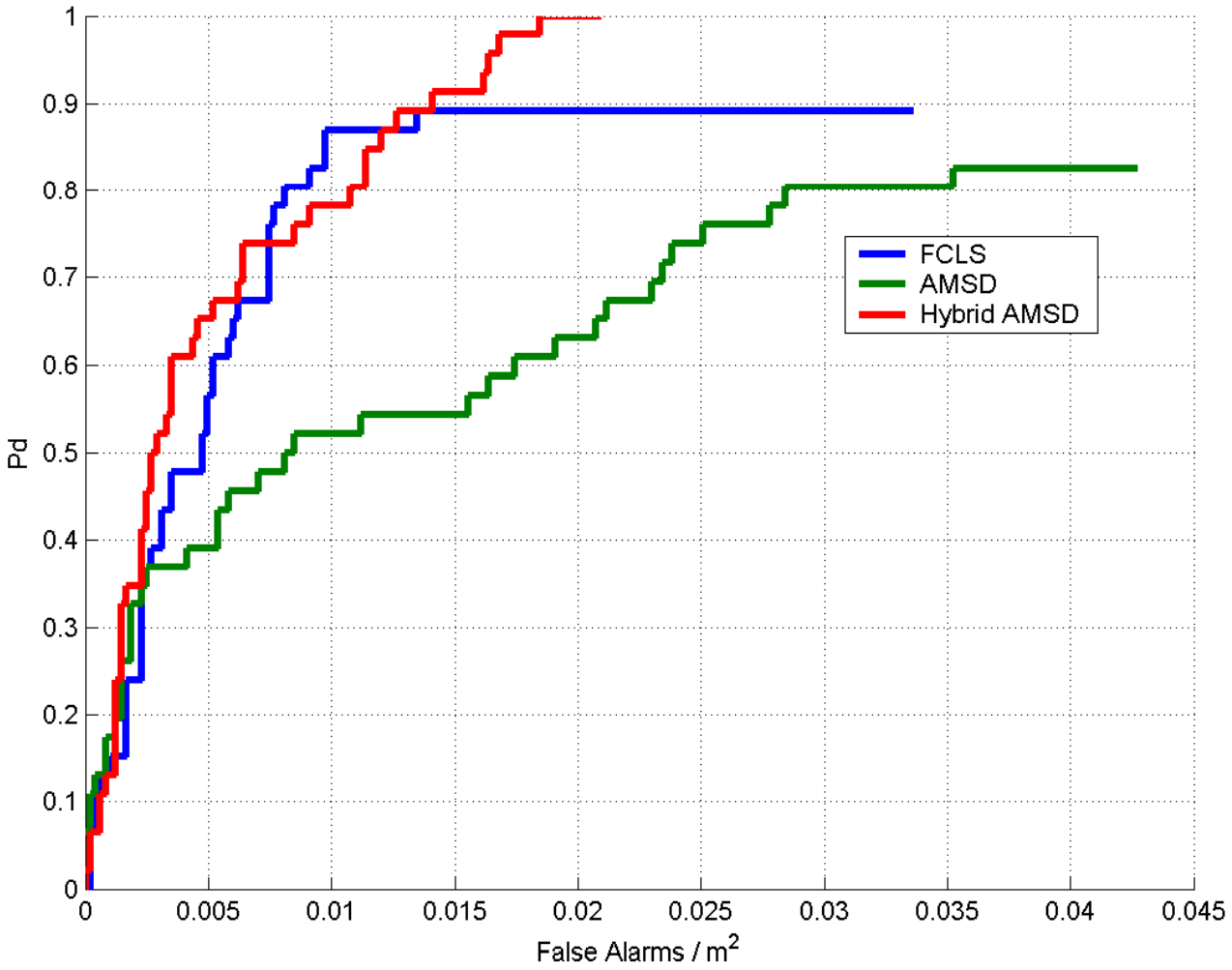
$$H_1 : x = Sa_t + Ba_b + w = Za + w$$

- Solution:
  - Estimate the abundances *a* given a target and background subspace such that
    - The abundances sum to one (Can use Lagrange multipliers)
    - The abundances are non-negative (Have to use quadratic programming)
  - Estimate the noise variance using an MLE.
  - Form a GLRT using the above estimates such that

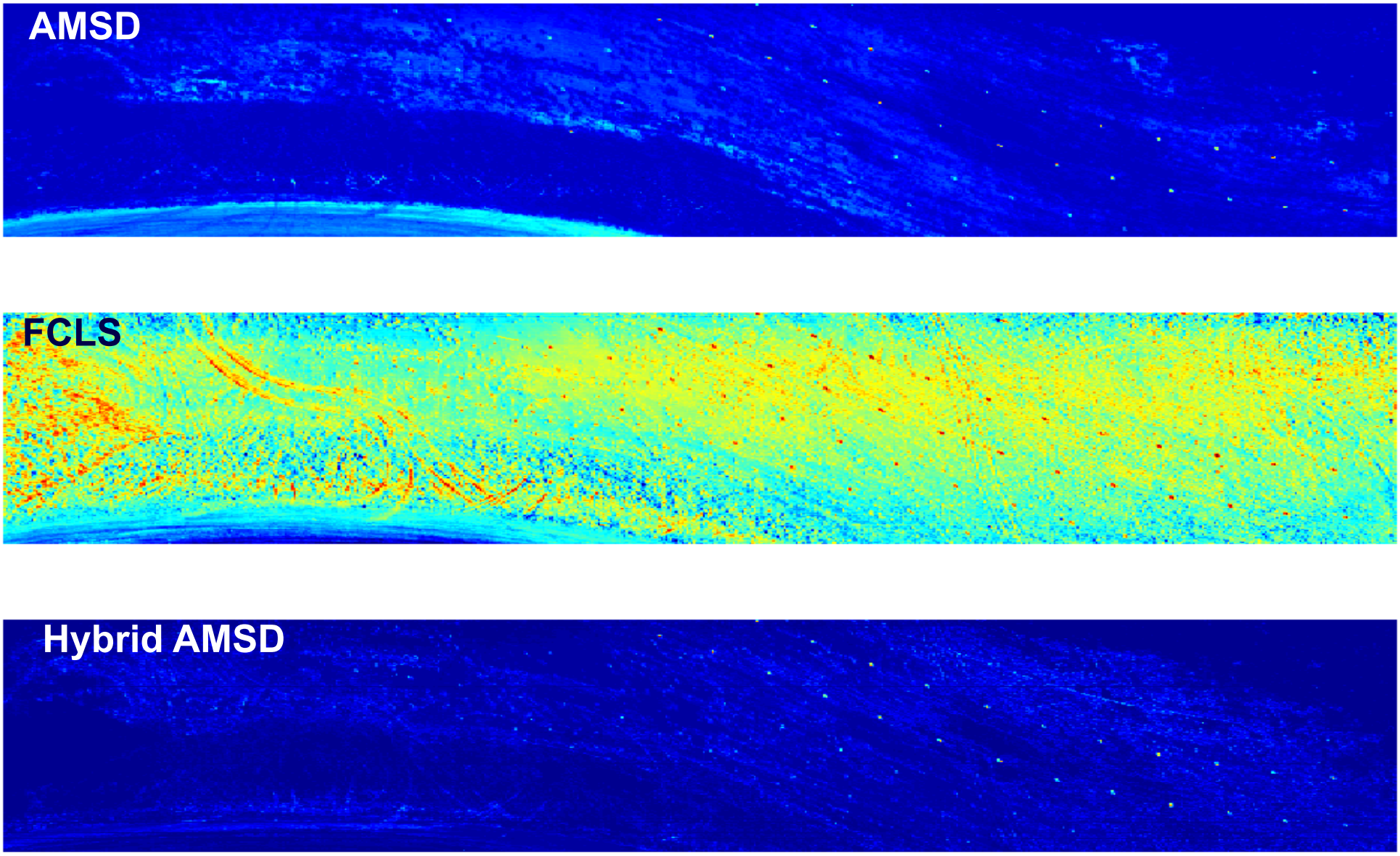
$$D(x) = \frac{(x - B\hat{a}_b)^T (x - B\hat{a}_b)}{(x - Z\hat{a})^T (x - Z\hat{a})}$$

- Issues:
  - Processing Load: Computationally expensive due to the quadratic programming necessary to estimate the abundances with non-negativity constraints
  - Optimality: The statistic has not been proven to be CFAR; however, the abundance and noise variance estimates are optimal in MSE sense

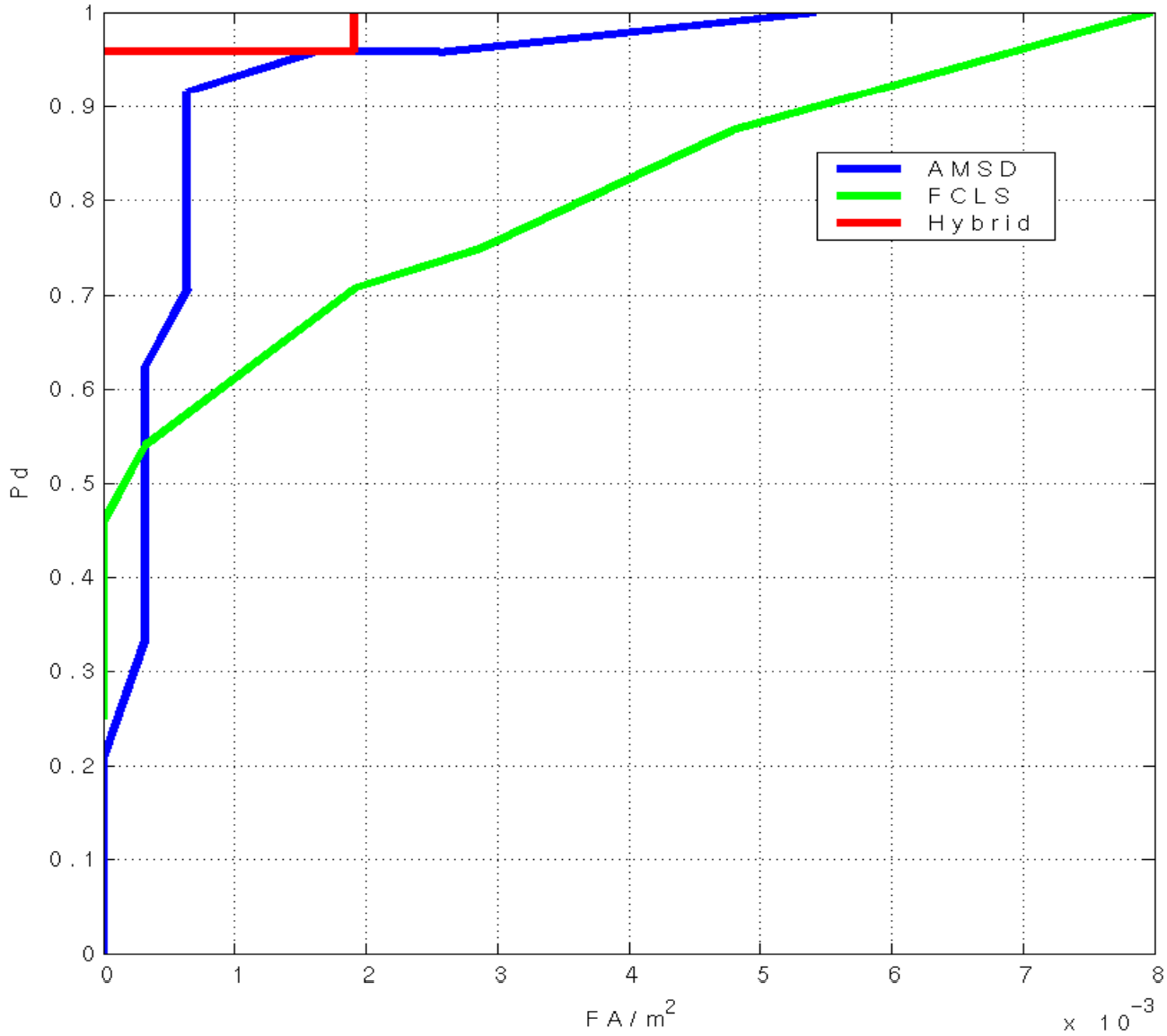
Hybrid AMSD Example: AHI Image 2535



COMPASS Example



COMPASS ROC Example



Adaptive Cosine Estimate (ACE) Algorithm

- Idea: Create a detector based on GLRT that models the local background as a single multivariate Gaussian distribution (Kelly and Scharf, 1999)
- Model:

$$H_0 : x = b, \qquad b \sim N(0, \Sigma)$$

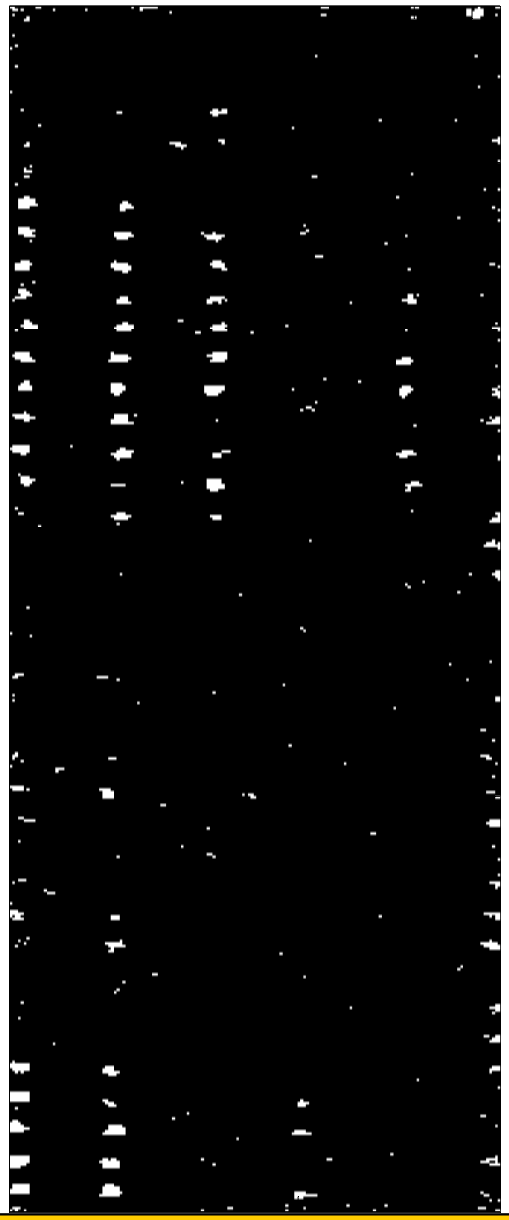
$$H_1 : x = Sa_t + b \qquad b \sim N(0, \sigma^2 \Sigma)$$

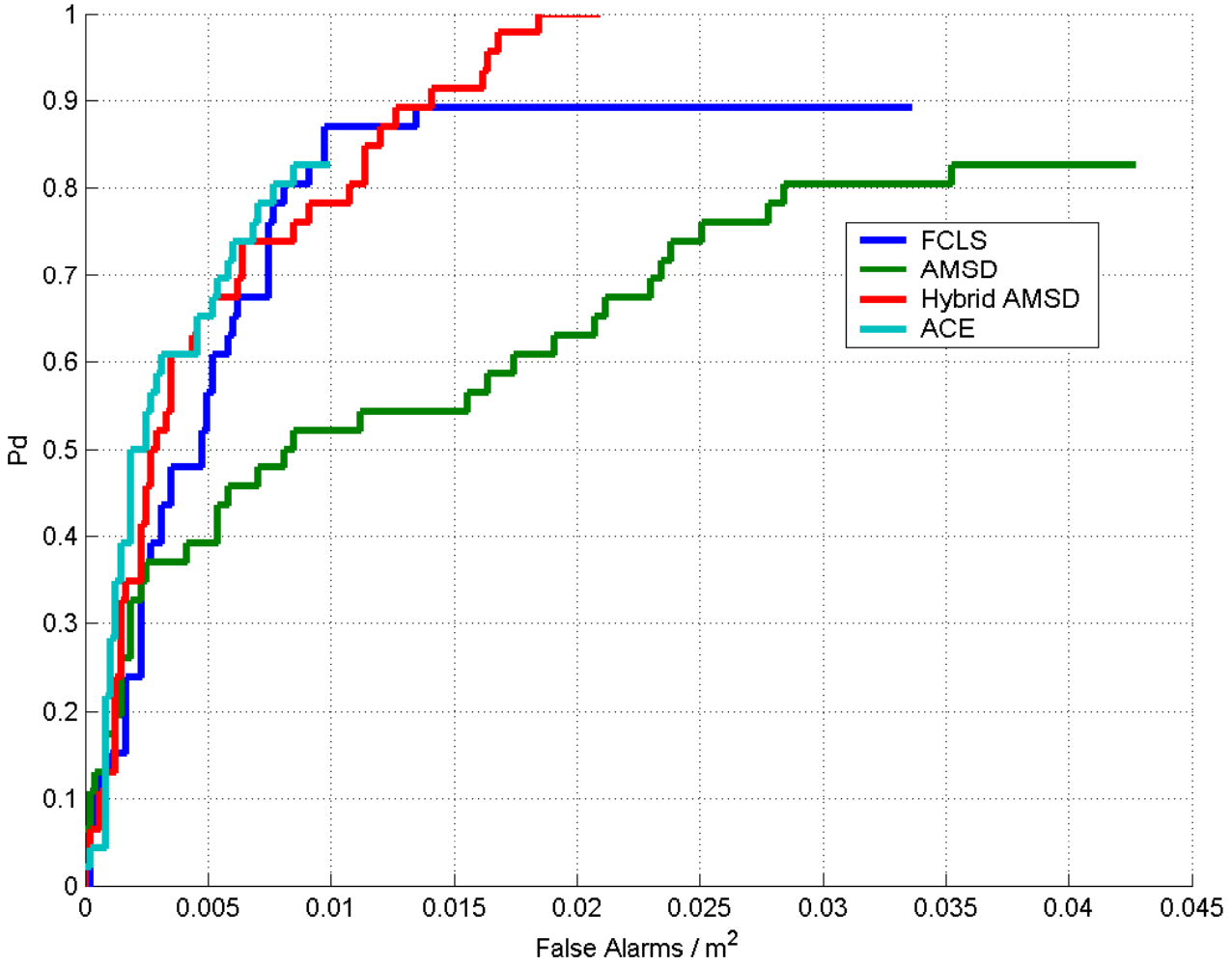
- Solution:
  - Estimate the unknown parameters using MLE
    - Abundances without non-negativity or sum-to-one constraint
    - Noise variance
    - Covariance is estimated in a local window around the pixel under test
  - Form a GLRT using the estimates to create the detection statistic

$$D(x) = \frac{x^T \Sigma^{-1} T (T^T \Sigma^{-1} T) T^T \Sigma^{-1} x}{x^T \Sigma^{-1} x}$$

- Issues:
  - Phenomenology: Violates the physical model that the abundances need to be non-negative and sum-to-one
  - Models background as a single Gaussian distribution

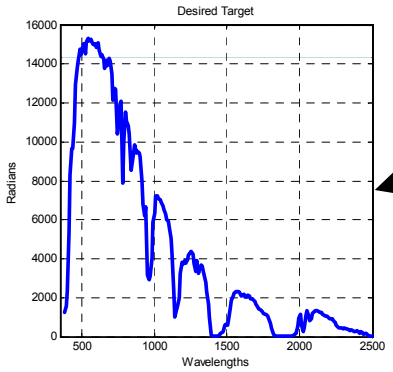
ACE Example: AHI Image 2535





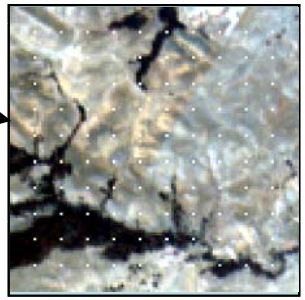
- **Top Pd for ACE is .82 due to limited performance in vegetation and riverbed areas**
- **Hybrid AMSD detects all targets**

Detectors in Development



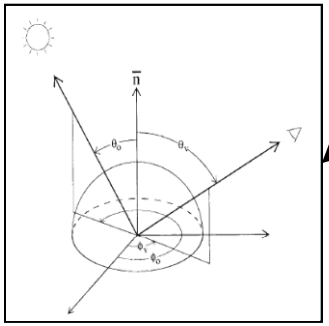
Target Signatures

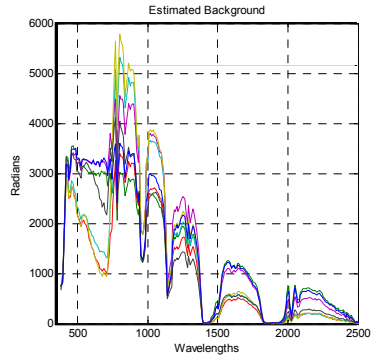
HSI



Forward Atmospheric Projection

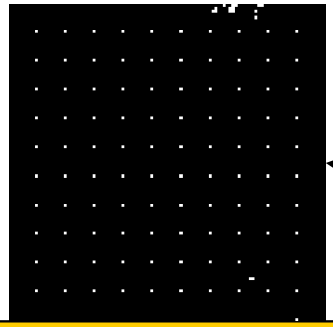
Background Estimation

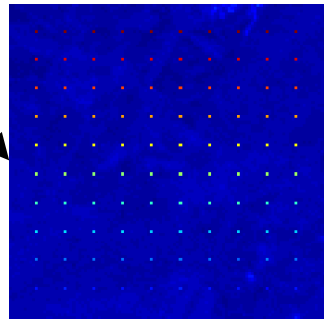




Detector

Post Processing





Output

Hybrid ACE Algorithm

- Idea: Create a detector based on GLRT that models the local background as a single multivariate Gaussian distribution that incorporates the abundance constraints
- Model:

$$H_0 : x = Ba_b + b, \qquad b \sim N(0, \Sigma)$$

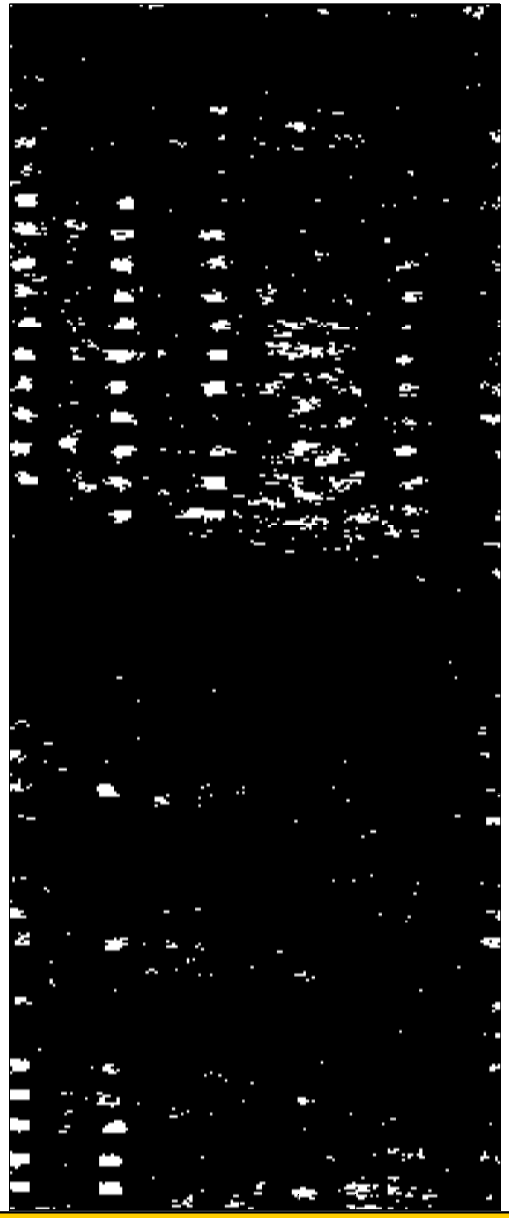
$$H_1 : x = Sa_t + Ba_b + b \qquad b \sim N(0, \sigma^2 \Sigma)$$

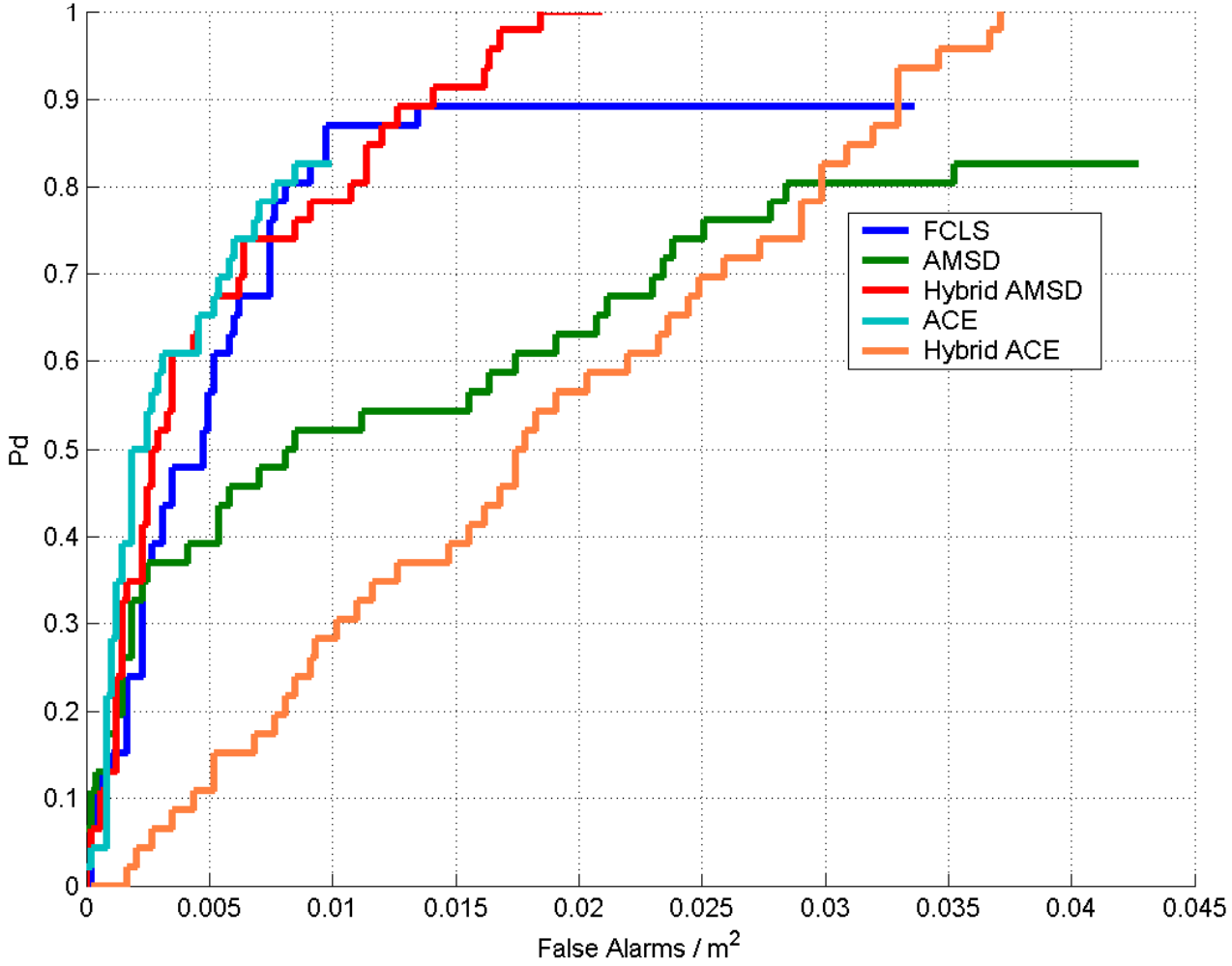
- Solution:
  - Estimate the unknown parameters
    - Abundances estimated using FCLS
    - Noise variance estimated using MLE
    - Covariance is estimated in a local window around the pixel under test
  - Form a GLRT using the estimates to create the detection statistic

$$D(x) = \frac{x^T \Sigma^{-1} T \hat{a}_w}{x^T \Sigma^{-1} x}, \quad (w \text{ refers to whitened space})$$

- Issues:
  - Models background as a single Gaussian distribution
  - Computationally expensive

Hybrid ACE Example: AHI Image 2535

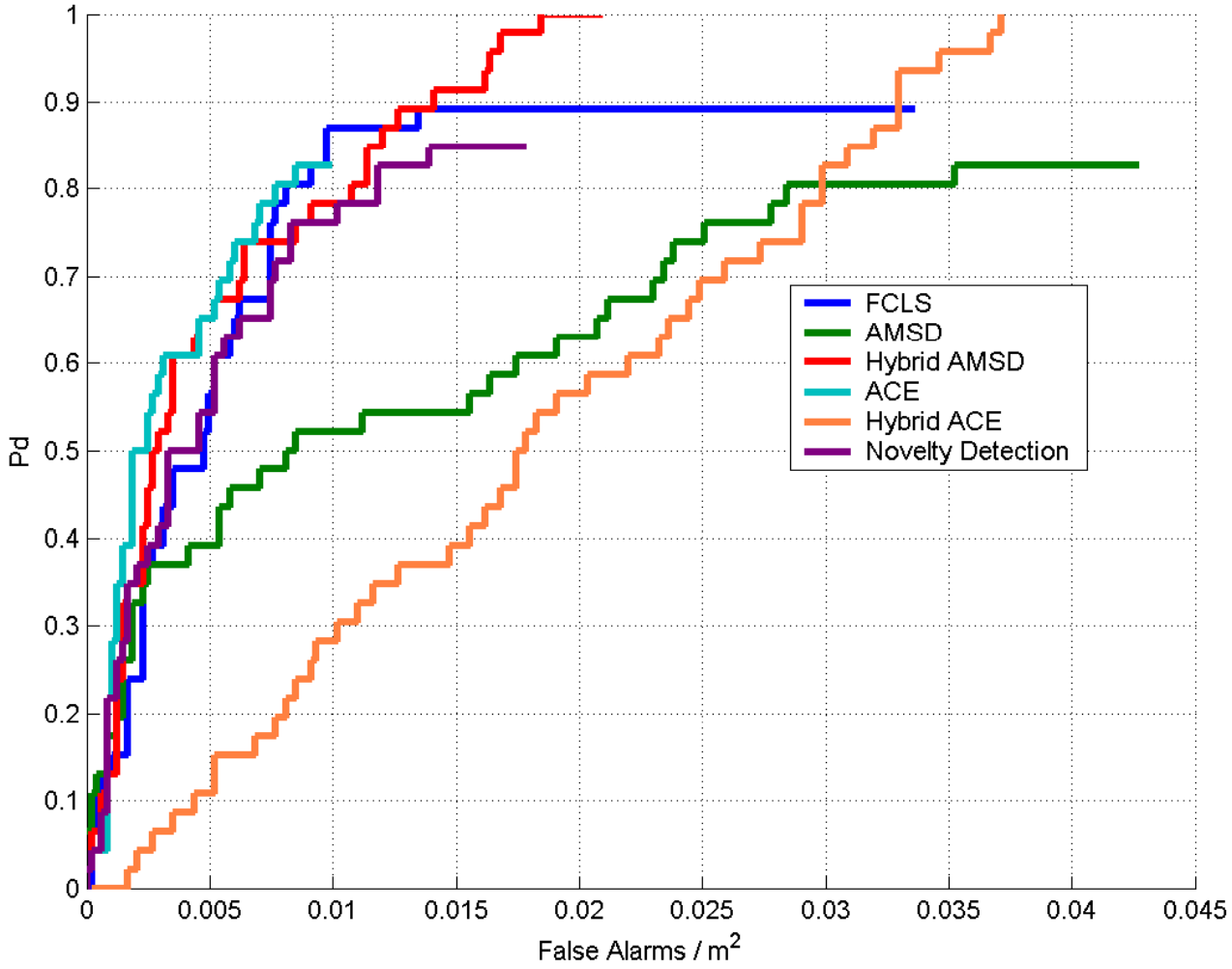




SVM Novelty Detection

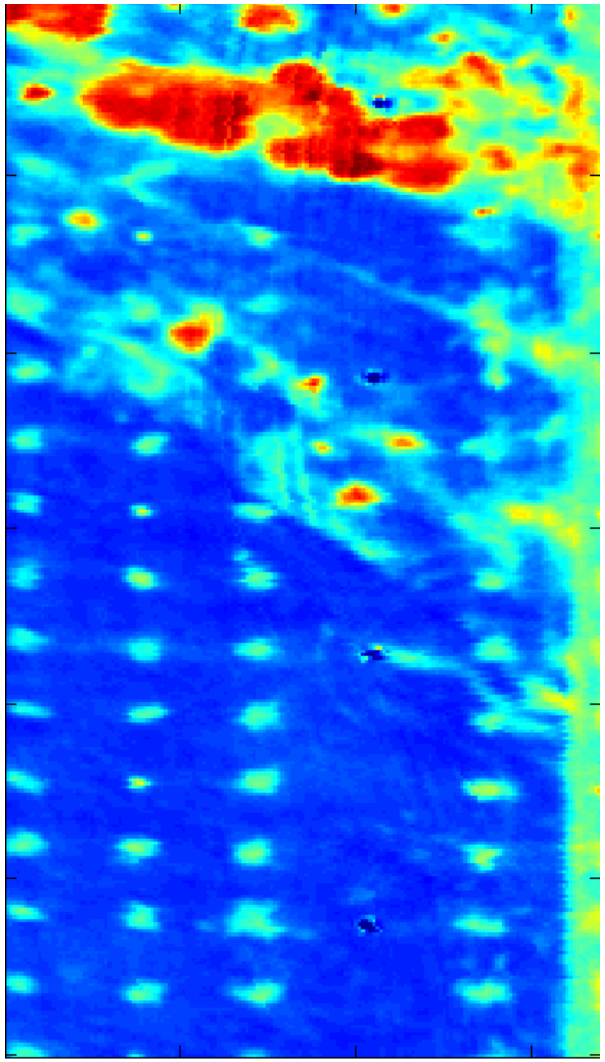
- Idea: Build a detector based on anomaly detection that is focused on the target space instead of the background
- Why SVM method?
  - Does not assume any statistical distribution
  - Can handle small training sample size
- Pros:
  - Quick implementation
  - Does not require many target signatures to form an acceptable solution
  - May be applied to forward projected or invariant subspace signatures
- Cons:
  - May be more sensitive to atmospheric models that generate the target radiance signatures

Novelty Detection Example: AHI Image 2535



- **Only three target signatures used**
- **Enhanced performance with additional training**
- **May use signatures from different grazing angles with different absorbed radiation (with subsequent forward modeling in radiance space of image of wide variety of radiance spaces) to enhance training**

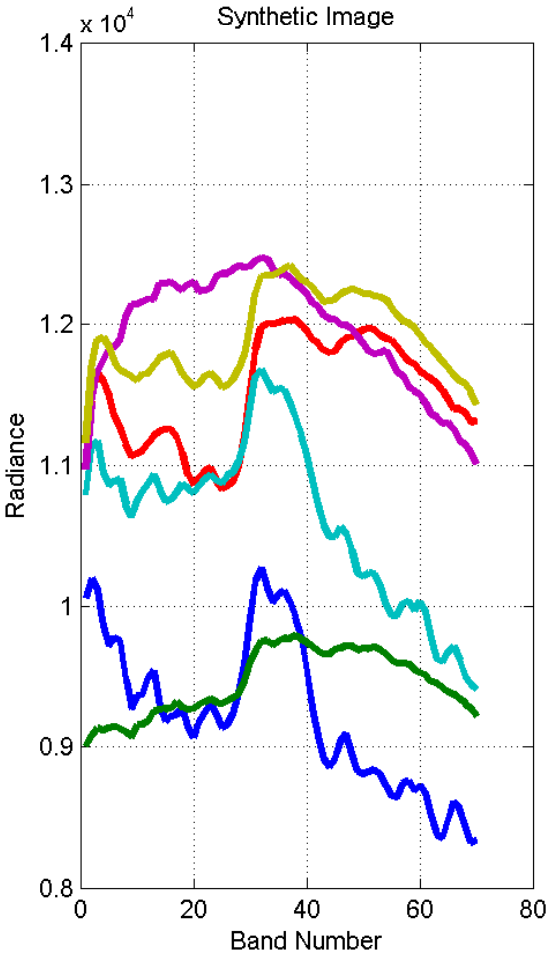
SVM Novelty Detection Example

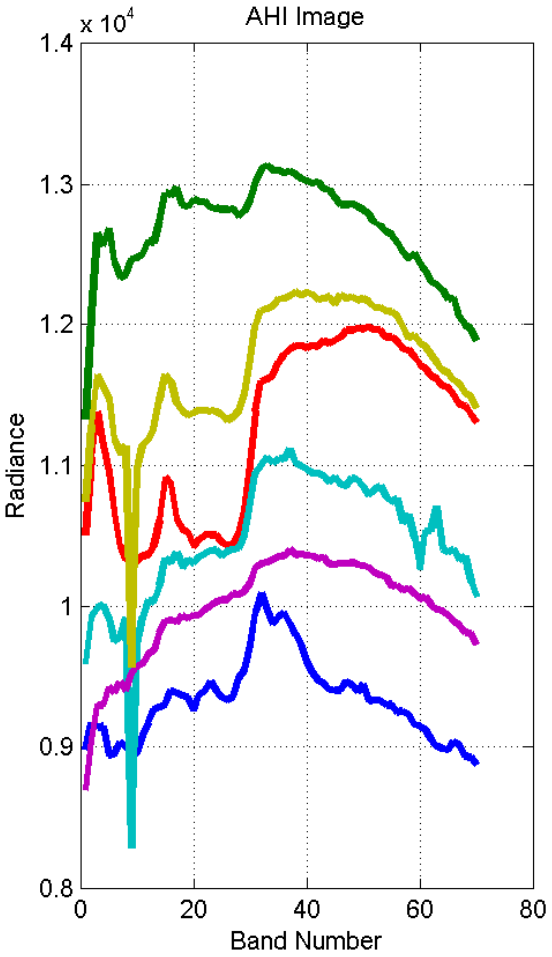


Vegetation signatures under 17344 different conditions provided by G. Healey

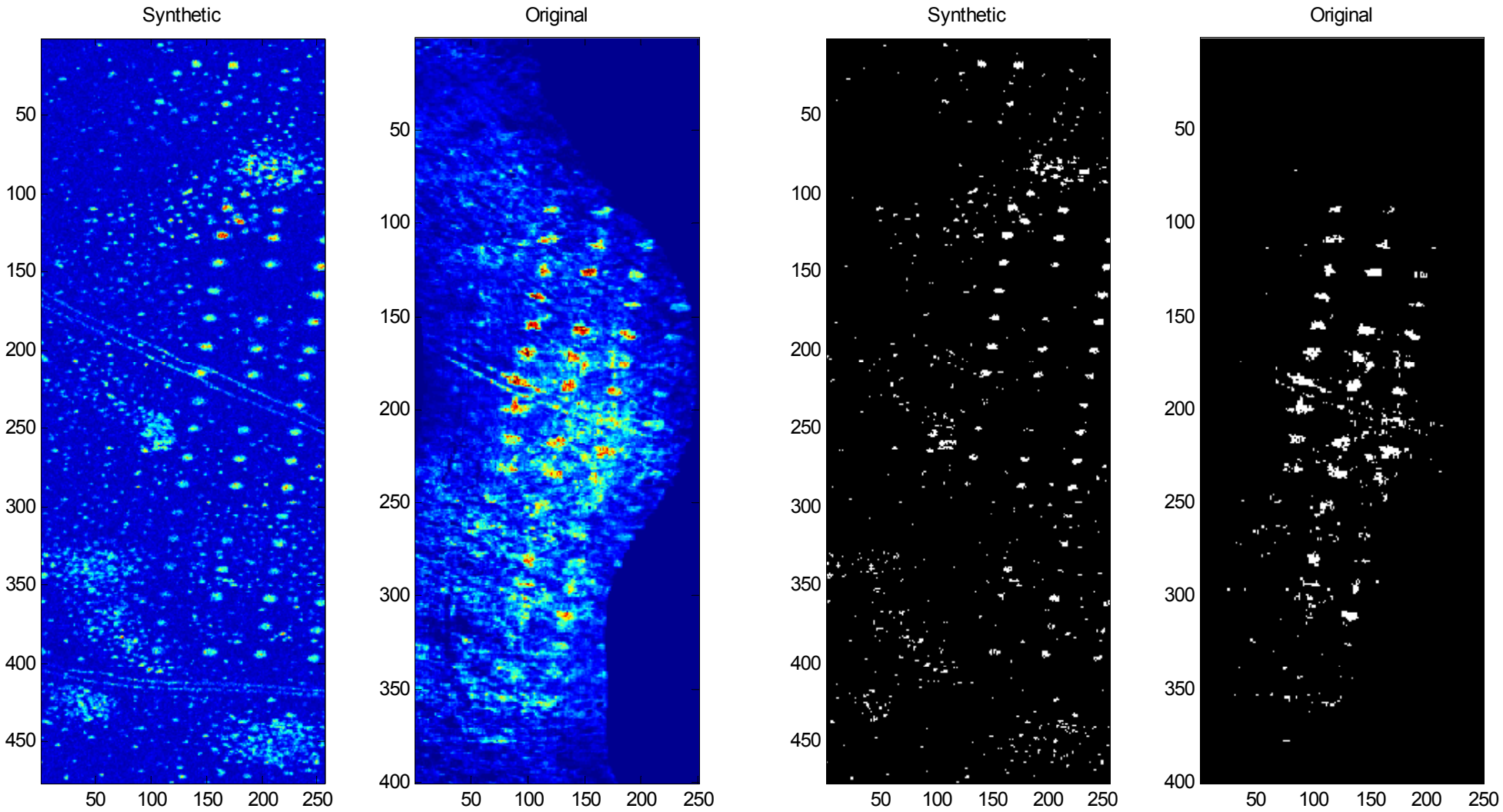
Synthetic Imagery Analysis

- Received synthetic LWIR imagery from RIT to compare to real AHI image
- Comparison was done relative to detector performance
- Experiment:
  - Estimate background subspace using MDL criterion
  - Draw one target pixel from the synthetic image for the target subspace
  - Characterize detector performance
  - Compare endmembers extracted
- Results: (Noon 700 ft)
  - Synthetic image background was capture in 6 endmembers instead of 3 for AHI image
  - Detectors had much higher FAR in synthetic image
  - Synthetic image had “cleaner” target detections
  - Significant improvement in updated synthetic scene over initial version





Synthetic Imagery

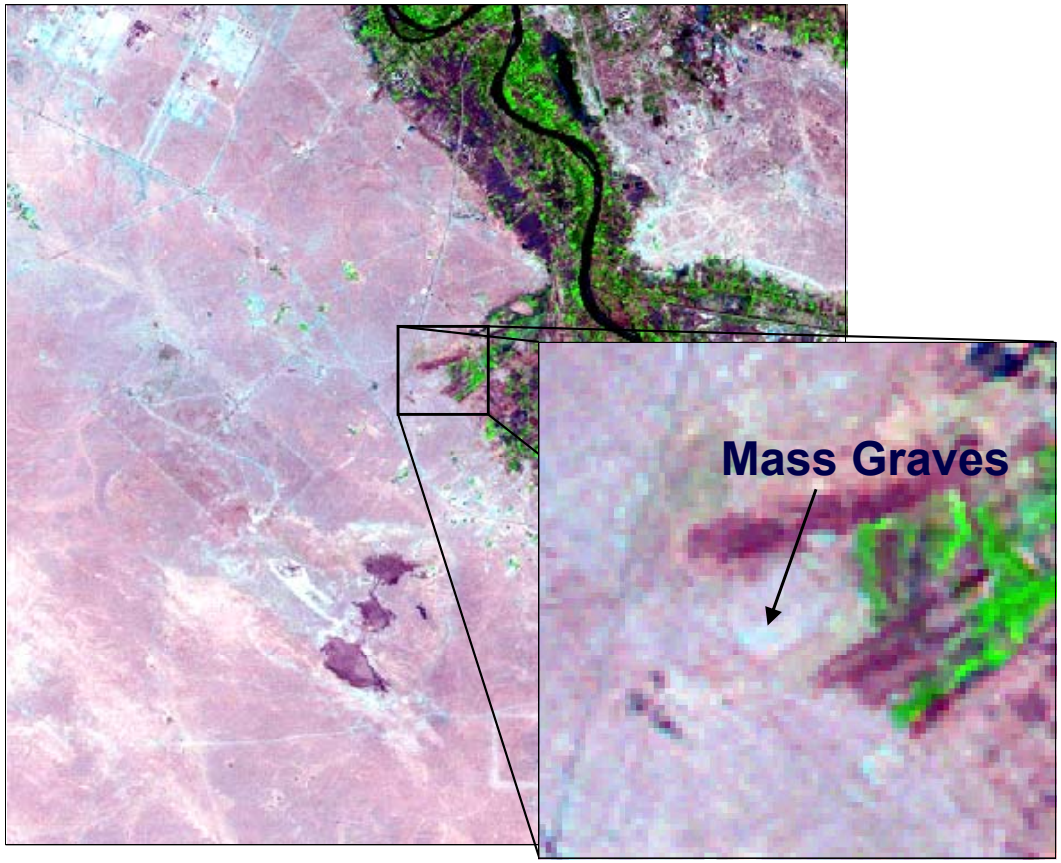


Initial Change Detection Work

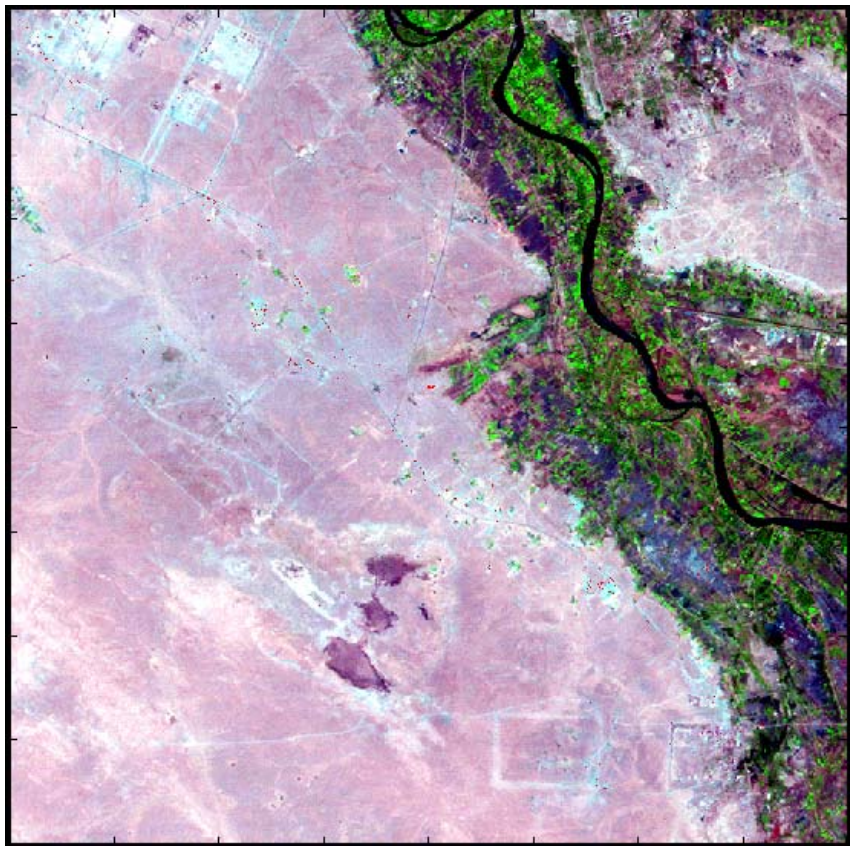
- Using the ATD/R methods developed, applied these techniques to change detection in LANDSAT imagery (under work for The Johns Hopkins University Applied Physics Laboratory)
- Motivation:
  - Identify specific change (e.g. undisturbed soil to disturbed soil)
  - Remove nuisance changes (FAs) from consideration
- Current work:
  - Registered imagery at sub-pixel level (thanks to Mr. Myron Brown)
  - Used Chronochrome algorithm to “equalize” the images
  - Applied the Hybrid AMSD method to the resulting difference image
  - Results improved upon original work performed by the Spectral Operations Research Center (SORC)
- Future work:
  - Utilize phenomenology and forward projection within process
  - Develop these techniques to use a site model as has been done in previous University of Maryland projects
  - Perform change detection without the need for sub-pixel registration
  - Use these same methods to combine detectors to reduce false alarms on the same image.

Change Detection Results

Original Image

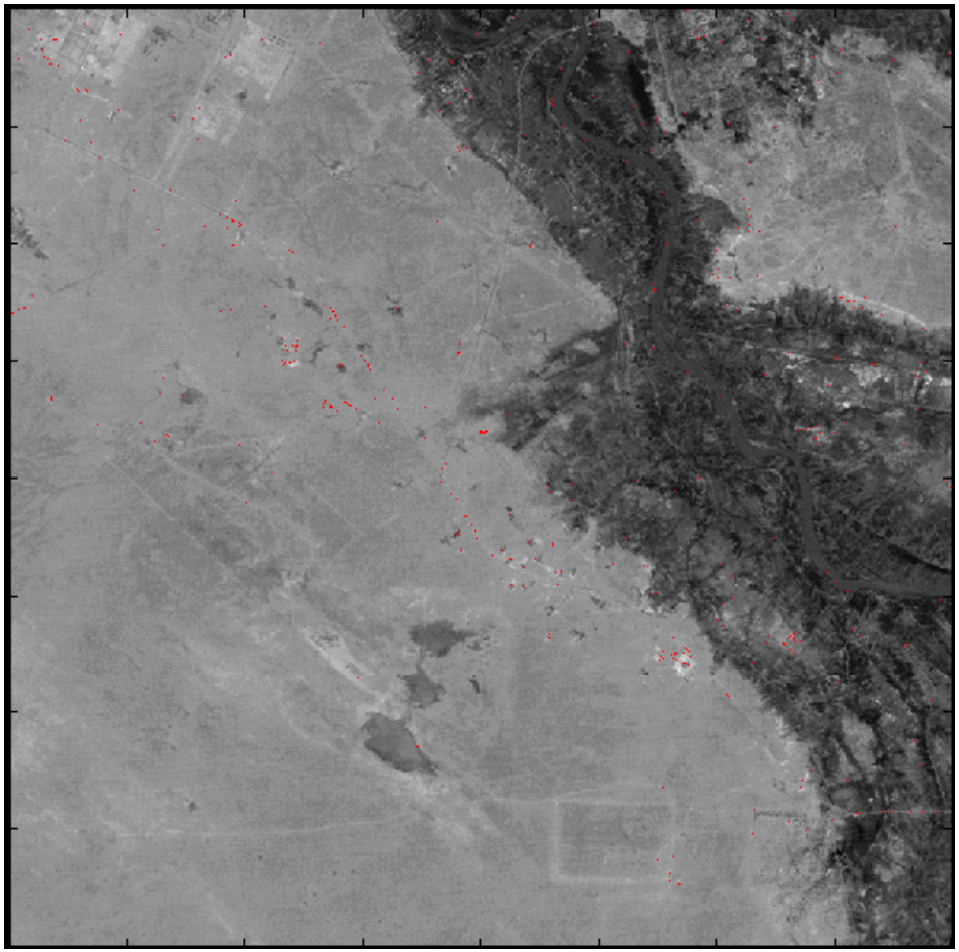


Detection Image

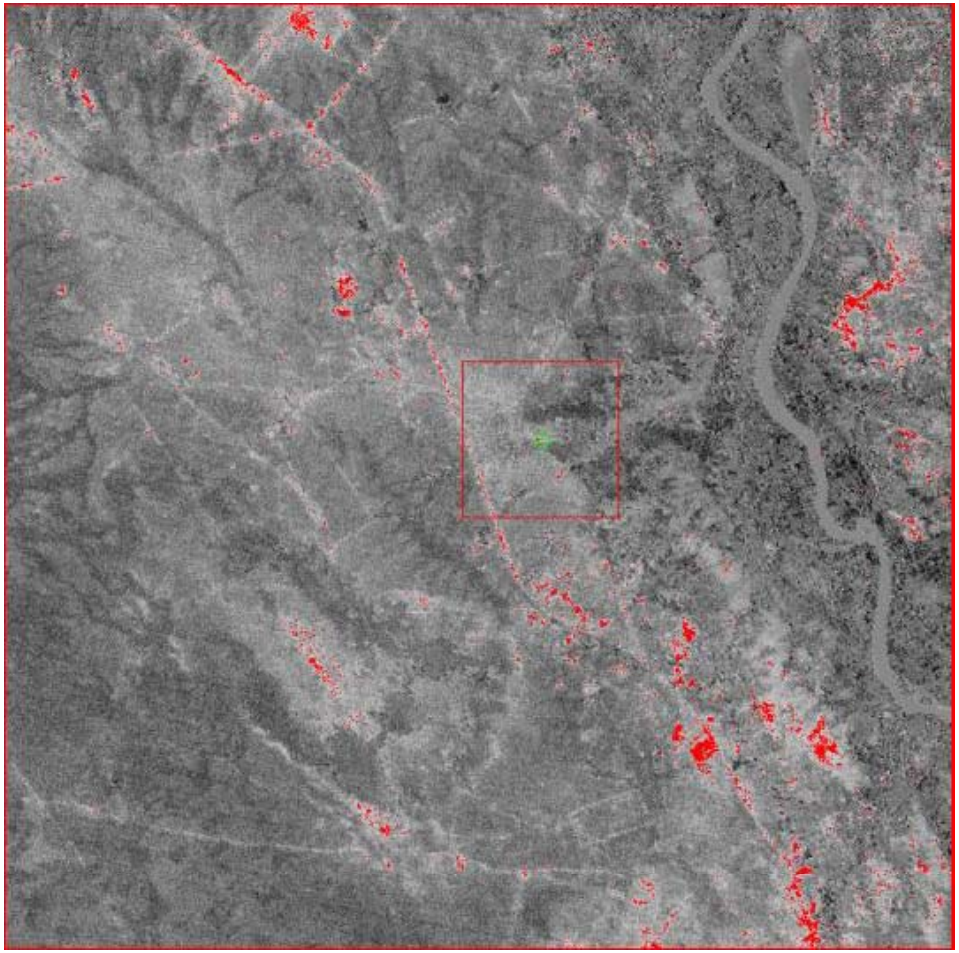


Change Detection Comparison

Hybrid Change Detection



SORC Detection



2004 Research Accomplishments

- Nearly automated detection system developed (inputs: target signatures, HSI image; output: detections)
- Development of phenomenologically-based subpixel target detection algorithms
  - Hybrid AMSD detector
  - Hybrid ACE detector
  - SVM based novelty detector
  - MDL-based criterion for automatic determination of background subspace dimension
- Hybrid AMSD detector developed (IGARSS 2004)
- Demonstrated use of MDL criterion for estimating number of endmembers in subspace (ASC 2004)
- Demonstrated effectiveness of forward-projection detection methodology
  - Generated forward-projection target signatures for AP Hill COMPASS data
- Demonstrated Hybrid detector to exhibit enhanced performance relative to current state-of-the-art detectors
  - Processed AHI and COMPASS data using FCLS, AMSD, and Hybrid detectors
- Delivered detection images to University of Florida in support of their fusion work
- Developed nonparametric invariant detector using 1-Class Support Vector Machines (Novelty detection)
- Scene anomaly detection at the ground level using HSI (D. Rosario)

Spectral Automatic Target Detection/Recognition (ATD/R)

2005 Research Efforts

This section contains ITAR information

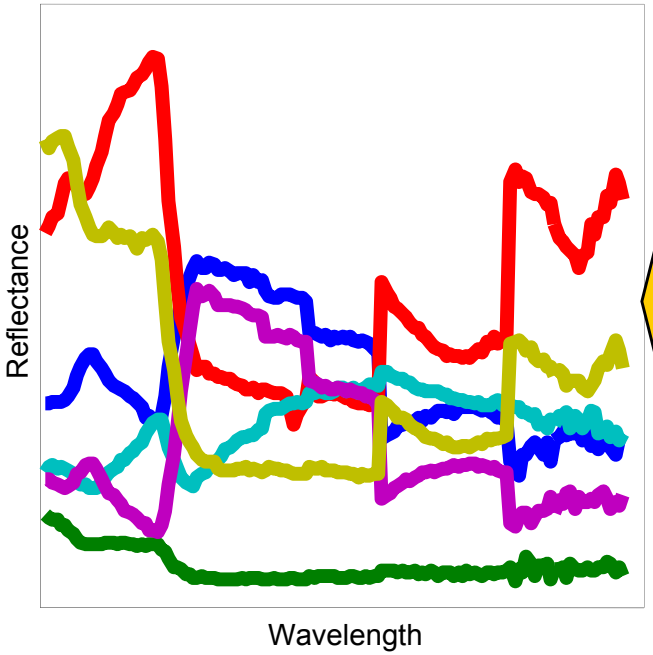
2005 Research Topic Outline

- In-Scene Estimation of Target Signatures
- Comparison of Parametric Detectors
  - Standard Detectors
  - Whitened Detectors
- Additional Topics
  - Non-Parametric Detector Investigation
  - Adaptive Thresholding
  - Phenomenologically-Constrained Kernels
  - Spectral Object Level Change Detection

In-Scene Estimation of Target Signatures

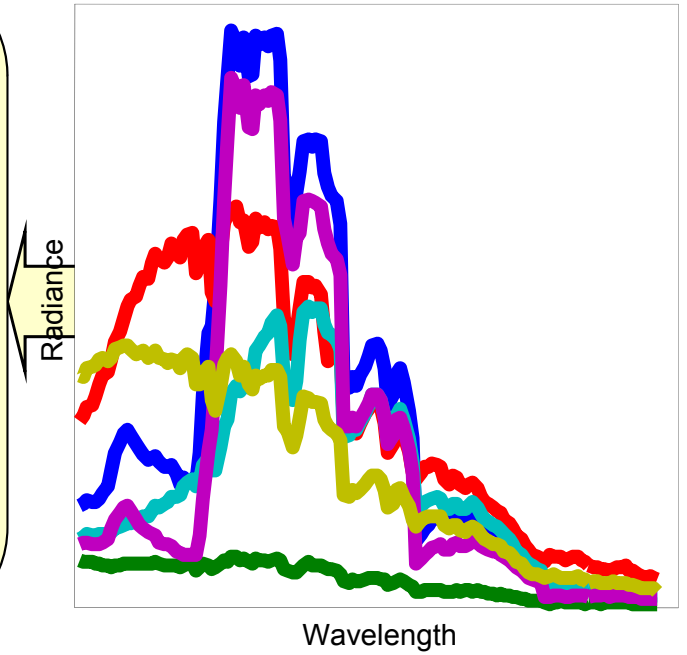
Radiance Inversion Methods

Radiance inversion methods were originally developed spectral analysis purposes. These methods were designed to produce products such as mineral maps which is a classification problem requiring all materials in the image to be identified.



Algorithms:

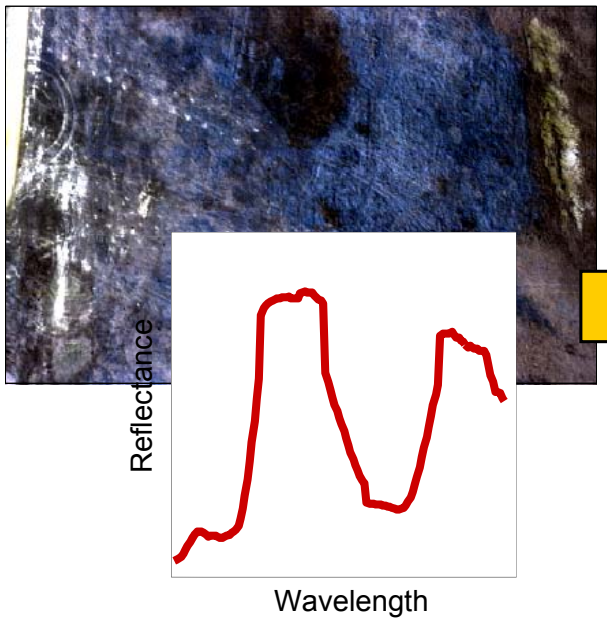
- **Internal Average Relative Reflectance (IARR)** uses divides the image by its mean radiance to get an estimated reflectance.
- **FLAASH** uses an atmospheric model with some simplifying assumptions to invert radiance back to reflectance.



These are well developed algorithms that have proven very useful for analysis and classification. However, these algorithms require increased processing time to invert every pixel and have to make simplifying assumptions to perform the inversion.

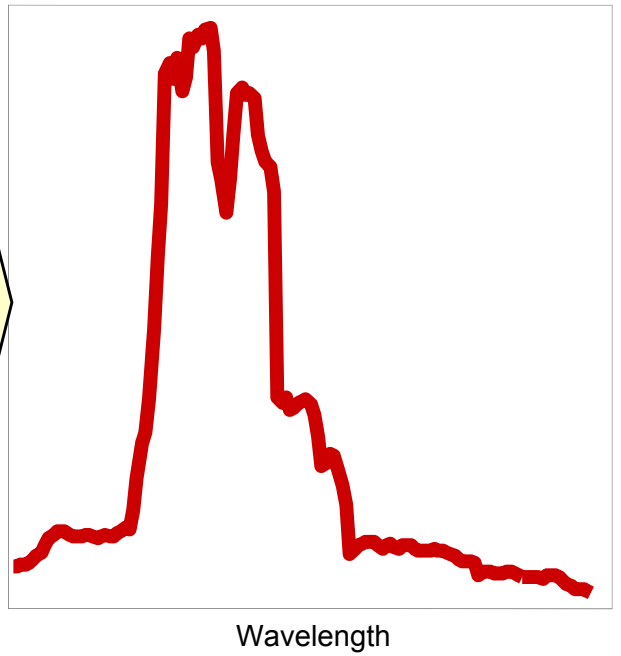
Radiance Projection Methods

Radiance projection methods were designed for target detection applications. Instead of inverting every pixel in the image, the algorithms try to predict what the target reflectance signature would be in radiance values seen at the sensor.



Algorithms:

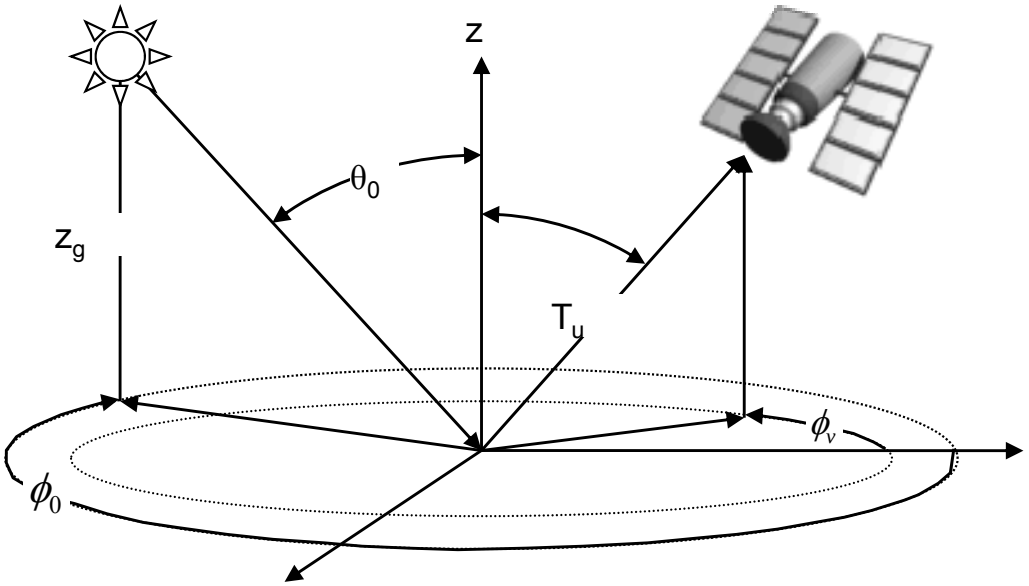
- **Spectral Radiance Generator** uses weather and source-target-receiver geometry to predict the target radiance signature.
- **Invariant Subspace Method** generates an invariant subspace using PCA applied to over 17,000 differently simulated environments.

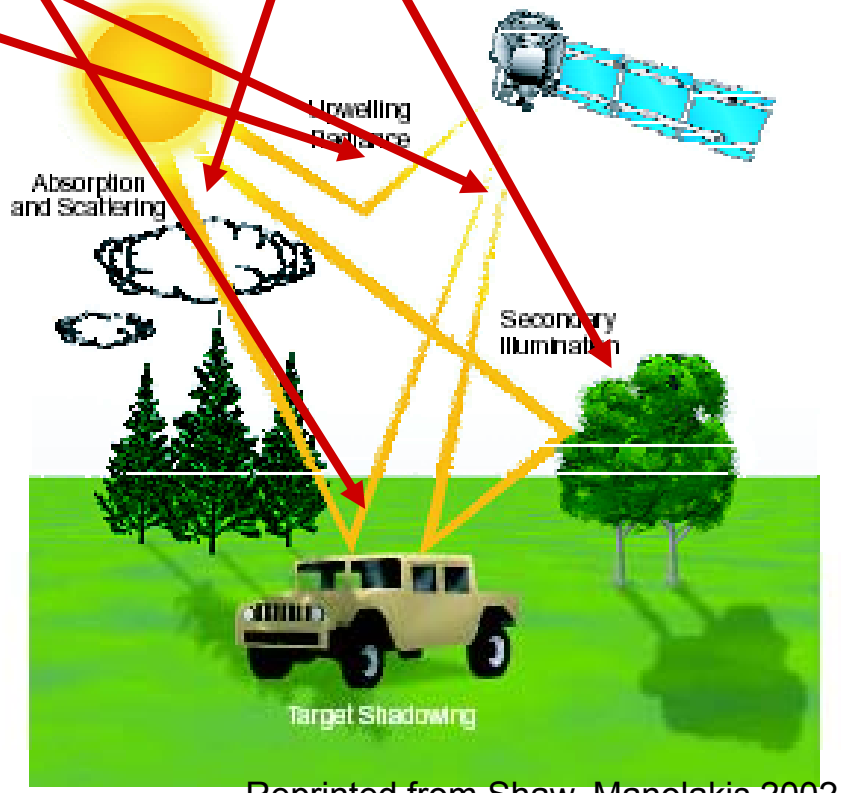


The algorithms mentioned here have had great success, but have significant overhead. SRG requires detailed weather information and precise source-target-receiver geometries while ISM requires significant processing.

Phenomenological Effects

$$L(x,y,\lambda) = T_u(z_g, z_u, \theta_v, \phi_v, \lambda)R(x,y,\lambda)\left[KT_d(z_g, \theta_0, \phi_0, \lambda)E_0(\lambda)\cos\theta_0 + \int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi/2} E_s(\theta,\phi,\lambda)\cos\theta\sin\theta d\theta d\phi\right] + P(z_g, z_v, \theta_v, \phi_v, \lambda)$$





Reprinted from Shaw, Manolakis 2002

Atmospheric Effects

The radiance value seen at the sensor can be modeled by the equation below. This equation shows how reflectance signatures are related to radiance values through transmittance, scattering, and absorption effects.

$$L(x,y,\lambda)=T_u(z_g,z_u,\vartheta_v,\phi_v,\lambda)R(x,y,\lambda)[KT_d(z_g,\vartheta_0,\phi_0,\lambda)E_0(\lambda)\cos\vartheta_0+$$

$$\int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi/2}E_s(\theta,\phi,\lambda)\cos\vartheta\sin\theta d\theta d\phi]+P(z_g,z_u,\vartheta_v,\phi_v,\lambda).\ [1]$$

$$\hat{L}=RA+RB+P$$

Direct Solar Path Effects:
Path from sun to target to sensor

Sky Effects:
Light from Scattering to target to sensor

Path Effects:
Path from sun to atmosphere to sensor

The equation is broken down into its three respective parts: solar path effects, sky illumination effects, and direct path effects. The strongest of these is the solar path effects followed closely by the sky illumination effects.

Estimation of Atmospheric Effects

Each of the three atmospheric effects can be estimated with varying levels of complexity from pure in-scene methods to pure model based methods. We developed an in-scene method called Average Relative Radiance Transform (ARRT) to estimate these effects.

- Average Relative Radiance Transform (ARRT) Version 1:
  - Estimates only A using the spectral mean of the image based on the IARR algorithm.
- ARRT Version 2:
  - Estimates only A using the spectral mean of the spectrally flat endmembers. Will identify when these are not available.
- ARRT Version 3:
  - Currently being tested to incorporate an estimate of B as well as A. This is important for imagery taken from high altitudes.

Having estimated the atmospheric effects, ARRT multiplies the desired target reflectance signature by the atmospheric estimate to obtain a target radiance signature. This is much more efficient than inverting the entire image back to reflectance.

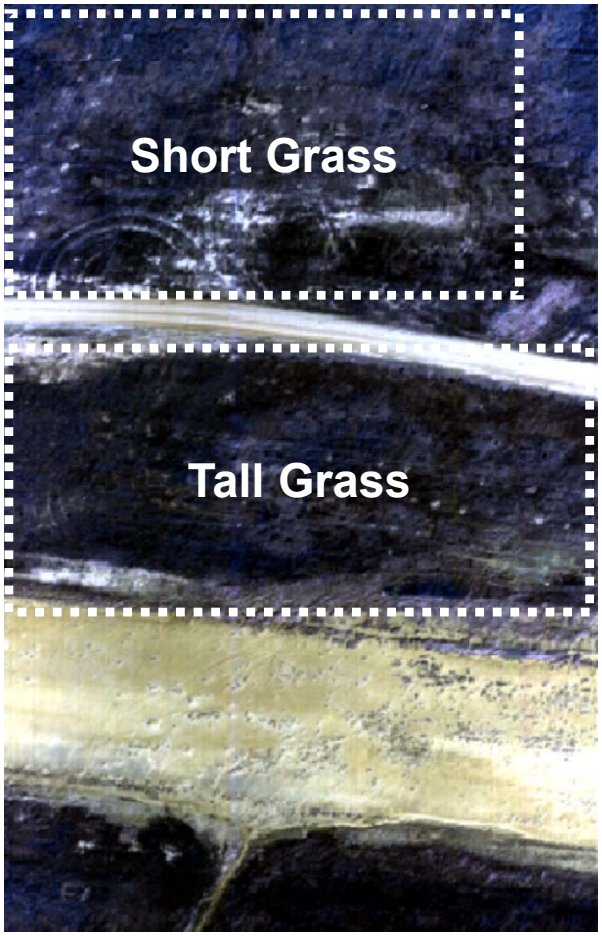
Data Background

To evaluate the usefulness of ARRT, we evaluated seven different images at varying altitudes with varying targets from the Wide Area Airborne Minefield Detection (WAAMD) program. Target fill factors vary from 100% to as low as 10%.

Data Facts:

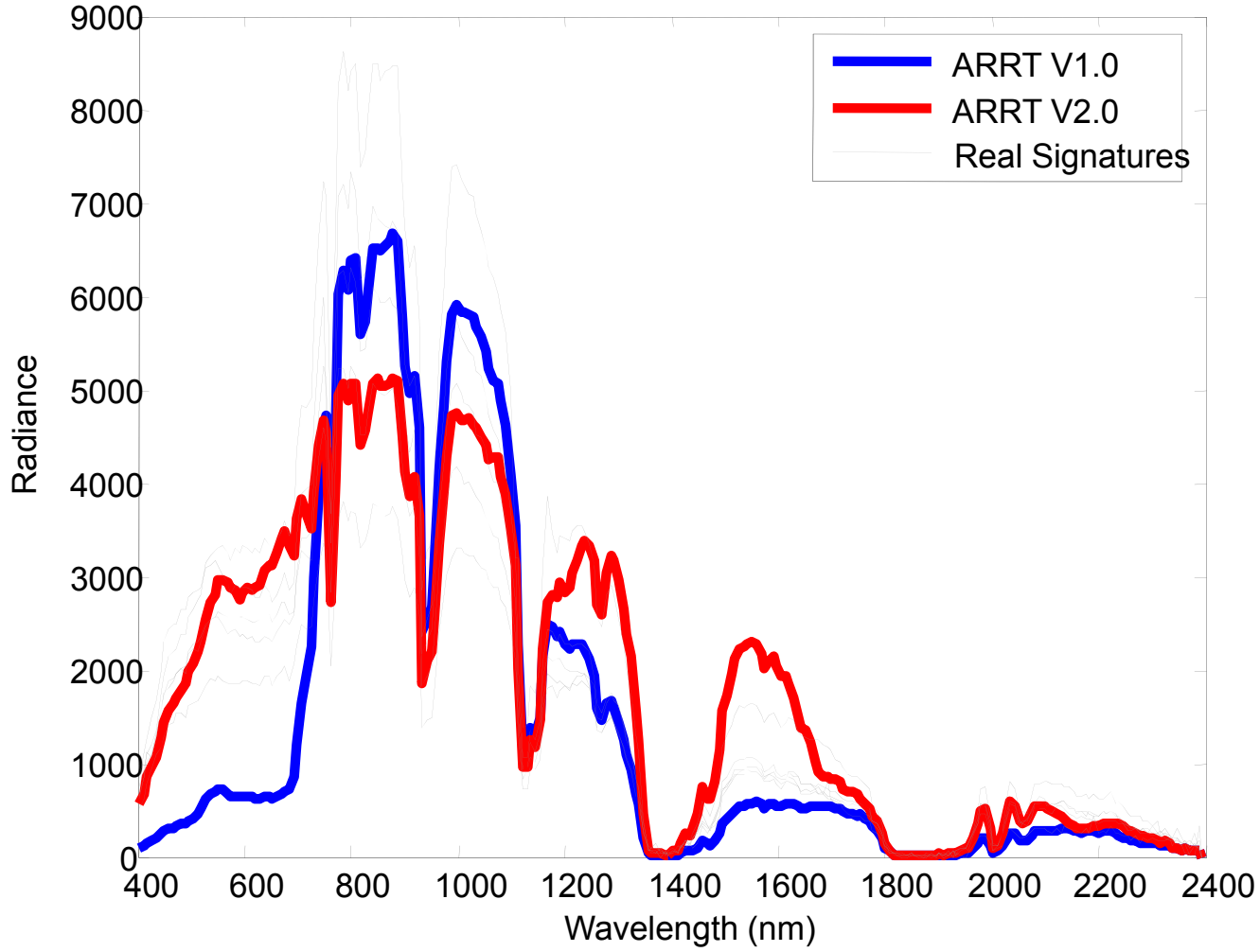
- Targets Used
  - Target 1: Round, white plastic, 15.24 cm.
  - Target 2: Round, green metal, 30.48 cm.
  - Target 3: Square, green metal, 30.48 cm.
- Seven images were used as listed below

Image	Description	Alt (m)	GSD (m^2)	Area (m^2)
1	Short Grass	272	0.0181	8803.8
2	Short Grass	315	0.0243	8709.1
3	Short Grass	272	0.0181	3243.5
4	Short Grass	298	0.0217	3333.1
5	Short Grass	309	0.0233	2982.4
6	Both	1220	0.1823	18668.0
7	Short Grass	1216	0.1815	18586.0



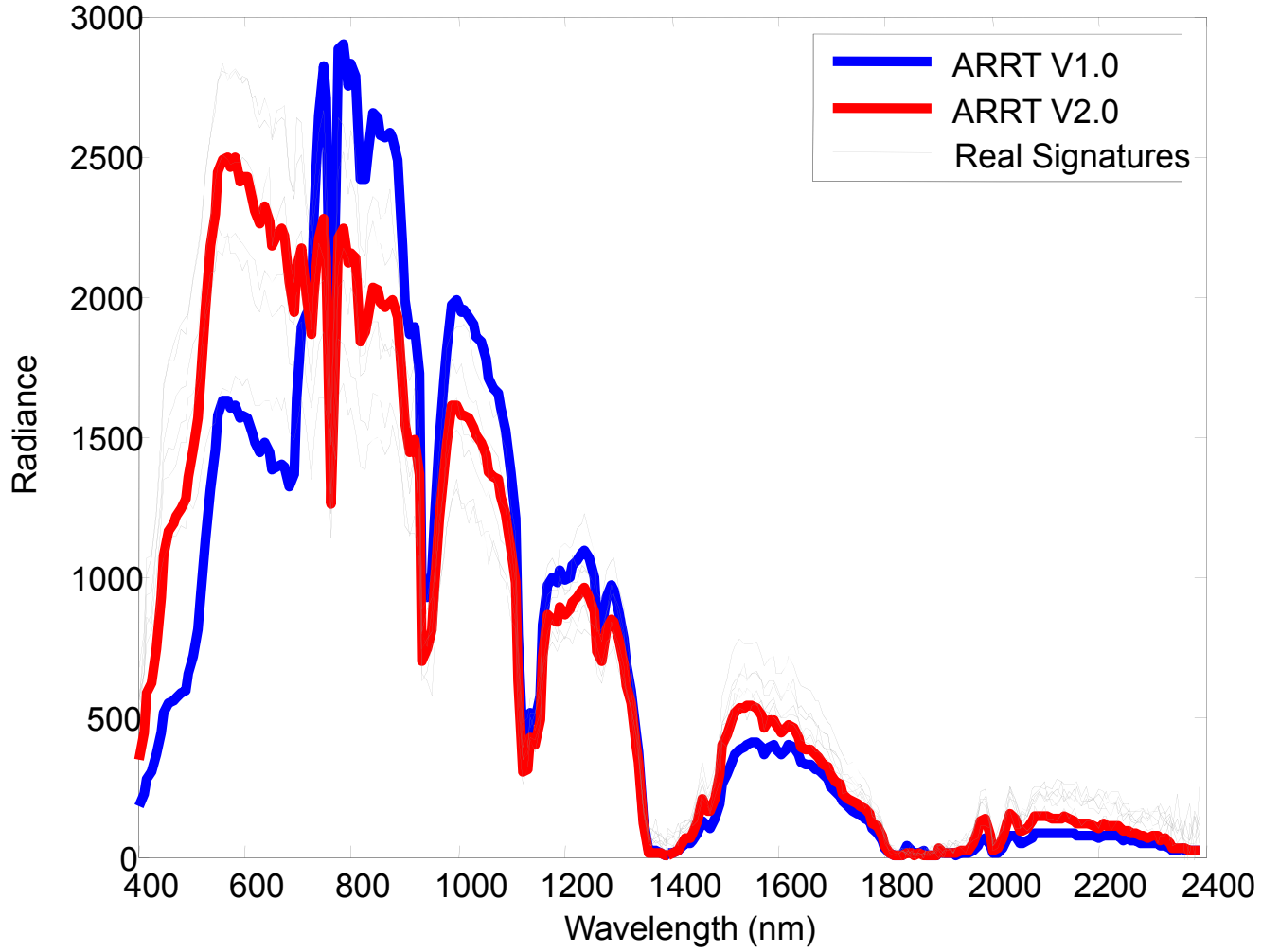
Comparison of ARRT to Real Signatures

The new version of ARRT produces a much more realistic signature. The ARRT V2.0 signature has only minor differences compared to the actual signatures in the image.



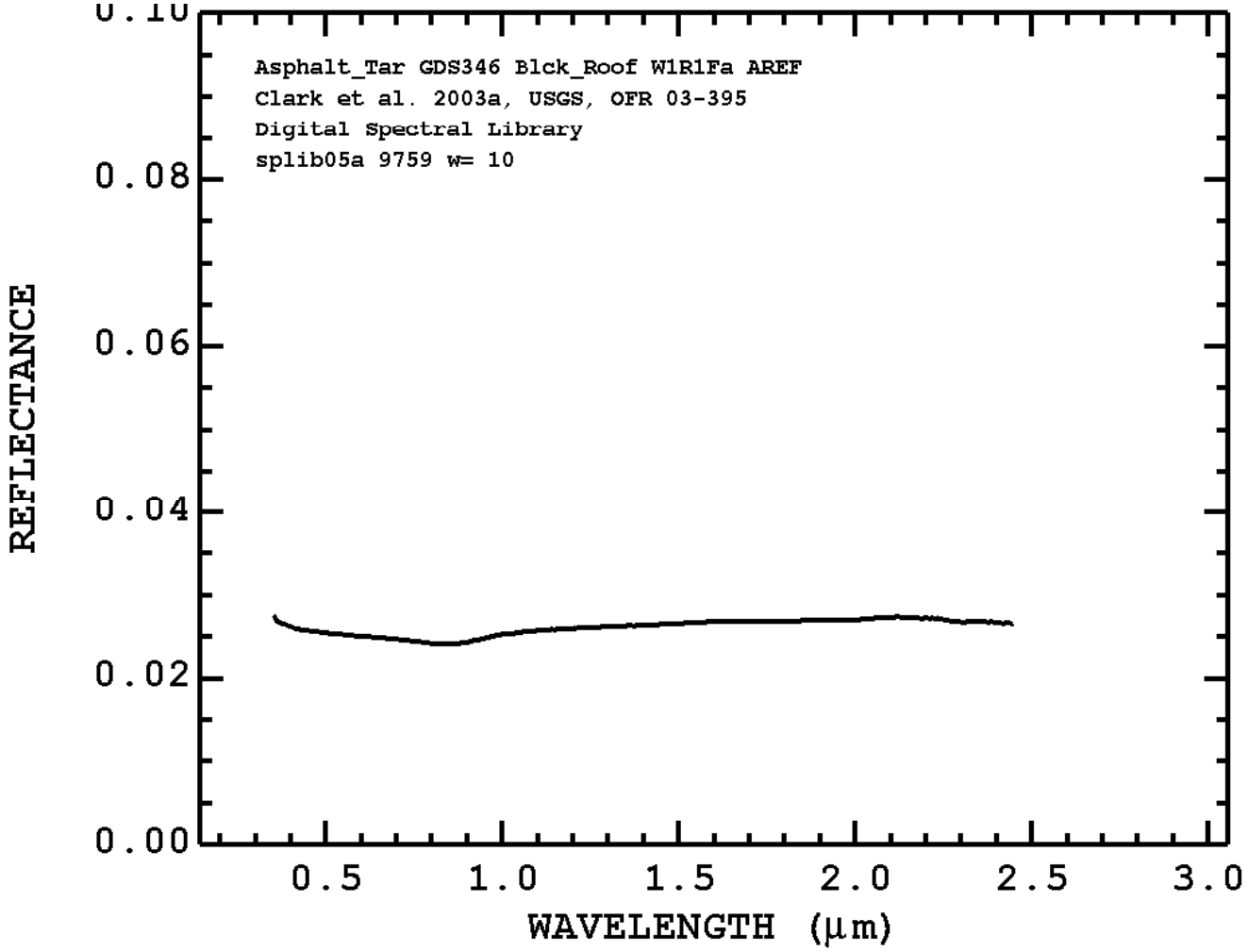
Comparison (cont.)

Unlike the previous version of ARRT, the new version produces a much better target signature. Note that the mismatch in the VIS/NIR bands have been corrected.



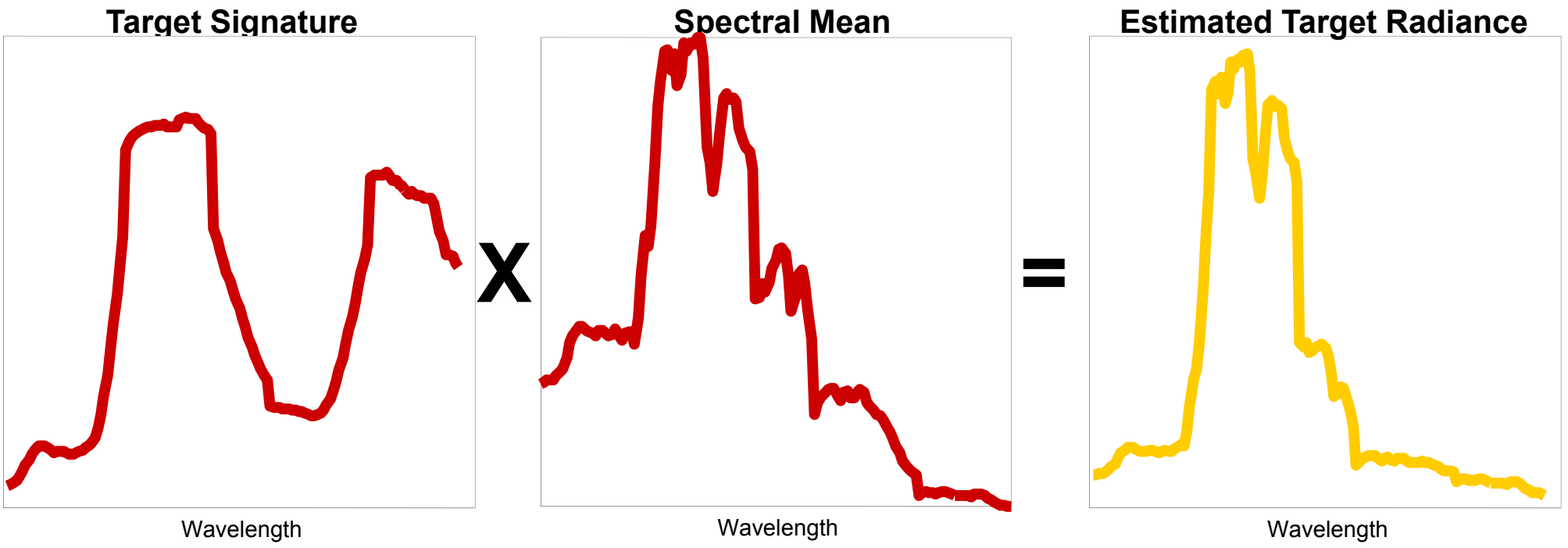
Approximately Flat Reflectances

One of the primary endmembers in a number of the WAAMD images is asphalt. As seen in the USGS spectral database, this is a signature with nearly flat reflectance.



Average Relative Radiance Transform

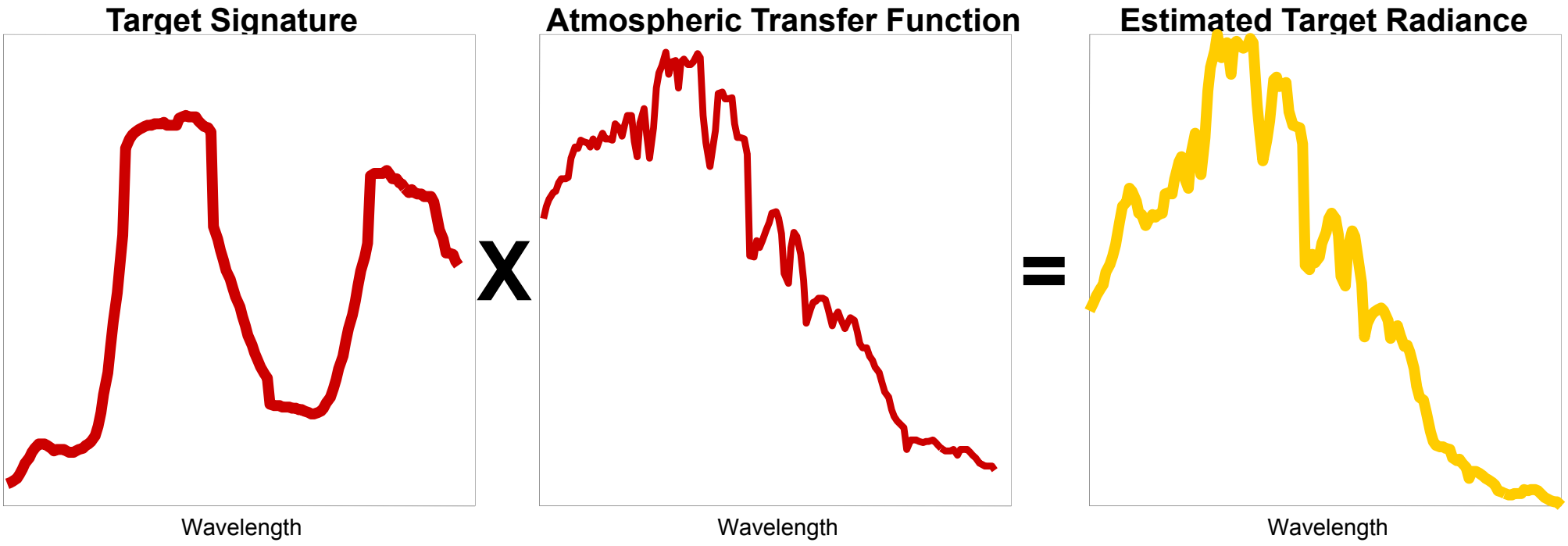
This method was developed based on the IARR algorithm. It fills the niche of needing an atmospheric compensation algorithm that is computationally simple and can be done in near real-time for operational subpixel detection systems.



The first variant of this algorithm works in many cases, but there are limitations with it as well. The algorithm (like IARR) works very well in desert areas, but has trouble handling some target types in other areas that have dense vegetation.

ARRT Version 2

This method is similar to the original ARRT formulation except that it chooses only those endmembers that have nearly flat reflectance signatures. These signatures are typified by having "smooth" signatures that tend to decline with increasing wavelength.



This variant seems to work in all of the WAAMD images. However, we know this method will not work in images that contain solely vegetation. Of course, the Empirical Line Method can be used in that case.

Subpixel Detector

The subpixel detector we used in this analysis is the Adaptive Cosine/Coherent Estimate (ACE). This detector is a standard detector in the literature and has demonstrated consistently good performance with its simple architecture.

Derivation:

- Assume the following hypothesis:

$$H_0 : x = b$$

$$H_1 : x = Sa_t + b$$

$$b \sim N(0, \sigma^2 \Sigma)$$

- Using maximum likelihood estimates in a generalized multivariate analysis of variance (GMANOVA), we arrive at the following detection statistic

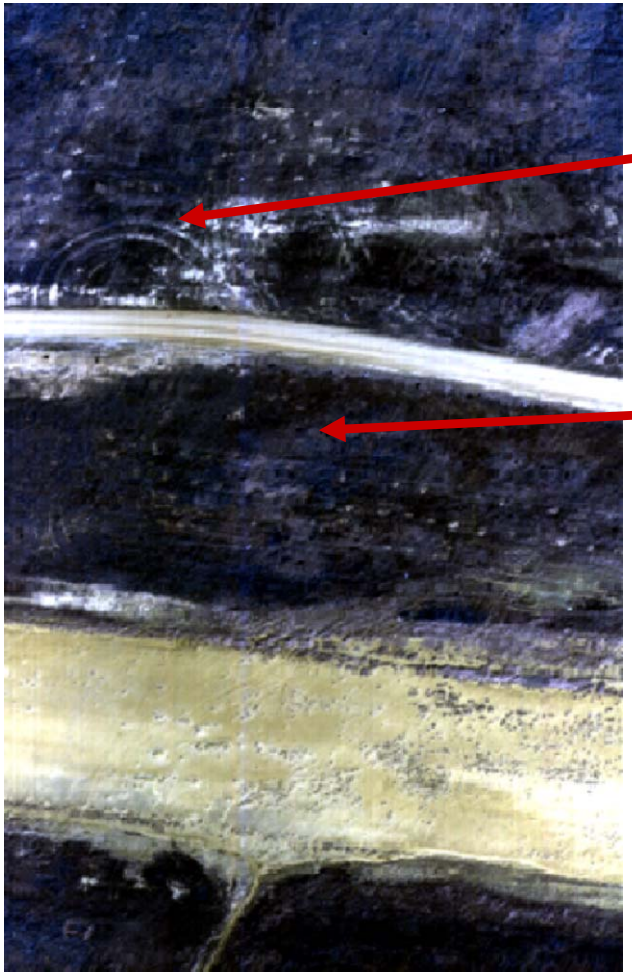
$$D(x) = \frac{x^T \Sigma^{-1} T (T^T \Sigma^{-1} T) T^T \Sigma^{-1} x}{x^T \Sigma^{-1} x}$$

Notes:

- This detector uses an unstructured background We chose to go this way to minimize any processing artifacts due to endmember estimation.
- The detector's parameters can either be locally or globally estimated. In this case, we used global estimation since local did not improve performance. This also gives us the benefits of real-time processing.

Data Results

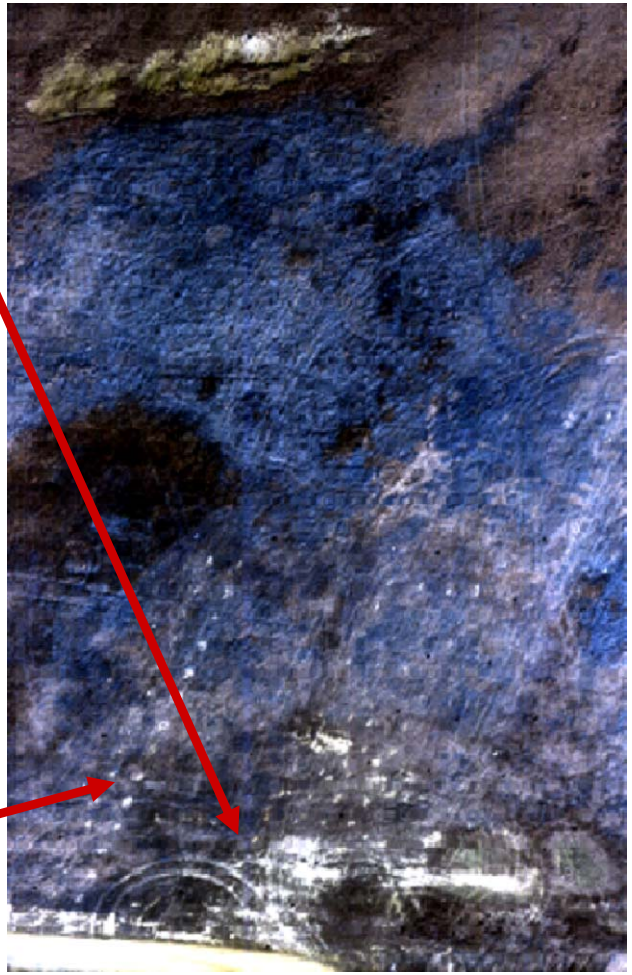
These images collected at an altitude of 1.2 km contain subpixel targets that have fill factors as small as 10%. These are the most difficult images for subpixel detection.



Short Grass Sites
Easier to detect in these areas since the targets are mostly uncovered.

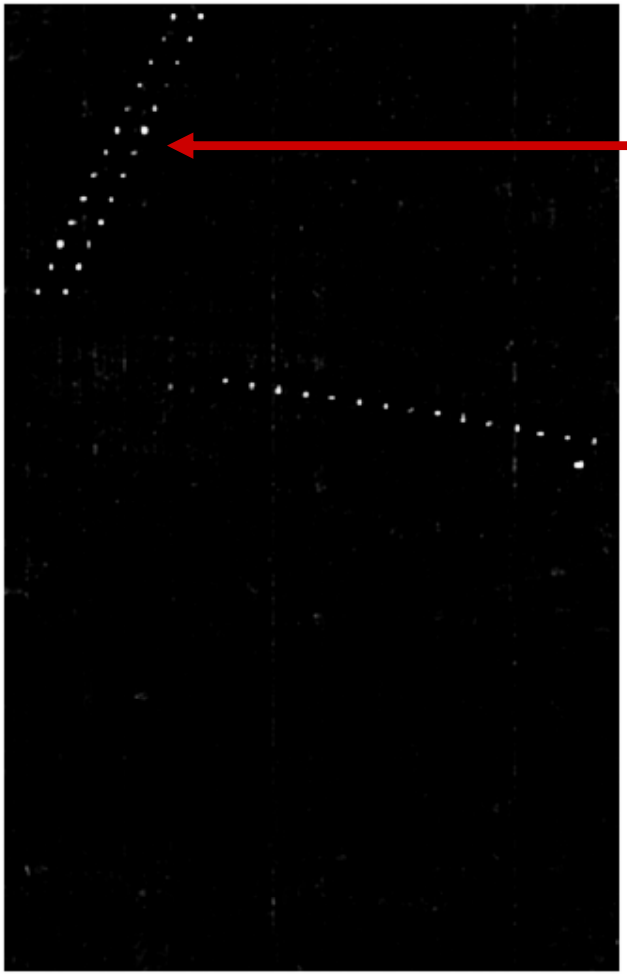
Tall Grass Site
Harder to detect in this area since the targets are mostly covered by tall grasses.

Pseudo-Targets
These round marks in the image are not the targets we are detecting in this analysis.



Data Results

These images collected at an altitude of 1.2 km contain subpixel targets that have fill factors as small as 10%. These are the most difficult images for subpixel detection.

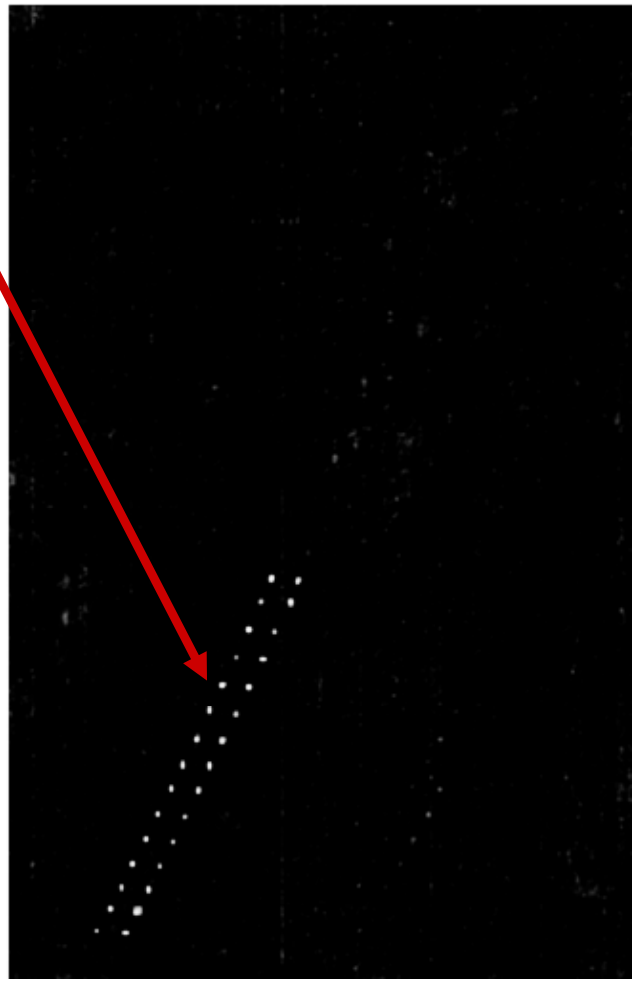


Detections

Note how the target detections (in bright white) stand out. These images are the raw detection scores without threshold. The background has been almost entirely suppressed.

WAAMD Goal

Detect targets with 50% Pd at a false alarm density of $1 \cdot 10^{-4}$ false alarms / m^2



Data Results (cont.)

Fixing the detection threshold across all seven images causes the performance to degrade slightly; however, the performance is still well above WAAMD goals.

False Alarm Density at 100% Pd Using ACE and ARRT

Image	Target 1	Target 2
1	$1.5546 \cdot 10^{-5}$	$5.1301 \cdot 10^{-4}$
2	0.0000	0.0000
3	0.0000	0.0000
4	0.0000	0.0000
5	0.0000	0.0000
6	0.0000	0.0000
7	0.0000	0.0000
Total	**$1.5546 \cdot 10^{-5}$**	**$5.1301 \cdot 10^{-4}$**

Note that Image 1 causes all of the issues when using a single threshold. Images 6 & 7 require the threshold to be low which in turn causes the false alarms in Image 1.

Comparison of Parametric Detectors

– Standard Detectors

– Whitened Detectors

Data Used in Our Analyses

To evaluate the different detectors in our study, we used WAAMD imagery collected at 4000 ft that contains some of the most difficult subpixel targets to detect. The Target fill factors vary from 10% to 50%.

Data Facts:

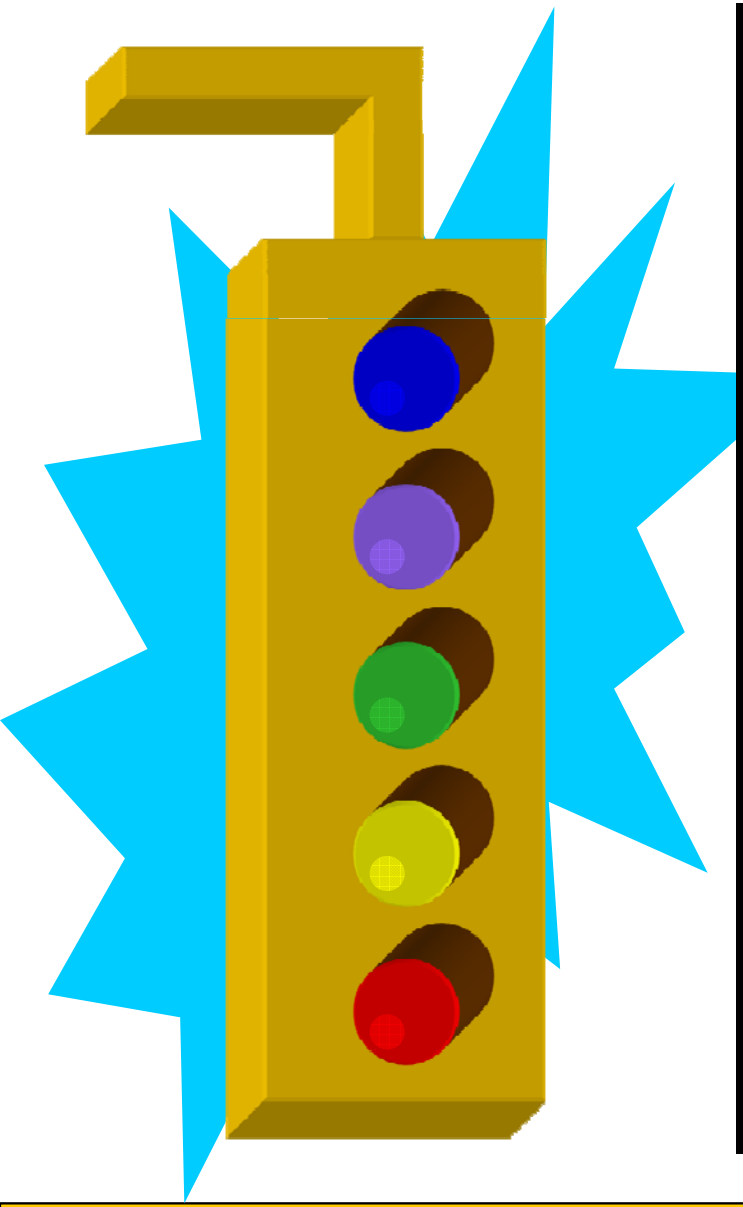
- Targets Used
  - Target 1: VS16 - Round, white plastic, 15.24 cm.
  - Target 2: M19 - Round, green metal, 30.48 cm.
  - Target 3: M20 - Square, green metal, 30.48 cm.
- Six images were used as listed below

Image	Description	Alt (m)	GSD (m^2)	Area (m^2)
1	Sparse Grass	1220	0.1823	18811
2	Short Grass	1220	0.1823	18811
3	Short Grass	1220	0.1823	19464
4	Sparse Grass	1216	0.1815	18815
5	Short & Tall Grass	1215	0.1806	18542
6	Short Grass	1213	0.1806	19097

Targets by Image

Image	VS16	M19	M20	All
1	20	42	0	62
2	0	0	12	12
3	0	0	24	24
4	20	30	0	50
5	0	0	16	16
6	0	0	28	28
All	40	72	80	192

WAAMD Performance Stoplight Criteria



Color	Pd	FAR $1/m^2$	Detection Performance	Basis for Criteria
Blue	50-60%	10^{-4}	"Blue Ribbon"	Detect scatterable surface MF using density
Purple	50-60%	10^{-3}	Very Good	Typical pattern MF detection need
Green	50-60%	10^{-2}	Satisfactory	Notional limit for pattern MF detect & FBC scatterables
Yellow	30-50%	$> 10^{-2}$	Marginal	Needs improvement for MF detect
Red	< 30%	$> 10^{-2}$	Unsatisfactory	

Detectors

The Matched Filter (MF), Normalized Orthogonal Subspace Projection (NOSP), Fully Constrained Least Squares (FCLS), Adaptive Matched Subspace Detector (AMSD), and Hybrid Detectors comprise the original set of detectors we analyzed.

Derivation:

- Assume the following hypothesis:

$$H_0 : x = Ba_b + w, \quad w \sim N(0, \sigma_w^2 I)$$

$$H_1 : x = Sa_t + Ba_b + w = Za + w$$

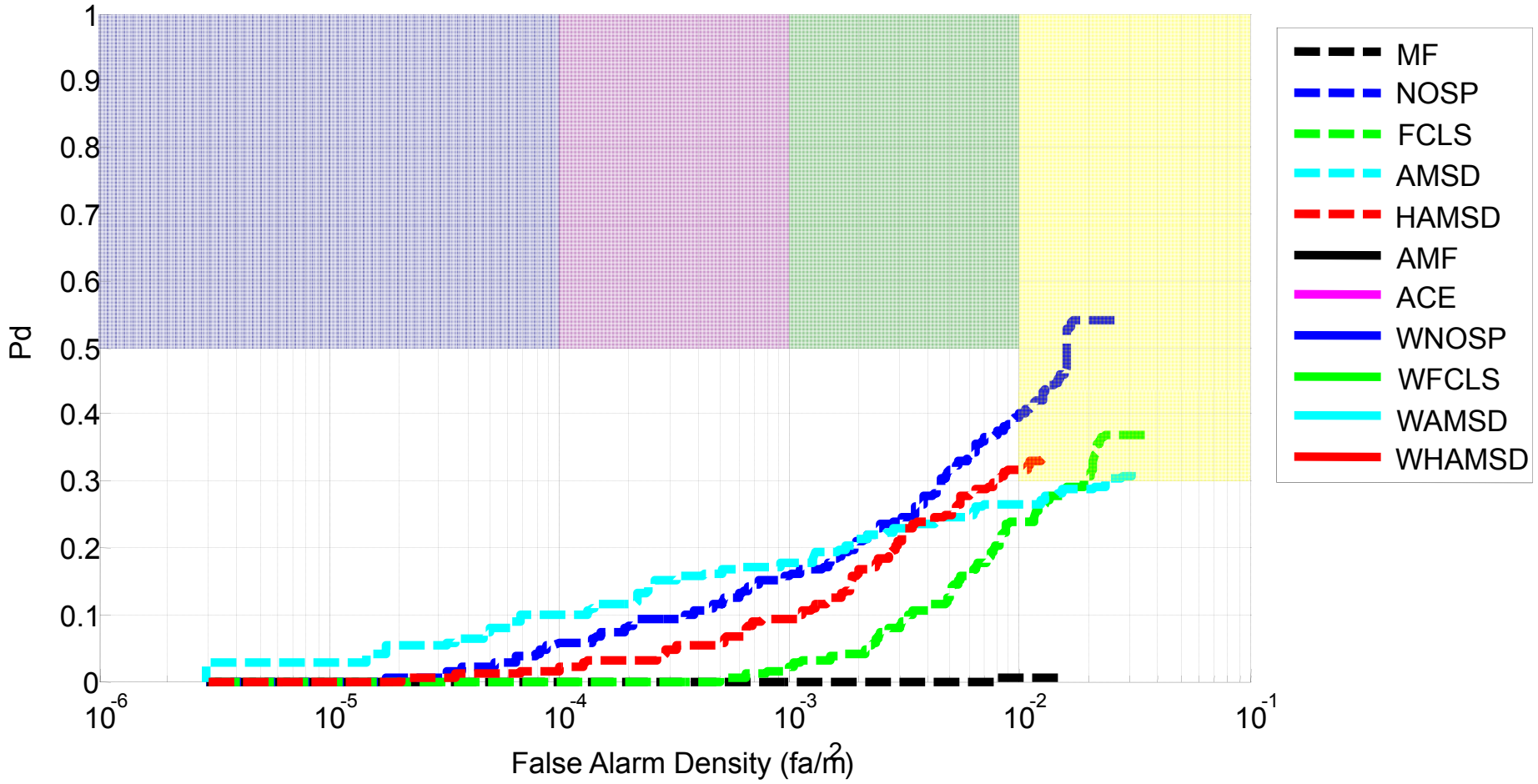
- MF is a GLRT assuming no background endmembers.
- NOSP is a Least Squares Solution that provides an unbounded abundance estimate
- AMSD is a GLRT assuming no bounds on the abundance estimates
- Hybrid is a GLRT assuming both sum-to-one and non-negativity bounds on the abundances.

Notes:

- These detectors represent some of the most well documented detectors used for hyperspectral imagery.
- The hybrid detector was our first “improvement” that incorporated known phenomenological effects into the detector structure – namely the sum-to-one and non-negativity constraints.

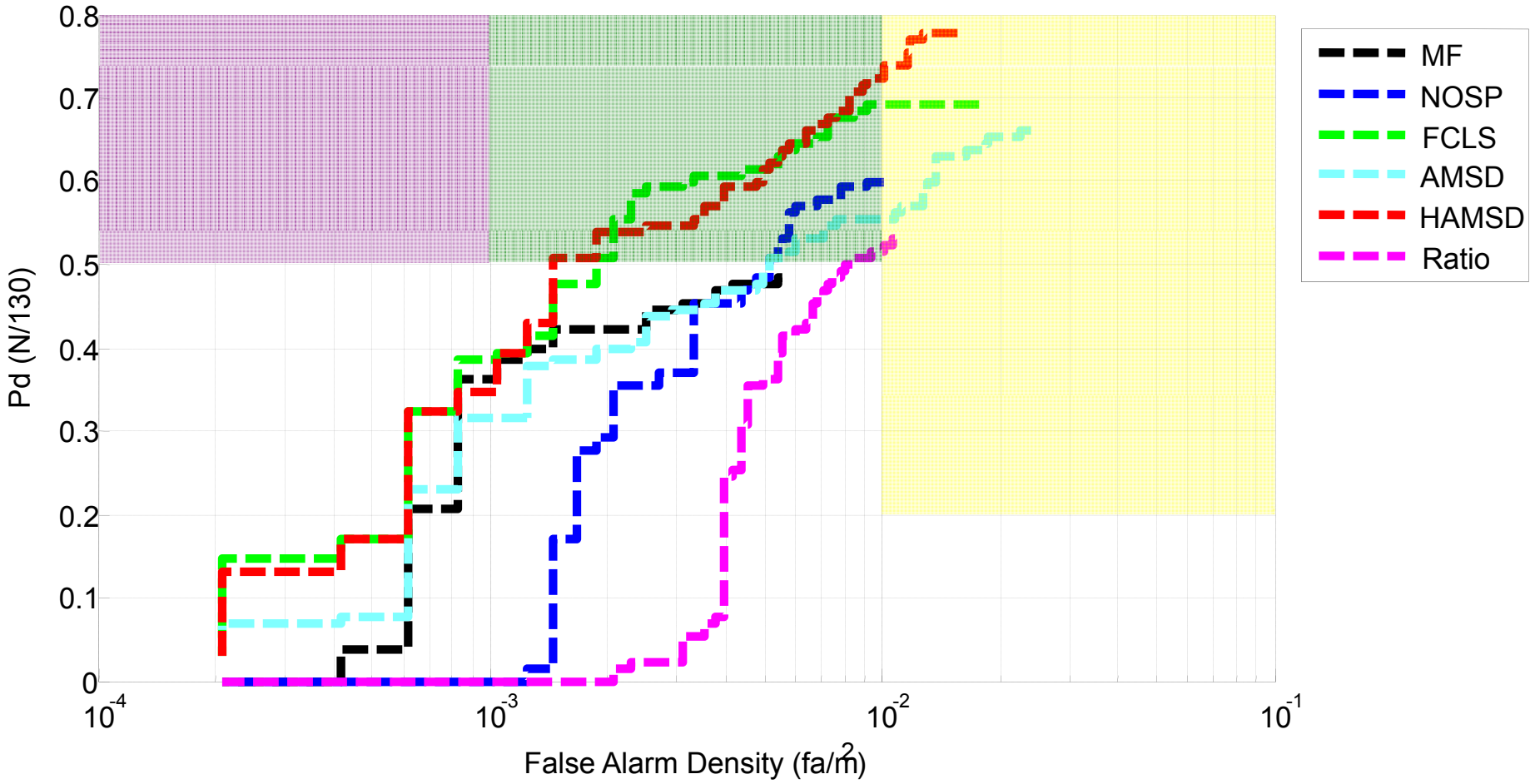
Initial Detector Results

Given the difficulty of these images, our initial detectors did not perform well. However, nearly all detectors were able to meet the "marginal" criteria established by WAAMD. Only the matched filter did not perform to WAAMD standards.



Initial Detector Results (LWIR Data)

The same detectors were applied to the LWIR data from Yuma using an average of 10 target signatures taken from the image. All detectors were able to meet the “satisfactory” criteria established by WAAMD.



Adaptive Cosine/Coherent Estimate

The Adaptive Cosine/Coherent Estimate (ACE) is considered to be one of the best performing detectors in the literature. It is mathematically equivalent to the RX detector when a target signature is not available.

Derivation:

- Assume the following hypothesis:

$$H_0 : x = b$$

$$H_1 : x = Sa_t + b$$

$$b \sim N(0, \sigma^2 \Sigma)$$

- Using maximum likelihood estimates in a generalized multivariate analysis of variance (GMANOVA), we arrive at the following detection statistic

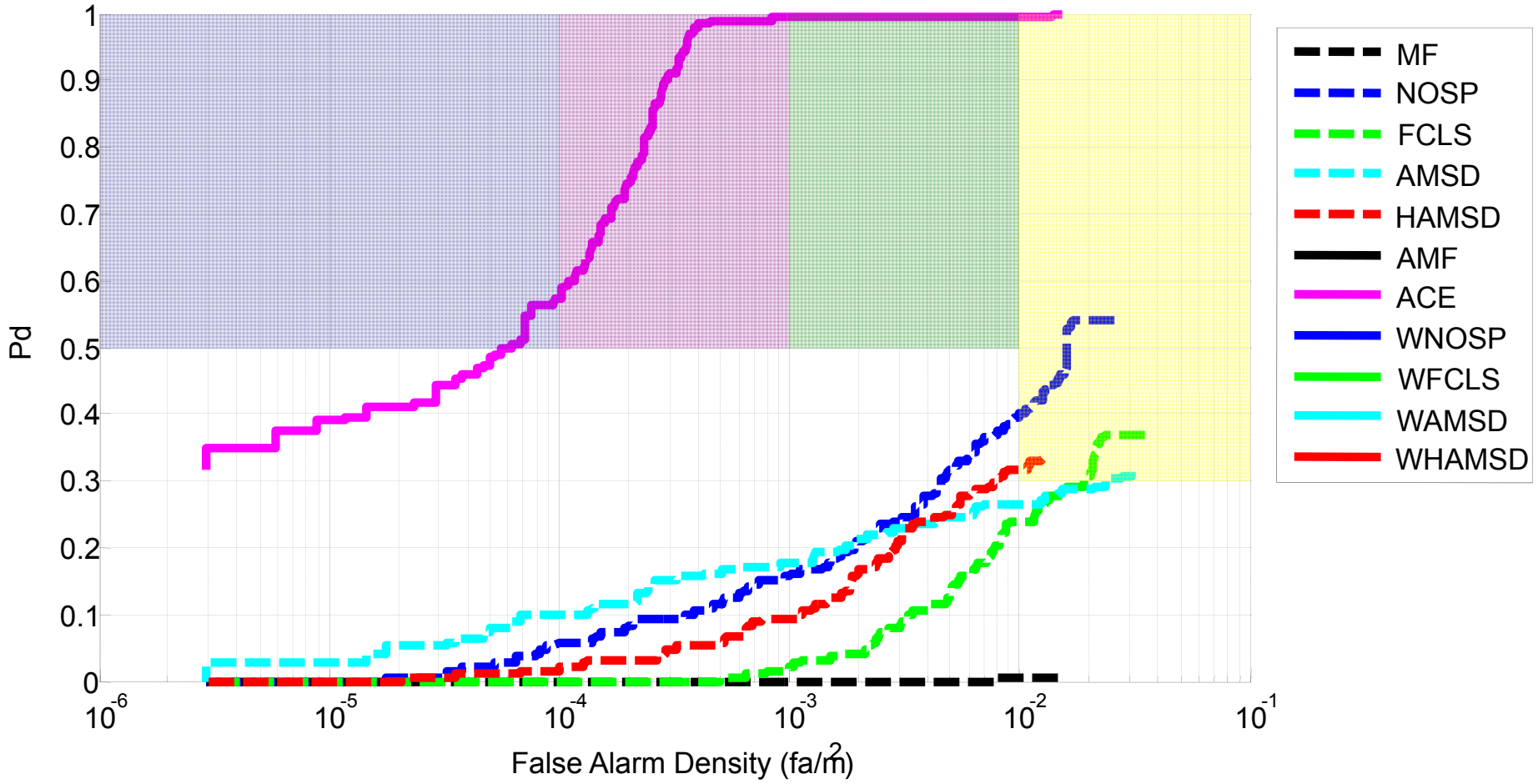
$$D(x) = \frac{x^T \Sigma^{-1} T (T^T \Sigma^{-1} T) T^T \Sigma^{-1} x}{x^T \Sigma^{-1} x}$$

Notes:

- This detector uses an unstructured background We chose to go this way to minimize any processing artifacts due to endmember estimation.
- The detector's parameters can either be locally or globally estimated. In this case, we used global estimation since local did not improve performance. This also gives us the benefits of real-time processing.

ACE Results

The ACE detector provides significantly improved detection performance over the other detectors mentioned so far. The performance is good enough to be considered “blue ribbon” by the WAAMD program.



Whitened Detectors

The major difference between the ACE detector and other detectors is the use of a full covariance matrix. Therefore, we updated all of the other detectors assuming a full covariance background model.

Derivation:

- Assume the following hypothesis:

$$H_0 : x = Ba_b + w, \quad w \sim N(0, \sigma_w^2 \Gamma)$$

$$H_1 : x = Sa_t + Ba_b + w = Za + w$$

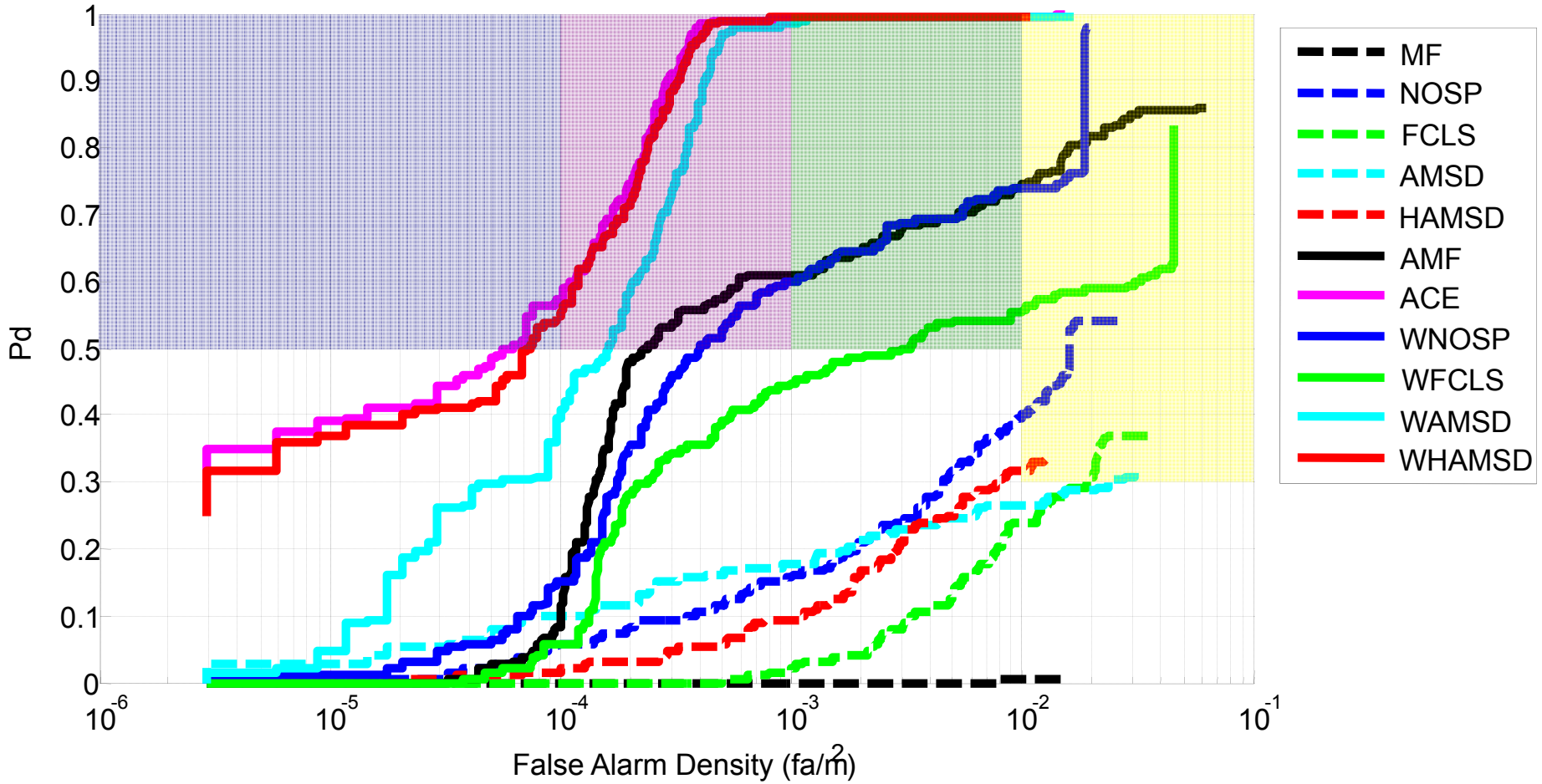
- MF is a GLRT assuming no background endmembers.
- NOSP is a Least Squares Solution that provides an unbounded abundance estimate
- AMSD is a GLRT assuming no bounds on the abundance estimates
- Hybrid is a GLRT assuming both sum-to-one and non-negativity bounds on the abundances.

Notes:

- These detectors are derived the same way as before except with a full covariance matrix Γ.
- The covariance matrix maps all of the detectors into a whitened space where the spectral bands are uncorrelated and have equal weight in the detector statistic.
- We call these “whitened detectors” because they are equivalent to the standard detectors on whitened data.

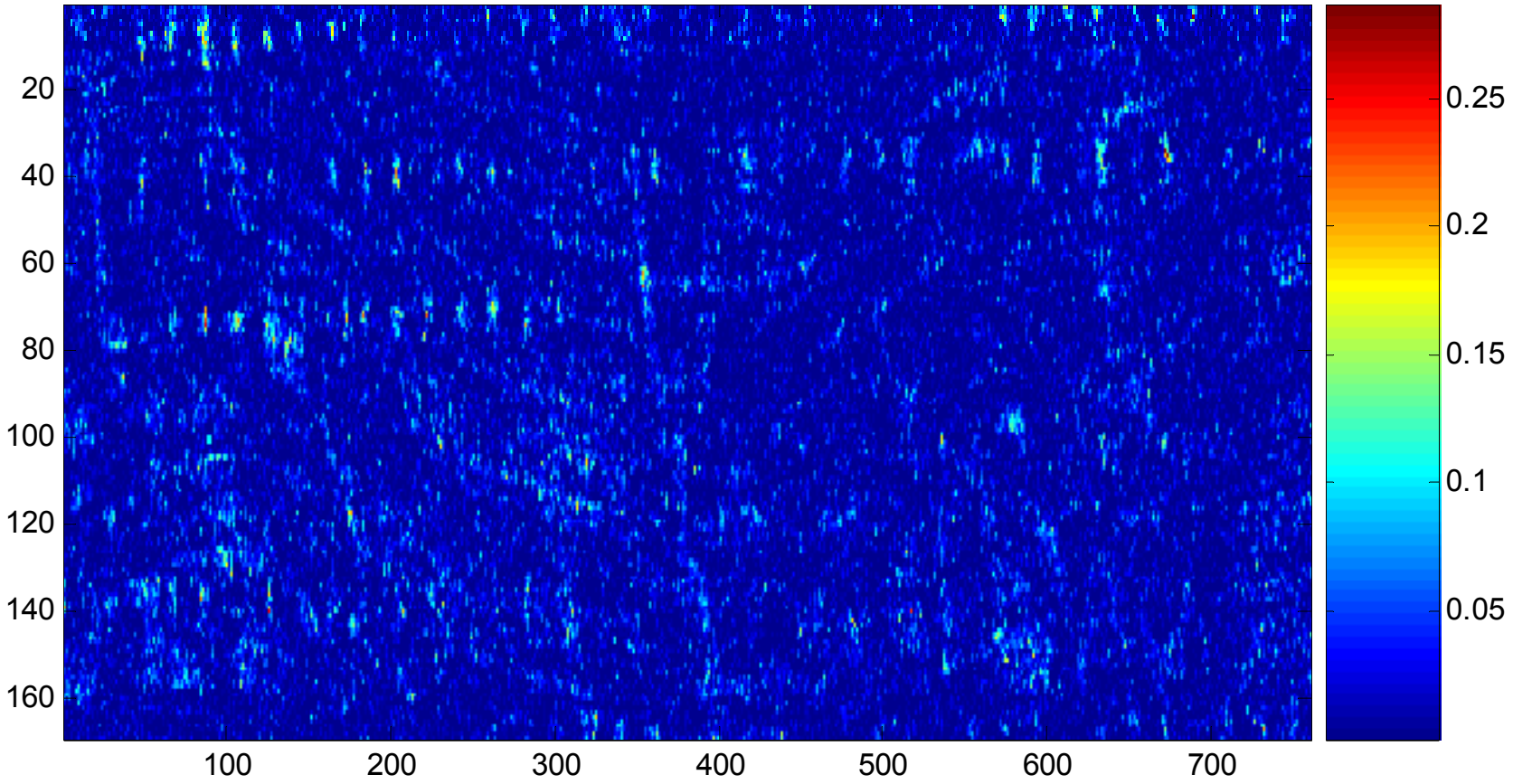
Whitened Detector Results

By using the covariance matrix, the whitened detectors all show a significant improvement in performance with the worst being considered “very good.” The whitened hybrid detector performs as well as ACE.



To Whiten or Not To Whiten ...

What happened? In the LWIR bands, the use of a covariance matrix actually degrades performance due to low SINR values. Therefore, whitened detectors on LWIR imagery tend to underperform.

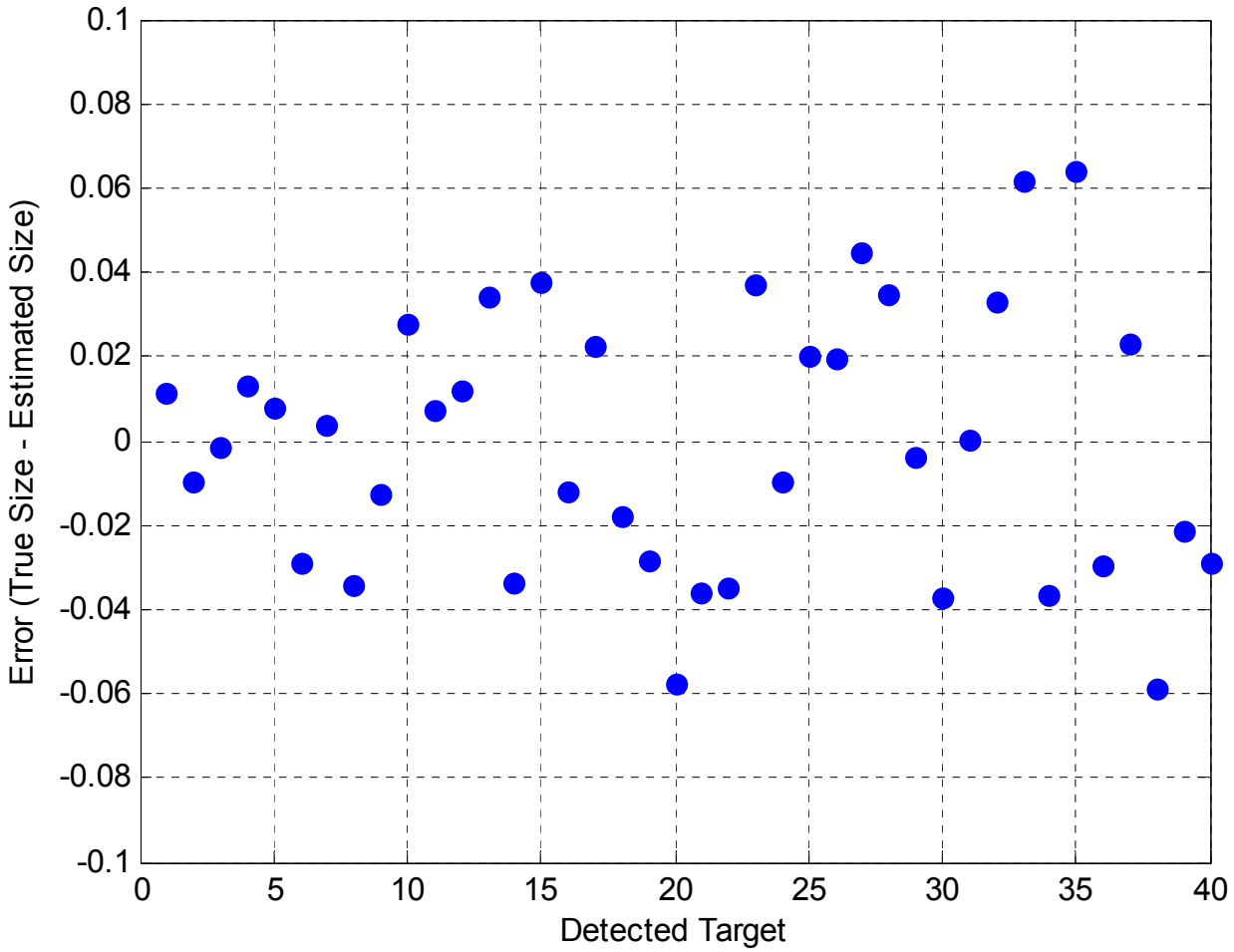


Size Estimation

One of the side benefits of the whitened hybrid detector is the ability to provide subpixel size estimates. This information could be used to provide a spatial filter to help remove false alarms in future implementations.

Notes:

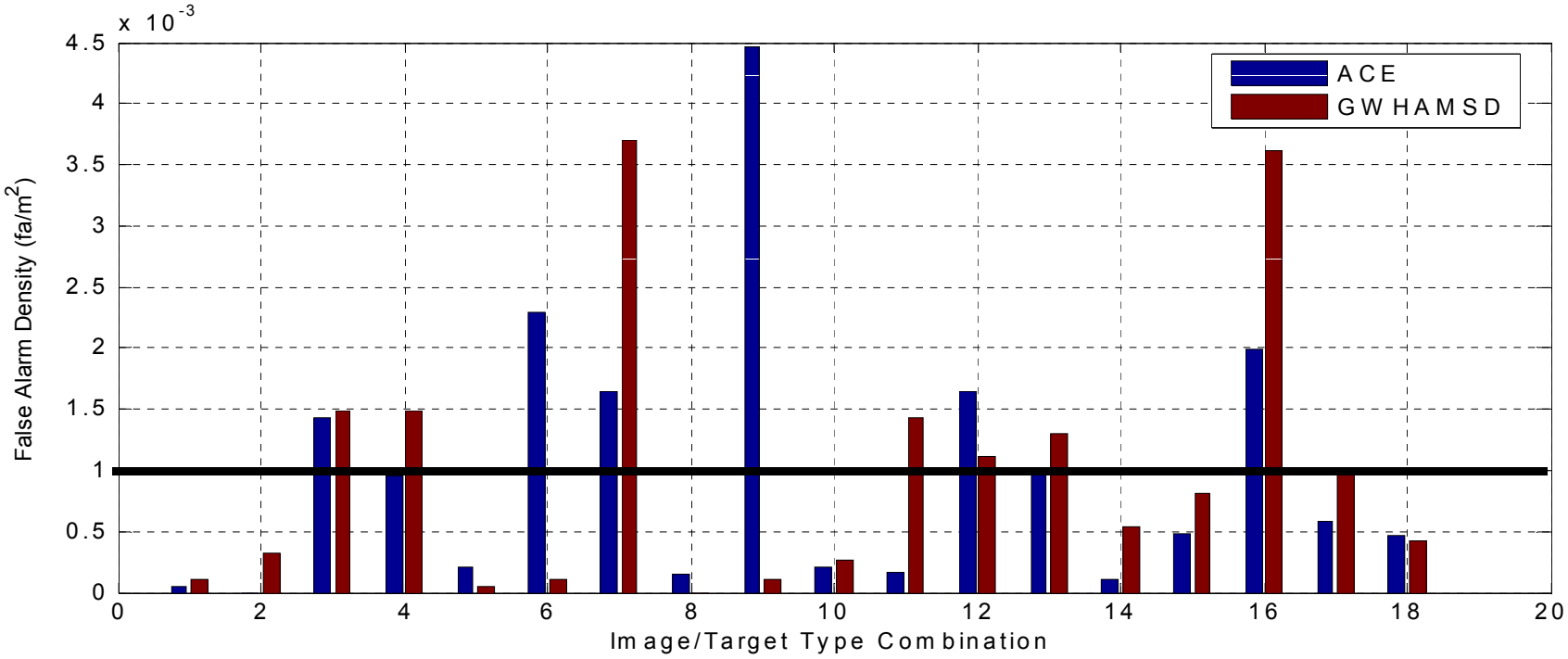
- The size estimates are calculated by summing the abundance estimates for a given detection.
- The graph shows the error between the VS16 fill factors and the estimated target abundances from the whitened hybrid detector.
- Statistics:
  - Bias = 0
  - Std = 0.0313



Comparison of WHAMSD to ACE:

Empirical demonstration of HAMSD's CFAR-like properties

To identify whether the whitened hybrid algorithm has CFAR-like properties, a comparison was done between the known CFAR detector ACE and WHAMSD to compare their performance at a fixed threshold across multiple images.



The results show that the WHAMSD and ACE algorithms have nearly identical performance with a difference in mean of 6.7x10^-6 and identical variances. Therefore, the WHAMSD has similar CFAR-properties to ACE.

Additional Topics

- Non-Parametric Detector Investigation
- Adaptive Thresholding
- Phenomenologically-Constrained Kernels
- Spectral Object Level Change Detection

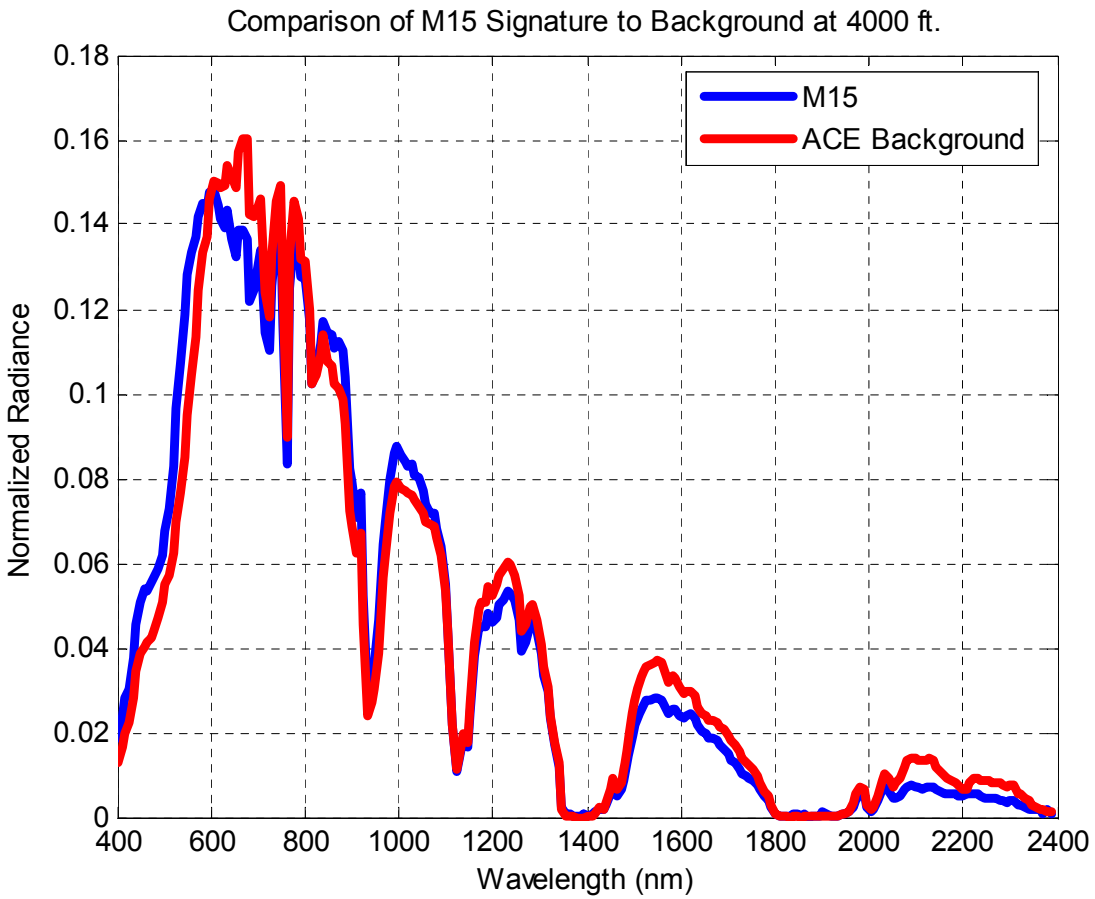
The Need to Continue

The results so far have been excellent; however, all of these results are on targets with high Signal to Interference and Noise Ratios (SINR). As SINR decreases, the parametric techniques degrade as in the case of buried targets in LWIR and targets such as M15s.

- The example shown is a comparison of the M15 signature to the background signature implicit in ACE taken from the 4000 ft. COMPASS imagery.
- Mathematically, the similarities can be expressed as $SINR = aS^T \Sigma^{-1} Sa$

$$= 4.37dB$$

- Previous SINR was over 30 dB.



SVM Approach

Problem: **Given a two-class classification problem, identify a linear boundary that separates the data such that performance on both training and testing data is good.**

- Identify the data that needs to be classified using a training set

$$T = \{\mathbf{x}_i, y_i | i = 1, \ldots, N\}$$

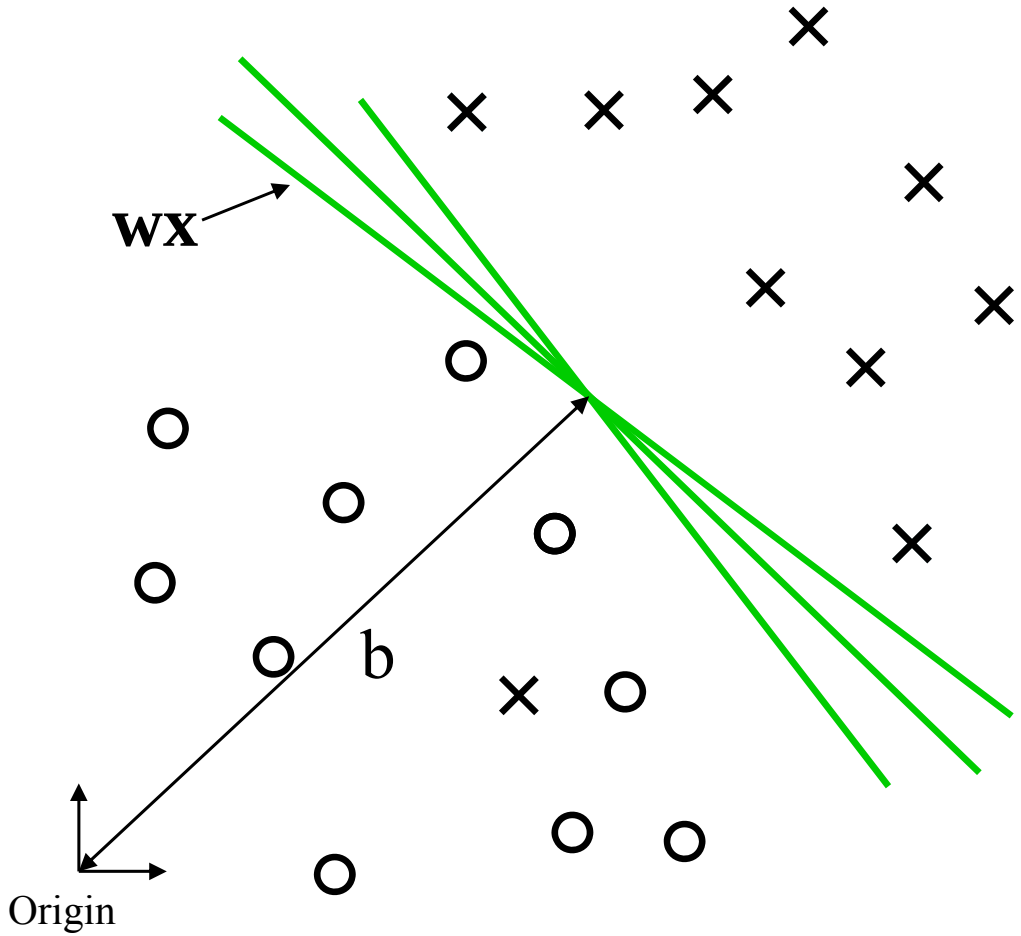
$$\mathbf{x}_i \in R^n$$

$$y_i \in \{-1, 1\}$$

- Want to find a linear boundary such that

$$\mathbf{w}\mathbf{x} + b = 0$$

- The question is which line do we use?



SVM Approach (Continued)

Solution: **Recall that we are trying to find a boundary that works for both training and testing data; therefore, we want to define a linear boundary that maximizes the distance between the two classes.**

- We can model the training set such that

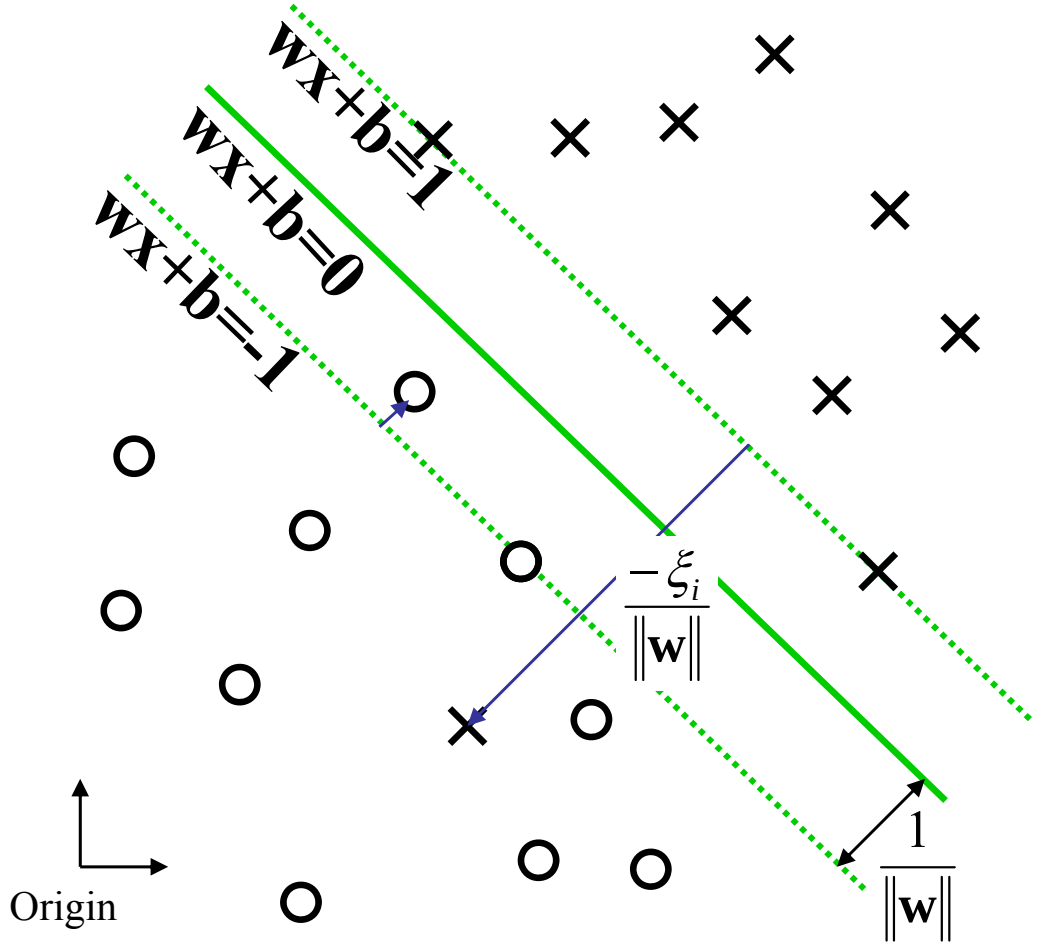
$$\mathbf{wx}_i + b \geq 1 - \xi_i \quad \forall y_i = 1$$

$$\mathbf{wx}_i + b \leq -1 + \xi_i \quad \forall y_i = -1$$

- This model can be rewritten so

$$y_i(\mathbf{wx}_i + b) - 1 + \xi_i \geq 0 \quad \forall i$$

- Therefore, we want to maximize the distance between the two classes by minimizing ||w||.



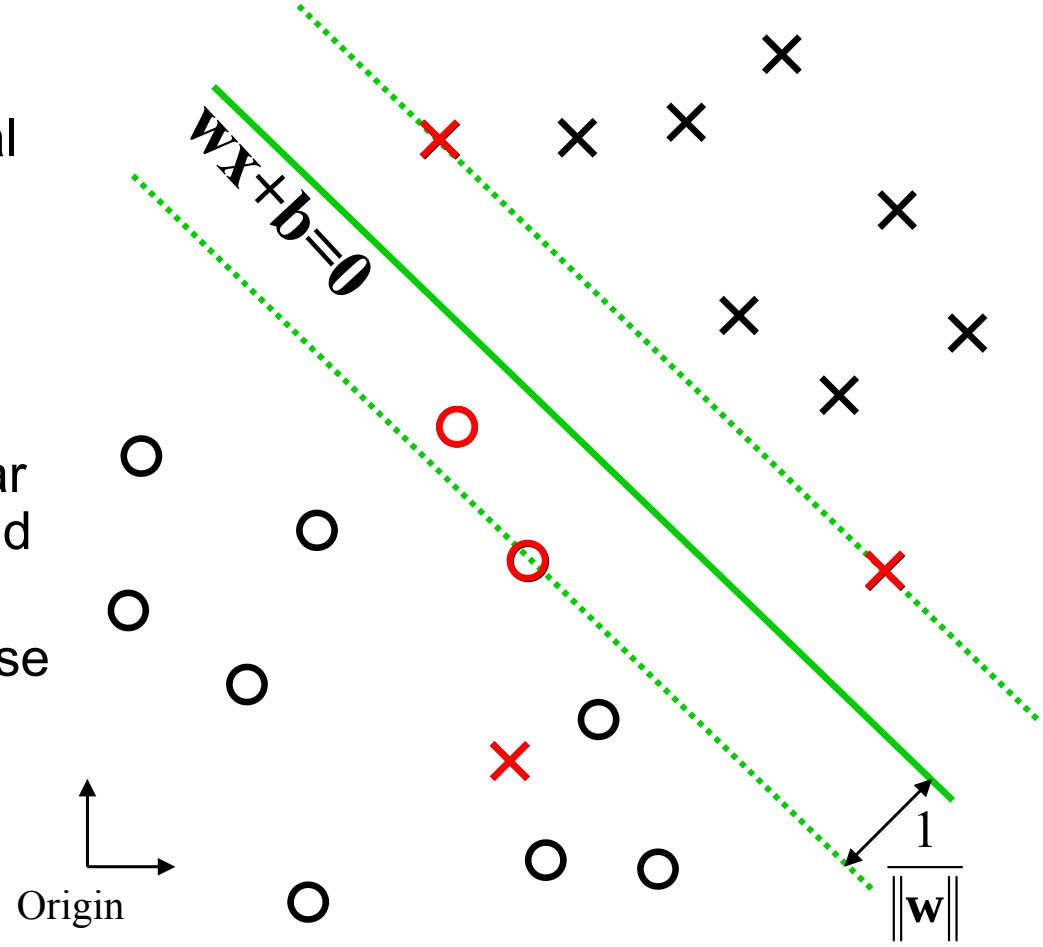
SVM Solution

The solution is a reduced set of training samples that are used in a linear combination to classify new points.

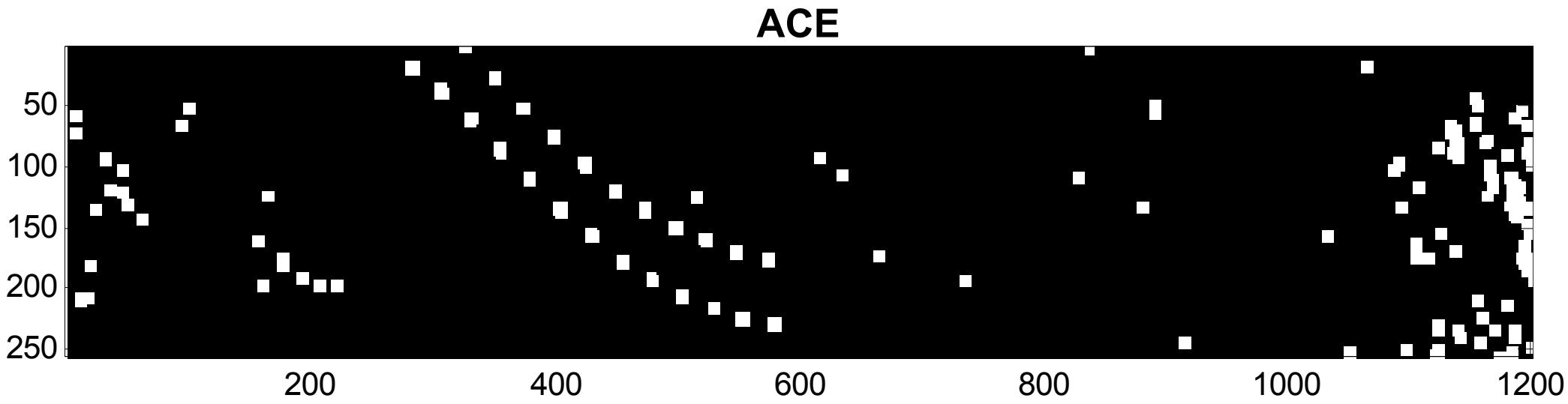
- Using quadratic programming techniques, the Lagrange multiplier values can be found to create our final solution such that

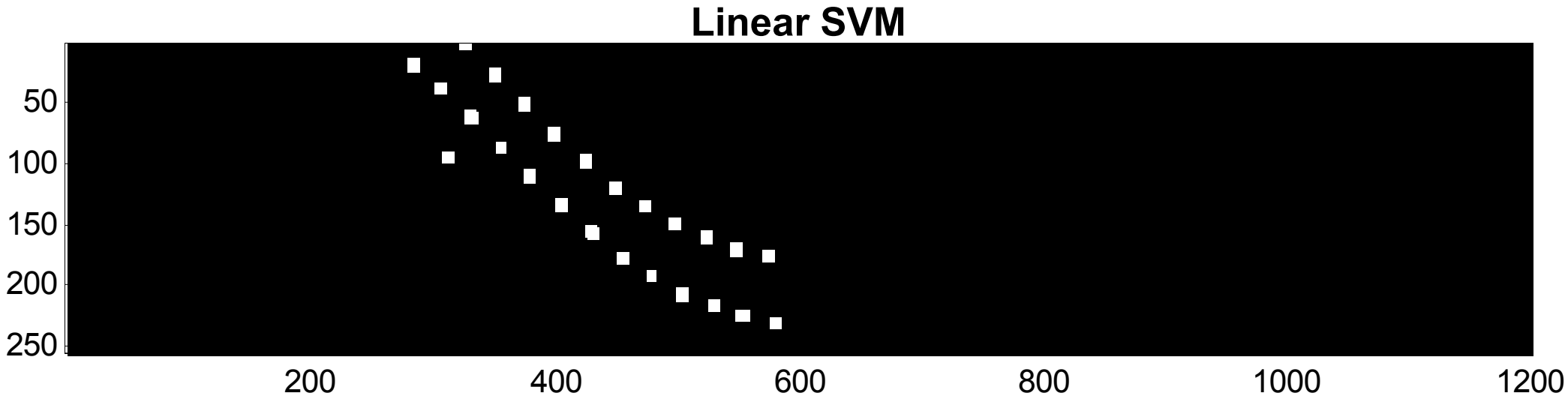
$$y = sign\left(b + \sum_{i \in S} \alpha_i y_i \mathbf{x}_i^T \mathbf{x} \right)$$

- Note that the solution is simply a linear combination of the support vectors and the point to be classified.
- The "support vectors" in set S are those training points that
  - Define the boundary
  - Margin errors
  - Class errors

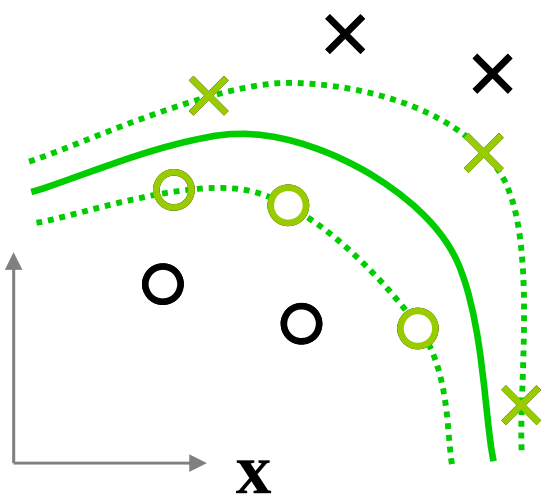


Comparison of ACE and Linear SVM



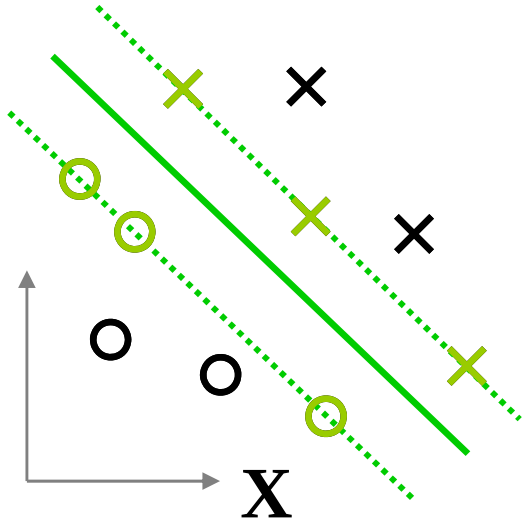


Use of Kernels for Nonlinear Classification



$$\Phi(\cdot)$$

$$\mathbf{X}_i = \Phi(\mathbf{x}_i)$$



$$\mathbf{X}_i \cdot \mathbf{X} = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x})$$

$$y = \text{sign}\left(\sum_{i \in S} \alpha_i y_i \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}) + b \right)$$

$$y = \text{sign}\left(\sum_{i \in S} \alpha_i y_i \mathbf{X}_i \cdot \mathbf{X} + b \right)$$

$$K(\cdot,\cdot)$$

$$\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}) = K(\mathbf{x}_i, \mathbf{x})$$

Mercer's Condition

$$y = \text{sign}\left(\sum_{i \in S} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b \right)$$

Any classifier that uses a dot product can then use these kernels to create a nonlinear boundary!

Physics-Based Kernels

"Specifically, the kernel ***is*** the prior knowledge we have about a problem and its solution." [Scholkopf, Smola 2002]

- Typically, a standard set of kernels is used when creating non-linear detectors/classifiers
  - There are a standard set of kernels used in the literature such as the RBF kernel which have good general properties across many applications
  - However, these kernels do not assume any information about the current application.
- Noting the quote above though, kernels are a direct way of incorporation a-priori knowledge of an application domain into the non-linear solution
  - We know certain properties of hyperspectral imagery that we can use to design kernels such as the reststrahlen effect in LWIR imagery.
  - We can design kernels as either mathematical combinations of the standard kernels that stress certain known spectral bands, or
  - We can design adaptive kernels such as R-Convolutional kernels that incorporate a natural feature selection within the training process.
- Currently, work is focusing on adaptive spline kernels and ANOVA kernels which are combinations of kernels for different spectral bands (VIS, NIR, SWIR).

Adaptive Threshold Selection

Select appropriate threshold to provide a fixed false alarm density even when the image has too few pixels to adequately model the tail distribution.

- Importance Sampling (IS) is a technique to simulate rare events from some distribution f().
  - Classical IS assumes some distribution f() and then applies a weighting function s.t.

$$\hat{p}_t = \frac{1}{K}\sum_{i=1}^{K} 1(X_i \geq t) W(X_i), \quad X_i \sim f_*$$

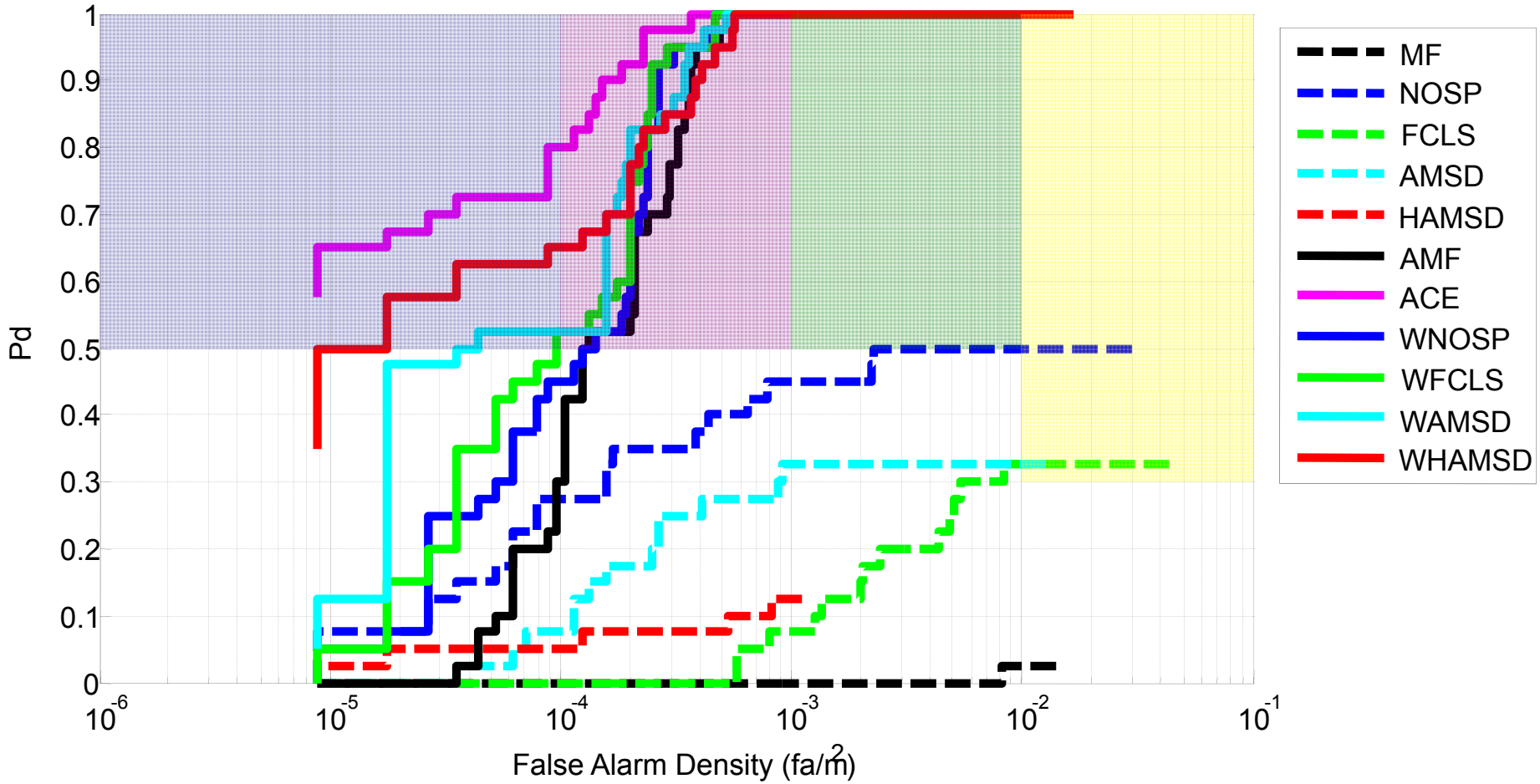
    where X are the samples, p is our estimated tail probability, W is the weighting function, K is number of samples, and f_* is the biasing density.
- In our application, we do not know f(); so, we resort to blind importance sampling based on CFAR detectors. Thus, we can theoretically identify the appropriate threshold from the data directly without needing a large number of samples.
- This last point is key because in our images, there are not enough samples to accurately estimate false alarm densities of 1*10^-4 fa/m2.

Theoretical results in [Srinivasan 2002] also note that this method can be used to adaptively select thresholds not only across images, but possibly within an image depending on how many samples are needed to estimate tail probabilities.

ROC Results

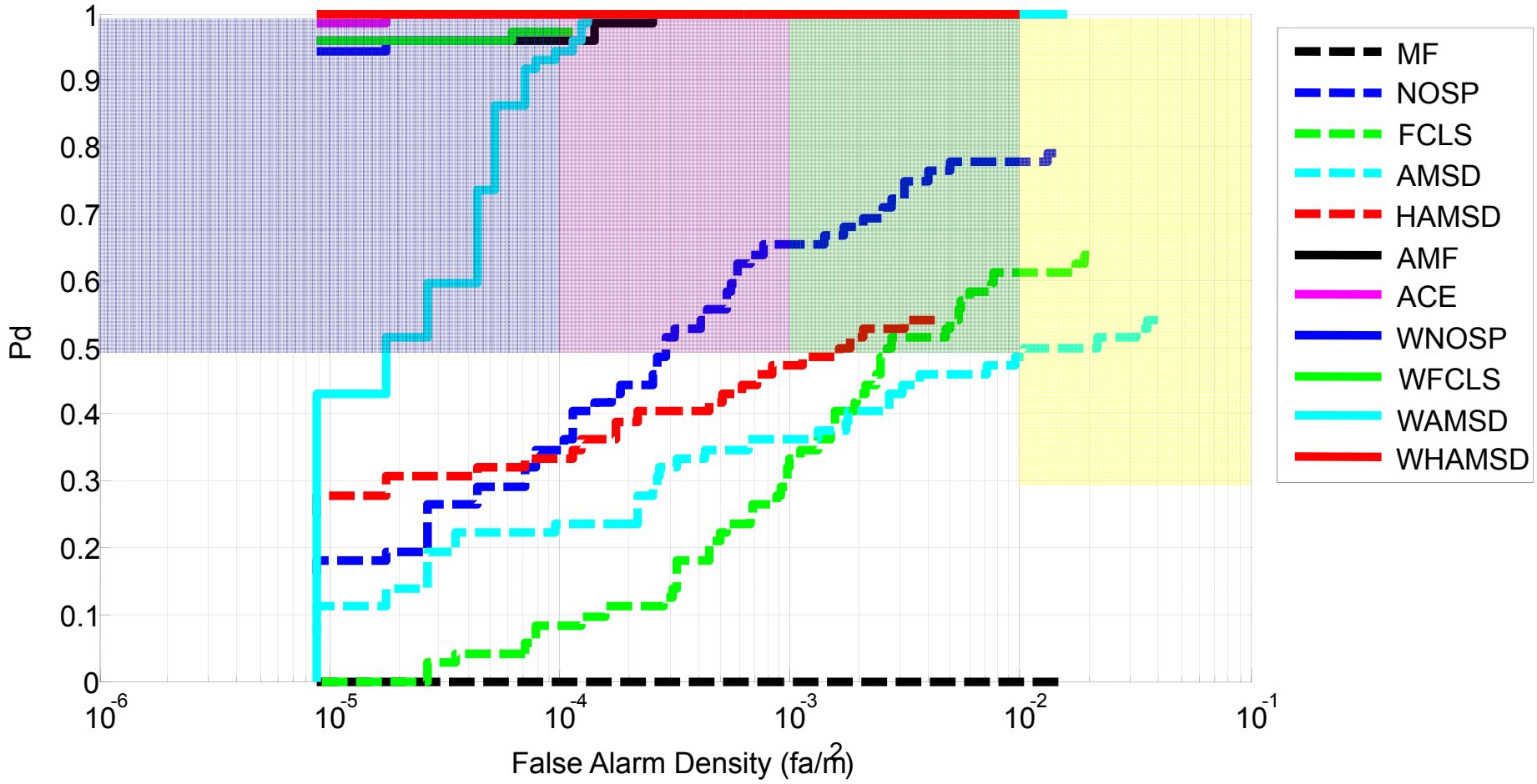
VS16 ROC Results

VS16 mines have the smallest fill factors of 10%. Of the standard detectors, NOSP is the only one that meets WAAMD criteria. All of the whitened detectors perform either very well or "blue ribbon." Most of the false alarms are caused by other mines.



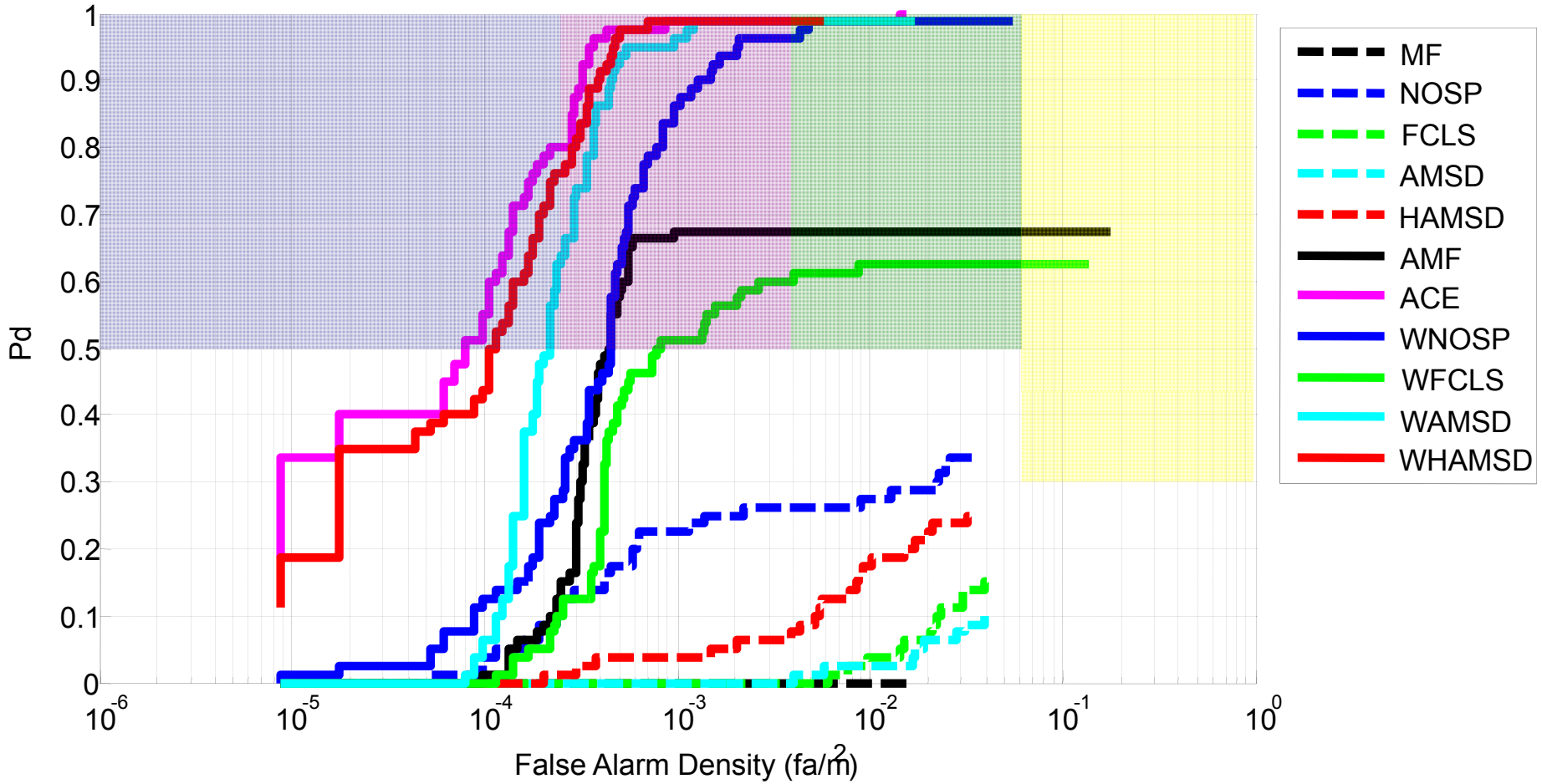
M19 ROC Results

M19 mines have some of the largest fill factors and hence some of the best detection results. All detectors except the matched filter meet WAAMD criteria with the whitened hybrid detector providing perfect separation between targets and clutter.



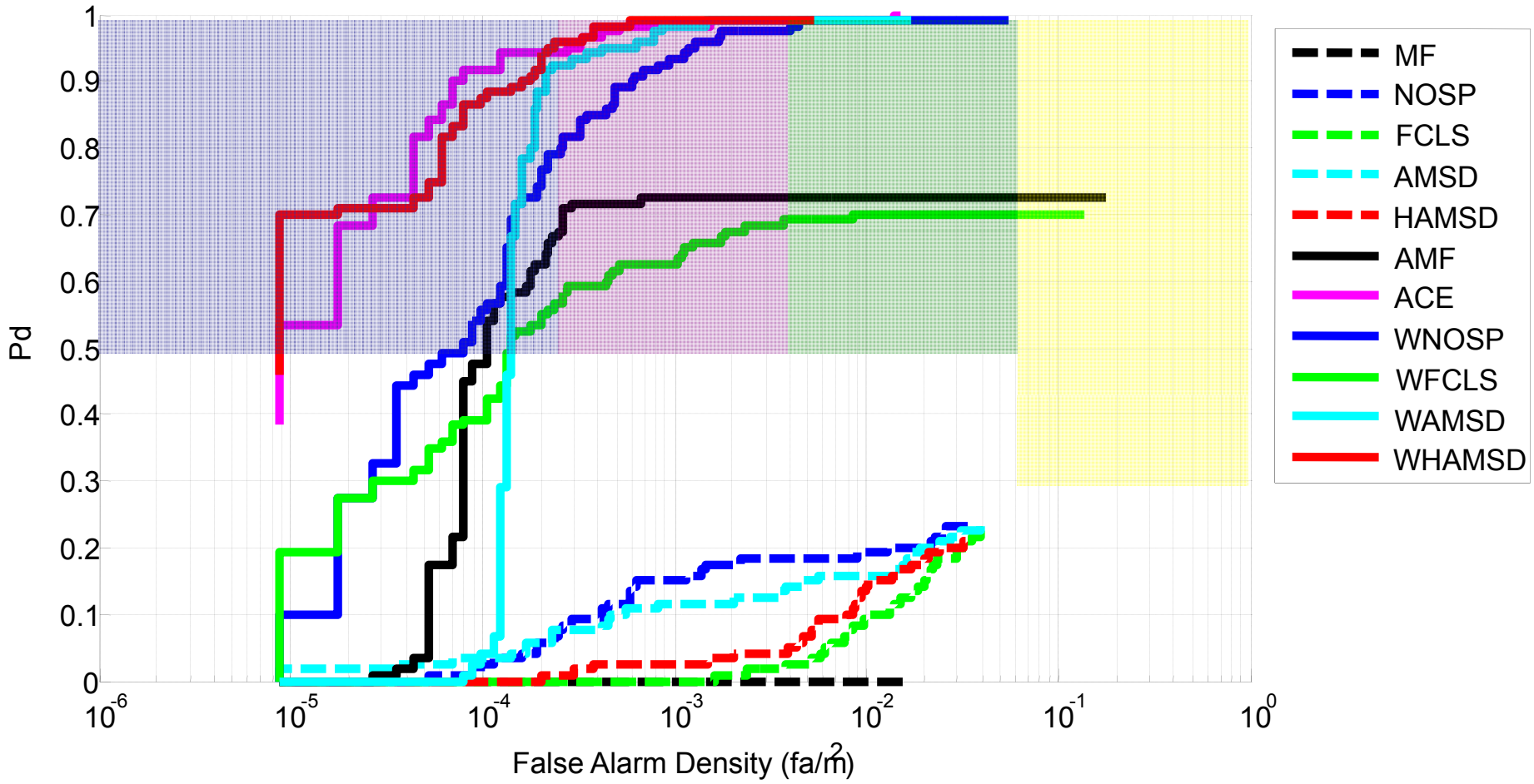
M20 ROC Results

The M20 mines have large fill factors as well, but are spectrally similar to the VS16 mines. This causes an increase in false alarms for the whitened detectors and makes nearly all the standard detectors under perform.



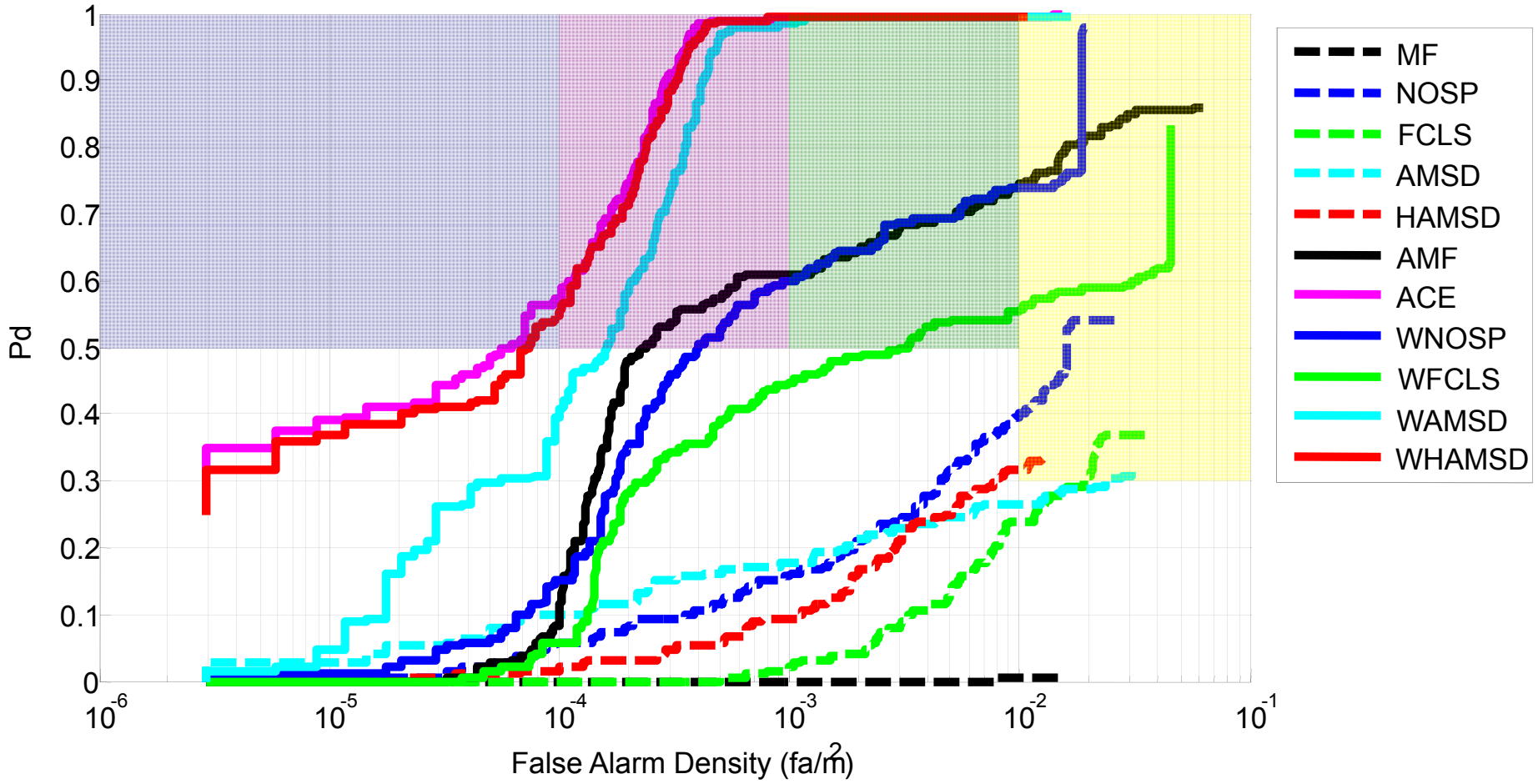
VS16 & M20 Combined Results

To show the spectral similarity between the VS16 and M20 mines, the M20 signature is used to detect both mine types. The results show that all the whitened detectors become "blue ribbon" as false alarms are reduced.



Combined ROC Results

When using all target signatures, only the NOSP detector performs within WAAMD criteria. The whitened detectors all meet WAAMD criteria with the ACE and whitened hybrid considered “blue ribbon.”



Summary

Work developed so far has provided detectors that meet “blue ribbon” performance on WAAMD data. This represents some of the best results on this data set, but more work remains to test it against multiple environments and other mine types.

ARRT is a simple algorithm based on IARR that uses the spectral mean to generate a target radiance signature given an HSI data cube and desired target reflectance signature. Results show that this algorithm is both quick and useful in providing signatures for subpixel target detection algorithms. This work meets our goal of providing phenomenological methods to improve detection performance.

The whitened detectors show that the main phenomenological constraint is the need to estimate a covariance matrix to decorrelate the spectral bands and give the bands equal weight in the detectors. The whitened hybrid detector performs as well as the ACE detector but does not surpass it; however, the detector does provide accurate and precise size measurements which could lead to improved detector performance.

2005 Research Accomplishments

- Adapted an algorithm to map target reflectance signatures to radiance using only in-scene information.
- Presented and published with A. Banerjee, “A Hybrid Method for Automatic Detection of Sub-pixel Targets,” in *Proceedings of the MSS CC&D Conference*, SPAWAR Charleston, SC, 17 February 2005, NOFORN/ITAR.
- Presented and published, “Average Relative Radiance Transform for Subpixel Detection,” in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium 2005*, Seoul, South Korea, July 2005.
- Processed 4000 ft COMPASS data from Forest Fusion 1 and various LWIR data from Yuma and Ft. Leonardwood.
- Delivered various detection results to University of Florida in support of their fusion work.
- Developed Nonparametric Detector based on Support Vector Machines

Spectral Automatic Target Detection/Recognition (ATD/R)

2006 Research Efforts

This section contains ITAR information

2006 Research Topic Outline

- Estimation of Target Radiance Signatures
- Hybrid Detectors
- Adaptive Threshold Method using Importance Sampling
- Estimation of correct number of endmembers
- Estimation of background using kernel methods
- Joint Spatial/Spectral Processing
- Spectral Object Level Change Detection

Estimation of Target Radiance Signatures

Atmospheric Effects

The radiance value seen at the sensor can be modeled by the equation below. This equation shows how reflectance signatures are related to radiance values through transmittance, scattering, and absorption effects.

$$L(x,y,\lambda) = T_u(z_g, z_u, \vartheta_v, \phi_v, \lambda) R(x,y,\lambda)[KT_d(z_g, \vartheta_0, \phi_0, \lambda) E_0(\lambda) \cos\vartheta_0 +$$

$$\int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi/2} E_s(\theta,\phi,\lambda) \cos\vartheta \sin\theta d\theta d\phi] + P(z_g, z_u, \vartheta_v, \phi_v, \lambda). \quad [1]$$

$$\hat{L} = RA + RB + P$$

Sun Light:
Path from sun to target to sensor

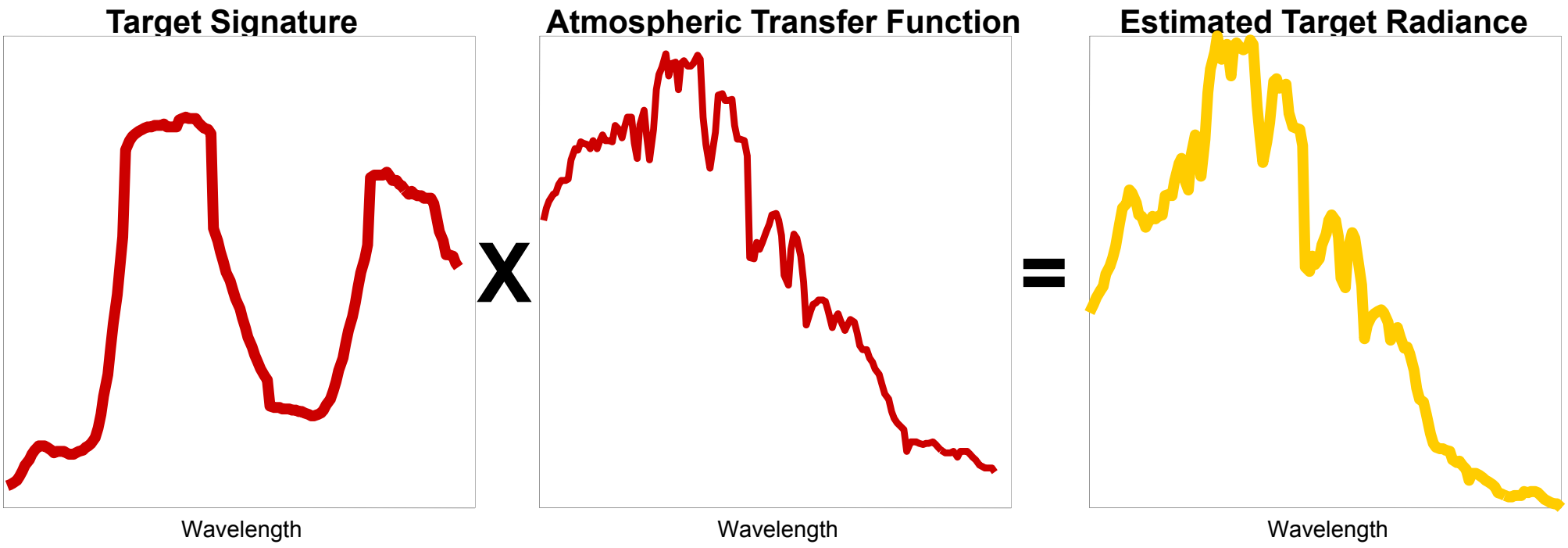
Sky Light:
Light from Scattering to target to sensor

Upwelled Radiance:
Path from sun to atmosphere to sensor

The equation is broken down into its three respective parts: sun light, sky light, and upwelled radiance. The strongest of these is the sun light followed closely by the sky light effects.

Improved ARRT

This algorithm finds estimates for the combined sun and sky shine components using a difference equation to find spectrally flat radiance signatures in the image. It then multiplies the average of these signatures by the desired target signature.

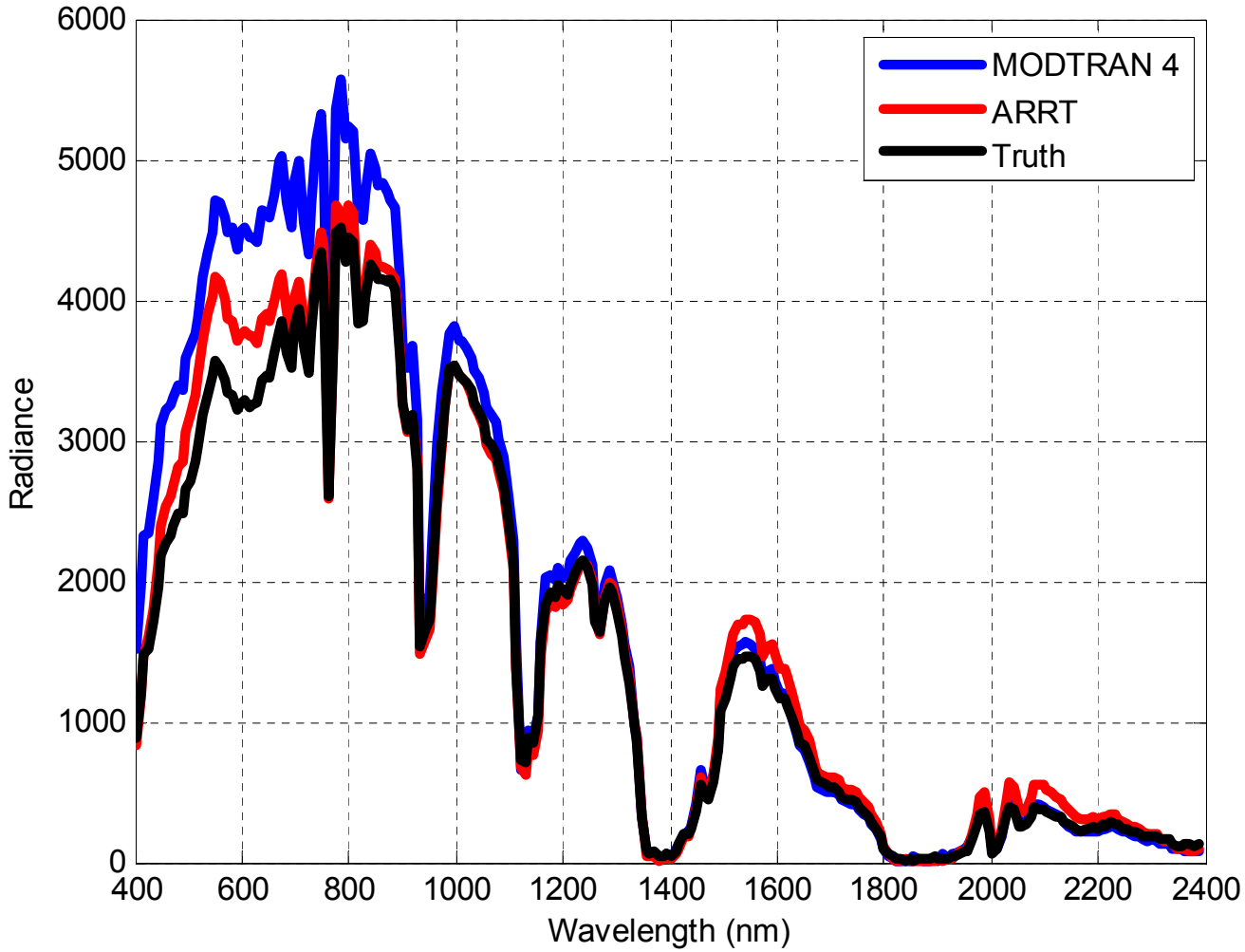


This variant seems to work in all of the WAAMD images. However, we know this method will fail or have degraded performance if flat signatures are not available or if the camera is at a high altitude (upwelled radiance not modeled).

Target Comparisons

ARRT, MODTRAN4, Reality

The improved ARRT algorithm can provide a better match to the true target signature in the image than model-based methods.



Hybrid Detectors

Subpixel Detection

Hyperspectral imagery provides the means to detect targets smaller than the size of a pixel using spectral unmixing techniques.

Unstructured Detectors:

- Assume the following hypothesis:

$$H_0 : x = w, \quad w \sim N(0, \sigma_w^2 \Gamma)$$

$$H_1 : x = Sa_t + w$$

- Note that this method models the background as a statistical distribution (e.g. Adaptive Coherent Estimate – ACE)

Structured Detectors:

- Assume the following hypothesis:

$$H_0 : x = Ba_b + w, \quad w \sim N(0, \sigma_w^2 I)$$

$$H_1 : x = Sa_t + Ba_b + w = Za + w$$

- This method uses the linear mixing model, but does not typically use any ancillary physical information (e.g. Adaptive Matched Subspace Detector – AMSD)

We developed a hybrid detector that takes a semi-structured approach to subpixel detection while including the known physical constraints of the linear mixing model

.

Hybrid Structured Detector

- Idea: Create a semi-structured detector that utilizes all physical constraints of the linear mixing model
- Hypothesis: $H_0 : x = Ba_b + w, \quad w \sim N(0, \sigma_w^2 \Gamma)$

  $H_1 : x = Sa_t + Ba_b + w = Za + w$
- Models the background as a linear mixing model and statistical distribution to account for unknown sensor effects
  - Background endmembers model the bulk of the background
  - A covariance estimate is used to model sensor effects and other unknown effects in the data
  - The fully constrained least squares (FCLS) estimate is used to estimate the abundances while maintaining their physical constraints

$$D(x) = \frac{(x - B\hat{a}_b)^T \Gamma^{-1} (x - B\hat{a}_b)}{(x - Z\hat{a})^T \Gamma^{-1} (x - Z\hat{a})}$$

Hybrid Unstructured Detector

- Idea: Create an unstructured detector that utilizes all physical constraints of the linear mixing model
- Hypothesis: $H_0 : x = w, \quad w \sim N(0, \sigma_w^2 \Gamma)$

  $H_1 : x = Sa_t + w, \hat{a} = f(S, B)$
- Models the background as a statistical distribution much like ACE
  - The fully constrained least squares (FCLS) estimate is used to estimate the abundances while maintaining their physical constraints
  - These estimates are then used directly in the final solution

$$D(x) = \frac{x^T \Sigma^{-1} T \hat{a}}{x^T \Sigma^{-1} x}$$

Data Used in Our Analyses

To evaluate the different detectors in our study, we used WAAMD imagery collected at 4000 ft that contains some of the most difficult subpixel targets to detect. The Target fill factors vary from 10% to 50%.

Data Facts:

- Targets Used
  - Target 1: VS16 - Round, white plastic, 15.24 cm.
  - Target 2: M20 - Square, green metal, 30.48 cm.
  - Target 3: M19 - Round, green metal, 30.48 cm.
- Six images were used as listed below

Image	Description	Alt (m)	GSD (m^2)	Area (m^2)
1	Sparse Grass	1220	0.1823	18811
2	Short Grass	1220	0.1823	18811
3	Short Grass	1220	0.1823	19464
4	Sparse Grass	1216	0.1815	18815
5	Short & Tall Grass	1215	0.1806	18542
6	Short Grass	1213	0.1806	19097

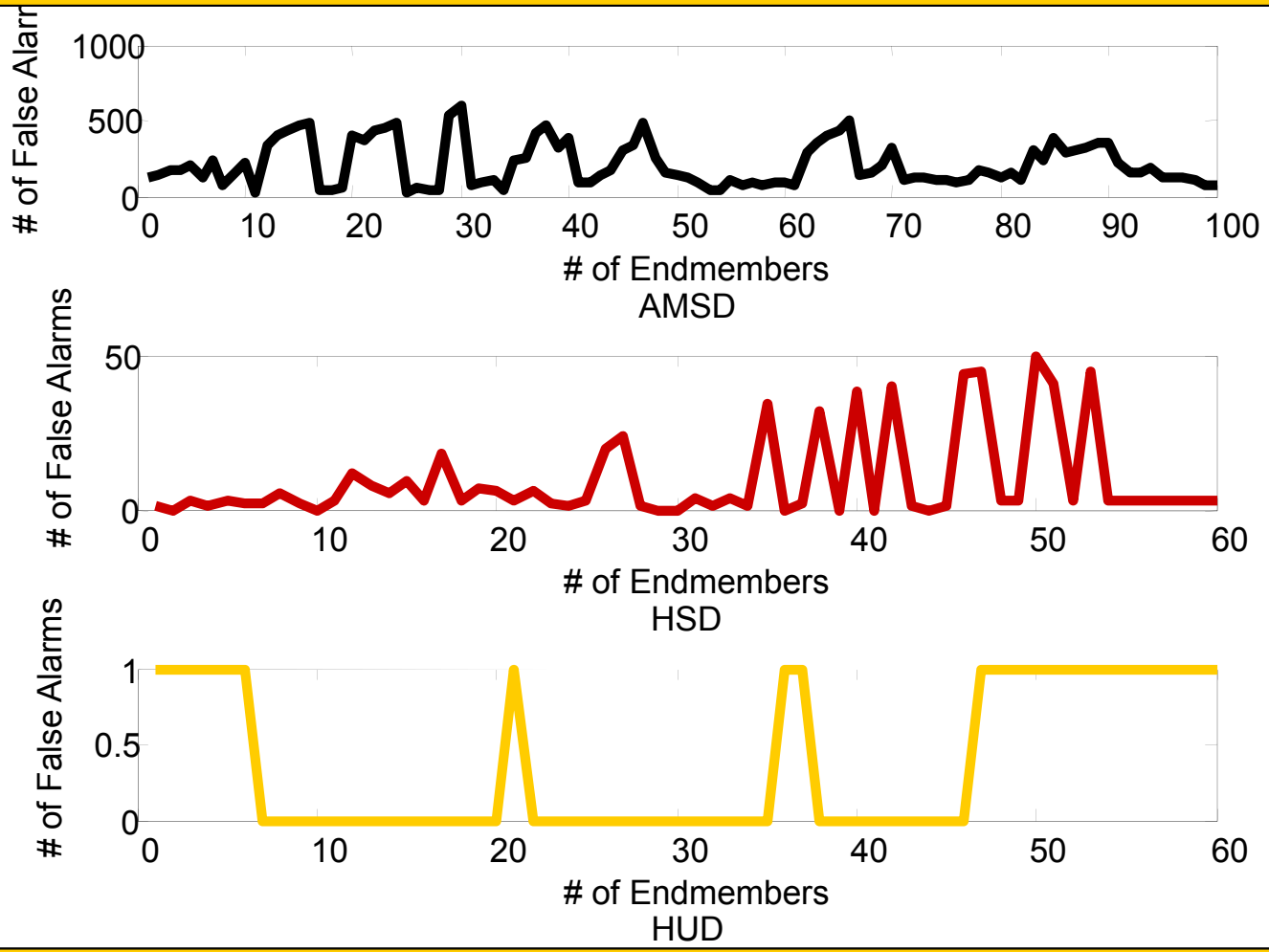
Targets by Image

Image	VS16	M20	M19	All
1	20	42	0	62
2	0	0	12	12
3	0	0	24	24
4	20	30	0	50
5	0	0	16	16
6	0	0	28	28
All	40	72	80	192

Endmember Analysis

Experiment using M19 Targets at 4000 ft.

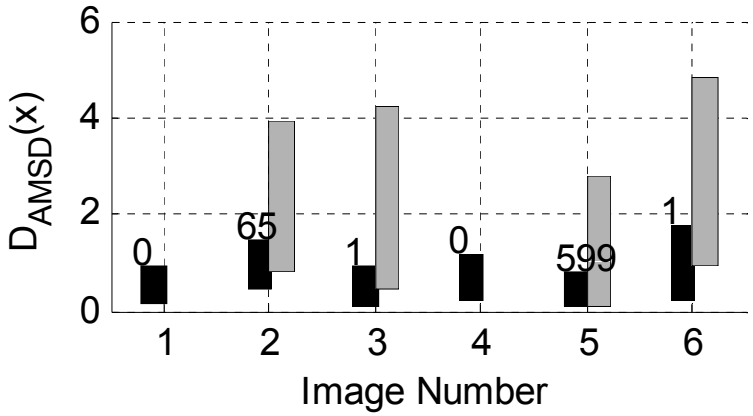
We designed this test to show the sensitivity of the structured detectors to the number of endmembers used. Ideally, the detectors should be partially insensitive because the number of endmembers is not known a-priori and must be estimated.

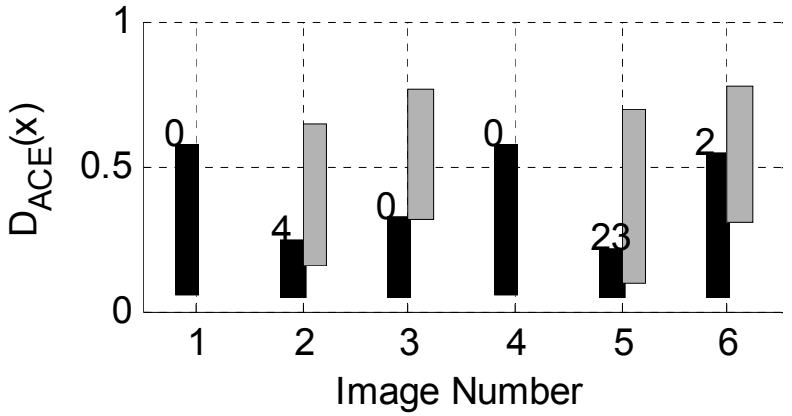


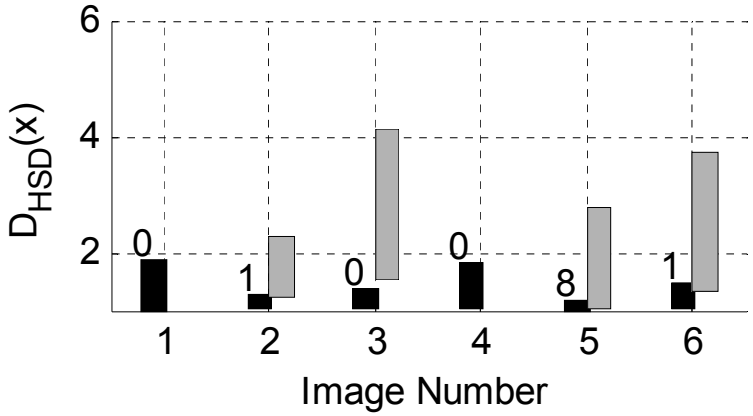
Separation Analysis

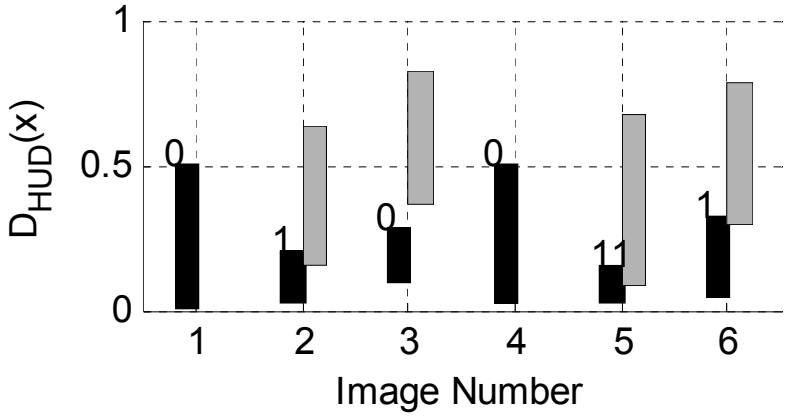
Comparison using M19 Targets at 4000 ft.

We used this experiment to show the separation between targets (gray bars) and clutter (black bars) for each image. The numbers above the black bars indicate the number of false alarms when 100% of the targets are detected.





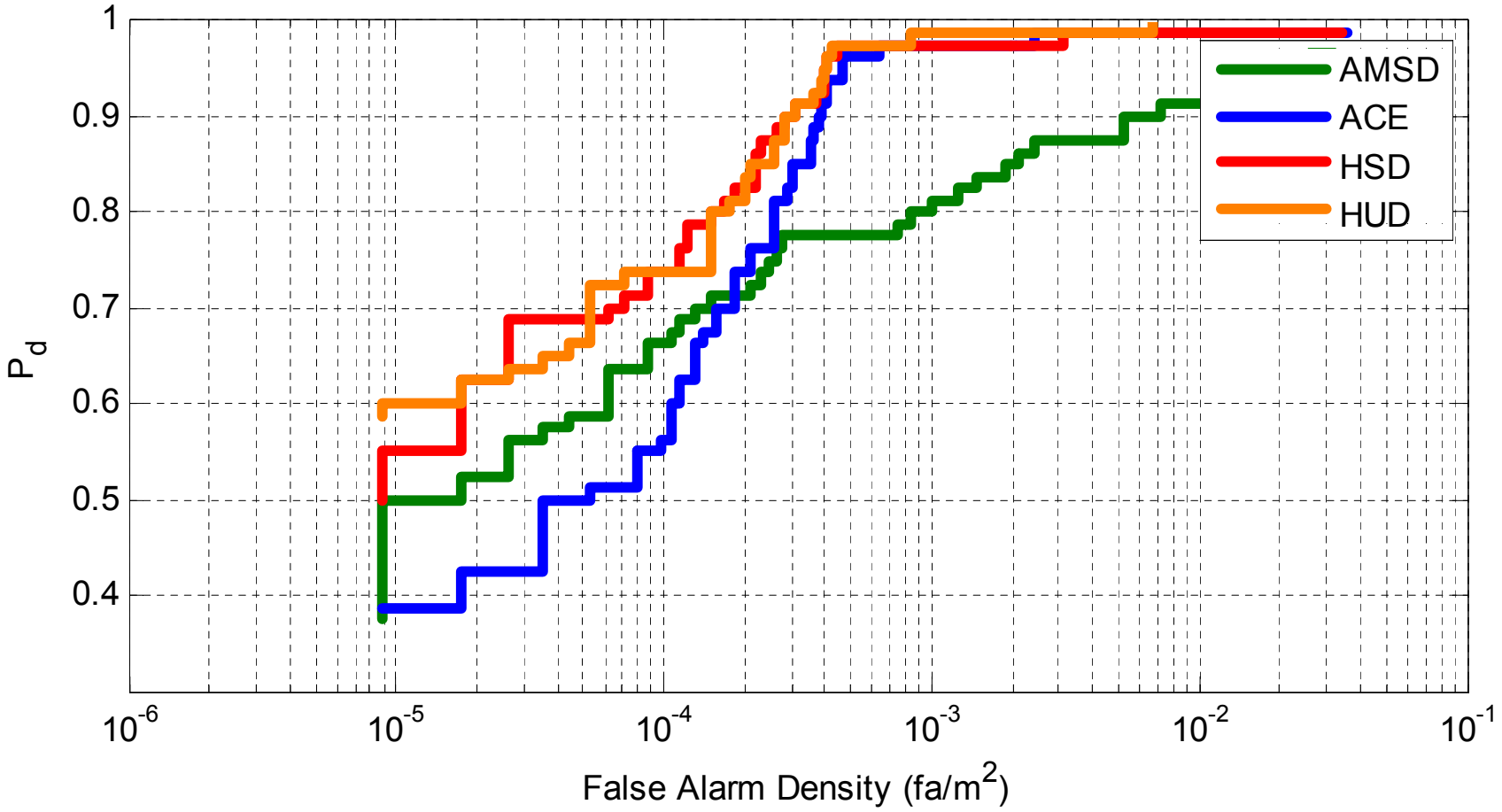




ROC Analysis

Comparison using M19 Targets at 4000 ft.

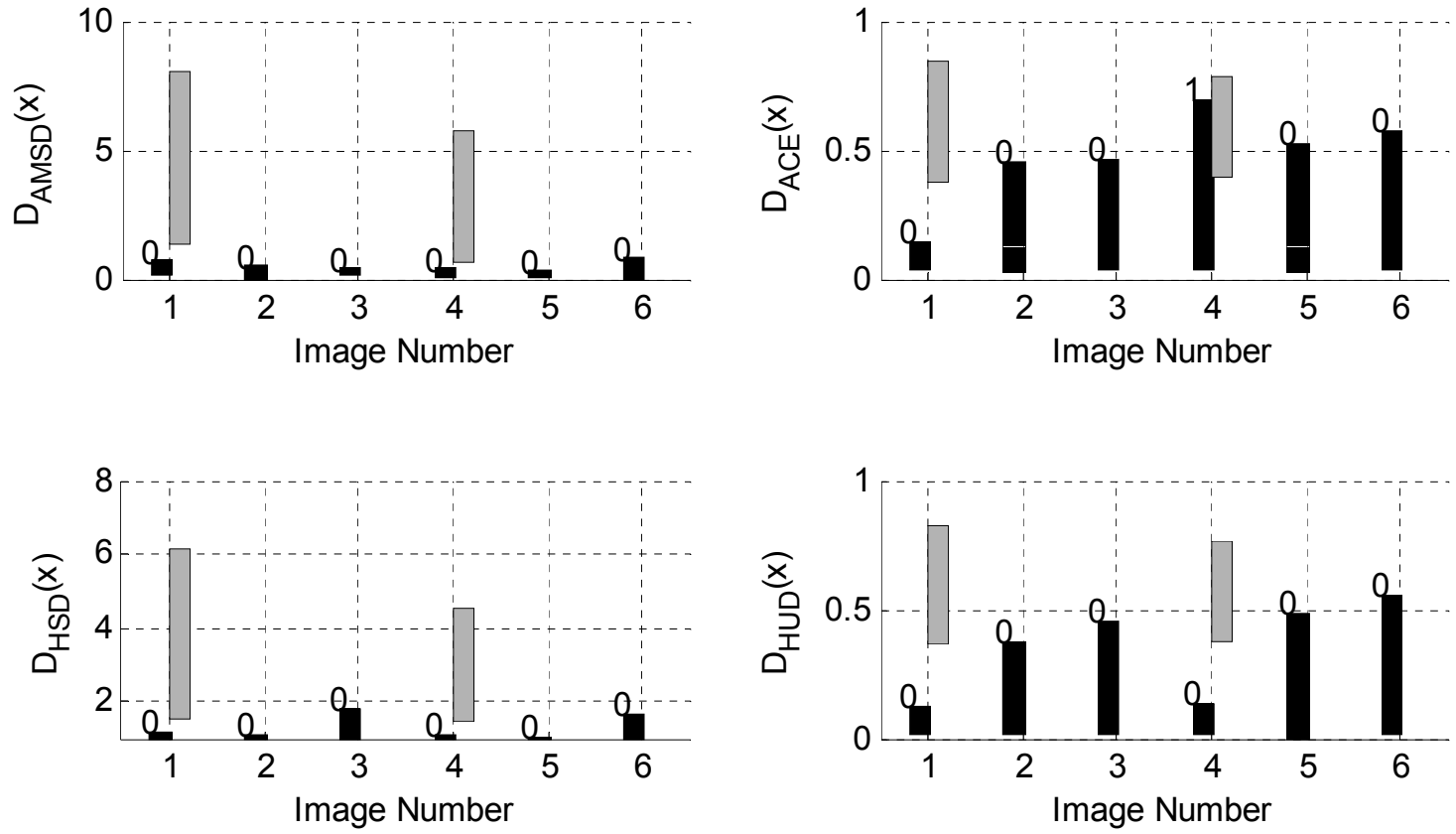
Receiver Operating Characteristic (ROC) curves show the average performance of the detectors for fixed thresholds across all images. The more consistent a detector performs, the better the ROC curve.



Separation Analysis

Comparison using VS16 Targets at 4000 ft.

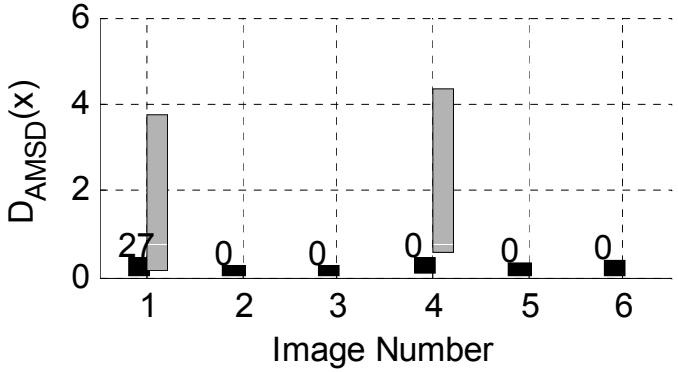
We used this experiment to show the separation between targets (gray bars) and clutter (black bars) for each image. The numbers above the black bars indicate the number of false alarms when 100% of the targets are detected.

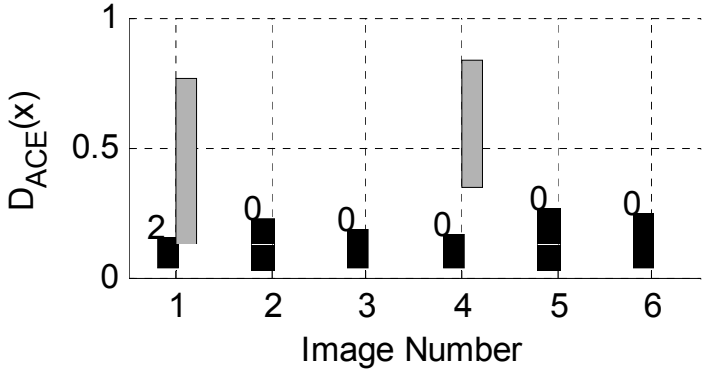


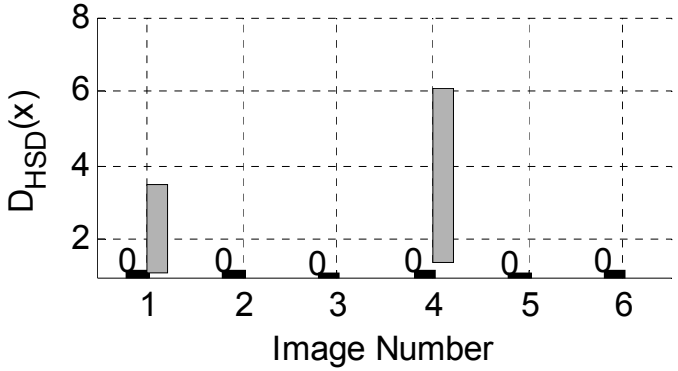
Separation Analysis

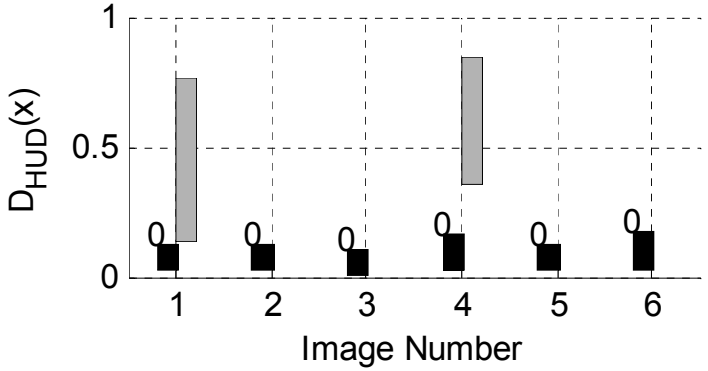
Comparison using M20 Targets at 4000 ft.

We used this experiment to show the separation between targets (gray bars) and clutter (black bars) for each image. The numbers above the black bars indicate the number of false alarms when 100% of the targets are detected.





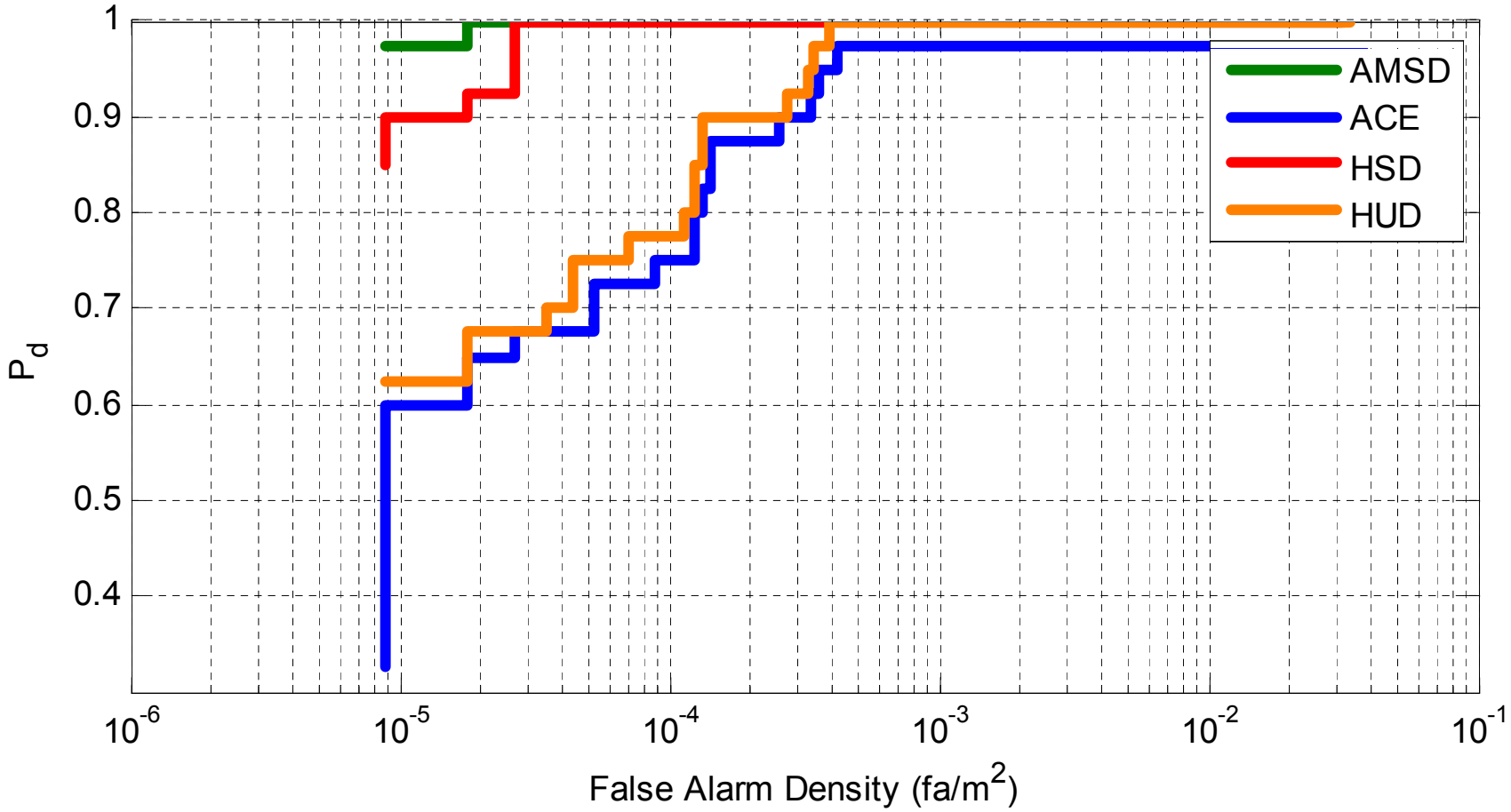




ROC Analysis

Comparison using VS16 Targets at 4000 ft.

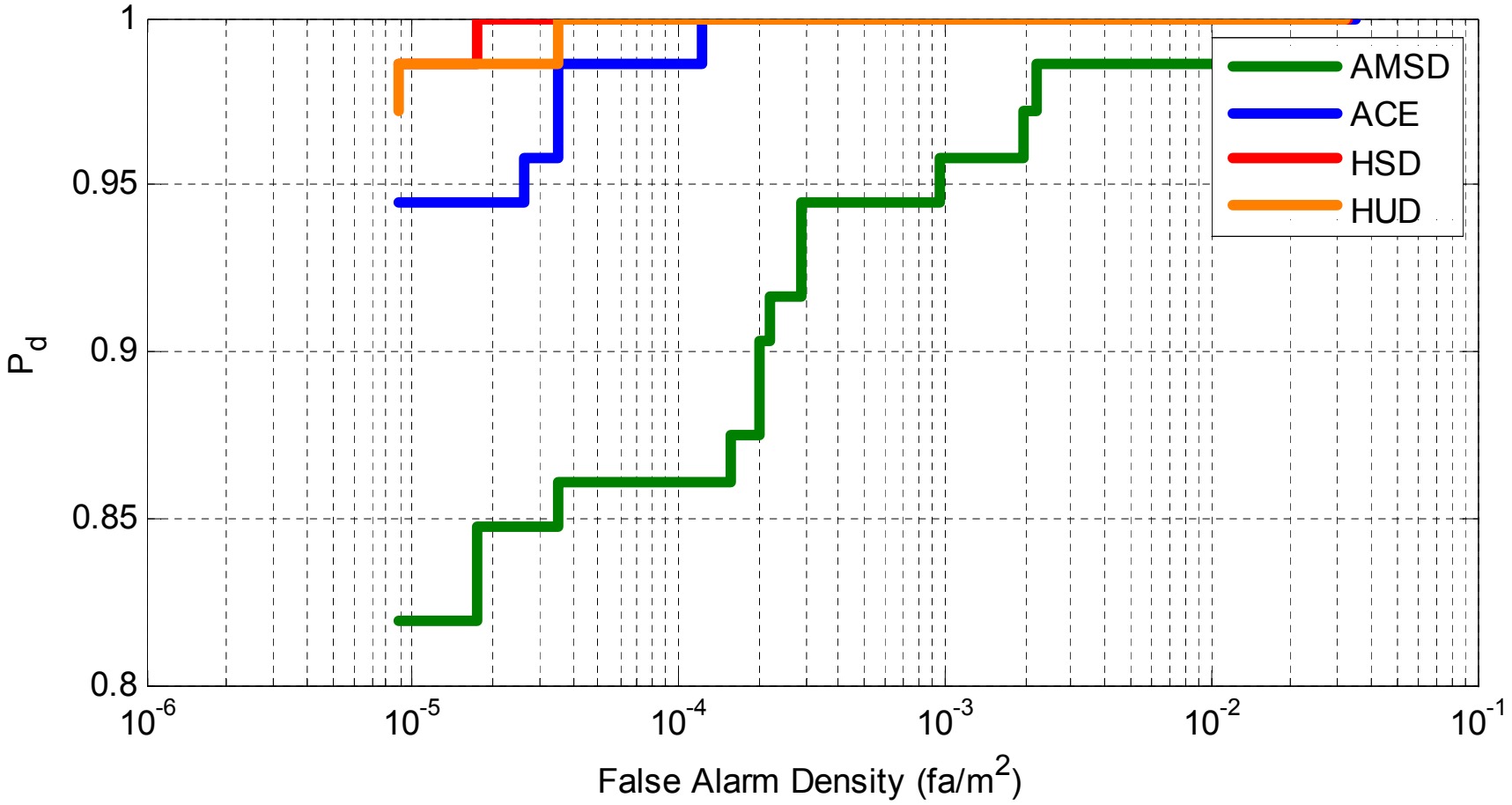
Receiver Operating Characteristic (ROC) curves show the average performance of the detectors for fixed thresholds across all images. The more consistent a detector performs, the better the ROC curve.



ROC Analysis

Comparison using M20 Targets at 4000 ft.

Receiver Operating Characteristic (ROC) curves show the average performance of the detectors for fixed thresholds across all images. The more consistent a detector performs, the better the ROC curve.



Adaptive Threshold Method using Importance Sampling

Detector Threshold Estimation

We need to identify a threshold to separate targets from clutter in hyperspectral data. The method needs to be as general as possible given the idiosyncrasies of HSI data and the availability of non-parametric detection algorithms.

- CFAR Methods
  - Thresholds are based on theoretical calculations from parametric distributions
  - HSI data typically does not conform to these parametric assumptions.
  - Thresholds are only valid for certain types of detectors
- Monte Carlo Methods
  - Truly blind method that requires no knowledge of underlying distribution.
  - Requires many samples to estimate a threshold. Typically about 2 orders of magnitude more samples than the threshold.
  - Example: For a false alarm density of 10^{-6}, we require 10^{8} samples.

- This led us to use a variant of importance sampling.

Importance Sampling

Importance sampling is a forced Monte Carlo method that is used simulate rare events. As with CFAR methods, this algorithm also requires detailed knowledge of the underlying probability density functions.

- Start with Monte Carlo estimate with K samples of X_i from distribution f, p_t is the tail probability, and $1(x)$ is an indicator function:

$$\hat{p}_t = \frac{1}{K}\sum_{i=1}^{K} 1(X_i \geq t)$$

- This can then be weighted with $W(X) = f(X)/f_*(X)$ to find the importance sampling tail probability estimate:

$$\hat{p}_t = \frac{1}{K}\sum_{i=1}^{K} 1(X_i \geq t)W(X_i), \quad X_i \sim f_*$$

- Most of importance sampling is concerned with the development of $W(X)$, and $W(X)$ requires knowledge of the underlying distribution.

Blind Importance Sampling

Bucklew noted that importance sampling could be made "blind." Blind methods require no knowledge of the underlying probability density function in their estimates.

- Start with *K* samples of X_i from unknown distribution *f*, and use an acceptance-rejection method to find a new set of samples:

$$\{Y_j\}_1^{K_r} = \{X_i \mid U_i \le h(X_i)\}_1^K$$

where $0 \le h(X) \le 1$ and *U* is drawn from the uniform distribution.

- This new set of samples *Y* are multiplicatively shifted random variables with distribution:

$$h_*(x) = \frac{1}{a_h} h(x) f(x)$$

- Using these new variables, we can estimate the weighting function without any knowledge of the true underlying distribution *f*.

Blind Importance Sampling (Weighting Function)

Bucklew noted that importance sampling could be made "blind." Blind methods require no knowledge of the underlying probability density function in their estimates.

- Using the new samples *Y*, the blind weighting function is:

$$W(x) \equiv \frac{a_h}{h(x)}$$

- The estimate a_h can be defined as:

$$a_h = P(U_i \le h(X_i)) = E\{h(X_i)\}$$

- This leads to the final estimate:

$$a_h = \frac{1}{K}\sum_{i=1}^{K} h(X_i)$$

Blind Importance Sampling (h-function)

Bucklew noted that importance sampling could be made "blind." Blind methods require no knowledge of the underlying probability density function in their estimates.

- What is a "good" h-function to use? Srinivasan suggested the following solution:

$$h(x) = e^{s(x-c)} 1(x \le c) + 1(x > c)$$

- The final blind tail probability can then be written as:

$$\hat{p}_t = \frac{1}{K_r} \sum_{i=1}^{K_r} 1(Y_i \ge t) \frac{a_h}{h(Y_i)}$$

- This solution still requires estimation of a few constants; namely, the *s* and *c* variable.

Blind Importance Sampling (Adaptive Estimation)

Bucklew noted that importance sampling could be made "blind." Blind methods require no knowledge of the underlying probability density function in their estimates.

- To find the variable *s*, we minimize the exponentially over bounded variance of the tail probability I_b.

$$\hat{I}_b = \hat{a}_h e^{-2st}\left(\frac{1}{K}\sum_{i=1}^{K}\frac{e^{2sx_i}}{h(x_i)}\right)$$

- Note the above solution has the following properties:
  - It does not require acceptance-rejection samples
  - Using the derivatives of I_b, we can numerically find *s*

$$s_{m+1} = s_m - \delta\frac{\hat{I}_b'}{\hat{I}_b''}$$

- Iterate the above equation until the minimum I_b is found.

Inverse Blind Importance Sampling

Inverse Blind Importance Sampling (IBIS) finds the threshold that produces a desired tail probability (e.g. false alarm density). We can use much of the same machinery from BIS to solve this problem.

- Assume we now have a tail probability p_0 we want to achieve.
- Using a similar numerical technique to the one for finding *s*, we can find a threshold *t* such that

$$t_{m+1} = t_m + \delta \frac{p_0 - \hat{p}_t}{\hat{p}_t'}$$

- Note that the above equation requires a derivative w.r.t *t*. This requires taking the derivative of the indicator function. An easier method is to use the following approximation (sigmoid nonlinearity).

$$1(\varepsilon) \approx \frac{1}{1+e^{-a\varepsilon}}, a > 0$$

RX Experiment

This final experiment applies the IBIS technique to the RX detector output on a VIS/NIR/SWIR hyperspectral image. The results are compared to Monte Carlo and CFAR methods with IBIS showing improved performance over the other methods.

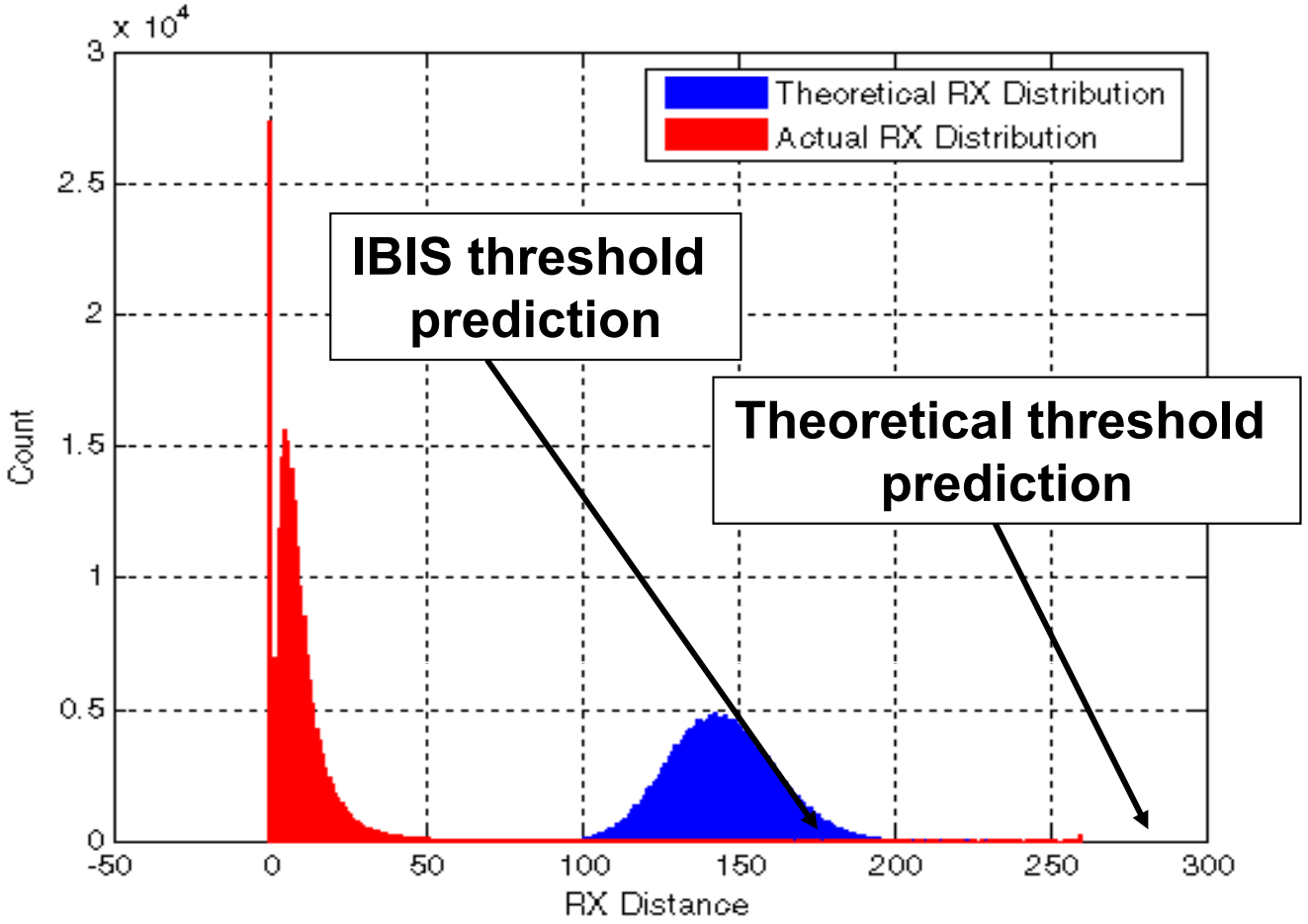
- Experiment particulars:
  - A single image with multiple "anomalies" was processed using the well-known RX algorithm.
  - Over 200,000 samples were in the image allowing us to calculate the ideal threshold for a given false alarm density of 0.001.
- Results:
  - RX statistic is not a good match to the data
  - Monte Carlo methods are prone to outliers
  - IBIS uses more samples than MC and hence out-performs MC

Table 1: RX Threshold Results

Estimator	Threshold	Pd	α
Theoretical RX	279.07	0.15	0.00039
Monte Carlo	269.92	0.16	0.00042
IBIS	176.65	0.23	0.00103
Ideal	180.59	0.22	0.00100

Comparison of Theoretical and Actual RX Statistics on HSI Data

The graph below shows the difference between expected RX performance and actual RX performance on HSI data. Because HSI data is typically non-Gaussian, the theoretical prediction is incorrect while IBIS performs well.



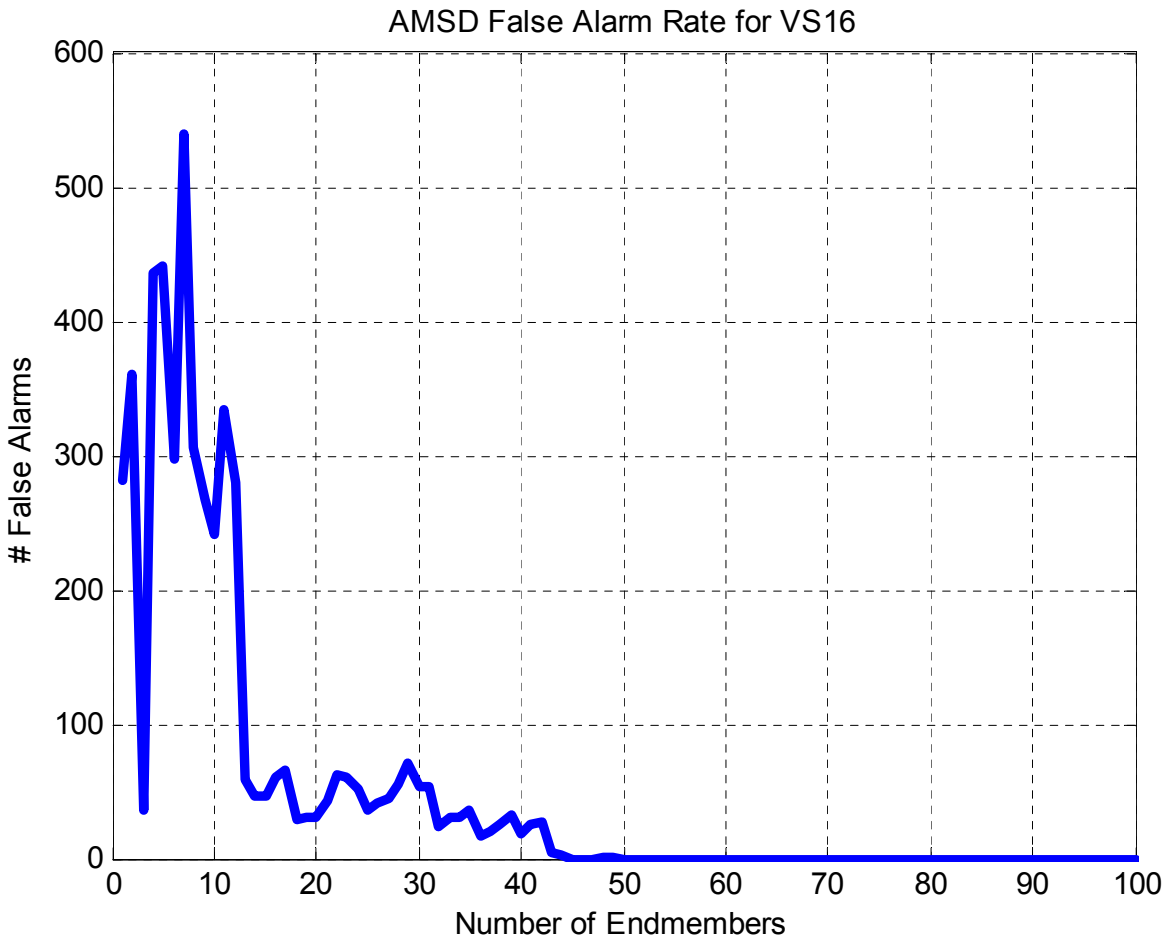
Additional Topics

- Estimation of correct number of endmembers
  - Estimation of background using kernel methods
  - Joint Spatial/Spectral Processing
  - Spectral Object Level Change Detection

Determining the Number of Endmembers

Our work with the hybrid detectors showed how the number of endmembers used can affect detection performance. Determining the correct number of endmembers is a critical part of detection.

- **In AMSD, the number of endmembers is selected by choosing 99.9% of the energy**
  - This method would choose 3 endmembers which gives 36 false alarms
  - Ideal is above 50 endmembers for 0 false alarms
- **We need another metric to identify the correct number for all detectors (AIC, BIC, MDL, etc.).**



Joint Spatial/Spectral Detection

The original RX algorithm was designed to find anomalies with a known spatial shape. We can extend this idea to the other detectors such as ACE. This work will focus on detectors that identify a target with a known shape and spectrum.

- The RX algorithm was designed to give the following detection statistic:

$$r(X) = \frac{(XS^T)^T(XX^T)^{-1}(XS^T)}{SS^T} \begin{matrix} \geq r_0, & \text{then } H_1 \\ < r_0, & \text{then } H_0. \end{matrix}$$

  where **X** is zero-mean data forming a L×N matrix, L is the number of spectral bands, N is the number of samples, **S** is a 1×N vector depicting the known shape of the object to be found.

- For example a new ACE detector with spatial filter and target **T** is:

$$D(x) = \frac{(XS^T)^T(XX^T)^{-1}T^T[T^T(XX^T)^{-1}T]^{-1}T(XX^T)^{-1}(XS^T)}{(XS^T)^T(XX^T)^{-1}(XS^T)}$$

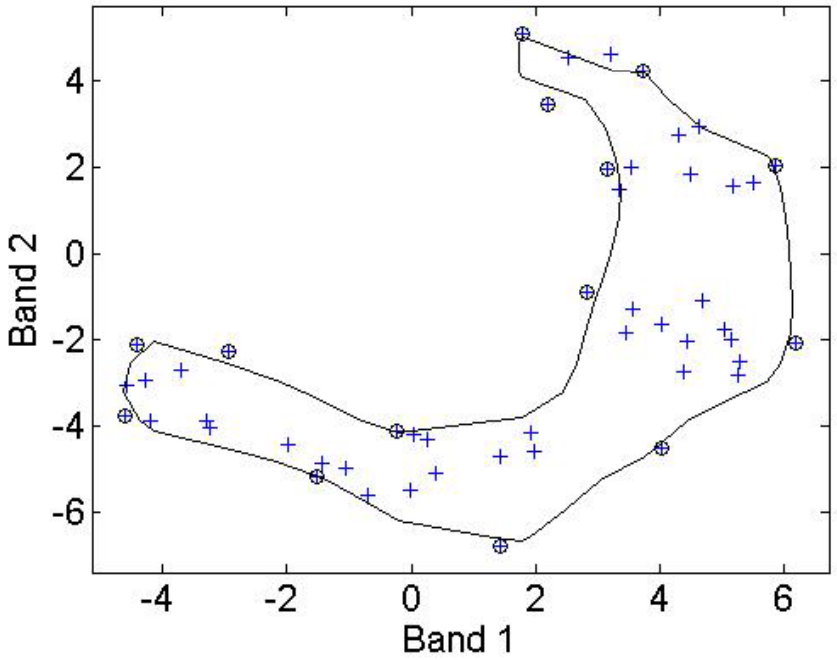
- Our research will focus on any possible advantages such a joint spectral/spatial detector may have over a spectral detector followed by a spatial filter.

Endmember Extraction using SVDD

Joint Work with Dr. Banerjee at JHU/APL

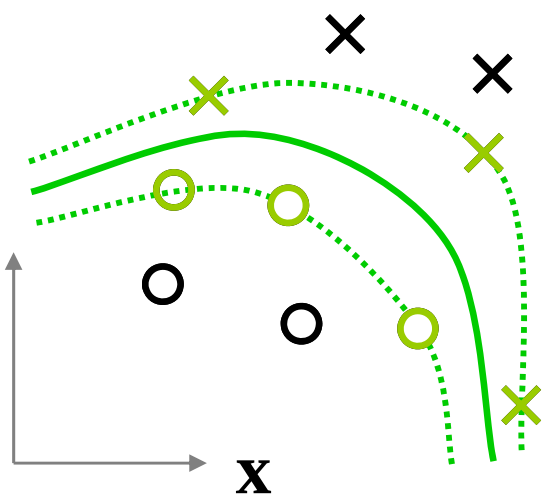
Extended work based on a Support Vector Data Description (SVDD) anomaly detector (Banerjee, Burlina, and Diehl, IEEE TGRS 2006).

- The SVDD attempts to characterize a class by enclosing it with the smallest hypersphere possible.
- Using kernel functions, a class can be enclosed by a complex boundary that better describes the class.
- The boundary of this class can be defined by points that can be interpreted as endmembers.



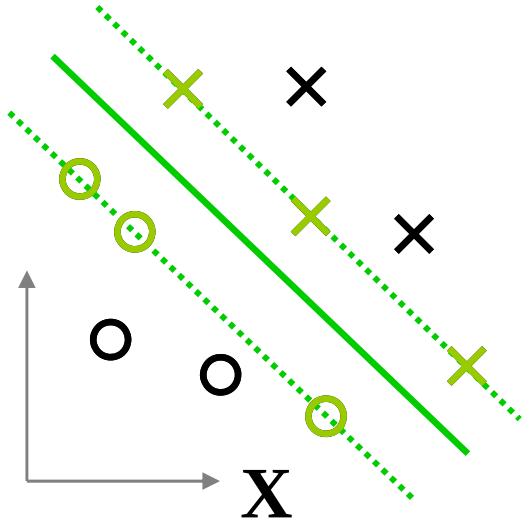
The endmembers define a convex set in the feature space. The endmembers are physically-meaningful and can be used in any kernel based detector. The work may also lead to nonlinear mixing effects that have not been previously modeled.

Use of Kernels for Nonlinear Classification



$$\Phi(\cdot)$$

$$\mathbf{X}_i = \Phi(\mathbf{x}_i)$$



$$\mathbf{X}_i \cdot \mathbf{X} = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x})$$

$$y = \text{sign}\left(\sum_{i \in S} \alpha_i y_i \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}) + b \right)$$

$$y = \text{sign}\left(\sum_{i \in S} \alpha_i y_i \mathbf{X}_i \cdot \mathbf{X} + b \right)$$

$$K(\cdot,\cdot)$$

$$\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}) = K(\mathbf{x}_i, \mathbf{x})$$

Mercer's Condition

$$y = \text{sign}\left(\sum_{i \in S} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b \right)$$

Any classifier that uses a dot product can then use these kernels to create a nonlinear boundary!

2006 Research Accomplishments

- Research:
  - Improved estimation of target radiance signatures
  - Finished development of the hybrid detectors
  - Developed an adaptive threshold method based on importance sampling
  - Began development of an algorithm to estimate the proper number of endmembers for an application
  - Began development of kernel-based endmember and abundance estimates
  - Began development of joint spectral/spatial detectors

- Publications:
  - J.B. Broadwater and R. Chellappa, “An adaptive threshold method for hyperspectral target detection,” in *Proceedings of the IEEE ICASSP*, vol. 5, May 2006, pp. V-1201 – V-1204.
  - J.B. Broadwater and R. Chellappa, “Hybrid Detectors for Subpixel Targets,” submitted to *IEEE TPAMI*, October 2006.
  - J.B. Broadwater and R. Chellappa, “Physics-based detectors applied to long-wave infrared hyperspectral data,” accepted at *Army Science Conference 2006*, November 2006.

UNIVERSITY OF FLORIDA

Information Fusion

University of Florida

University of Florida

Research Results
– Summary –

University of Florida

MURI Participants

- Dr. Paul Gader – Principal Investigator
- Alina Zare
- Jeremy Bolton
- Andres Mendez-Vasquez

University of Florida

Research Objectives

- **Goal: Develop robust, mathematically sound algorithms for fusing HSI and SAR imagery**

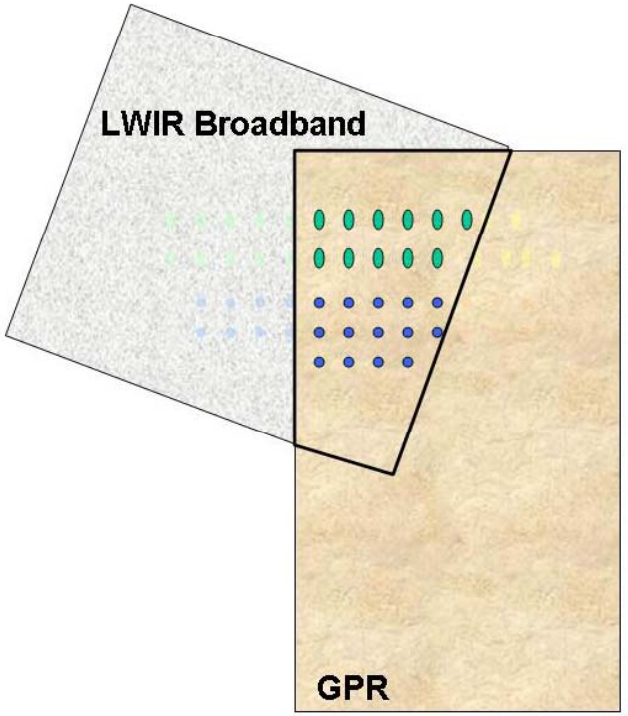
University of Florida

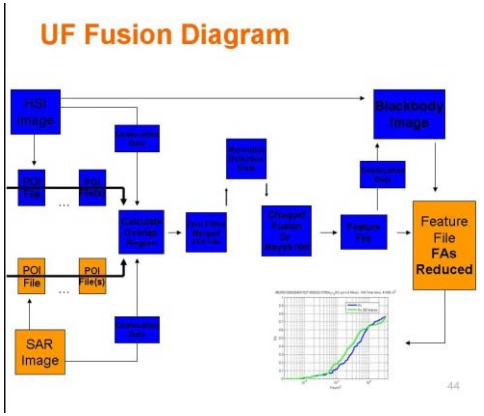
Research Activities

- Vegetation Mapping for Landmine Detection in Long Wave Hyperspectral Imagery.
- SAR and Hyperspectral Sensor Fusion method using minimum classification error of Choquet integrals
- SAR and Hyperspectral sensor fusion method using a Bayesian Sparsity Promoting estimation of Choquet integral parameters
- SAR and Hyperspectral sensor fusion with Bayesian network.
  - Investigated both Structure and Parameter learning
- Bayesian Sparsity Promoting techniques to identifying endmembers in Hyperspectral imagery.
- Continuous Choquet Integral advancements
  - Relationship to Random shapes, Dempster-Shafer, Capacity Functionals

WAAMD Support

- Supported WAAMD-Wide Area Airborne Minefield Detection
  - Provided Detector outputs to WAAMD community
    - Reststrahlen Feature (HSI)
    - Blackbody Feature (HSI)
      - SPICE
      - EM clustering with skewness feature
      - ICE
    - Choquet Confidence (Fusion)
    - Choquet Hit-Miss (SAR)
  - Perform decision level fusion of SAR and Hyperspectral data
    - Choquet Integral
      - **REDUCING FARs by up to 95%** over individual detectors
      - **INCREASING PDS by up to 30%** at reasonable FARs
      - Experimented w/ Multiple Measures
      - Experimented w/ Multiple Optimization Methods
      - During support: improved from “Marginal” to “Satisfactory-Very good” detection results: FAR from .01 to .001 FAs/m
    - Bayesian Network
    - AND/OR
    - Vegetation Masking
      - **Reducing FARs up to 88%**
  - Participated in Level 1 Evaluation
    - Met “Satisfactory” to “Very Good” objectives
  - Developed a fusion infrastructure with Raytheon and TRA
    - Software suite

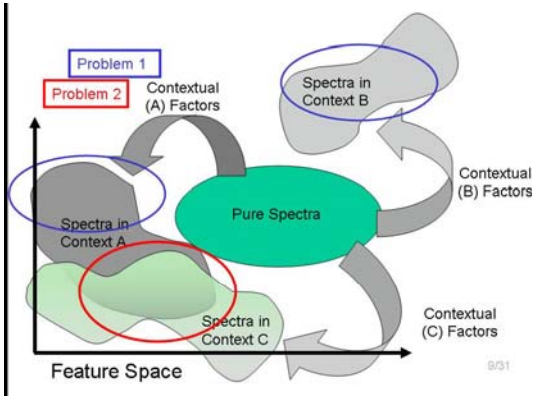




University of Florida

Random Set Method

- Contextual hindrance discovered in previous research
  - Hyperspectral Imagery (HSI)
  - Many unknown / unspecified factors transform the data on an image-by-image basis
  - Environmental conditions: Sun's intensity, Humidity, Atmosphere, Mineralogy …
    - Makes the classification problem difficult
  - Context: conditions or situations in which the data was collected
  - Determine a test image's context → Improve classification
- Possible Standard Statistical Approach
  - Require values for contextual variables
  - Curse of high dimensionality
  - Sparse densities
  - I.I.D. assumption for tractability
    - Does not capture correlated factors: context
    - Not robust to outliers
  - Existing context-based approaches
    - Ignore the true idea of context



- A random set model
  - Implicitly captures information encoded in the SET of samples
  - Does not require explicit specification of contextual factors
  - Avoid curse high dimensions, Avoid sparse densities
  - Solid mathematical foundation
  - Versatile model: allows for possibilistic, probabilistic, and evidential solutions
  - Automatically weights relevant contexts / models
  - Allows for optimization
  - Allows for context learning
- Related References:

- J. Bolton and P. Gader, "Application of Random Set Based Clustering to Landmine Detection with Hyperspectral Imagery," IEEE Geoscience and Remote Sensing, Barcelona, July 2007, pp. 2022-2025.
- J. Bolton and P. Gader, "Random Set Model for Context-Based Classification", IEEE World Congress on Computational Intelligence, Hong Kong, June 2008, Accepted.
- J. Bolton and P. Gader, "Application of Context-Based Classifier to Hyperspectral Imagery for Mine Detection", SPIE Defense and Security Conference , Orlando, March 2008, Accepted.
- J. Bolton and P. Gader, "The Benefits of Context Estimation for Target Spectra Detection in Hyperspectral Imagery," IEEE Geoscience and Remote Sensing, (Submitted).

University of Florida

Choquet fusion methods

- Choquet Integral Fusion
  - Applied Choquet fusion techniques for decision level fusion of HSI and radar imagery
    - Nonlinear integral that represents wide variety of fusion operations
  - Developed novel Fusion and Optimization methods
  - Experimented using different optimization methods
    - MCE optimization
    - Sparsity promotion models
    - Maximum a Posteriori EM MCE Logistic LASSO
  - Experimented using various fuzzy measures
    - Voting, averaging, Sugeno, OWA, and, or, order statistics, …
    - Integral wrt non-additive measure
    - Fully characterized the family of fuzzy measures that induce a metric w/in the Choquet integral

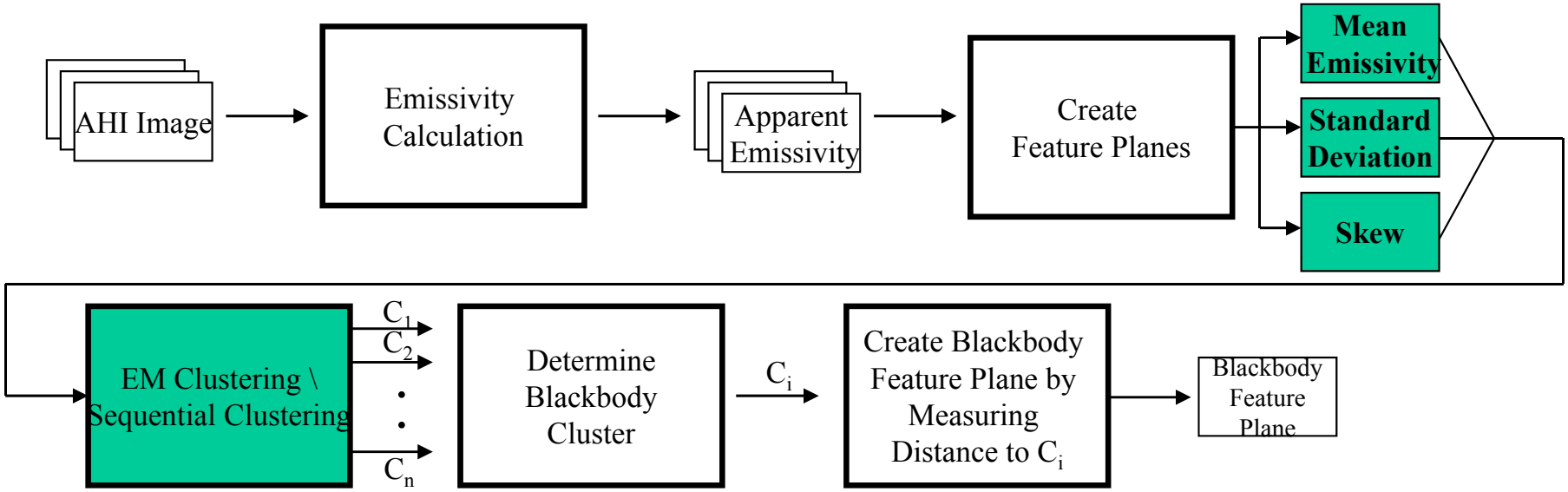
- Related References:

- A. Mendez-Vazquez, P. D. Gader, J. M. Keller, and K. Chamberlin, "Minimum classification error training for Choquet integrals with applications to landmine detection," in IEEE Transactions on Fuzzy Systems, pp. 225-238 ,February 2008, vol. 16, num. 1.
- A. Mendez-Vazquez and P. D. Gader, "Learning Fuzzy Measure Parameters by Logistic LASSO," in IEEE NAFIPS 2008, New York, May 2008.
- A. Mendez-Vazquez and P. D. Gader, "Maximum a Posteriori EM MCE Logistic LASSO for Learning Fuzzy Measures," in IEEE World Congress in Computational intelligence, Hong Kong, China, June 2008.
- A. Mendez-Vazquez and P. D. Gader, "Sparsity promotion models for the Choquet integral," Procs. of the IEEE Symposium on Foundations of Computational Intelligence, Honolulu, Hawaii, April 2007.
- P. Gader, A. Mendez-Vasquez, K. Chamberlin, J. Bolton, and A. Zare, "Multisensor and algorithm fusion with the Choquet integral: applications to landmine detection," Procs. of IEEE International Geoscience and Remote Sensing Symposium, September 2004, pp. 1605-1608, vol 3.
- M. A. Schatten, P. D. Gader, J. Bolton, A. Zare, and A. Mendez-Vasquez, "Sensor fusion for airborne landmine detection," Proceedings of SPIE, Vol. 6217, May 2006, CID 62172F.
- P. Gader, L. Wen-Hsiung, and A. Mendez-Vazquez, "Continuous Choquet integrals with respect to random sets with applications to landmine detection," in IEEE International Conference on Fuzzy Systems, Budapest, Hungary, July 2004, pp. 523-528, vol. 1.
- J. Bolton, P. Gader, and J. Wilson, "Discrete Choquet Integral as a Distance Metric," IEEE Transactions on Fuzzy Systems, In Press.

University of Florida

Vegetation Detection in LWIR

- Vegetation in the LWIR, acts like a blackbody, with a high mean emissivity and a low standard deviation of emissivity.
- Determined **skewness** of emissivity is an important feature for vegetation detection in the LWIR in addition to mean and standard deviation of emissivity.
- Developed vegetation detection algorithm for the LWIR based on clustering.



- Applied vegetation detection algorithm to WAAMD data set; the method provided a significant reduction in false alarms. **Obtained up to 88% reduction in false alarms.**

- Related Journal Article:

  A. Zare, J. Bolton, P. Gader, M. Schatten, "Vegetation Mapping for Landmine Detection Using Long Wave Hyperspectral Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 46, No. 1, pp. 172-178, Jan. 2008.

University of Florida

Endmember Detection using Sparsity Promoting Priors

- Developed the SPICE (Sparsity Promoting Iterated Constrained Endmembers) algorithm which autonomously determines **endmembers** and the **number of endmembers** for a given scene.

- Algorithm applies **sparsity promoting priors** to the ICE Algorithm to prune unnecessary endmembers.

- **Transferred algorithm** to NVSED and US Army Armament Research Development and Engineering Center by supplying code to Miranda Schatten and John M. Romano.

- Algorithm was **applied to WAAMD data set** and vegetation detection for false alarm reduction with **up to 60% reduction in false alarms**.

- Related Publications:

A. Zare, P. Gader, "Sparsity Promoting Iterated Constrained Endmember Detection for Hyperspectral Imagery," *IEEE Geoscience and Remote Sensing Letters*, Vol. 4, No. 3, pp. 446-450, July 2007.

A. Zare, P. Gader, "SPICE: a sparsity promoting iterated constrained endmember extraction algorithm with applications to landmine detection from hyperspectral imagery," *Proceedings of SPIE*, Vol. 6553, May 2007, CID: 655319.

University of Florida

Hyperspectral Endmember Detection with Simultaneous Band Selection

- Developed **B-SPICE** (Band Selecting Sparsity Promoting Iterated Constrained Endmembers)
- This algorithm autonomously performs endmember detection, band selection, determines number of endmembers required and number of bands required for a particular dataset.
- The algorithm applies sparsity promoting priors to band weights and endmember abundances to prune unnecessary bands and endmembers.
- Related Journal Article:

  A. Zare, P. Gader, "Hyperspectral Band Selection and Endmember Detection Using Sparsity Promoting Priors," *IEEE Geoscience and Remote Sensing Letters*, In Press.

  A. Zare, P. Gader, "Sparsity Promoting Iterated Constrained Endmember Detection with Integrated Band Selection," *Procs. of the IEEE Geoscience and Remote Sensing Symposium*, Barcelona, July 2007, pp. 4045-4048.

University of Florida

Technology Transfers and Collaboration

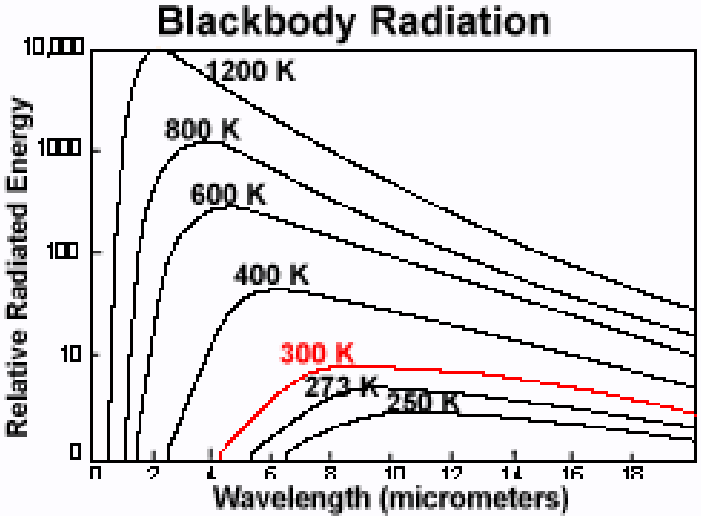
- NVESD Wide Area Airborne Minefield Detection
  - Groups
    - NVESD
    - IDA
    - ARL
    - Raytheon (Data Assimilation, Ground Truth, Geometric Transforms)
    - Veridian
    - STI
    - TRA
    - SRI
    - SAIC
    - Duke
    - …
  - Other University of Florida Faculty / Students (Gerhard Ritter)
- NVESD Forward Looking Program
  - Reststrahlen + Thermal for Mine Detection

University of Florida

Vegetation Detection in LWIR

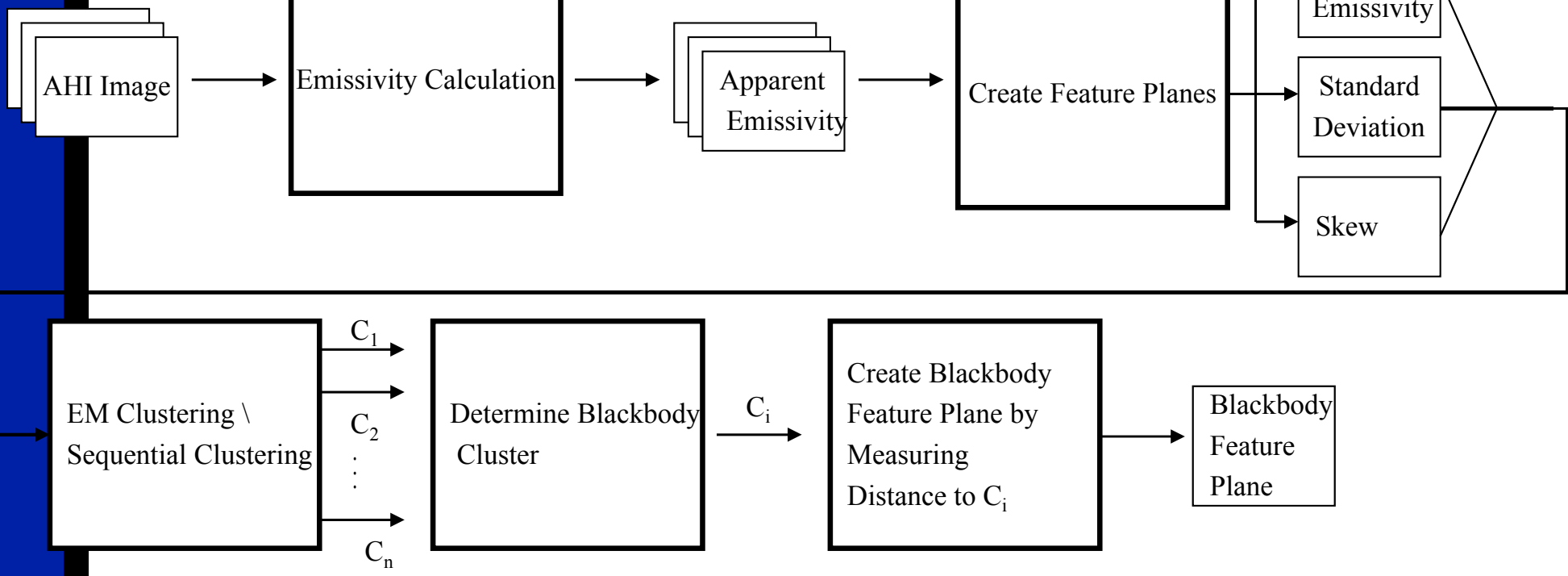
Vegetation Detection in LWIR

- Goal – Reduce False Alarm Rate
- Method - Detect Vegetation in LWIR
- Foundation
  - Vegetation a blackbody over LWIR
  - High emissivity over all LWIR wavelengths
    - High mean, low standard deviation
- Using emissivity normalization method currently
- Statistics of apparent emissivity used as features



University of Florida

Blackbody Feature



University of Florida

LWIR Vegetation Detection: Literature Review

- The literature contained very little about vegetation detection in the LWIR
- Most of the literature focused on the calculation of emissivity and surface temperature
- Some of the Apparent Emissivity Calculation Methods that were Reviewed:
  - Emissivity Normalization
    - The method we are currently using
    - *Advantage:* Provides both apparent emissivity spectral shapes and apparent emissivity values
    - *Disadvantage:* Requires input of an assumed emissivity value
  - Alpha Derived Emissivity Method
    - Based on Wien's approximation of the Planck function
    - *Advantage:* Alpha residual values are independent of temperature which allows calculation without assuming any temperature values
    - *Disadvantage:* Only provides spectral shape of apparent emissivity not absolute apparent emissivity values
  - Taylor Expansion of Planck Function
    - Uses Taylor Series expansion to linearize the Planck Function
    - *Advantage:* Provides spectral shape of emissivity curve that is possibly more accurate than Wien's approximation
    - *Disadvantages:* Requires an estimate of surface temperature and only provides emissivity curve not the apparent emissivity value

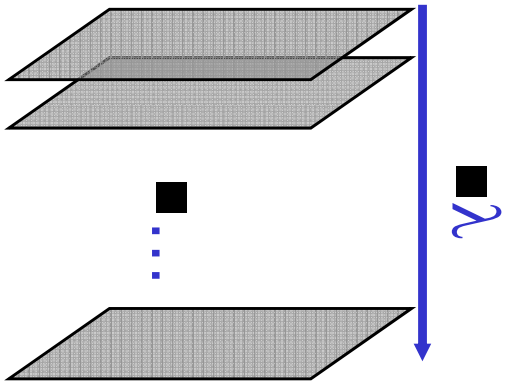
Create Apparent Emissivity Feature Planes

Calculate Apparent Emissivity Statistics across spectral bands

$$\mu = \bar{x} = \frac{1}{N}\sum_{j=1}^{N} x_j$$

$$\sigma = \sqrt{Variance}$$

$$k = \frac{1}{N}\sum_{j=1}^{N}\left(\frac{x_j - \bar{x}}{\sqrt{Variance}}\right)^3$$

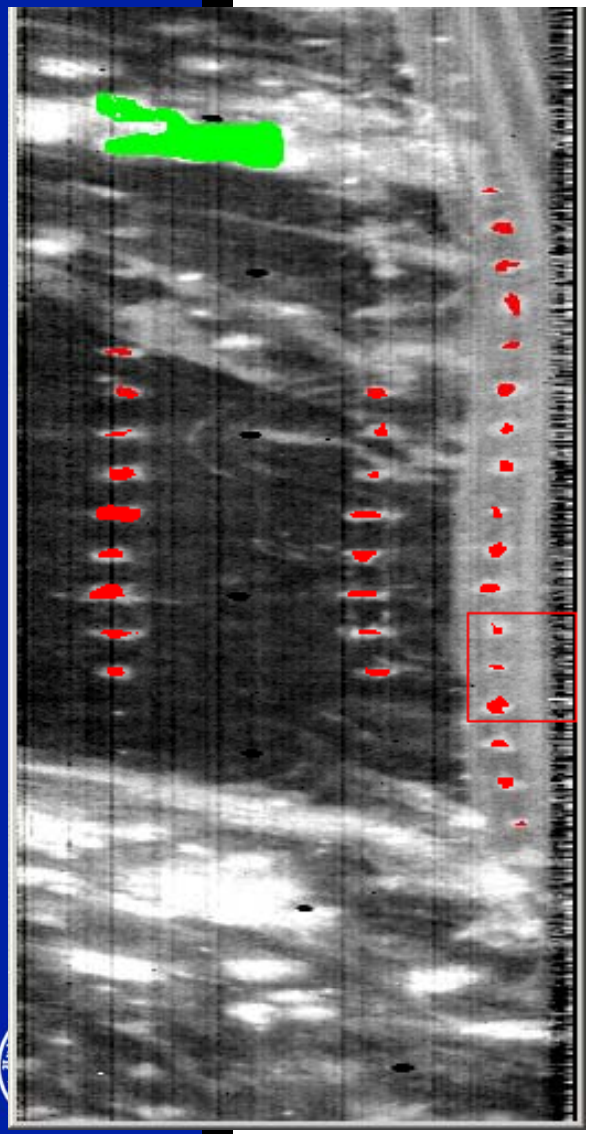


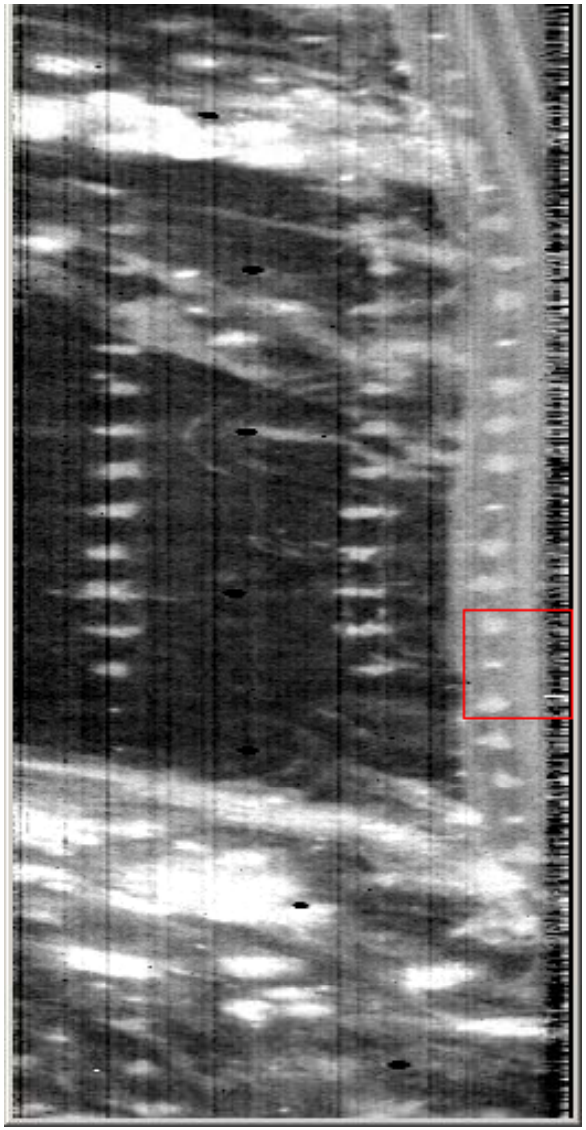
Create feature vectors at each pixel

$$\mathbf{f}_{(x,y)} = \begin{pmatrix} \mu_{(x,y)} & \sigma_{(x,y)} & k_{(x,y)} \end{pmatrix}^T$$

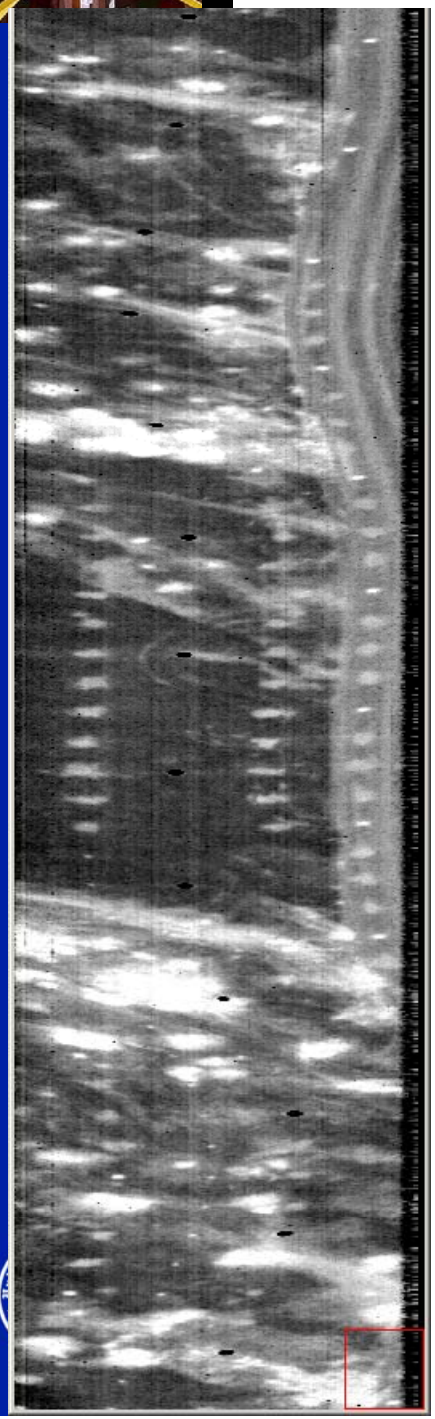
University of Florida

Regions of Interest

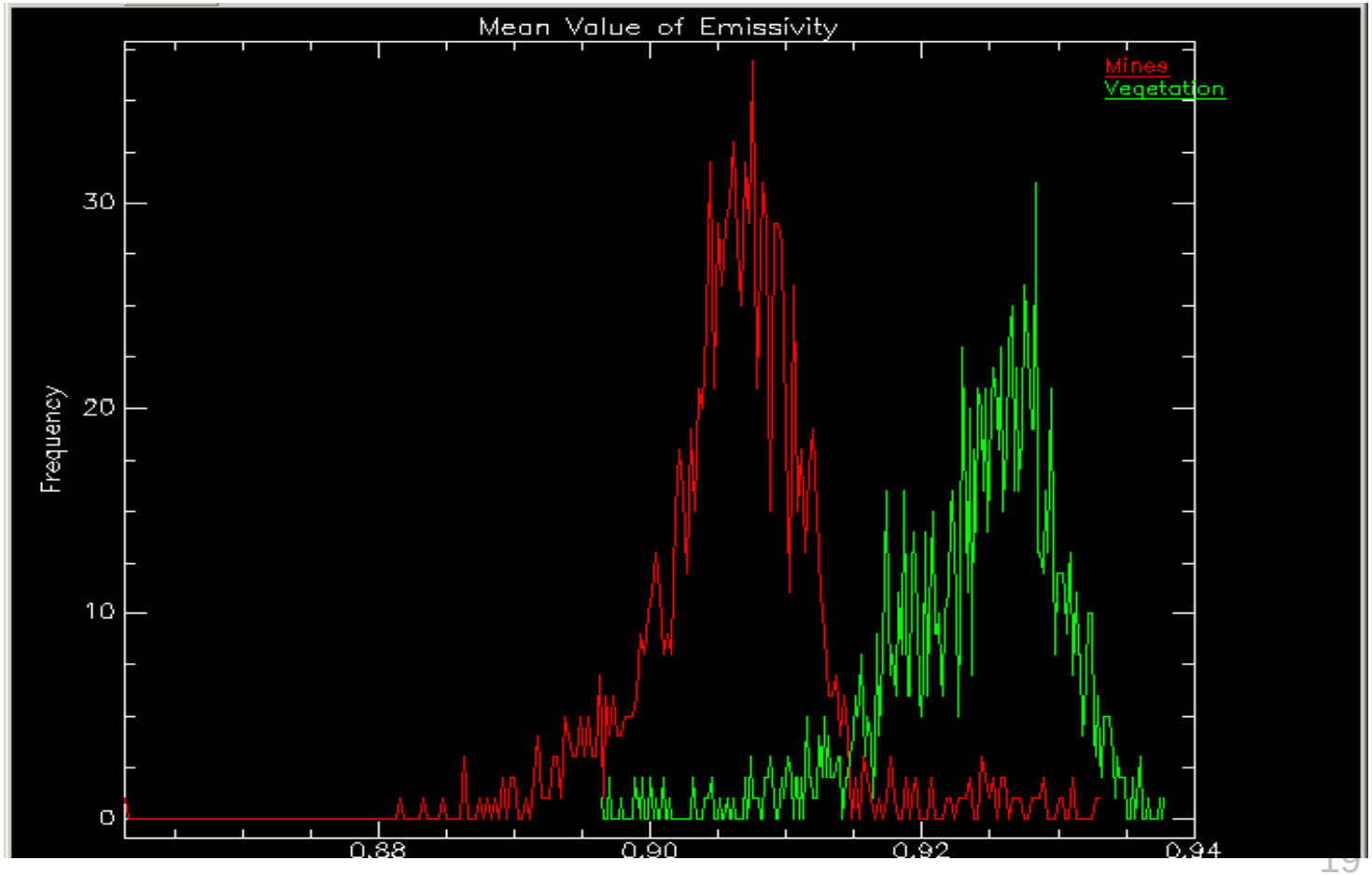


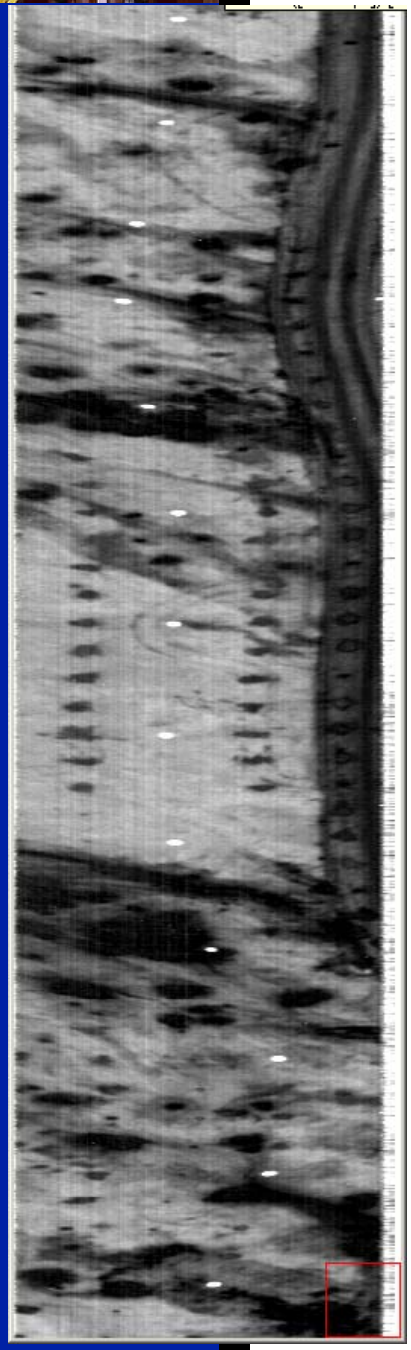


- Hand-elected ROIs to plot emissivity statistics
- Emissivity at 8.9 microns
  cm_am_atlt700_s300_r1_20030402_085640_r_bin3_s10e220.img
- Red: Mine (1184 points)
- Green: Vegetation (1516 points)

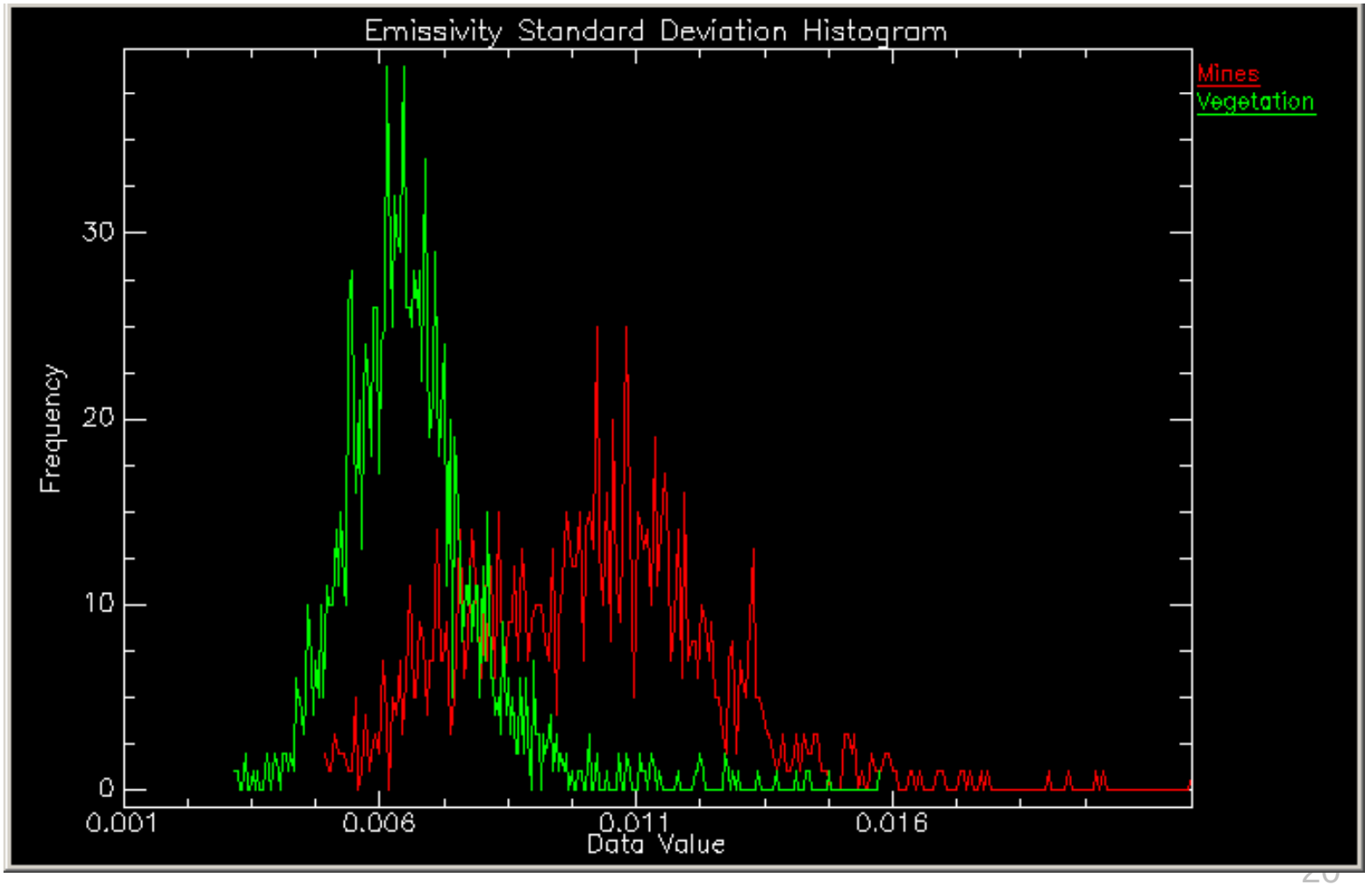


- Mean Emissivity at Mine/Background pixels
- Statistics calculated on wavelengths 7.88 – 9.92 microns of emissivity image

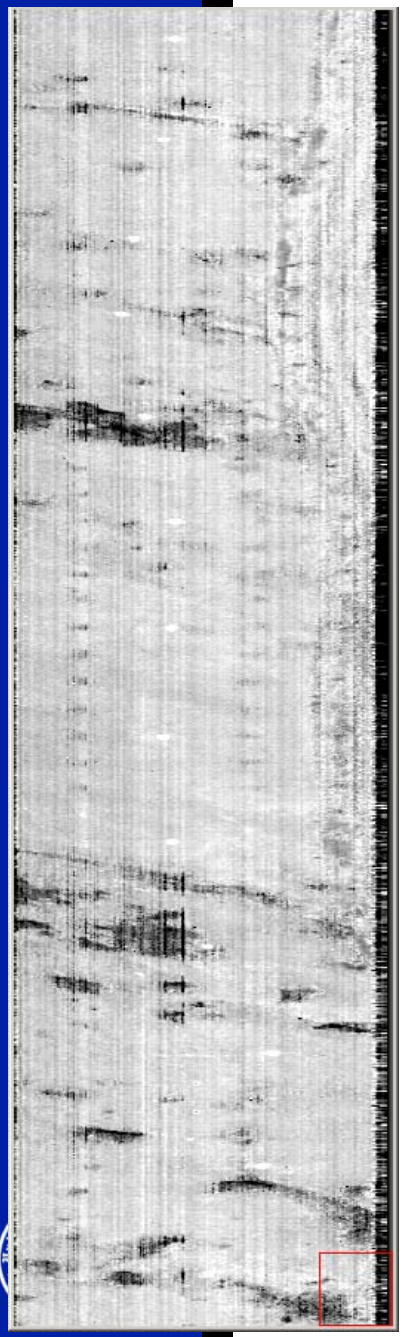




- Standard Deviation of Emissivity at Mine/Background pixels
- Statistics calculated on wavelengths 7.88 – 9.92 microns of emissivity image

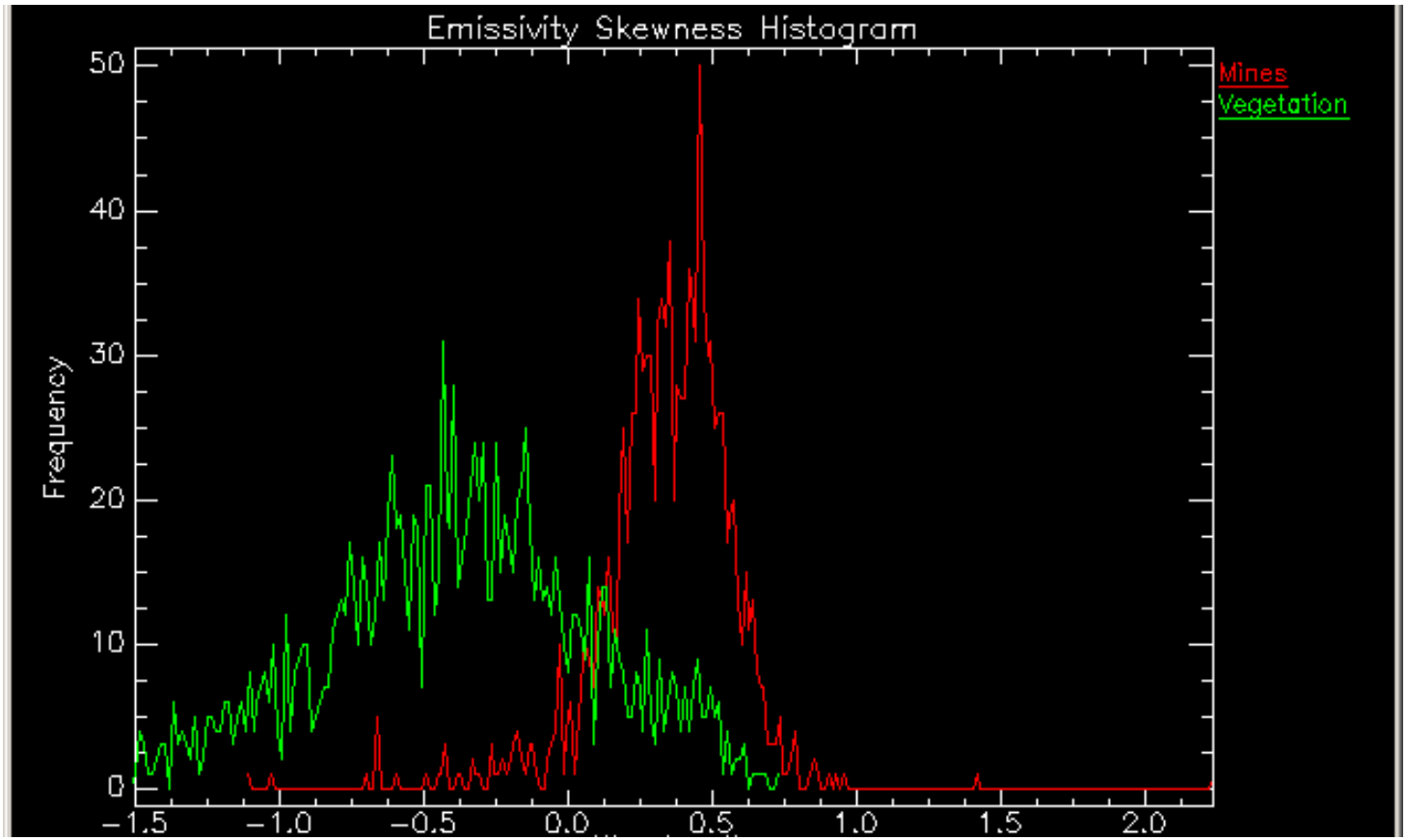


Standard Deviation = $\sqrt{Variance}$

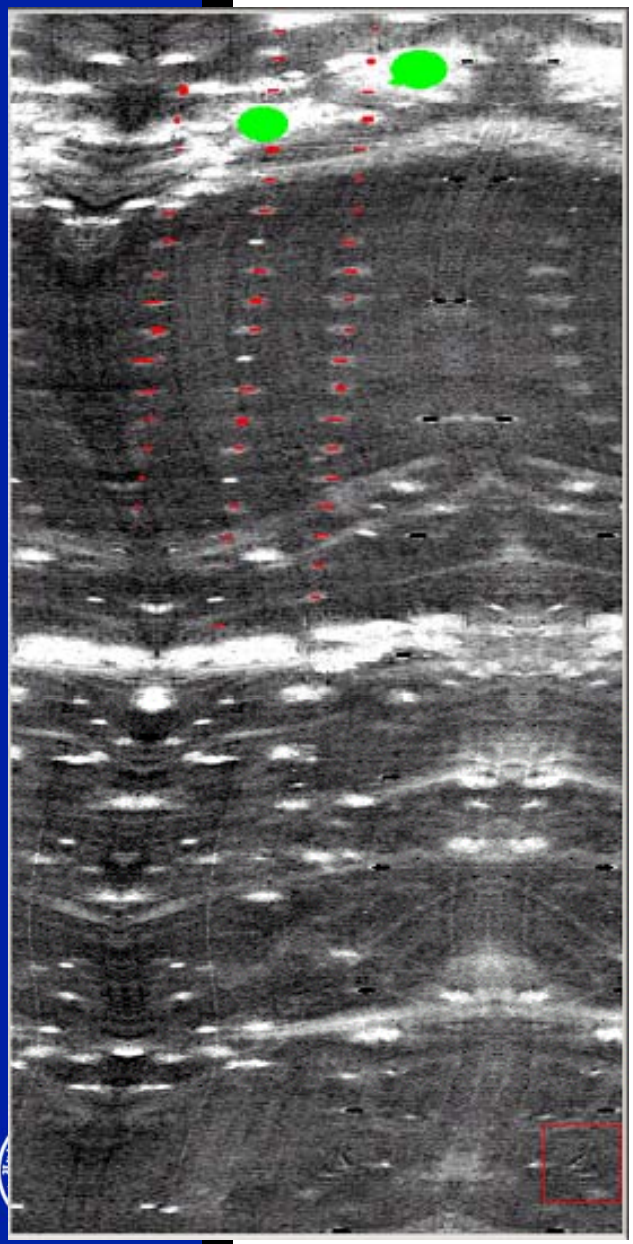


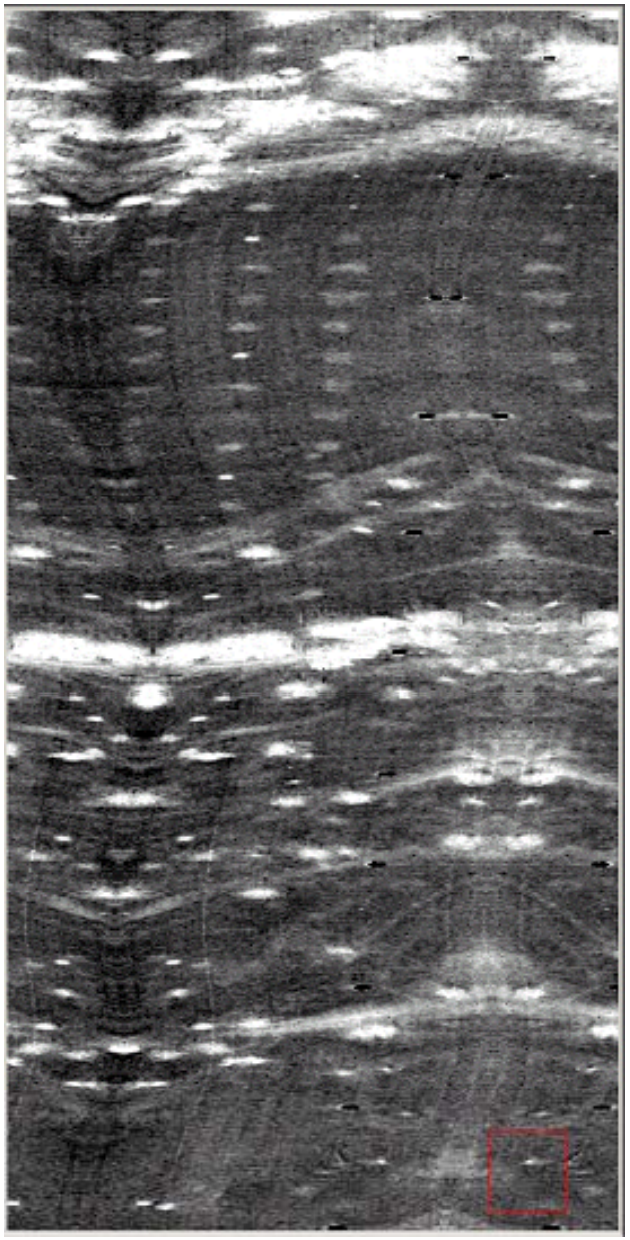
- Skewness of Emissivity at Mine/Background pixels
- Statistics calculated on wavelengths 7.88 – 9.92 microns of emissivity image

$$Skewness = \frac{1}{N}\sum_{j=1}^{N}\left(\frac{x_j - \bar{x}}{\sqrt{Variance}}\right)^3$$



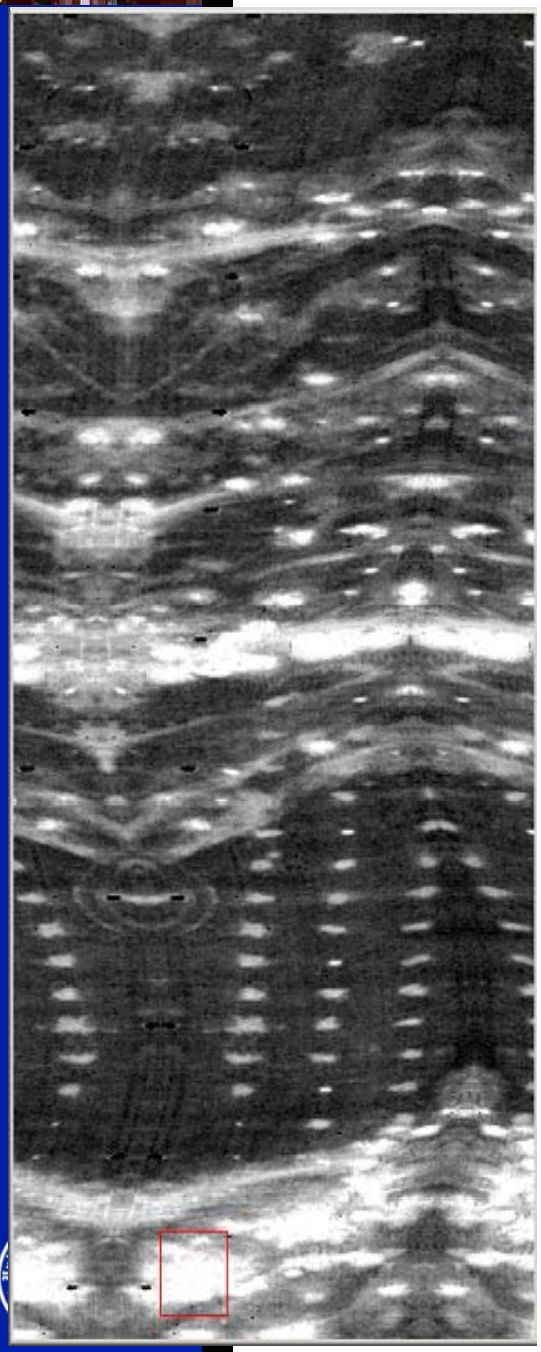
Regions of Interest – Different Image



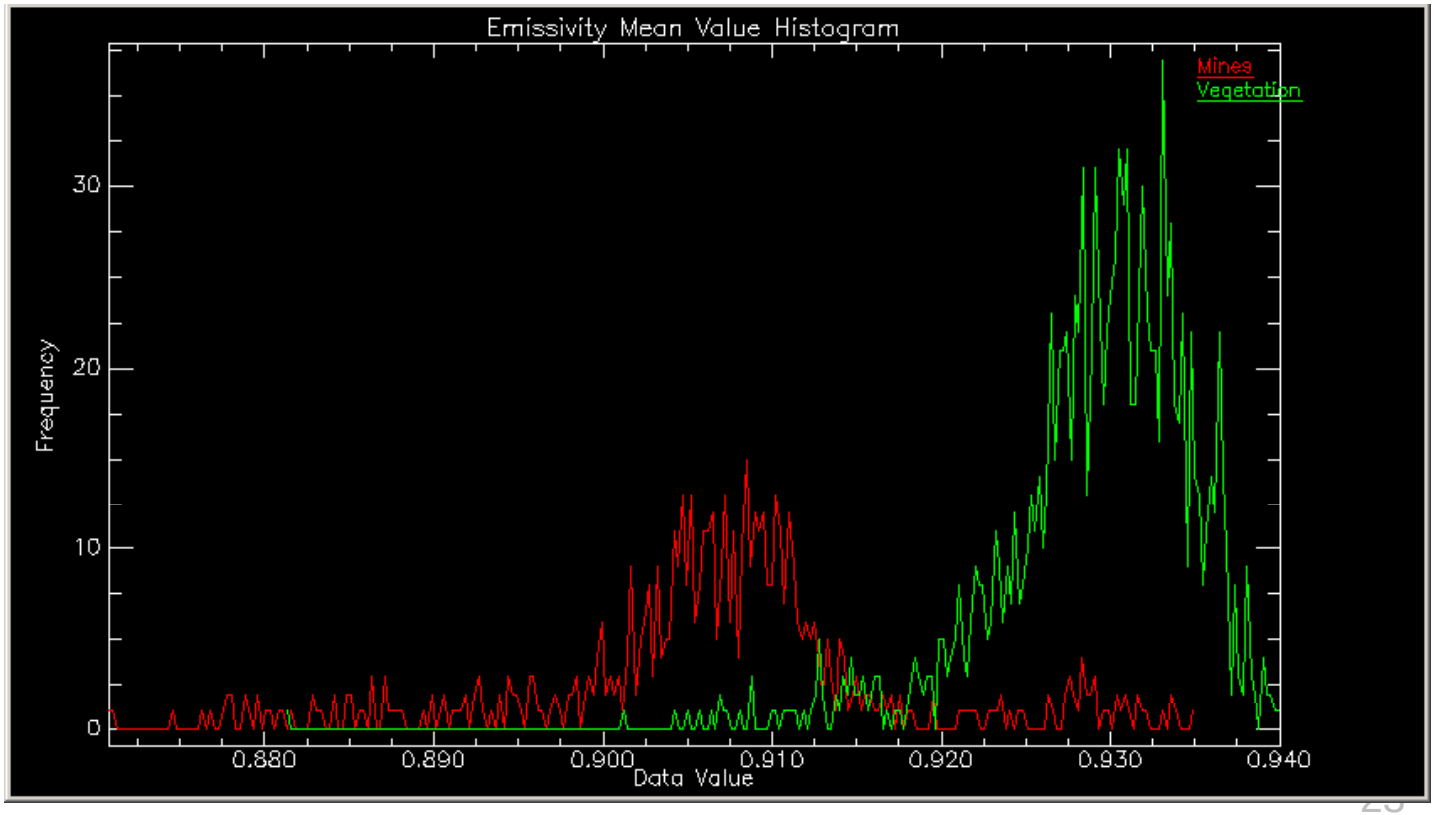


- Ground-truth selected mine ROIs
- Hand-selected vegetation ROIs
- Emissivity at 8.9 microns
  cm_am_atlt700_s300_r1_20030402_085640_r_bin3_s10e220.img
- Red: Mine (592 points)
- Green: Vegetation (1276 points)

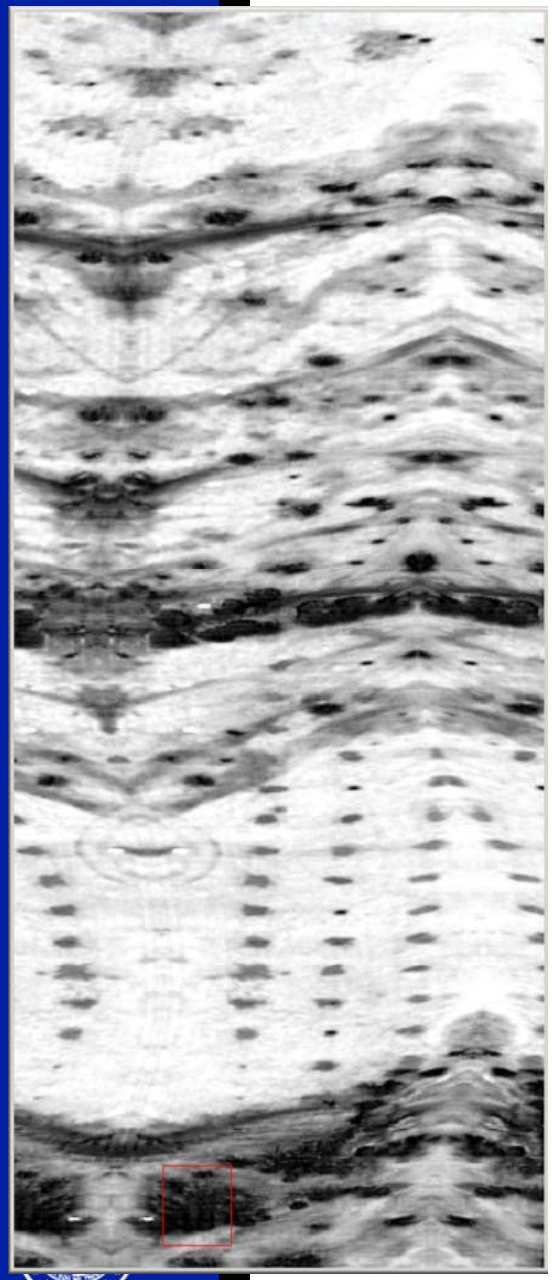
Mean Statistics



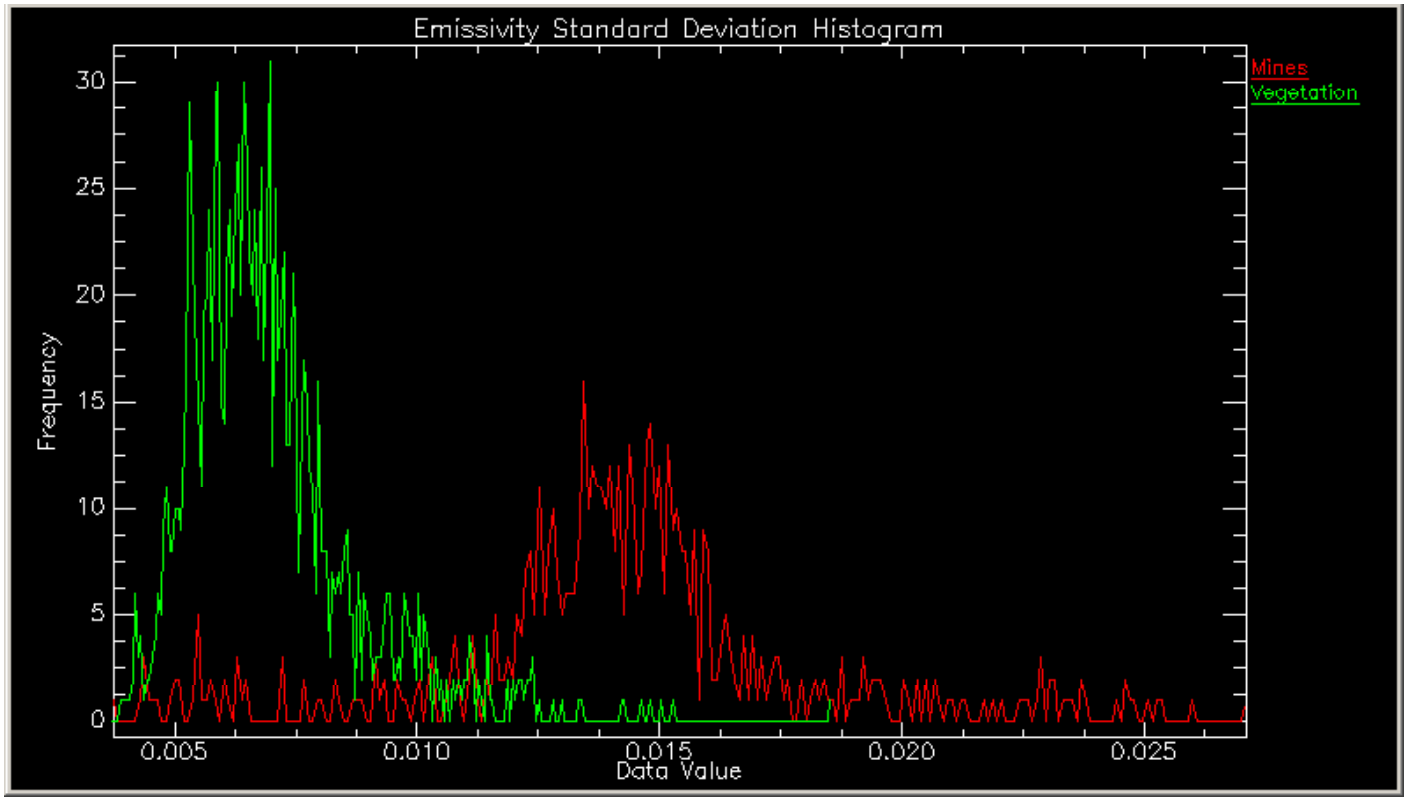
- Mean Emissivity at Mine/Background pixels
- Statistics calculated on wavelengths 7.88 – 10.5 microns of emissivity image



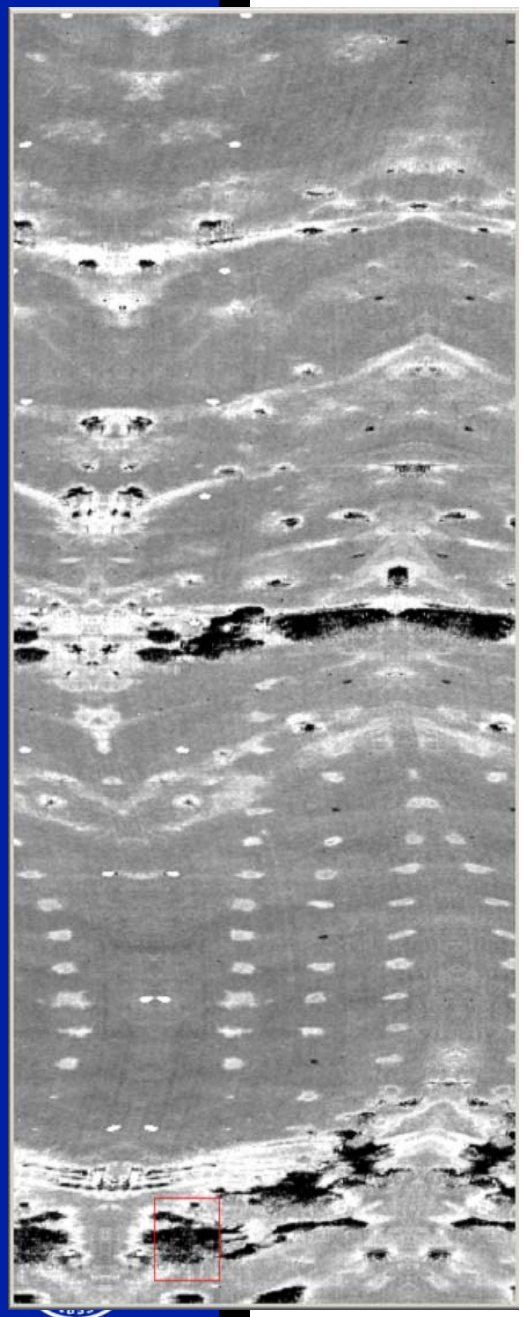
Standard Deviation Statistics



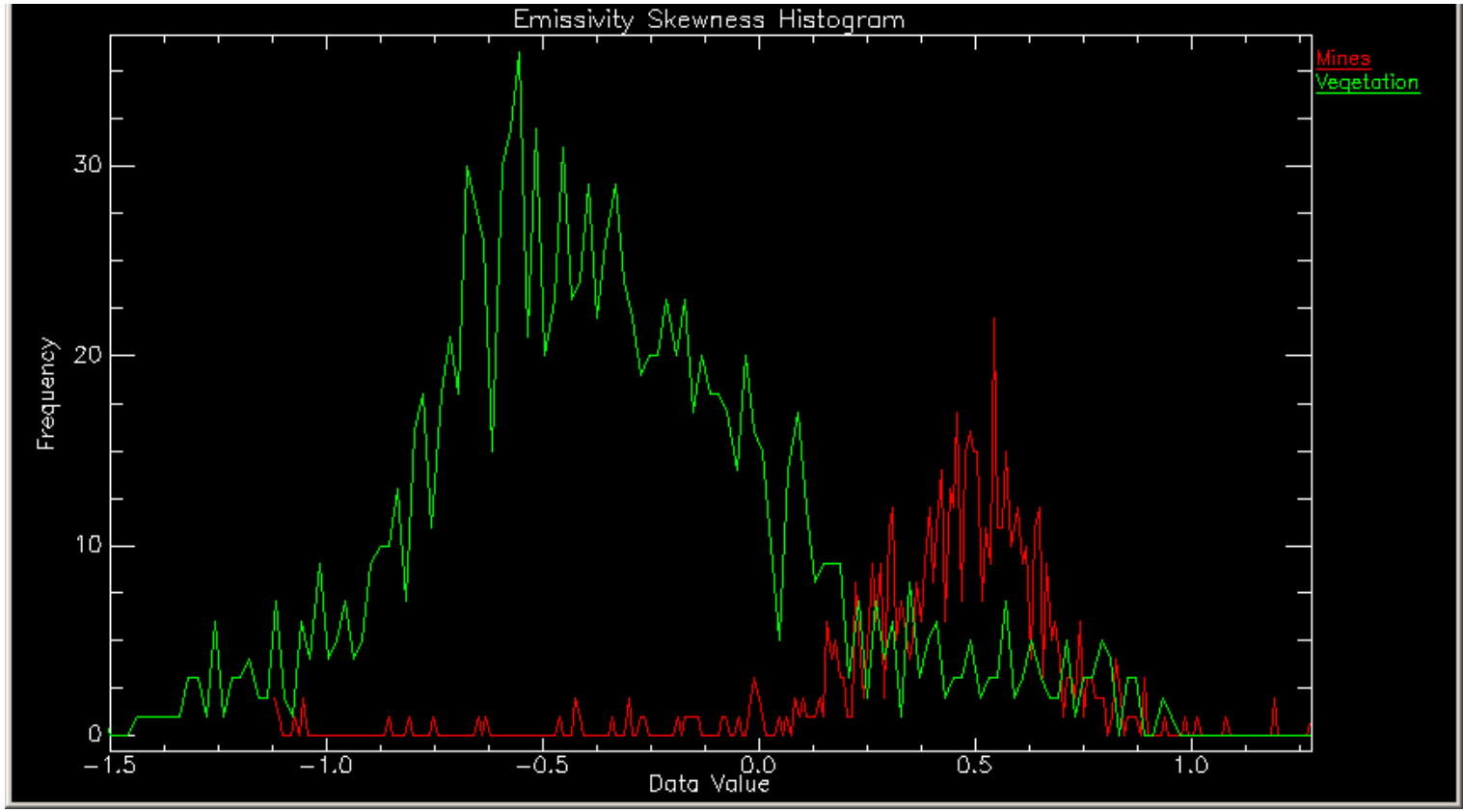
- Standard Deviation of Emissivity at Mine/Background pixels
- Statistics calculated on wavelengths 7.88 – 10.5 microns of emissivity image



Skewness Statistics



- Skewness of Emissivity at Mine/Background pixels
- Statistics calculated on wavelengths 7.88 – 10.5 microns of emissivity image



University of Florida

Blackbody Feature Plane

- Select Cluster with Highest Mean Emissivity and Lowest Emissivity Standard Deviation

$$C_V = \begin{cases} C_i & \text{if } \arg\max_{C_i}\left(\mu_\mu^i\right) = \arg\min_{C_i}\left(\mu_\sigma^i\right) \\ \varnothing & \text{otherwise} \end{cases}$$

- Calculate Distance Measure to the Mean of the Blackbody Cluster

$$v(x,y) = \begin{cases} 0 & \text{if } C_v = \varnothing \\ \dfrac{1}{1+\sqrt{\left(f(x,y)-\mu_v\right)^t \Sigma^{-1}\left(f(x,y)-\mu_v\right)}} & \text{otherwise} \end{cases}$$

- A Local Max Filter is Applied and the Values are Flipped to Indicate the Absence of Blackbodies

$$V(x,y) = 1 - v^{\max}(x,y)$$

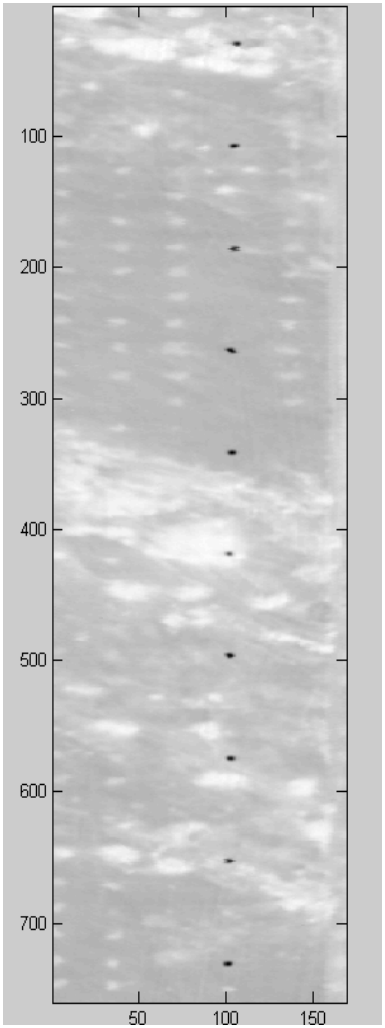
University of Florida

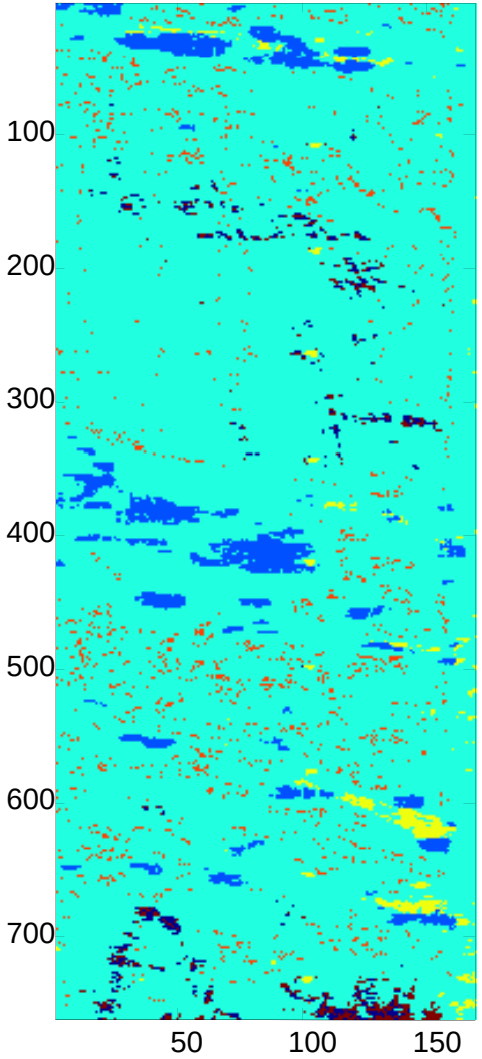
EM Clustering

- Emissivity Feature Vectors clustered to find pixels with blackbody characteristics

$$p_{\mathbf{f}} = \sum_k p_k G(\mathbf{f}(x,y); \boldsymbol{\mu}_k, \Sigma_k)$$

- 6 Clusters were used
  - The initial mean values are chosen from the image such that there are two means with small, two with medium, and two with large mean emissivity
  - All covariances for the Gaussians are initialized to the identity matrix





Cluster Membership Image for 2535

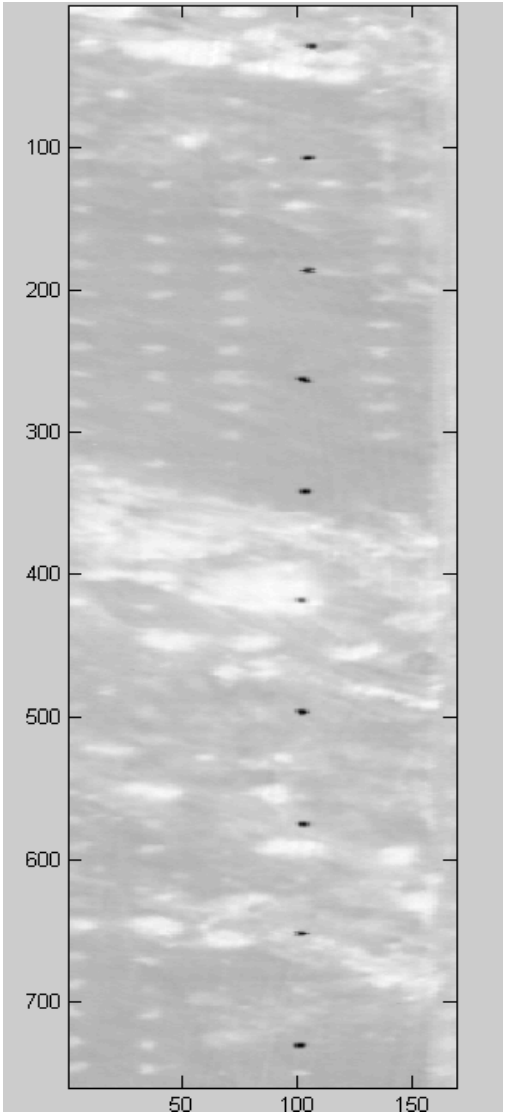
University of Florida

- **LWIR Blackbody Feature Plane**
- **Site:** Countermine, Yuma
- **Image:** 2535

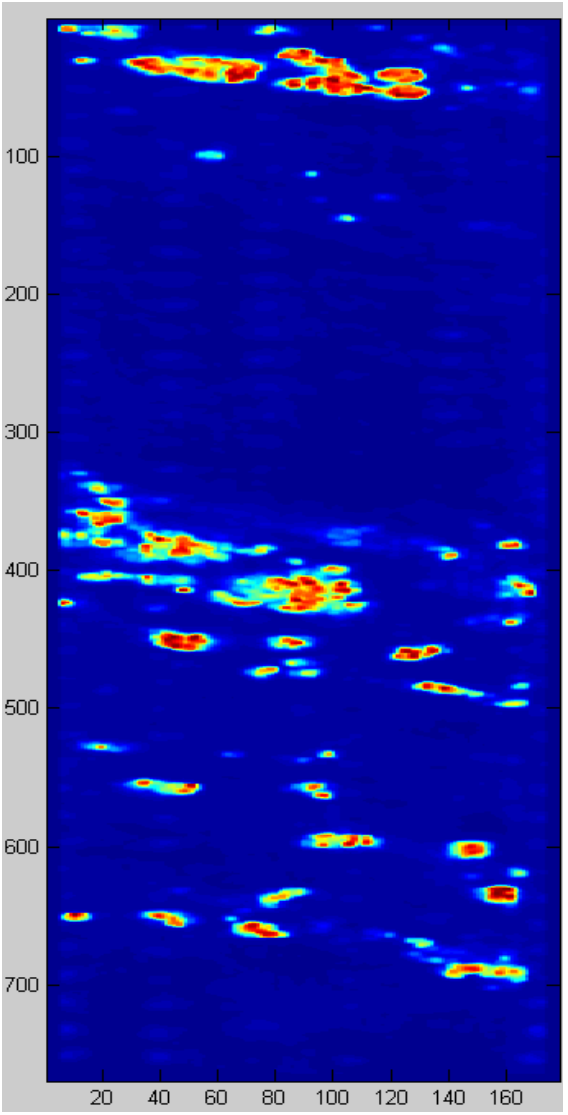
- This was found by clustering using the EM Algorithm. The cluster with the highest mean and lowest standard deviation was chosen. Membership was calculated on that cluster.

- **Feature Planes:**
  - **Band 1:** Mean of Emissivity
  - **Band 2:** Std Dev of Emissivity
  - **Band 3:** Skewness of Emissivity

- **EM Initialization:**
  - Pixels for EM were chosen by selecting the 1% of the pixels with the highest mean emissivity, 1% of the pixels with the lowest mean emissivity, and 1% of the pixels around the mean of the emissivity.
  - Means were initialized to those pixels with the highest and the lowest standard deviation of each group described above.

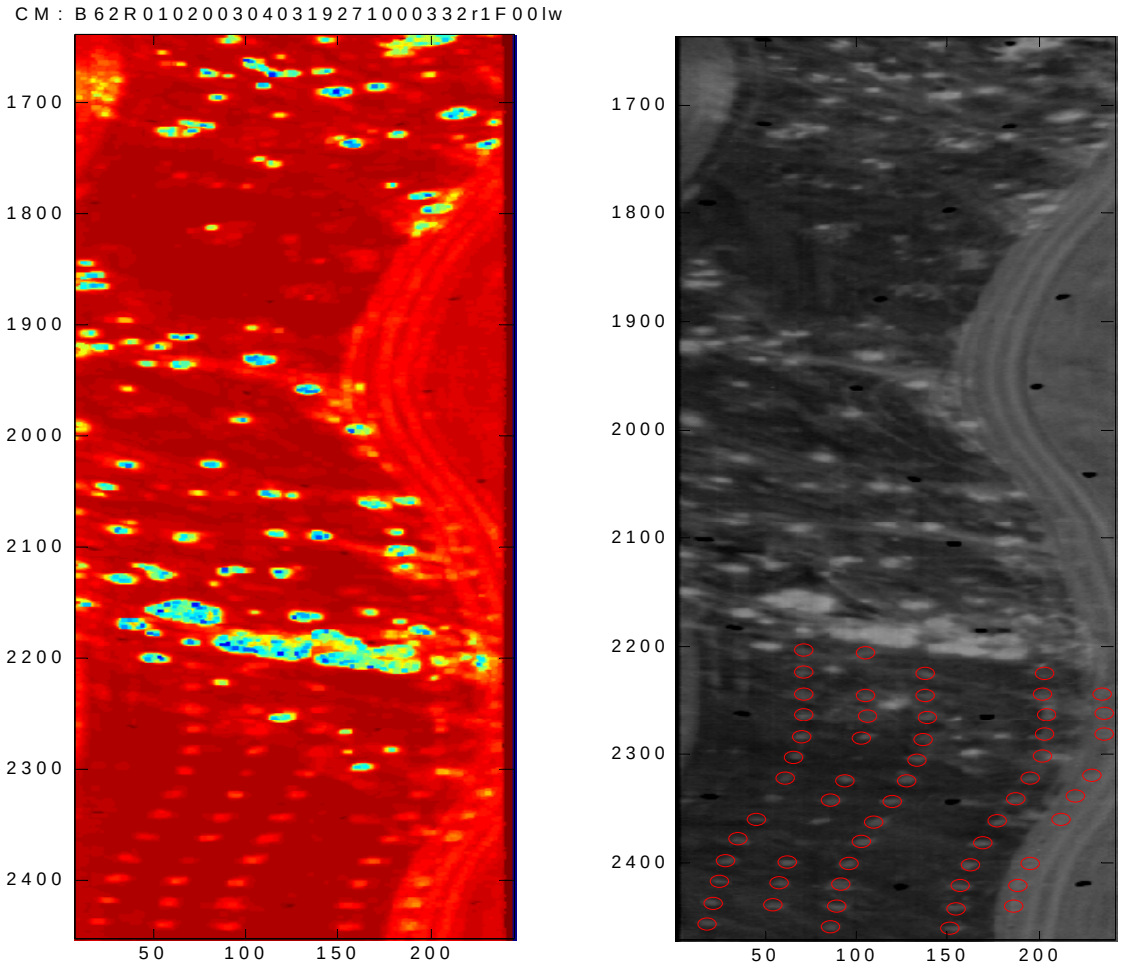


Original 2535



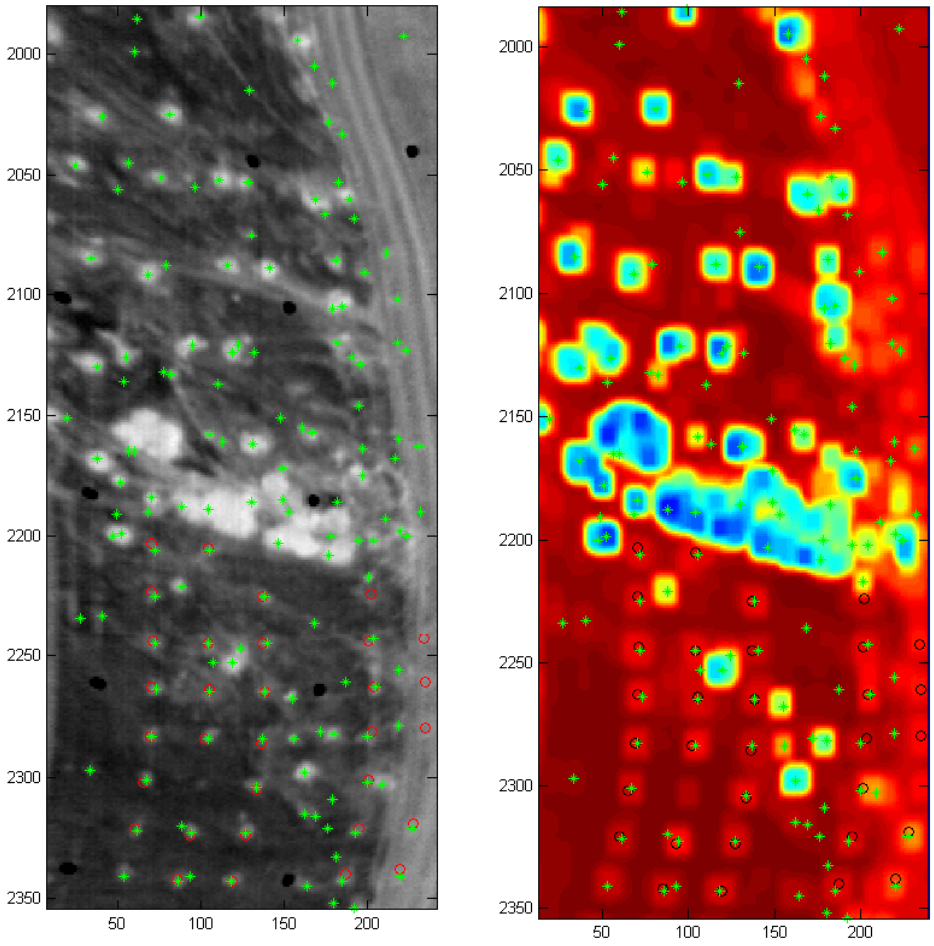
Distance Measure to BB Cluster

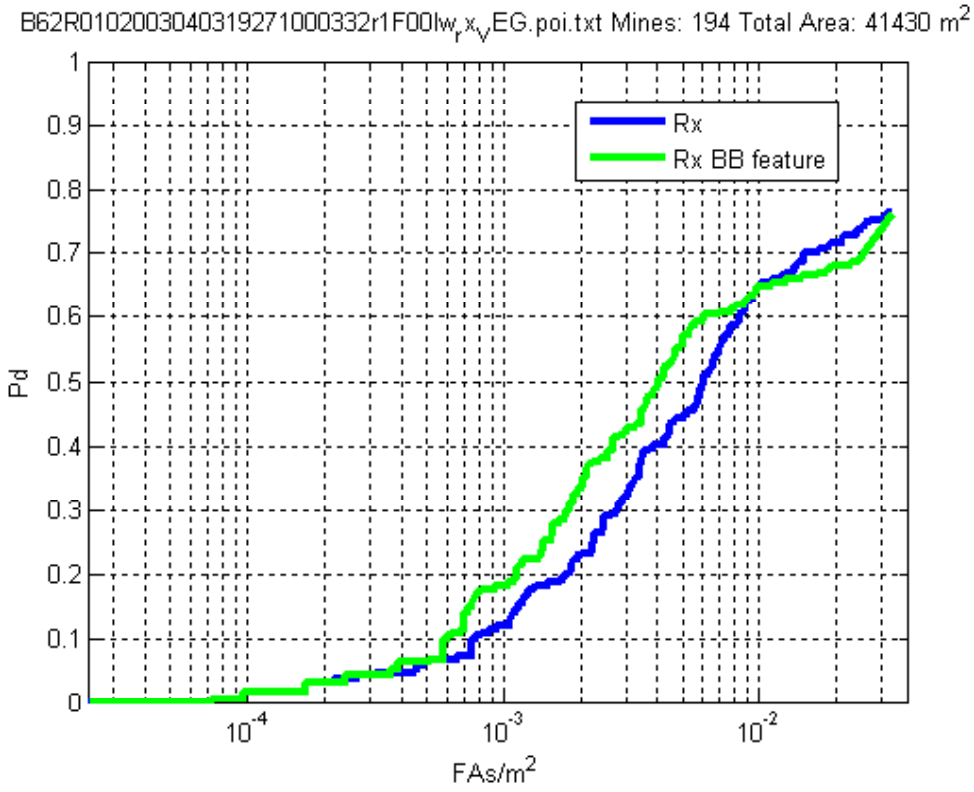
Countermine (CM)



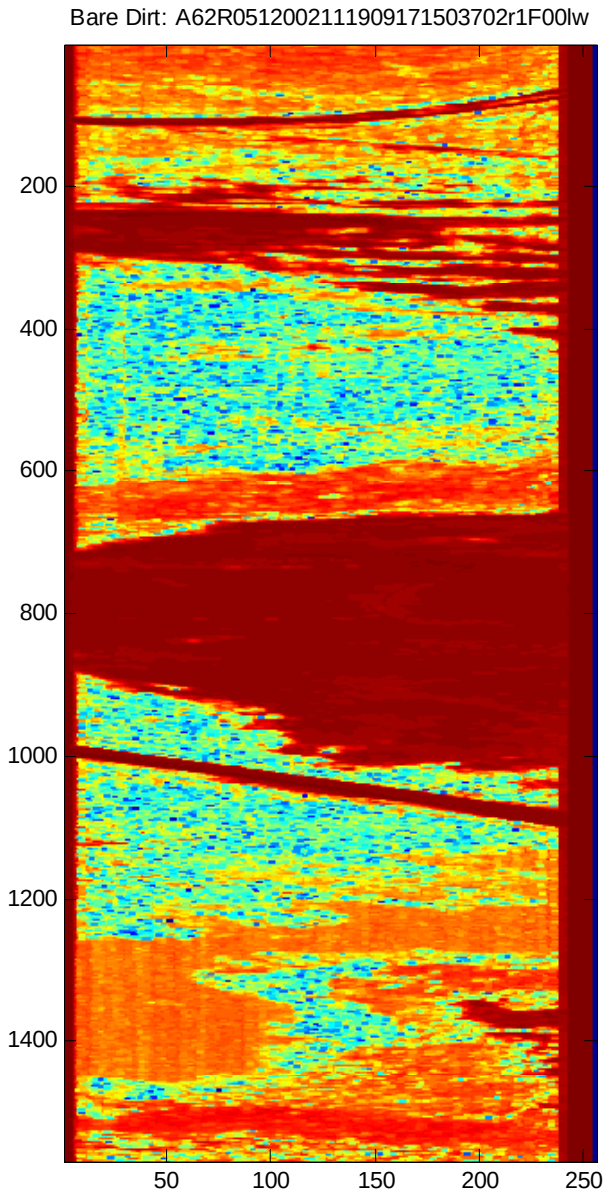
CM FA removal

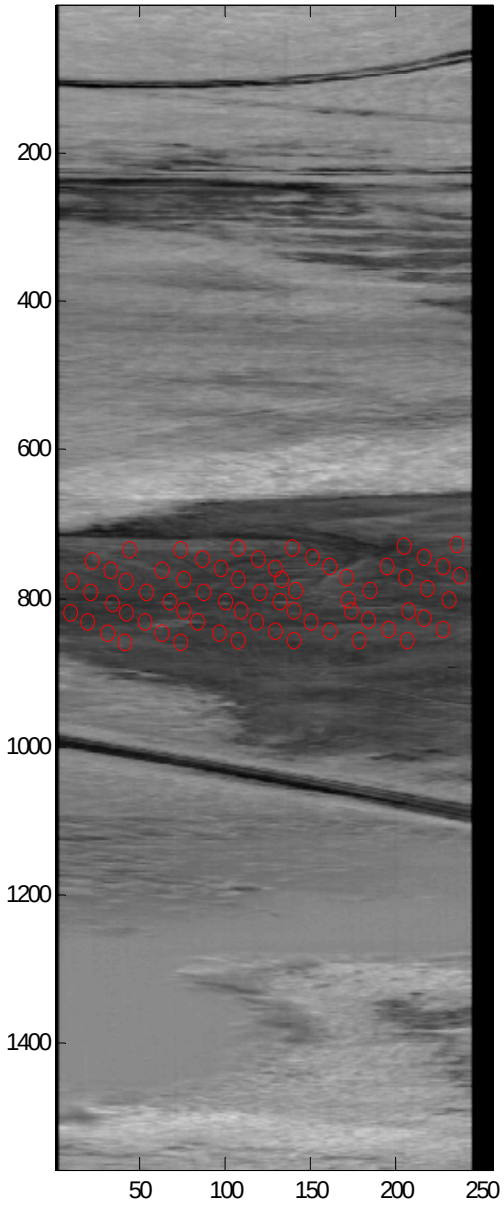
- B62R010200304031927100033 2r
- Site: Countermine, Yuma
- At 60% PD the FAR has been reduced 25%.





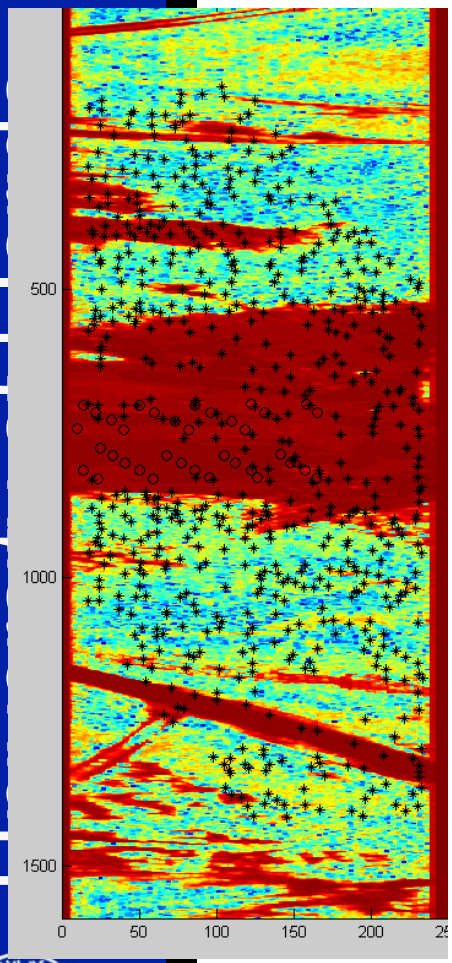
Forest Fusion 1: Bare dirt

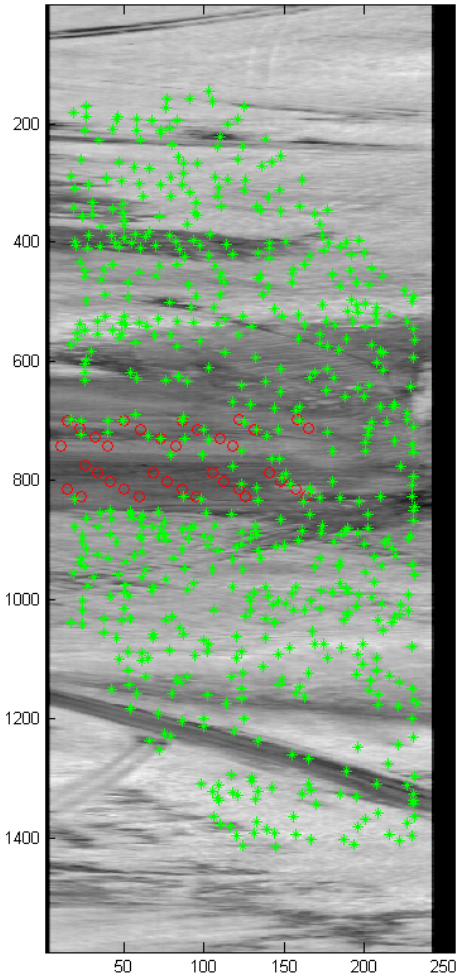


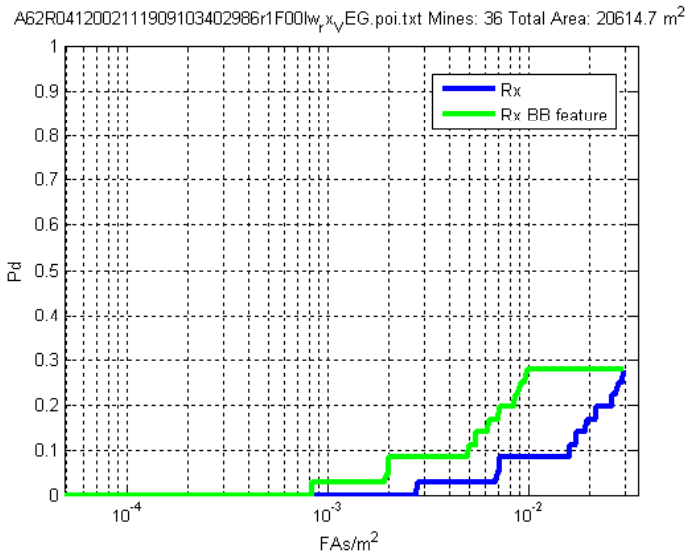


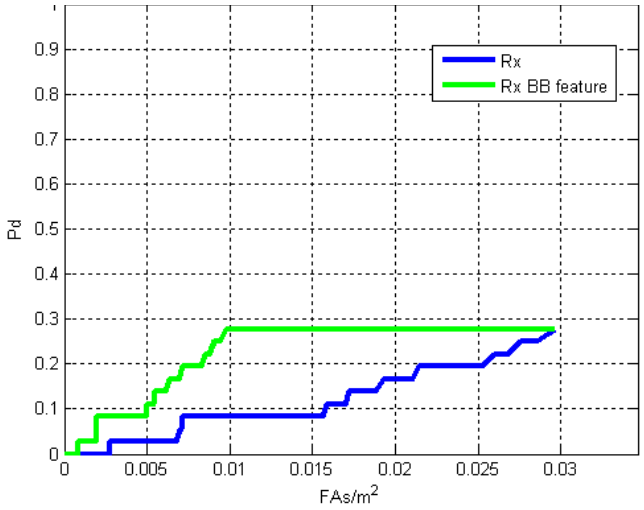
FF1 Bare Dirt FA Reduction

- A62R041200211190910340298 6r
- Site: Bare Dirt, FF1
- PD has tripled at a FAR of .01
- Has reached its top PD with 1/3 of the FAs

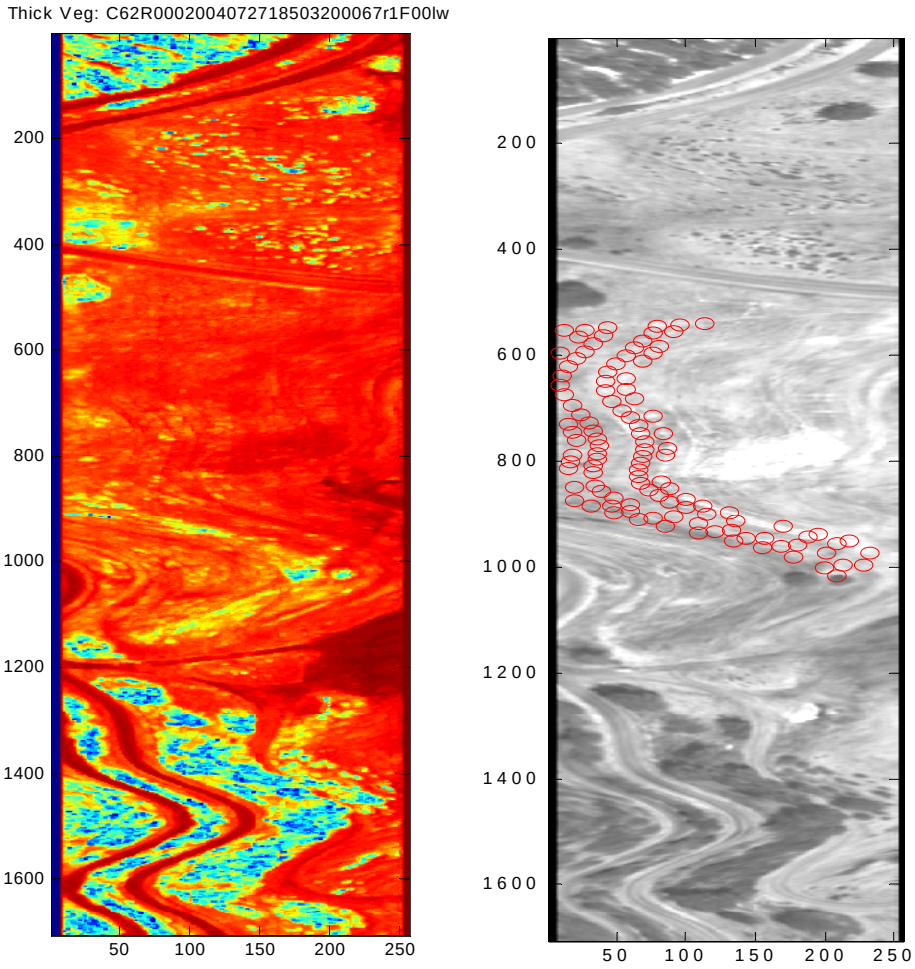






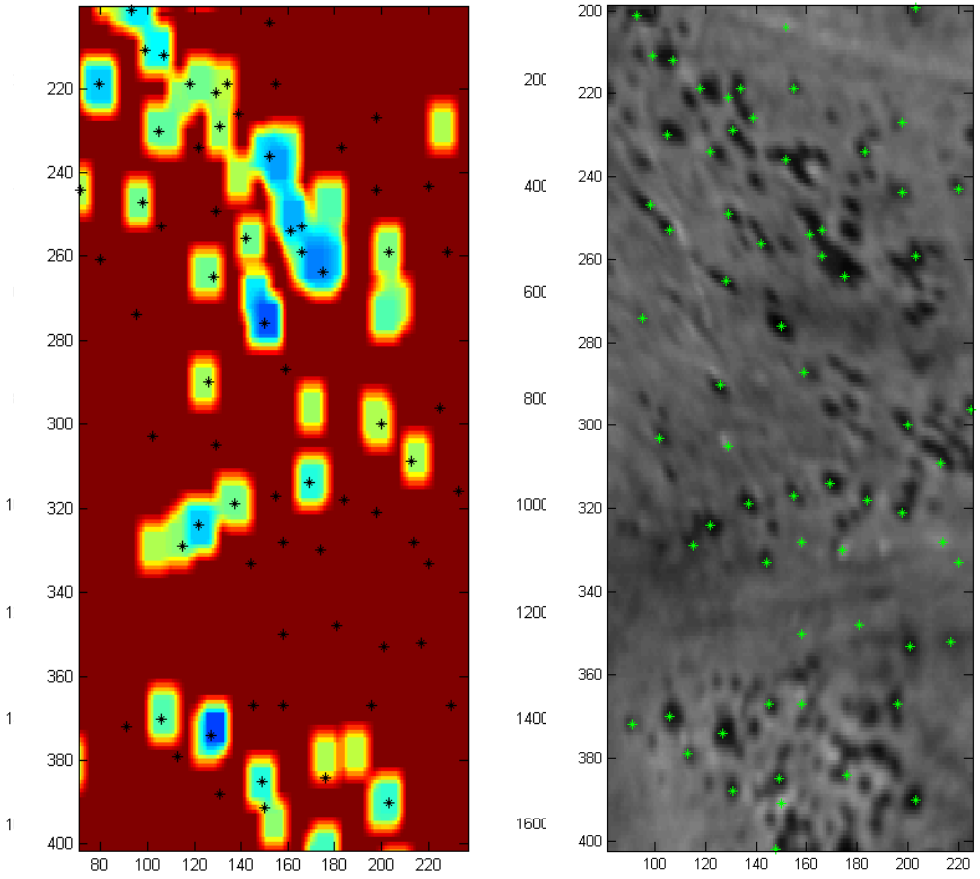


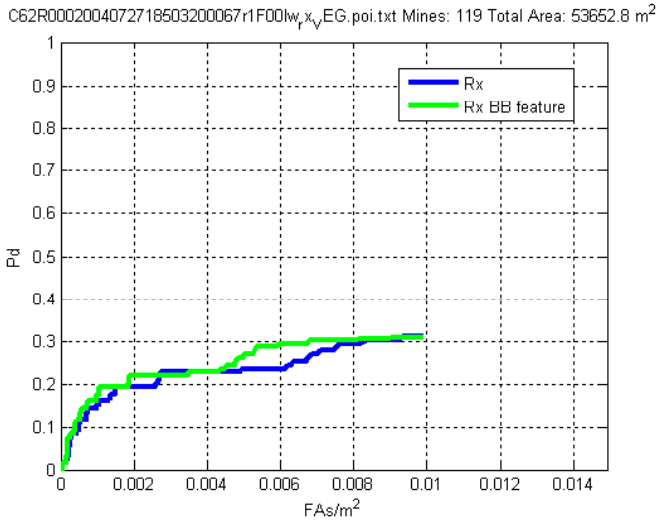
Forest Fusion 2: Thick Vegetation

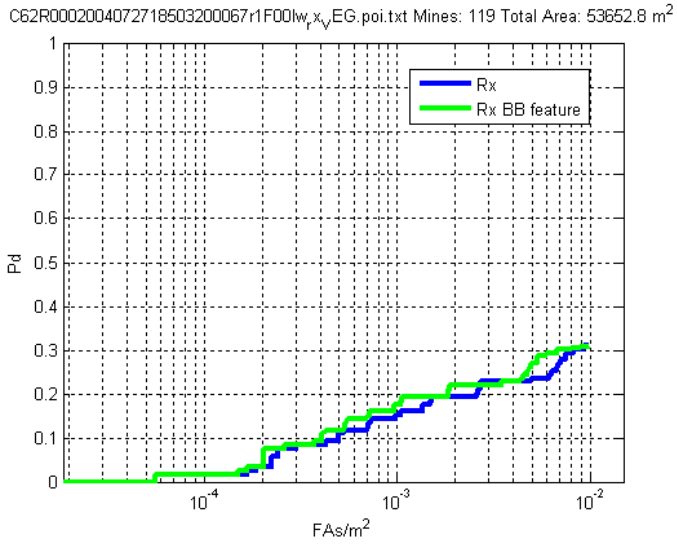


FA removal Thick Vegetation

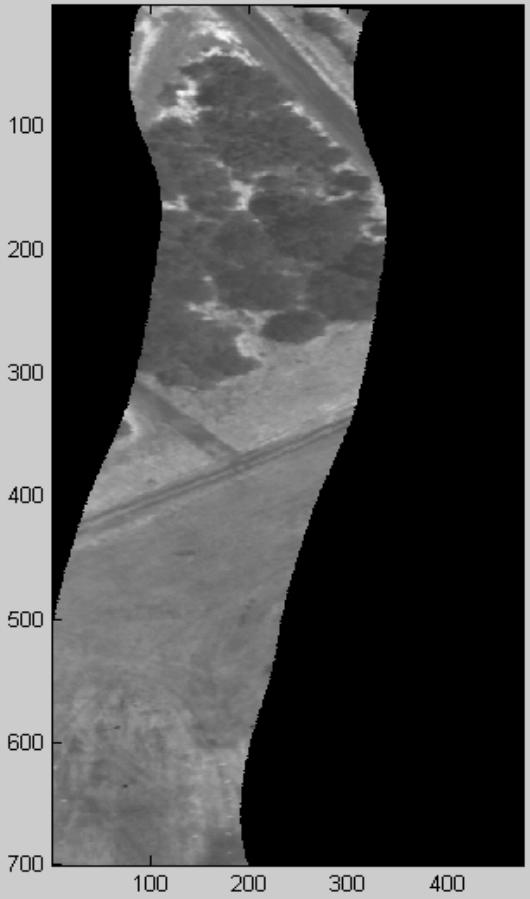
- C62R000200407271850320067r
- Site: Thick Veg, FF2
- Increased PD as much as 8 percentage points

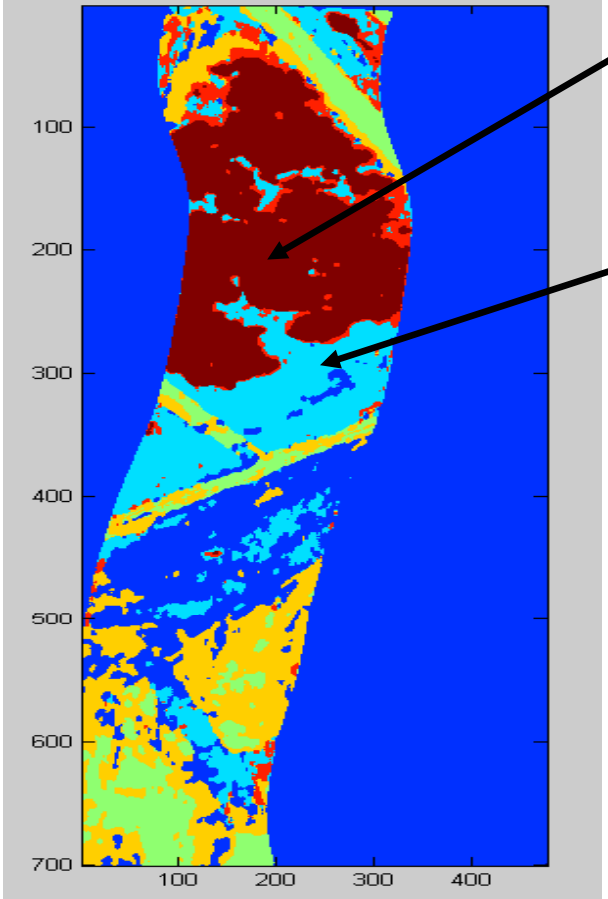






FF2 Cluster Results





Trees

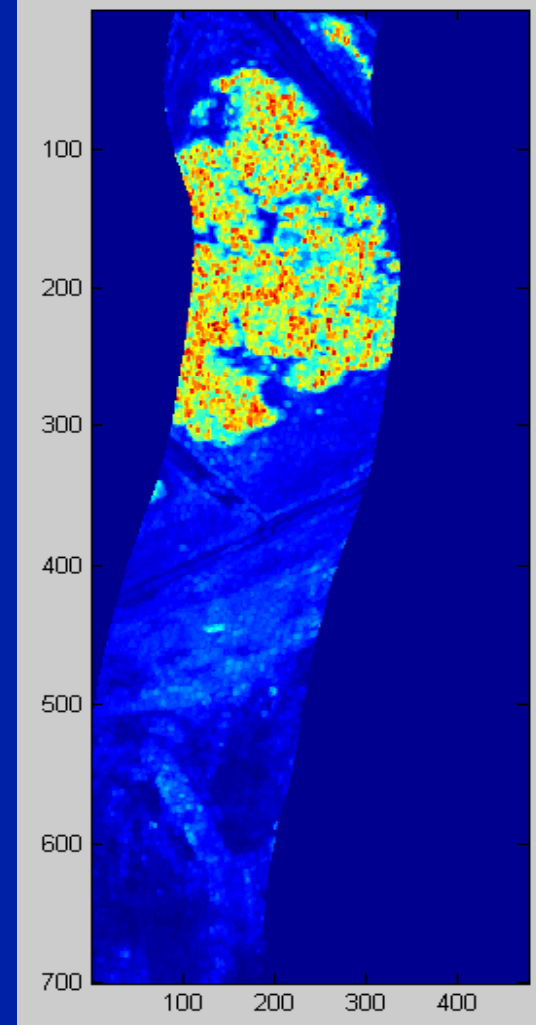
Grassy Area

■Mean	0.9357	0.9300	0.9281
■St dev	0.0144	0.0277	0.0256
■Skew	-1.9866	-2.5615	-2.3806
■			
■Mean	0.9164	0.9079	0.9259
■St dev	0.0257	0.0266	0.0253
■Skew	-1.3136	-0.1751	-2.1814

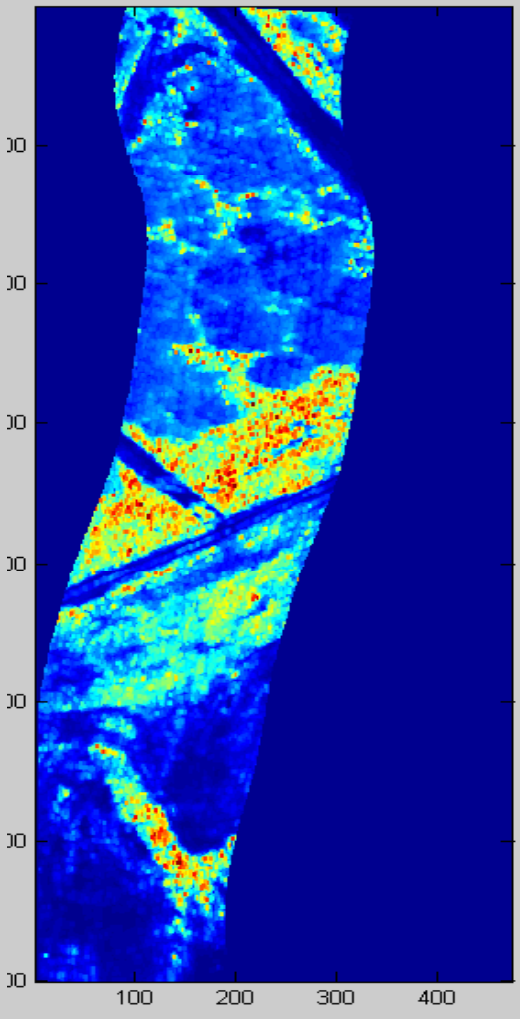
AHI Image:

C62R00020040726180343000001r1F00lw

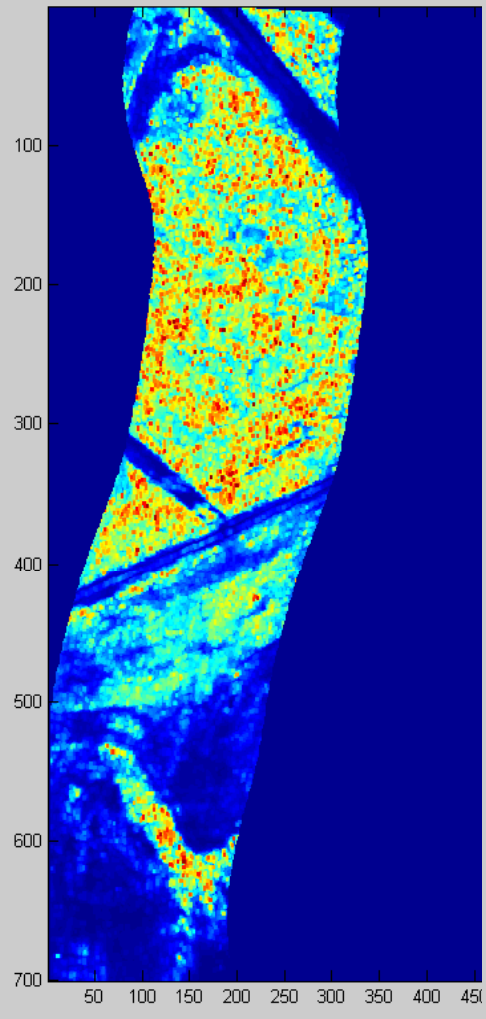
Use Distance to Closest Vegetation Cluster



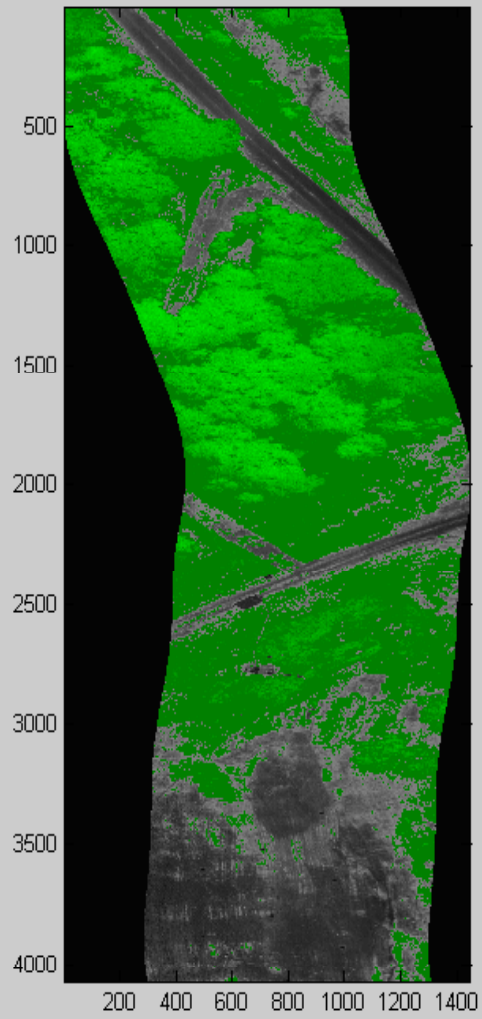
Distance to Tree Mean



Distance to Grass Mean

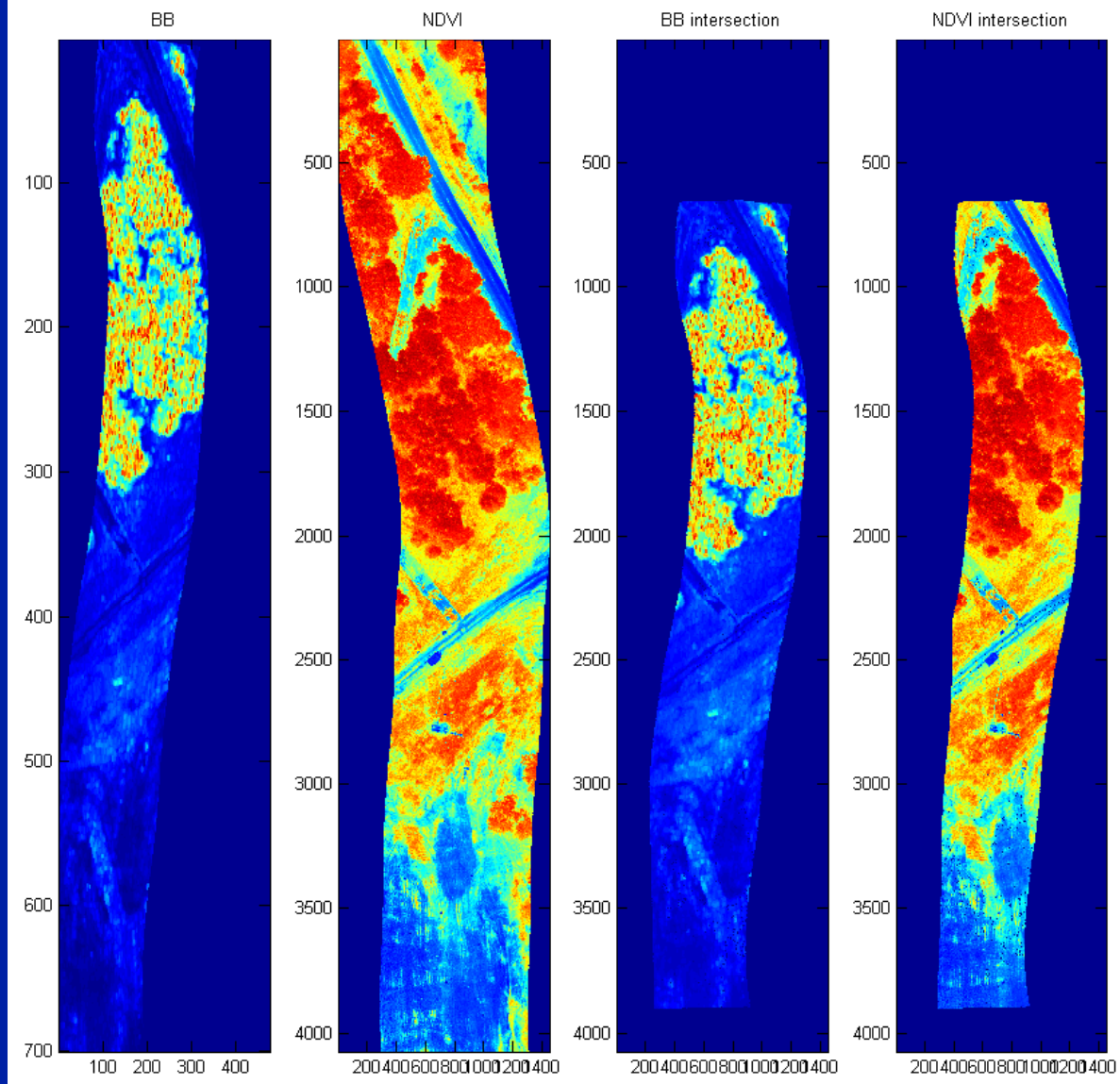


Min of the two distances



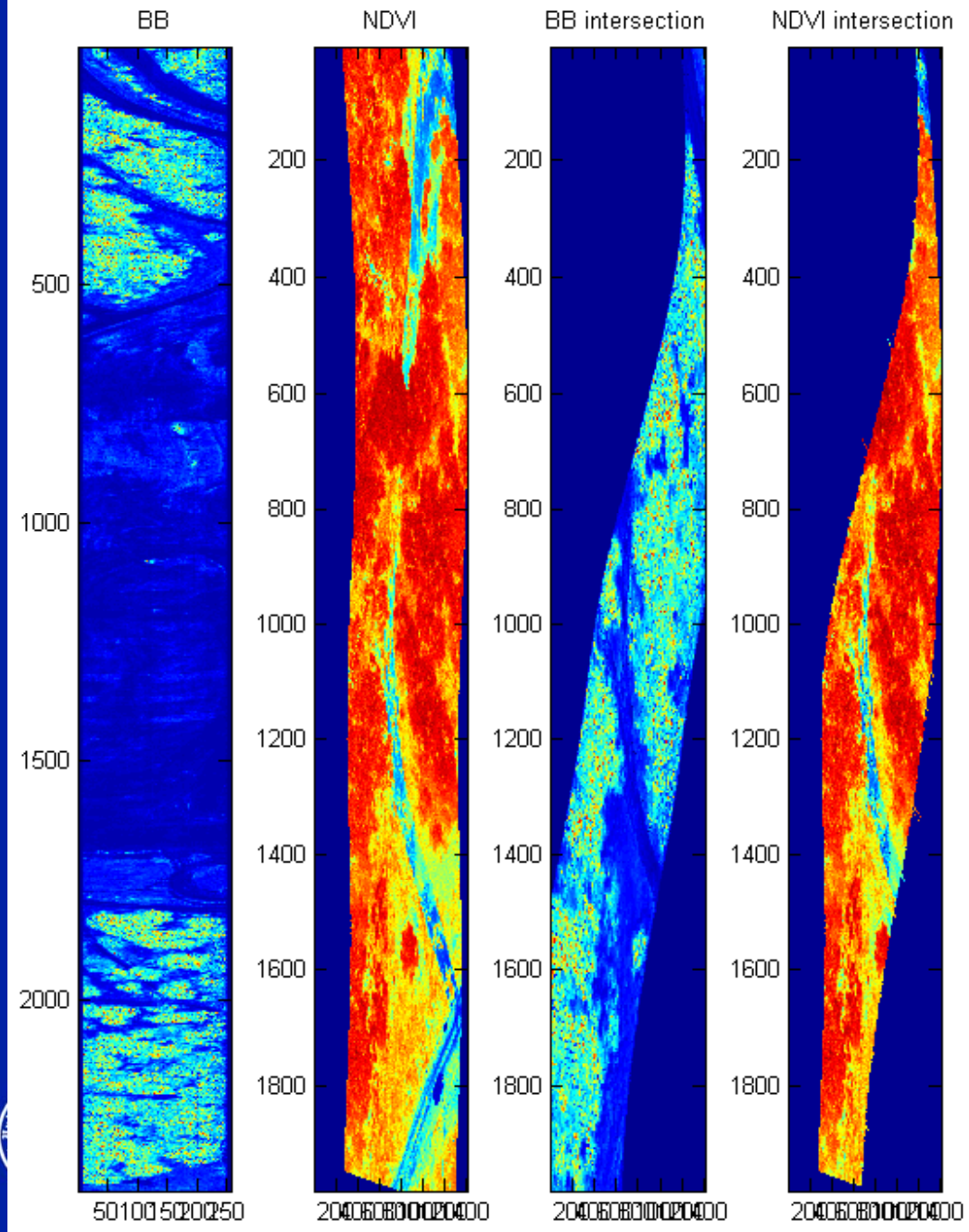
NDVI

Comparison with NDVI



- VSNIR allows for vegetation detection during the day
  - NDVI – gold standard
- LWIR: BB feature allows for vegetation detection at night
  - Use NDVI as ground truth for evaluation of LWIR derived BB feature
- Calculated NDVI from Wasabi Imagery
  - Wasabi is a 3-band line scanner
  - Bands at Red, Green and NIR
  - NIR-Red / NIR+Red

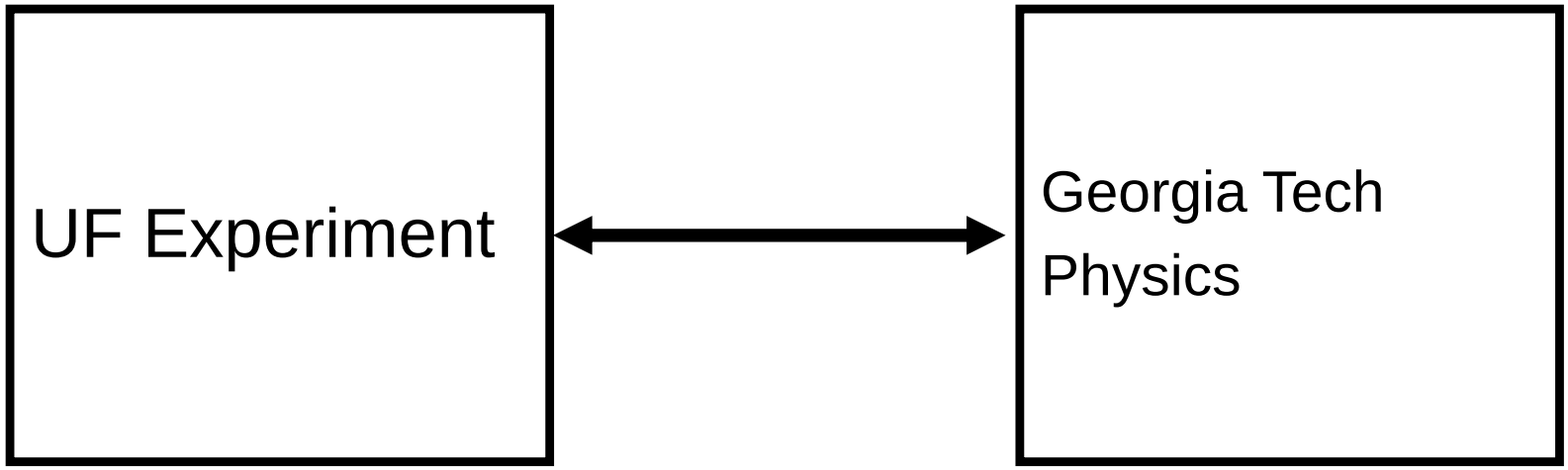
Comparison with NDVI



- VSNIR allows for vegetation detection during the day
  - NDVI – gold standard
- LWIR: BB feature allows for vegetation detection at night
  - Use NDVI as ground truth for evaluation of LWIR derived BB feature
- Calculated NDVI from Wasabi Imagery
  - Wasabi is a 3-band line scanner
  - Bands at Red, Green and NIR
  - NIR-Red / NIR+Red

University of Florida

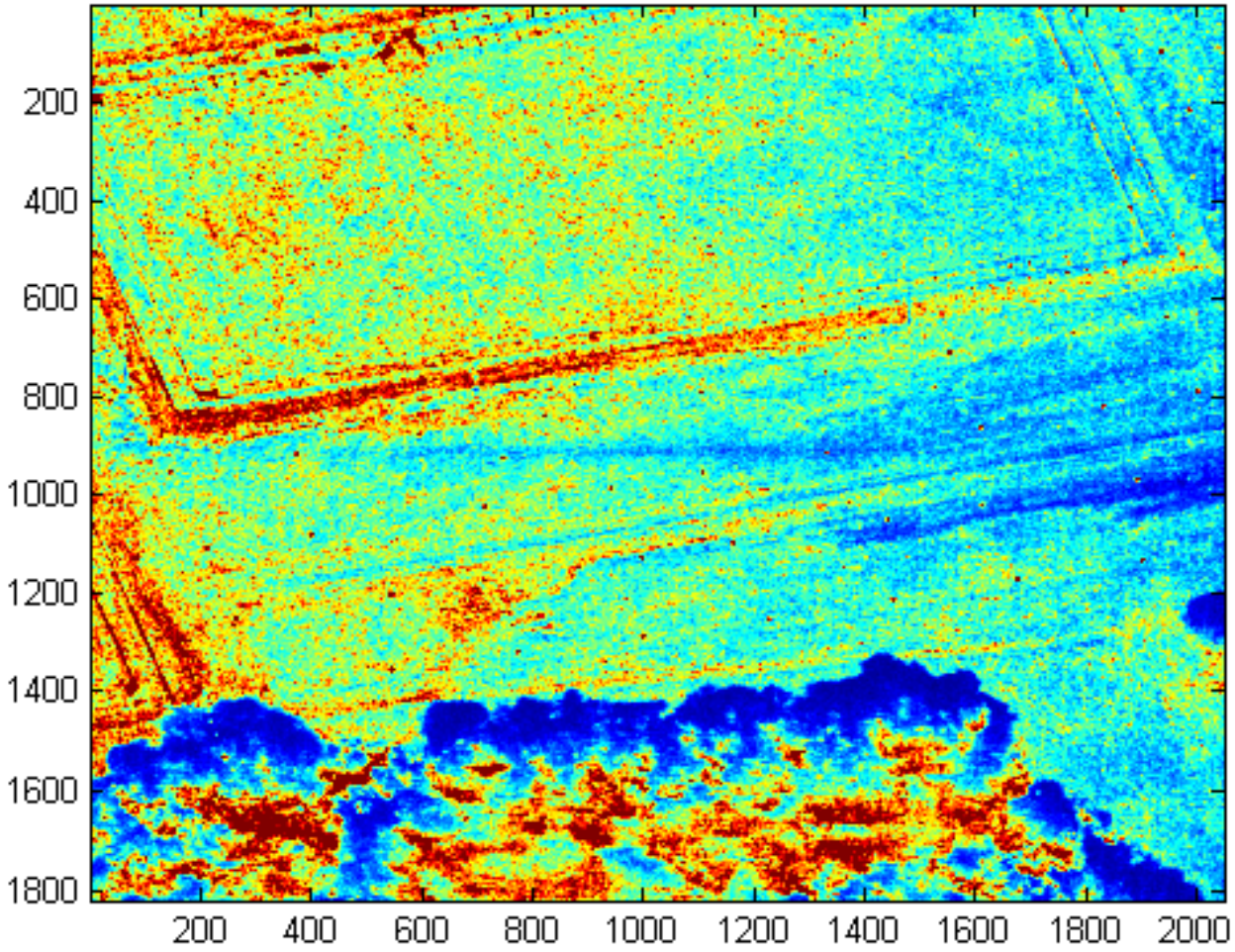
Interplay Between Experiment and Theory



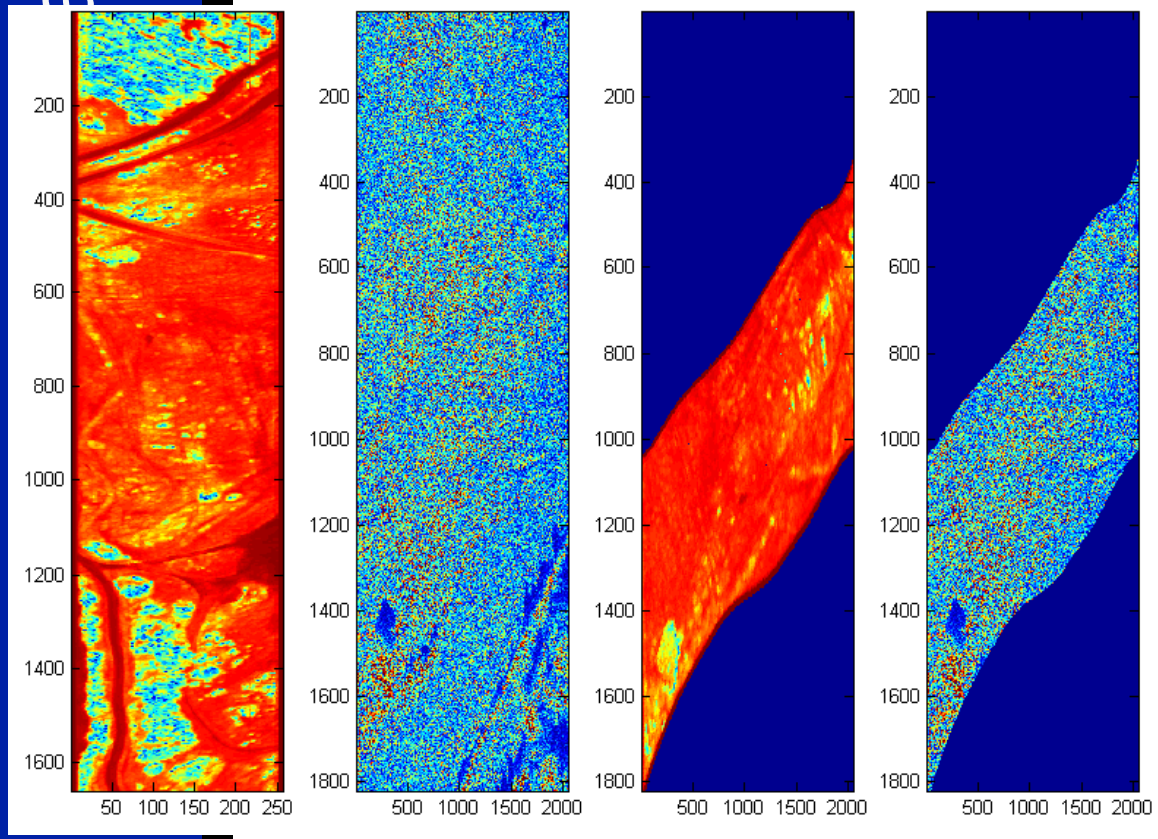
- ***Met with Georgia Tech***
- ***Discussed Emissivity Statistics***
- ***Georgia Tech investigated physical basis***

University of Florida

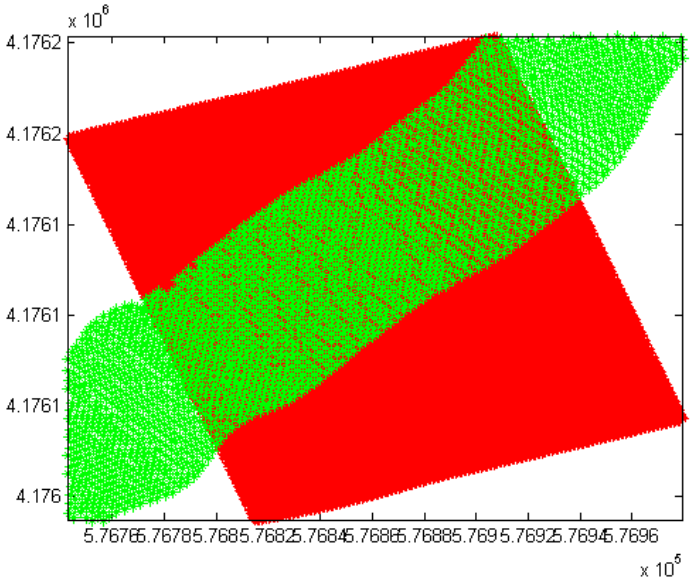
Lynx Imagery – Texture Features



Incorporation of SAR Texture Features w/ BB Feature

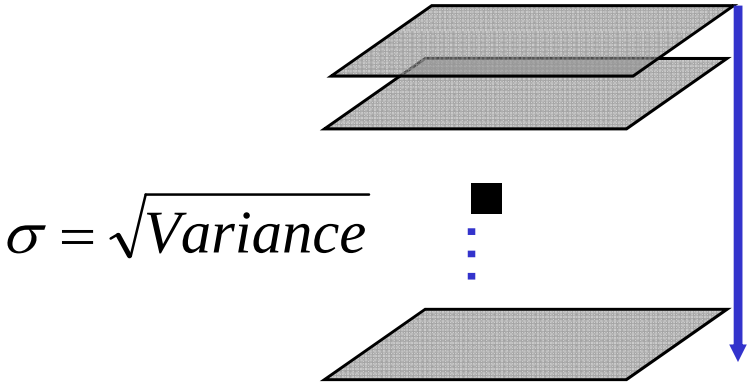


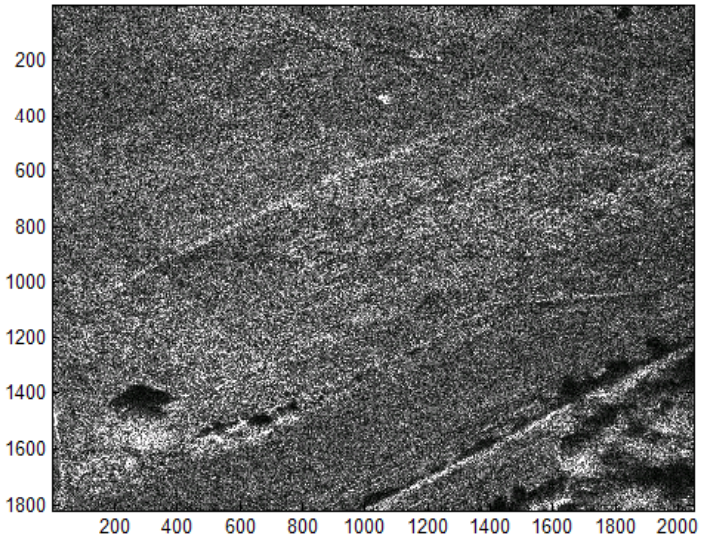
- Left:
  - BB feature image (AHI)
  - Lynx image (Lynx)
  - The previous images mapped onto the same grid
- Below:
  - Shows the overlap area of the Lynx and AHI images

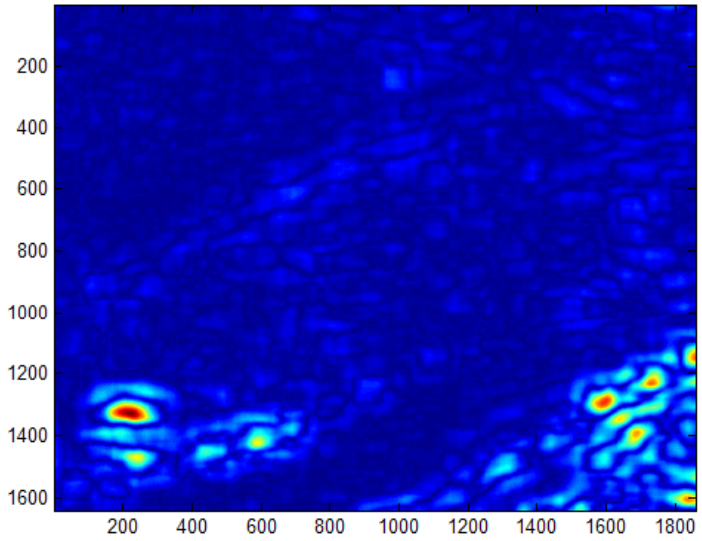


Texture Features: Discriminating between Grass and Trees

- "Growing Variance" texture feature
- Computed by taking the variance in a sliding window. This is repeated several times with an increasing window size.
- The standard deviation is calculated across the variance images







University of Florida

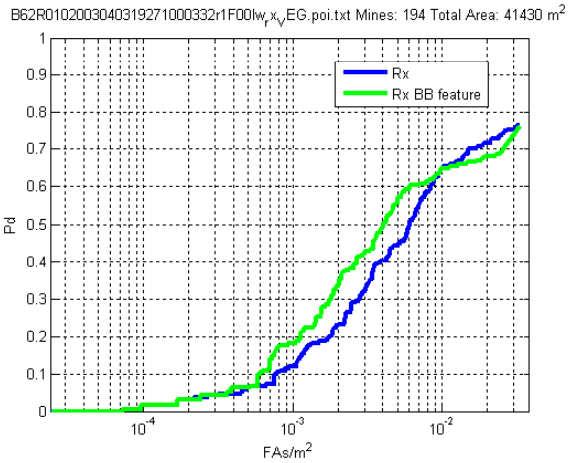
University of Florida

Vegetation Detection in LWIR

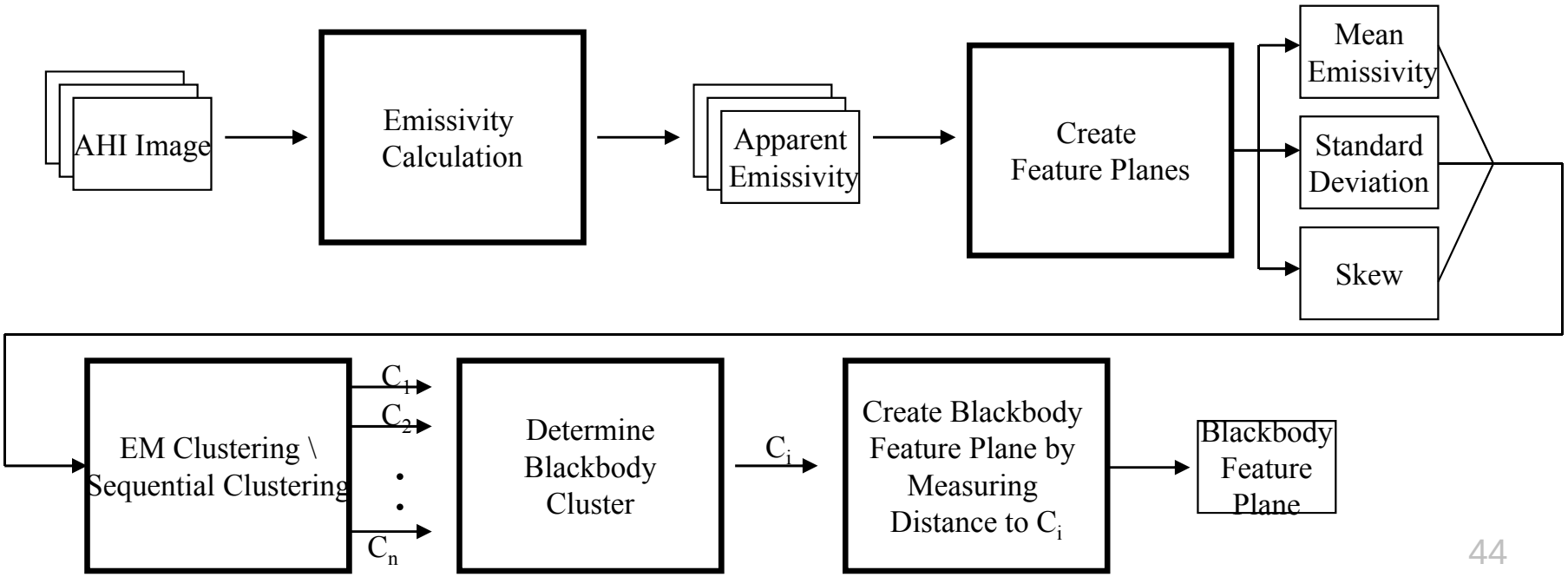
Using Sparsity Promoting Priors to determine Number of Endmembers

Vegetation Detection in the LWIR

- Goal – Reduce False Alarm Rate
- Method - Detect Vegetation in LWIR
- Foundation
  - Vegetation a blackbody over LWIR
  - High emissivity over all LWIR wavelengths
    - High mean, low standard deviation
- Using emissivity normalization method currently
- Statistics of apparent emissivity used as features



- B62R010200304031927100033 2r
- **Site**: Countermine, Yuma
- At 60% PD the FAR has been reduced 25%.



University of Florida

University of Florida

SPICE: Autonomous Endmember Detection with Sparsity Promotion

- Endmember Detection
  - Last meeting discussed LWIR Vegetation Detection for FA reduction
  - Exploring endmember detection algorithms to find vegetation endmembers
  - Determine number of endmembers required for a scene

- ICE Algorithm: Iterated Constrained Endmembers
  - M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Donne and J. F. Huntington, "ICE: A Statistical Approach to Identifying Endmembers in Hyperspectral Images," *IEEE Trans. On Geoscience and Remote Sensing*, vol. 42, Oct. 2004, pp. 2085–2095.

- Goal of ICE: To find the endmembers, this algorithm performs a least squares (LS) minimization based on the Convex Geometry Model.

University of Florida

ICE Algorithm : Comparison to N-FINDR

- Winter's N-FINDR Algorithm:
  - The algorithm performs a transformation to an M-1-dimensional subspace and finds a *maximal* simplex constrained to a subset of the scene.
  - Requires pure pixel representation of endmembers in the scene
  - Both ICE and N-FINDR require knowledge of the number of endmembers

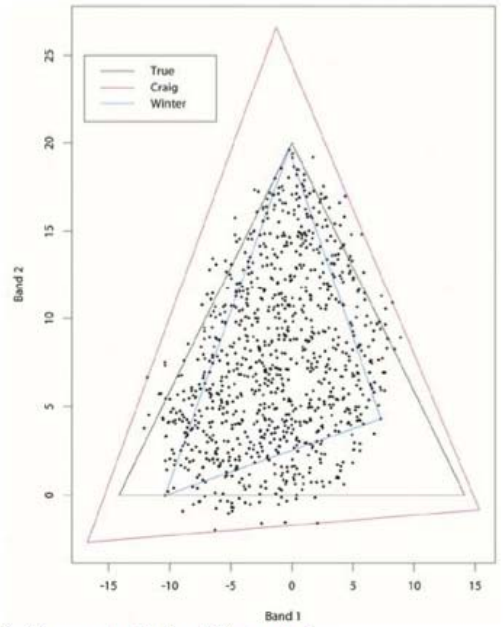


Fig. 3. Toy example. Craig and Winter solutions.

M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Donne and J. F. Huntington, "ICE: A Statistical Approach to Identifying Endmembers in Hyperspectral Images," *IEEE Trans. On Geoscience and Remote Sensing*, vol. 42, Oct. 2004, pp. 2085–2095.

University of Florida

ICE Algorithm

- Goal of ICE: To find the endmembers, this algorithm performs a least squares (LS) minimization based on the Convex Geometry Model.

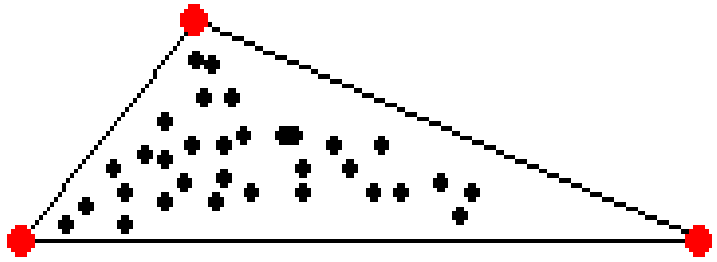
- Convex Geometry Model:
  - Assumes that the spectral response in each pixel is a linear combination of endmember spectra, with the weights being proportions.

$$\mathbf{X}_i = \sum_{k=1}^{M} p_{ik}\mathbf{E}_k + \varepsilon_i, \quad i = 1,\ldots,N \qquad p_{ik} \geq 0, k = 1,\ldots,M, \quad \sum_{k=1}^{M} p_{ik} = 1$$

where N is the number of pixels, M is the number of endmembers, $\mathbf{E}_k$ is the kth endmember, ε_i is an error term, and p_{ik} are mixing proportions.

ICE Algorithm

- If the Convex Geometry Model holds, then the data lie inside a simplex in d-dimensional space, and the *M* endmembers form the vertices of this simplex.



- Choose the Endmembers (E_k) that minimize the Residual Sum of Squares (RSS)

$$RSS = \sum_{i=1}^{N} \left(X_i - \sum_{k=1}^{M} p_{ik} E_k \right)^T \left(X_i - \sum_{k=1}^{M} p_{ik} E_k \right)$$

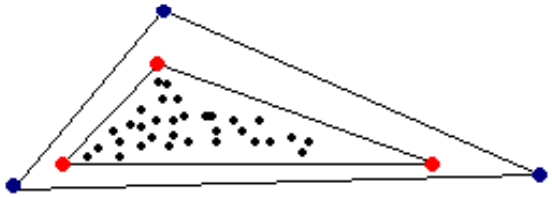
$$p_{ik} \geq 0, k = 1,\ldots,M, \quad \sum_{k=1}^{M} p_{ik} = 1$$

University of Florida

ICE Algorithm

- Any *M*-simplex that totally encloses the data points minimizes RSS

$$RSS = \sum_{i=1}^{N}\left(X_i - \sum_{k=1}^{M} p_{ik}E_k\right)^T\left(X_i - \sum_{k=1}^{M} p_{ik}E_k\right)$$



- To constrain the size, add a term to the objective function that is proportional to the size of the simplex.
- Sum of squared distances between all the simplex vertices will be used.

$$SSD = \sum_{k=1}^{M-1}\sum_{l=k+1}^{M}\left(E_k - E_l\right)^T\left(E_k - E_l\right)$$

- SSD can also be written as:

$$SSD = M(M-1)V$$

- Where V is the sum of the variances of the coordinates of the endmembers

University of Florida

University of Florida

ICE Algorithm

- The ICE objective function to minimize is:

$$RSS_{reg} = (1-\mu)\frac{RSS}{N} + \mu V$$

- $\mu \in (0,1)$ is the regularization parameter.
- The objective function needs to be minimized over both the endmembers and the proportion values subjected to their constraints.
- This minimization can be performed using alternating optimization using the following two steps.
  - 1. Endmembers are held constant and P is estimated.
    - Quadratic Programming.
  - 2. Given P, the endmembers are estimated.

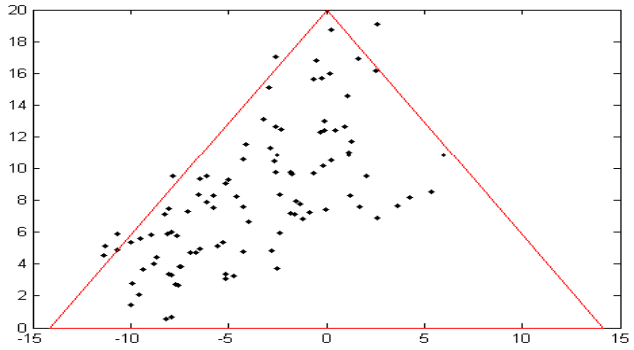
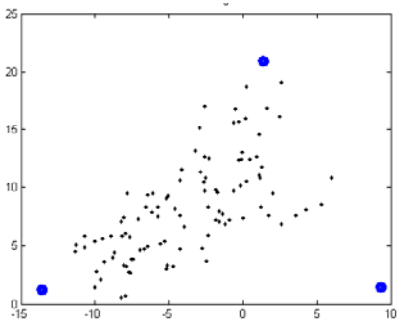
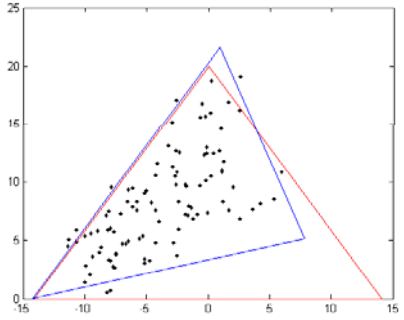
$$\mathbf{e}_j = \left\{ \mathbf{P}^T\mathbf{P} + \lambda\left(\mathbf{I}_M - \frac{11^T}{M} \right) \right\}^{-1} \mathbf{P}^T x_j$$

  - This is done iteratively until the RSS_{reg} value is below a tolerance threshold, *tol*.

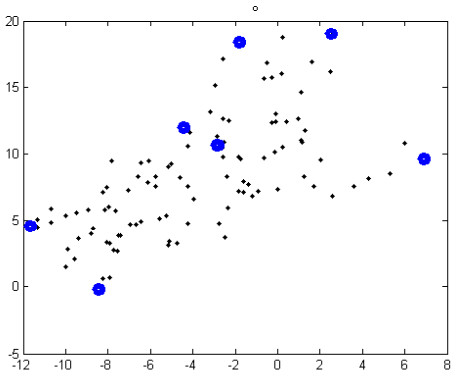
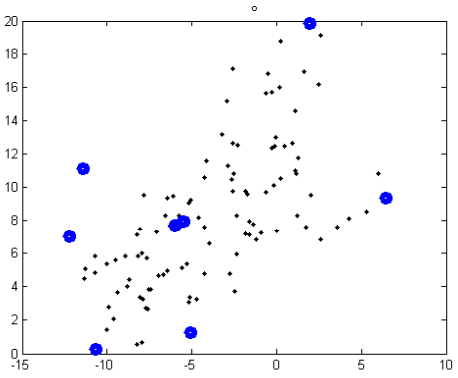
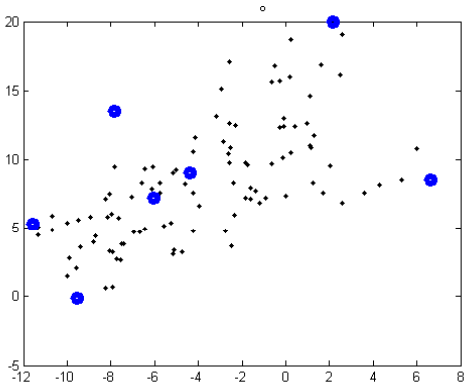
University of Florida

ICE Algorithm

- ICE does not have a method to autonomously determine the number of endmembers.
- The number of endmembers for a scene must be known before running ICE.
- “True” endmembers

- Using an incorrect number of endmembers:

University of Florida

ICE Algorithm

- Want a method to autonomously determine the number of endmembers required for a data set.
- Idea: Begin with a large number of endmembers and use sparsity promotion to encourage unnecessary endmembers to be pruned from the set.
- SPICE: Sparsity Promoting ICE

University of Florida

SPICE: Autonomous Endmember Detection with Sparsity Promotion

- RSS is Least Squares
- Generic least squares minimization objective function:

$$LS = \frac{1}{2}\sum_{i=1}^{N}\left(x_i - \boldsymbol{\beta}^{\mathbf{T}}\mathbf{y_i}\right)^2$$

- Find β and $\mathbf{y}$ that best approximate the data set, X.
- Minimizing LS equivalent to maximizing –LS.
- The LS objective function can be rewritten as

$$-LS = \ln e^{-\frac{1}{2}\sum_{i=1}^{N}\left(x_i - \boldsymbol{\beta}^{\mathbf{T}}\mathbf{y_i}\right)^2}$$

- We want to try and drive unnecessary coordinates of β values to zero.
  - For endmember detection, β corresponds to abundances and $\mathbf{y_i}$ the are the endmembers

University of Florida

Sparsity Promotion

- Weight decay is a standard method to promote small parameter values during a least squares minimization.
- The weight decay term attempts to prevent the β values from becoming large.

$$-LSWD = -\frac{1}{2}\sum_{i=1}^{N}\left(x_i - \boldsymbol{\beta}^{\mathbf{T}}\mathbf{y_i}\right)^2 - \gamma\sum_{k=1}^{M}\beta_k^2$$

$$-LSWD = \ln e^{-\frac{1}{2}\sum_{i=1}^{N}\left(x_i-\boldsymbol{\beta}^{\mathbf{T}}\mathbf{y_i}\right)^2-\gamma\sum_{k=1}^{M}\beta_k^2} = \ln\left[e^{-\frac{1}{2}\sum_{i=1}^{N}\left(x_i-\boldsymbol{\beta}^{\mathbf{T}}\mathbf{y_i}\right)^2} e^{-\gamma\sum_{k=1}^{M}\beta_k^2}\right]$$

- Both of the exponentials are proportional to a Gaussian:

$$N(\boldsymbol{\beta}^{\mathbf{T}}\mathbf{y_i},1) = \frac{1}{\sqrt{2\pi}}e^{-\frac{\sum_{i=1}^{N}\left(x_i-\boldsymbol{\beta}^{\mathbf{T}}\mathbf{y_i}\right)^2}{2}} \propto e^{-\frac{1}{2}\sum_{i=1}^{N}\left(x_i-\boldsymbol{\beta}^{\mathbf{T}}\mathbf{y_i}\right)^2}$$

- The objective function can then be seen as the log of the product of the probability of X given β and the prior of β :

$$p(X \mid \boldsymbol{\beta})\,p(\boldsymbol{\beta})$$

University of Florida

Sparsity Promotion

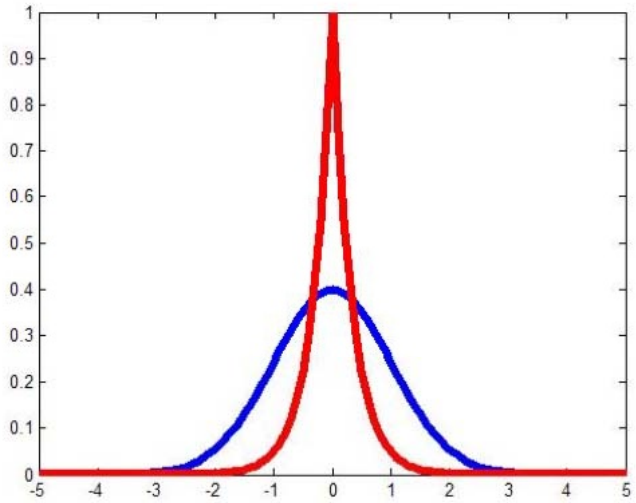
- The Gaussian prior is not effective at sparsity promotion.
- The Gaussian prefers the series of small β values instead of the preferred sparsity promoting values.
- Suppose

$$\boldsymbol{\beta}^1 = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

$$\boldsymbol{\beta}^2 = \begin{bmatrix} 1/2 & 1/2 \end{bmatrix}$$

$$\sum_{k=1}^{M} \left(\beta_k^1\right)^2 = 1^2 + 0^2 = 1$$

$$\sum_{k=1}^{M} \left(\beta_k^2\right)^2 = \frac{1}{2}^2 + \frac{1}{2}^2 = \frac{1}{2}$$



University of Florida

Sparsity Promotion

- In order to promote sparsity, instead of a Gaussian prior on β, we would like to use a zero-mean Laplacian prior.

$$-LSSP = -\frac{1}{2}\sum_{i=1}^{N}\left(x_i - \boldsymbol{\beta}^{\mathbf{T}}\mathbf{y_i}\right)^2 - \gamma\sum_{k=1}^{M}|\beta_k|$$

$$-LSSP = \ln e^{-\frac{1}{2}\sum_{i=1}^{N}\left(x_i - \boldsymbol{\beta}^{\mathbf{T}}\mathbf{y_i}\right)^2 - \gamma\sum_{k=1}^{M}|\beta_k|}$$

$$= \ln\left[e^{-\frac{1}{2}\sum_{i=1}^{N}\left(x_i - \boldsymbol{\beta}^{\mathbf{T}}\mathbf{y_i}\right)^2} e^{-\gamma\sum_{k=1}^{M}|\beta_k|}\right]$$

University of Florida

SPICE Algorithm

- The SPICE Algorithm adds a sparsity promoting term to ICE's objective function.

- Sparsity promotion is applied to the endmembers. Using sparsity promotion, if an endmember's proportion values are driven to zero, then that endmember can be pruned.

$$SPT = \sum_{k=1}^{M} \gamma_k \sum_{i=1}^{N} |p_{ik}| \quad \text{where} \quad \gamma_k = \frac{\Gamma}{\sum_{i=1}^{N} p_{ik}}$$

- As the sum of the proportions for an endmember become smaller, the associated constant becomes larger.
- Since the proportions have a non-negativity constraint, we can write:

$$SPT = \sum_{k=1}^{M} \gamma_k \sum_{i=1}^{N} |p_{ik}| = \sum_{k=1}^{M} \gamma_k \sum_{i=1}^{N} p_{ik}$$

SPICE Algorithm

- Incorporating the sparsity promotion term into ICE's objective function:

$$RSS^{*}_{reg} = (1-\mu)\frac{RSS}{N} + \mu V + SPT$$

- As was done with the basic ICE algorithm, this objective function can be minimized iteratively.
- After every iteration of the minimization process, the maximum proportion values for every endmember can be calculated.

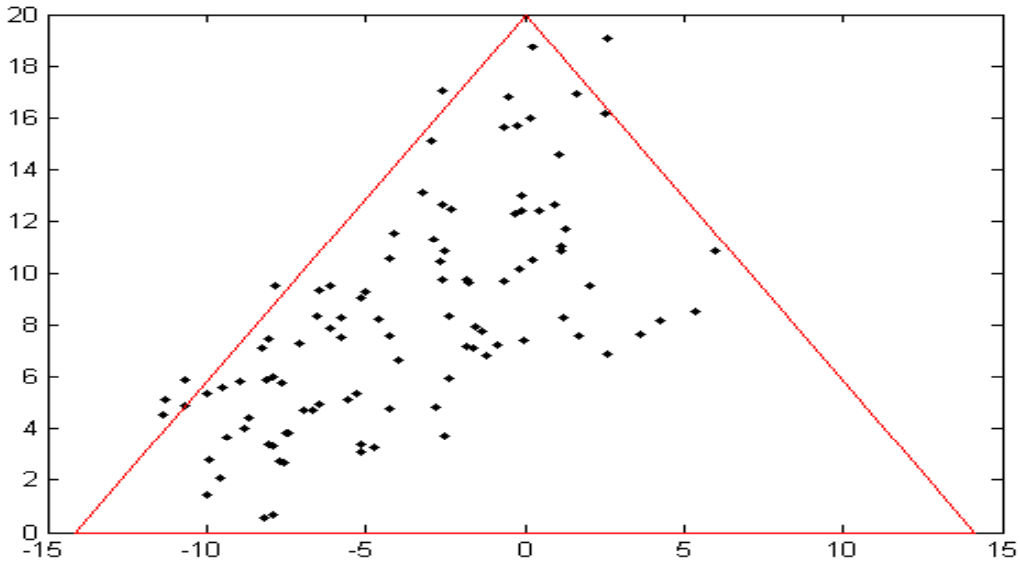
$$MP_k = \max_i \{p_{ik}\}$$

- If the maximum proportion for an endmember drops below a minimum threshold, then prune the endmember

University of Florida

2D Example

- True endmembers and data points used for testing:



- 100 points were generated from 3 endmembers:

$$(-10\sqrt{2},0),(10\sqrt{2},0),(0,20)$$

- Zero-mean Gaussian random noise was added to each pixel.

University of Florida

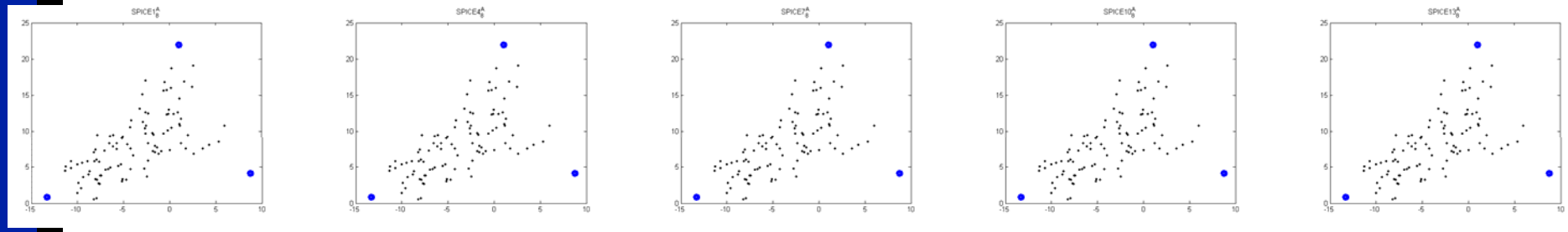
SPICE Results: 2D Example

Initial number of endmembers : 10

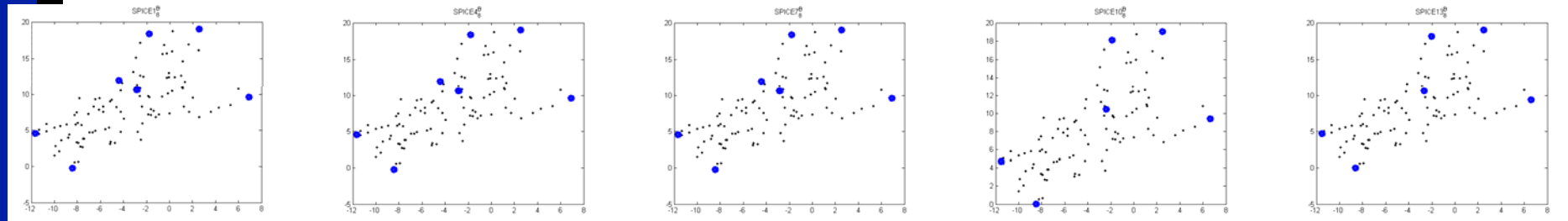
Objective function being minimized: $RSS^*_{reg} = (1-0.001)\frac{RSS}{N} + (0.001)V + SPT$

Prune Threshold =0.005	Prune Threshold =0.01	Prune Threshold =0.02	Prune Threshold =0.05	Prune Threshold =0.1

Results with Sparsity Promotion



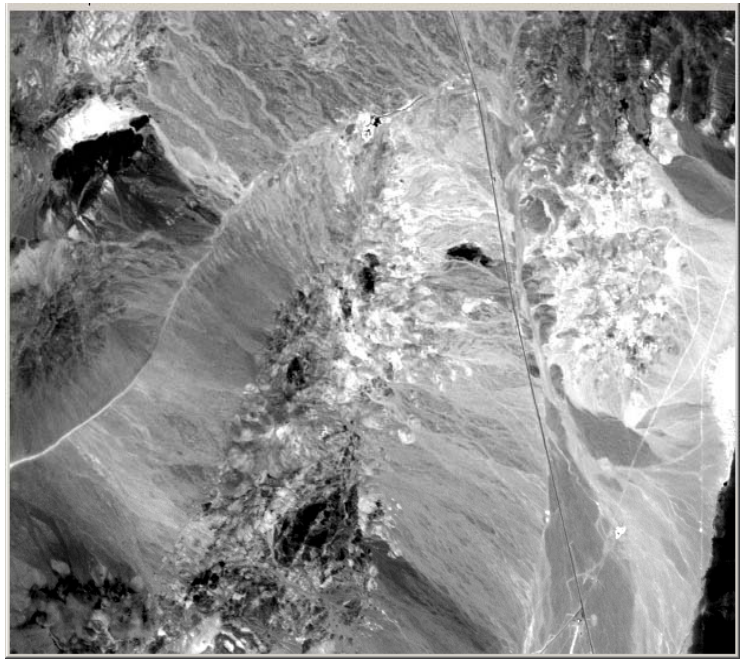
Results without Sparsity Promotion

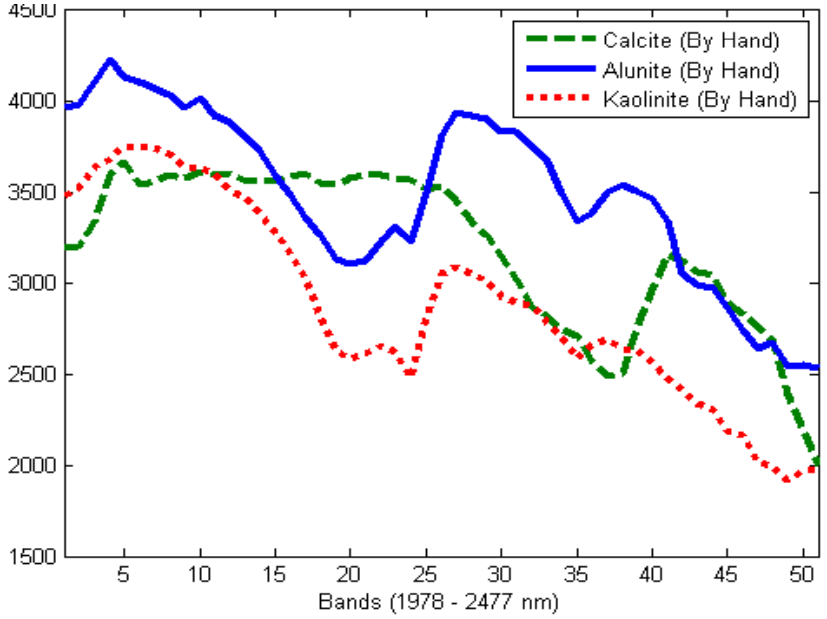


University of Florida

SPICE Results: Cuprite Data

- Wanted experiment with real image data with a known number of endmembers
- Selected 3 endmembers from the well-known Cuprite data set: Alunite, Calcite, Kaolinite

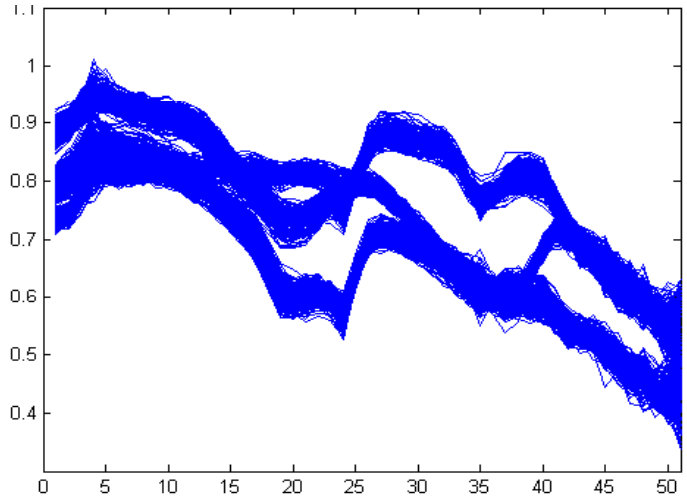


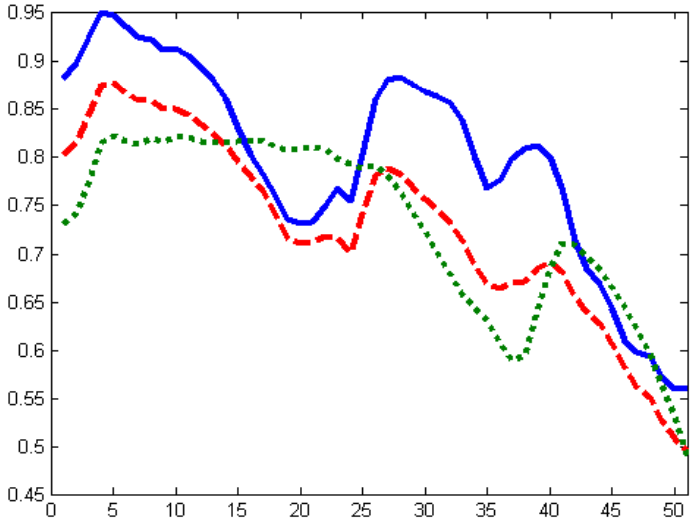


SPICE Results: Cuprite Data

- Collected pixels similar to the 3 endmembers
- Ran SPICE on this data

Experiment	Initial Number of Endmembers	Gamma Constant for SPICE	Number of endmembers found by SPICE	Number of endmembers found by ICE
1	5	1	3	5
2	10	0.5	3	9
3	10	0.5	3	8
4	10	10	3	9
5	10	10	3	8
6	15	1	3	12
8	30	1	3	12
9	40	1	3	13
10	50	1	3	11





University of Florida

University of Florida

SPICE Algorithm

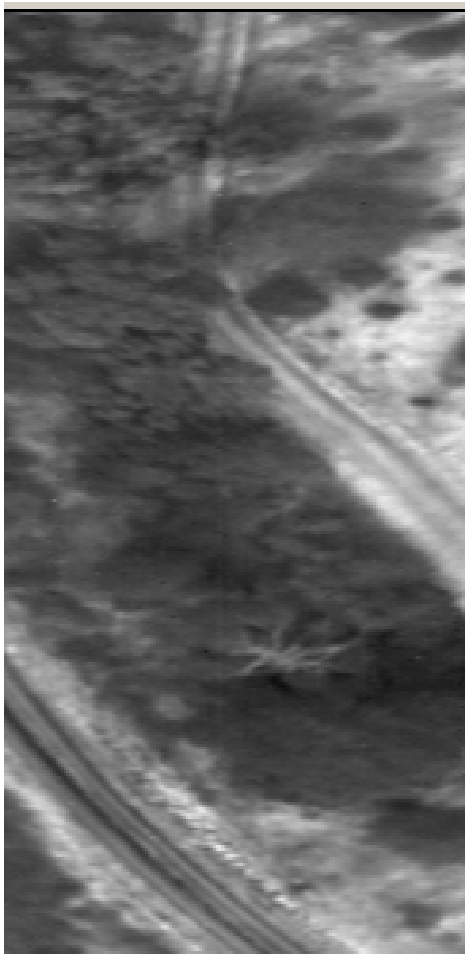
- SPICE was run on a subset of the image to reduce computation time.
- As was done by the authors of ICE, Pixels were chosen using the Pixel Purity Index (PPI).
- Most of the information required for endmember determination is found near the boundary of the data.
- Use a subset of the points that are found near the boundary of the data.
- PPI takes many random one-dimensional projections of the data. PPI counts the number of times each pixel is projected on or near one of the extreme points.

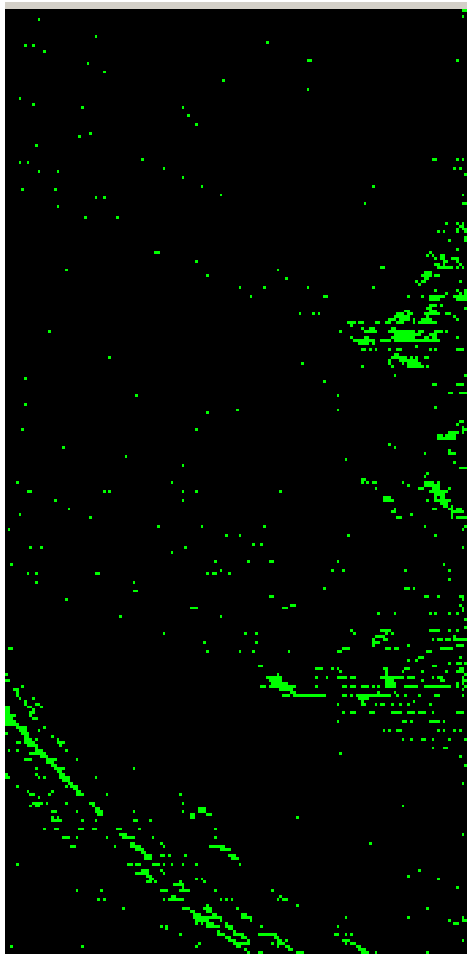
- J. Boardmann, F. Kruse, and R. Green, “Mapping Target signatures via partial unmixing of AVIRIS data,” in *Summaries of the 5th Annu. JPL Airborne Geoscience Workshop,* vol. 1, AVIRIS Workshop, R. Green, Ed., Pasadena, CA 1995, JPL Publ. 95-1 pp23-26

University of Florida

SPICE Algorithm

- Example of PPI results:
  - 1445 Points were chosen from this image (r0_20040726_203751_00001.nonaligned.img)
  - For this image, PPI was run for 20,000 iterations. Pixels within a 3 pixel threshold from either end of each random projection were counted.
  - The pixels in green on the right are the pixels that were chosen with a PPI of at least 1.

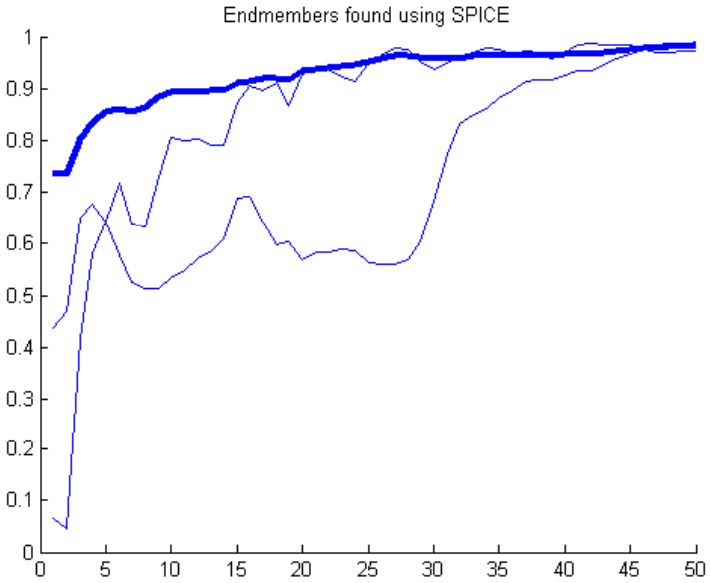


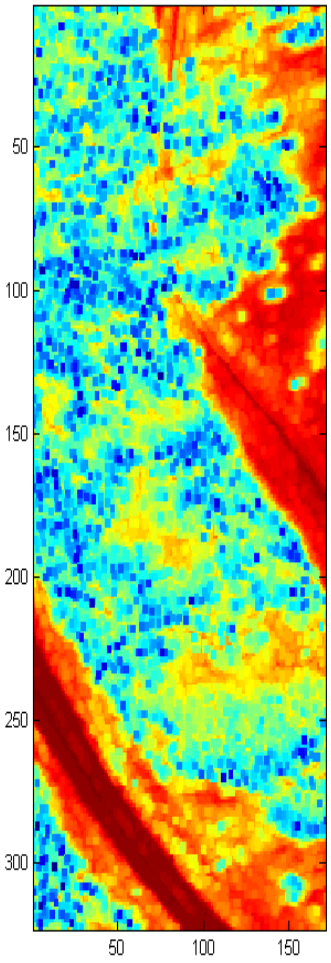


University of Florida

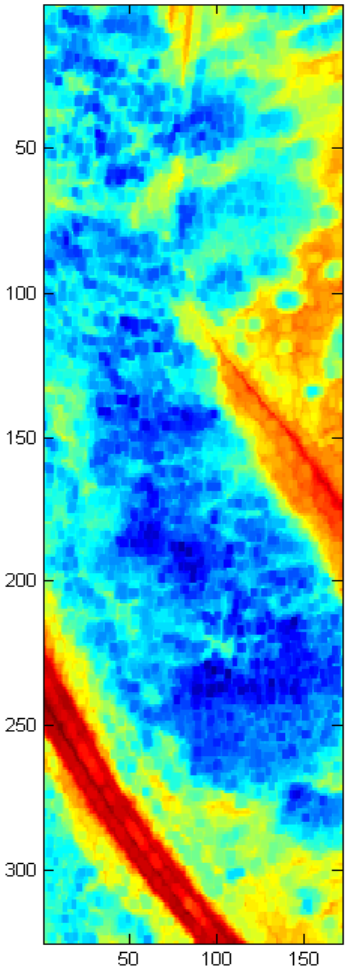
WAAMD Results

- Image: r0_20040726_203751_00001.nonaligned.img
  FF2, Thick Vegetation Region
- Sparsity Coefficient (Gamma): 5
- Regularization Parameter: 0.01
- Initial Number of Endmembers: 20
- Final Number of Endmembers: 3

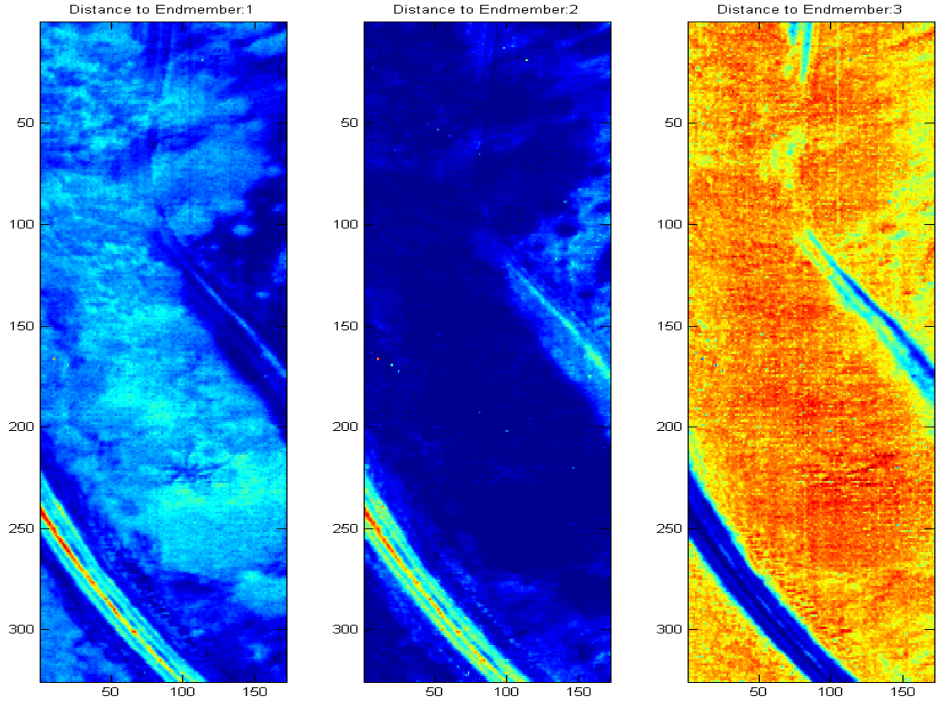




Previous Blackbody Feature from Clustering

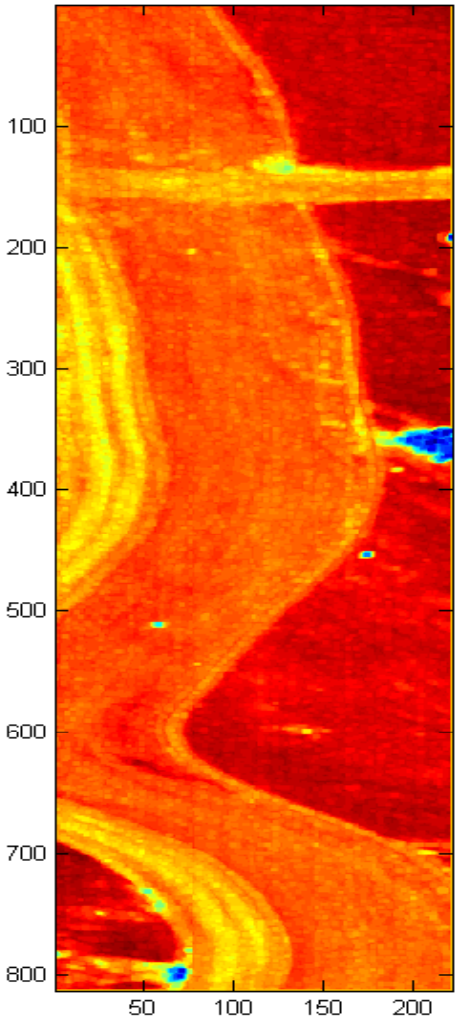


New Blackbody Feature from SPICE

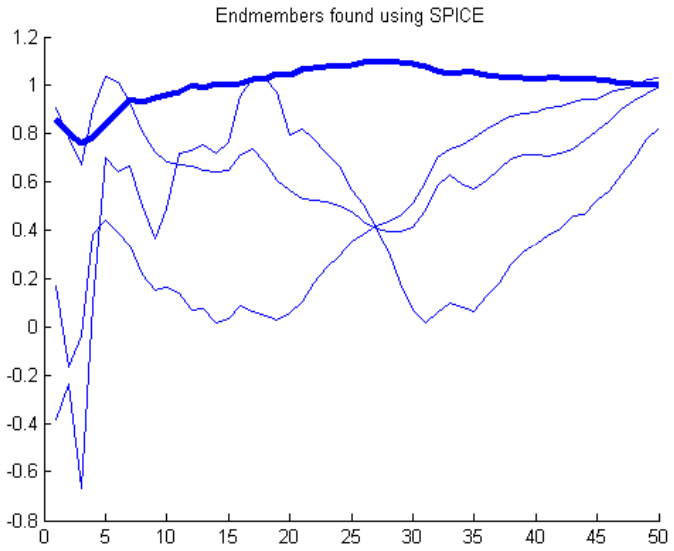


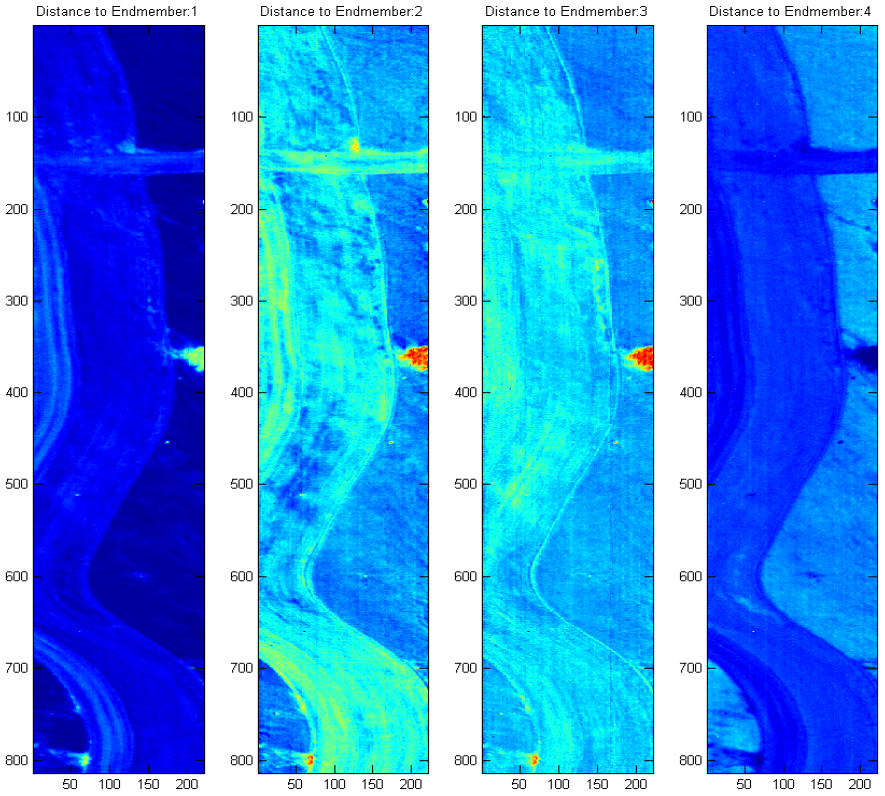
WAAMD Results

- Image: r4_20030404_120906_00001.nonaligned.img
  Countermine Region
- Sparsity Coefficient (Gamma): 5
- Regularization Parameter: 0.001
- Initial Number of Endmembers: 20
- Final Number of Endmembers: 4
- **Previous BB Feature did not find a blackbody cluster!**



New Blackbody Feature from SPICE

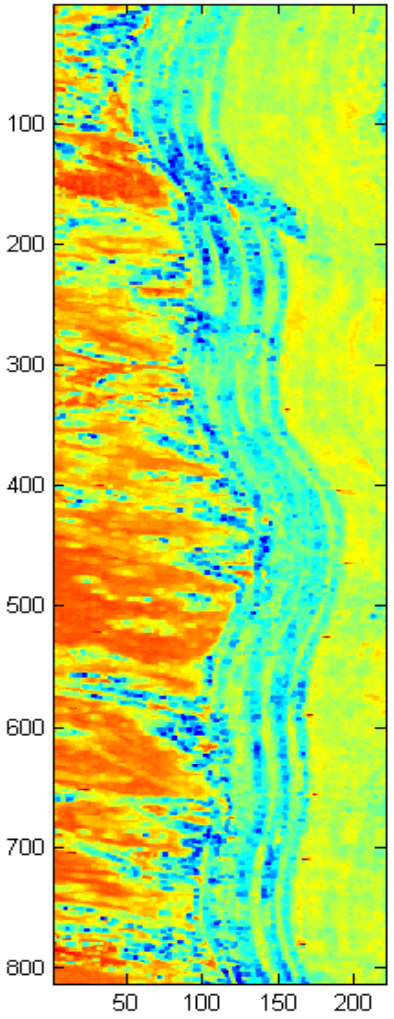




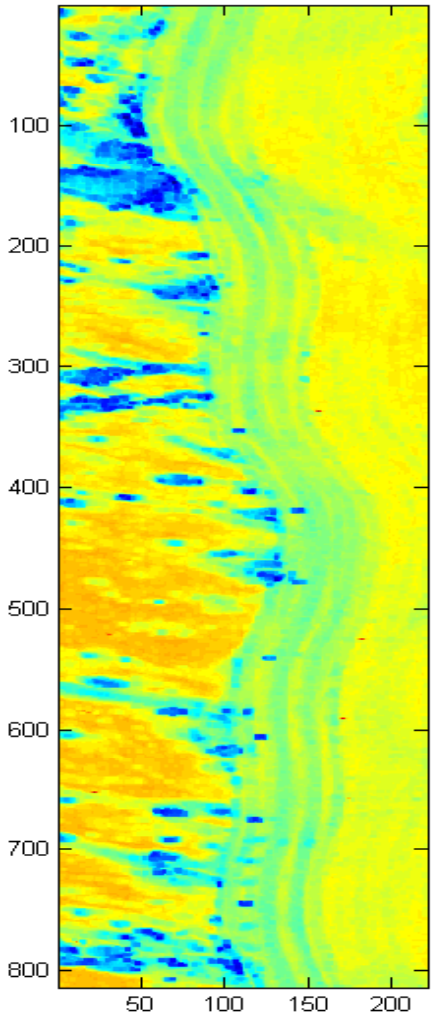
University of Florida

WAAMD Results

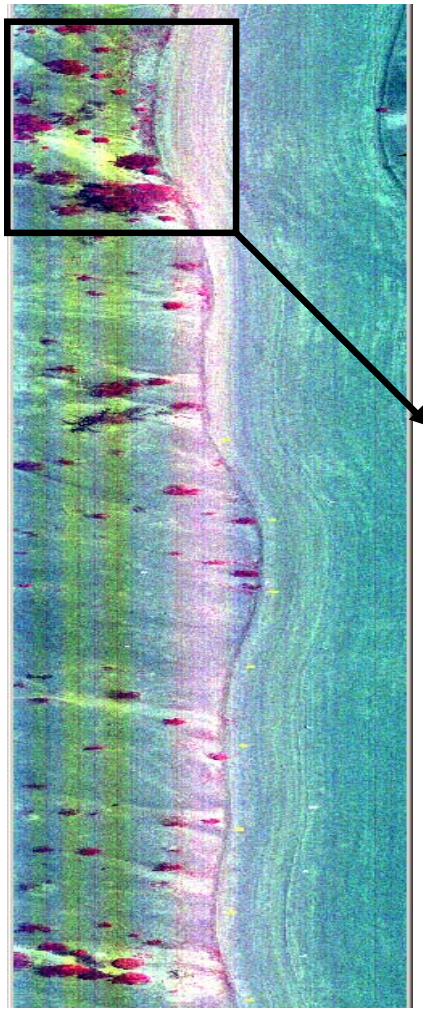
- Image: r4_20030404_120906_00001.nonaligned.img
  Countermine Region
- Sparsity Coefficient (Gamma): 5
- Regularization Parameter: 0.001
- Initial Number of Endmembers: 20
- Final Number of Endmembers: 4

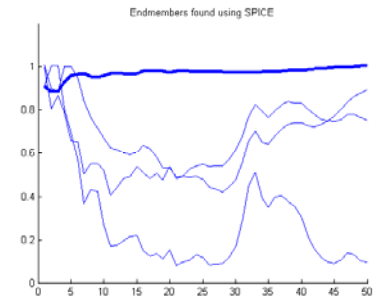


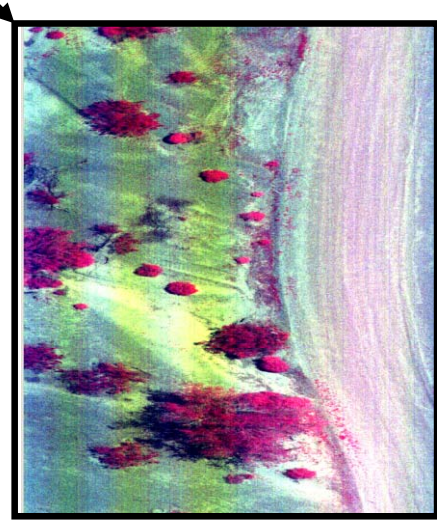
Previous Blackbody Feature from Clustering



New Blackbody Feature from SPICE



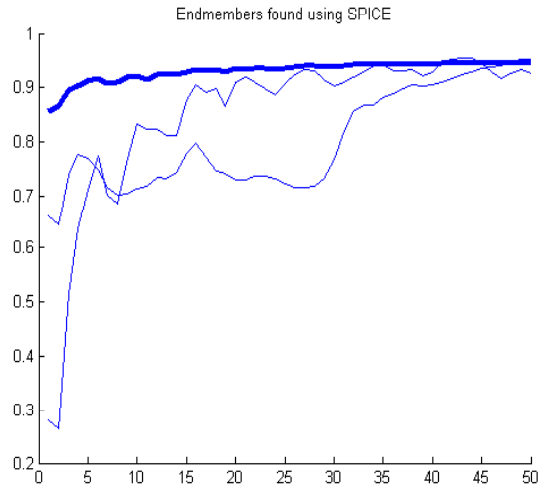


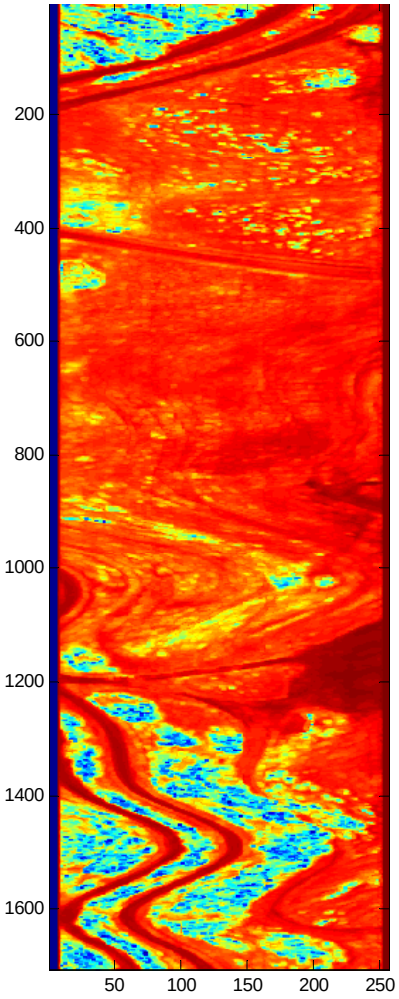


Wasabi Imagery

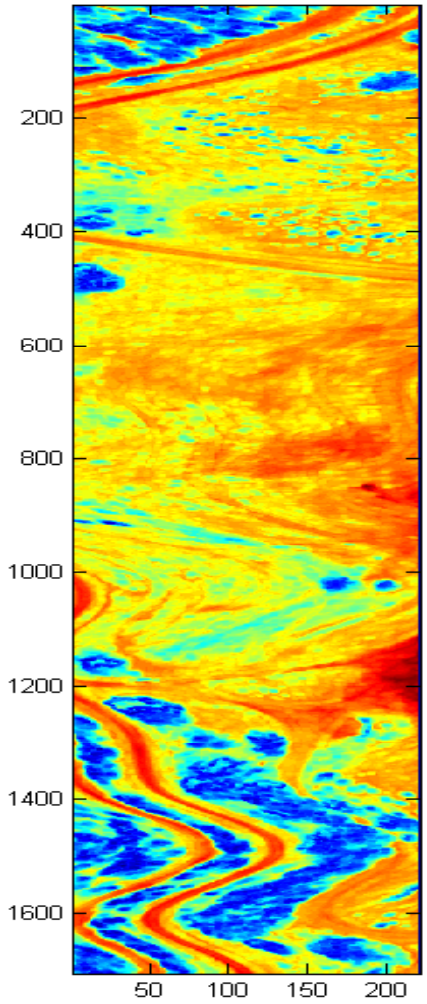
Level 1 Evaluation Results

- Image: C62R00020040727185032000 67r1F00lw.env, Thick Veg
- Sparsity Coefficient (Gamma): 10
- Regularization Parameter: 0.01
- Initial Number of Endmembers: 20
- Final Number of Endmembers: 3





Previous Blackbody Feature from Clustering



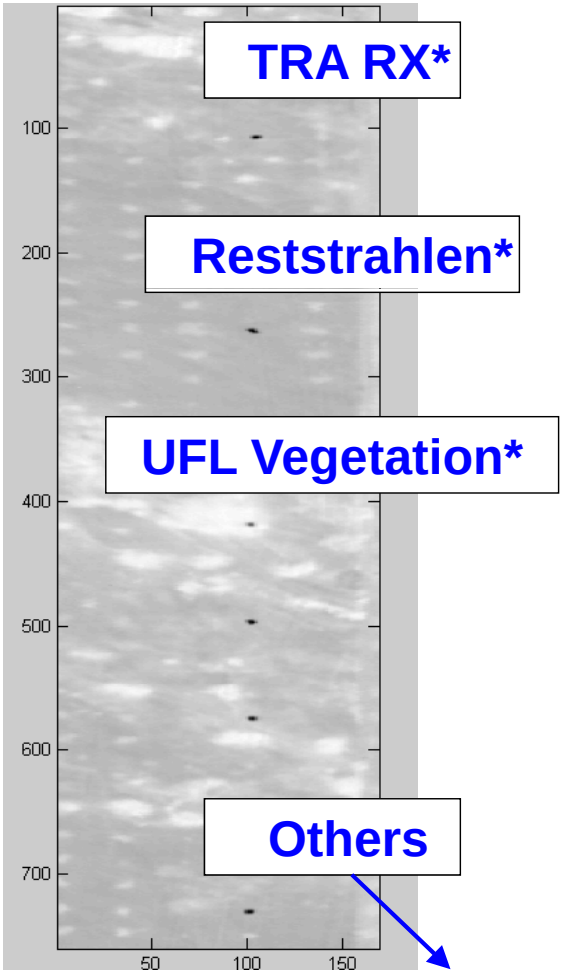
New Blackbody Feature from SPICE

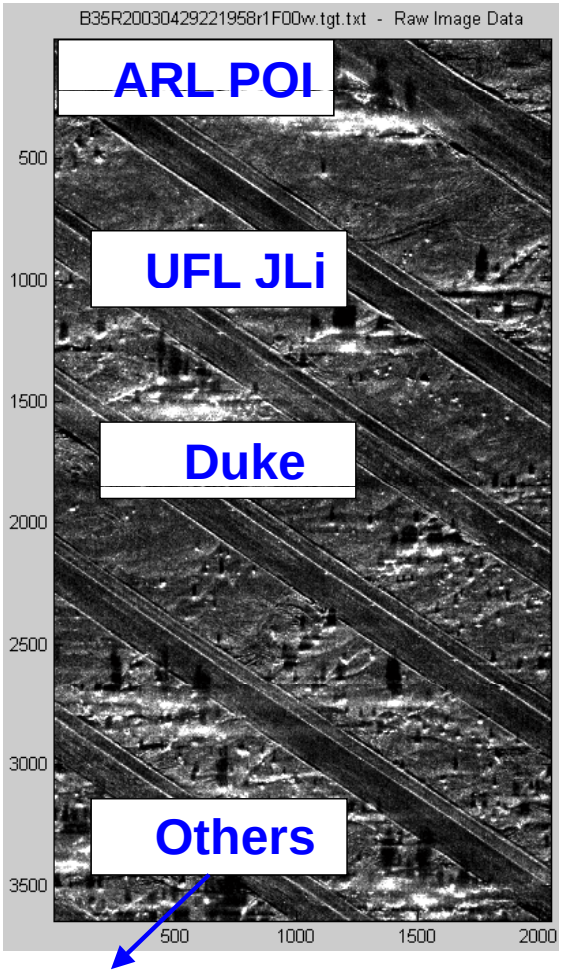
University of Florida

University of Florida

Overview of Fusion Framework

Fusion Research Setting





CHOQUET Integral Fusion

Nonlinear integral that represents wide variety of fusion operations

Voting, averaging, and, or, order statistics, Dempster-Shafer, …

Integral wrt non-additive measure – can optimized by estimating measure

University of Florida

Reststrahlen Ratio Feature

- Reststrahlen Ratio Calculation:

$$R(x,y) = \frac{1}{10}\sum_{b=20}^{30} f^{b}(x,y) \qquad N(x,y) = \frac{1}{10}\sum_{b=60}^{70} f^{b}(x,y)$$

$$S(x,y) = \frac{1}{1+e^{C(x,y)}} \quad \text{where C(x, y)} = \frac{R(x,y) - N(x,y)}{\frac{1}{P}\sum_{(u,v)} R(u,v) - N(u,v)}$$

and P is the number of pixels in the image

- Bands 20 to 30 are within the Reststrahlen region of the input image and correspond to **wavelengths of 8.87 to 9.40 microns**.
- Bands 60 to 70 are bands outside of the Reststrahlen region and correspond to **wavelengths of 10.9 to 11.5 microns**.
- Local background removal and normalization is applied

University of Florida

Reststrahlen Ratio Feature

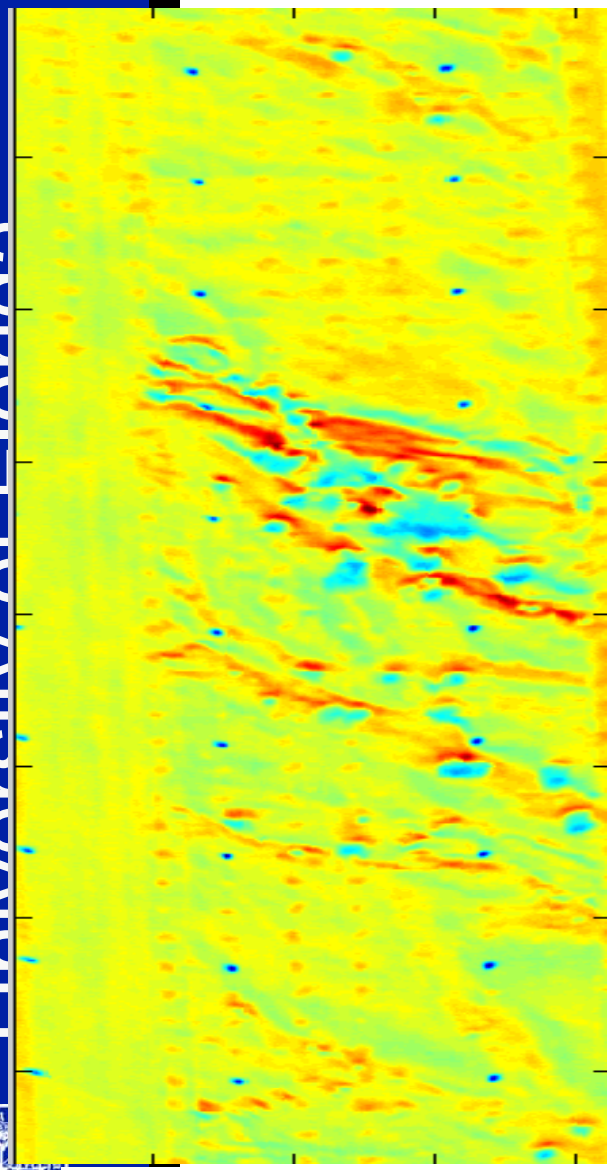
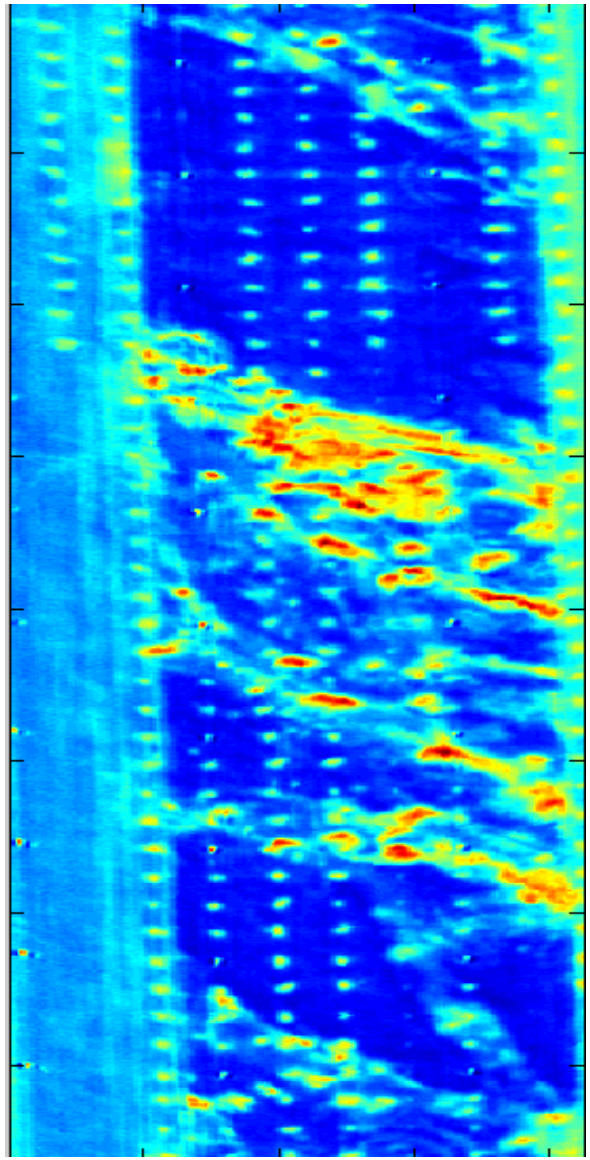
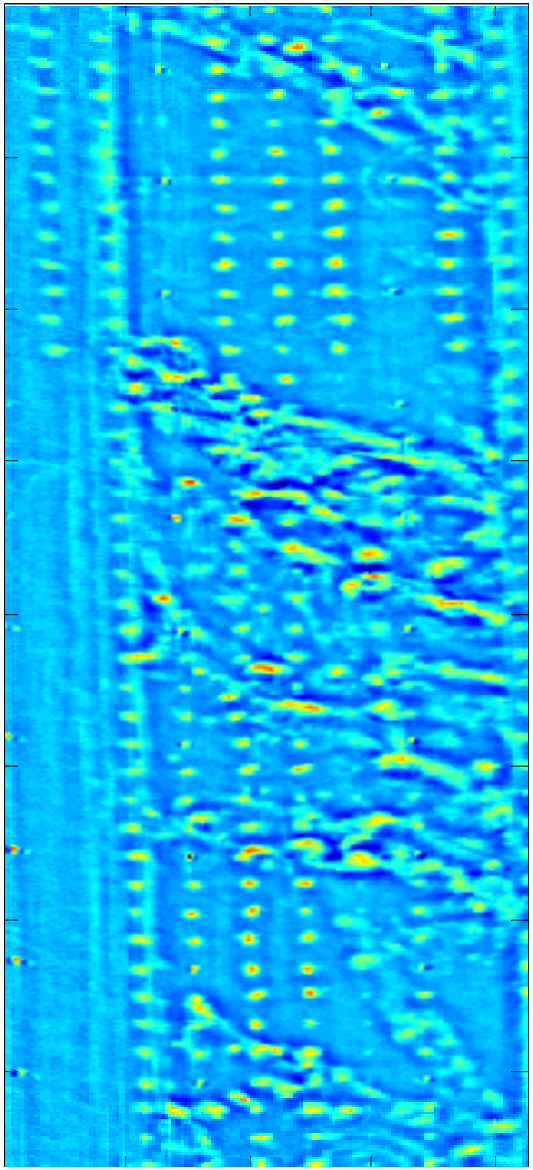


Image 1946: Band 20



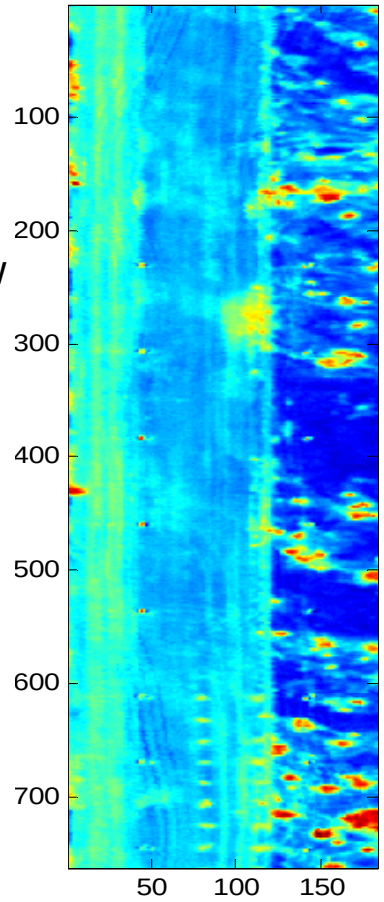
Reststrahlen Ratio Feature



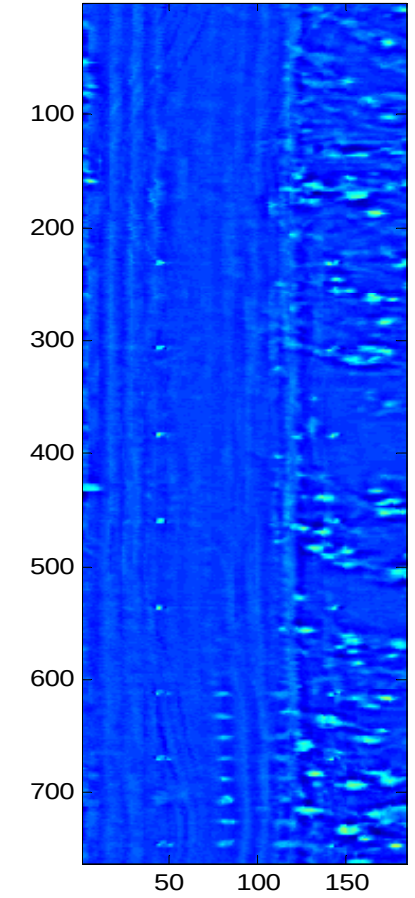
After Background Subtraction

Reststrahlen Ratio

- Reststrahlen Ratio Results on Image with No Mines
- Image:
  B62R020200304031931540072 7r1F00lw
- Contains
  - Holes
  - Fiducials
  - IR Panels



Reststrahlen Ratio Feature

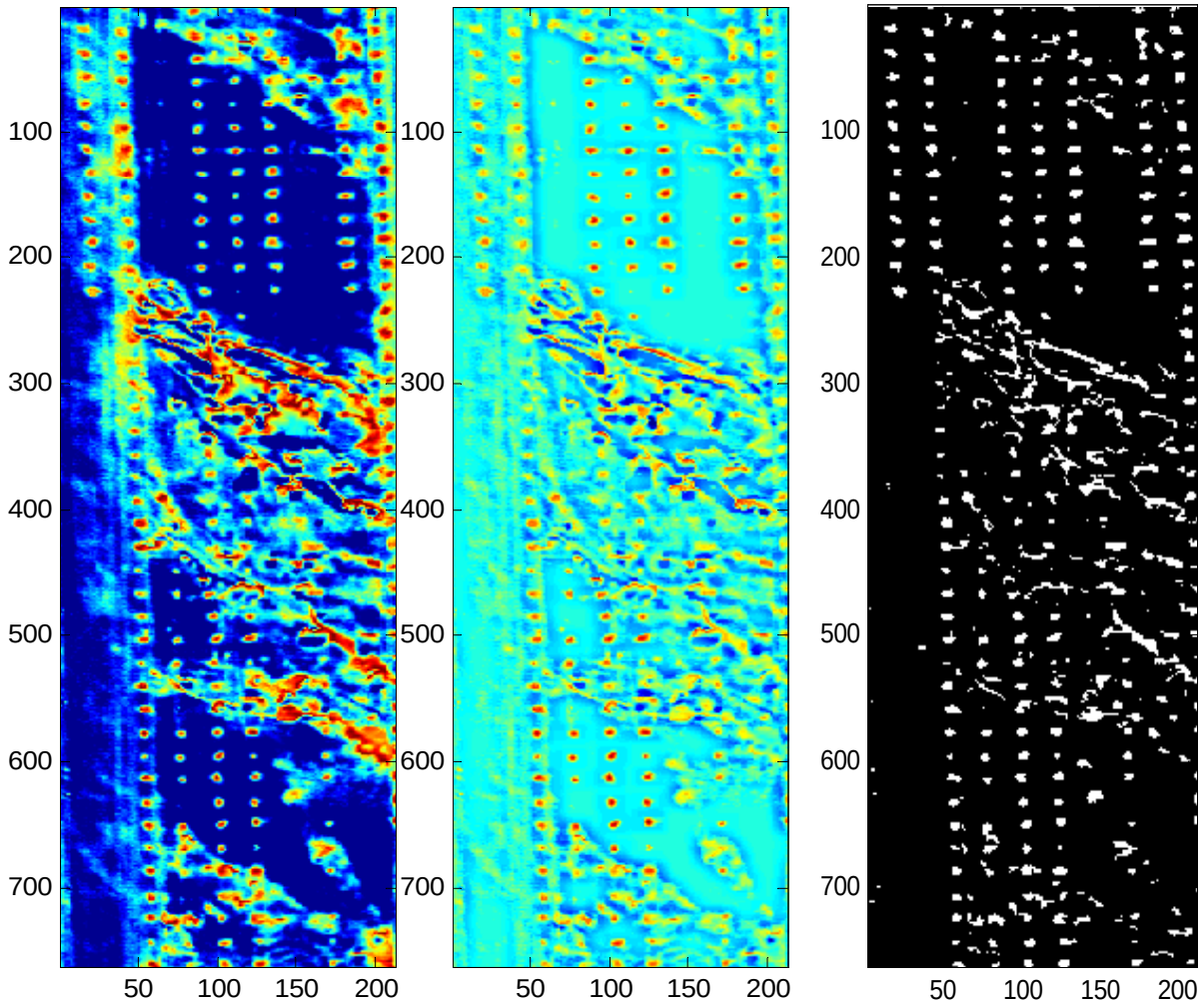


After Background Subtraction

University of Florida

Fully Constrained Least Squares (FCLS) from UMD

- Pixel by pixel spectral analysis using Linear Mixing Model
  - Produces a detection image
- Creating POI file
  - Local background removal
  - Connected component analysis
  - Normalization



Detection Image

Local Background Removal

Threshold

University of Florida

University of Florida

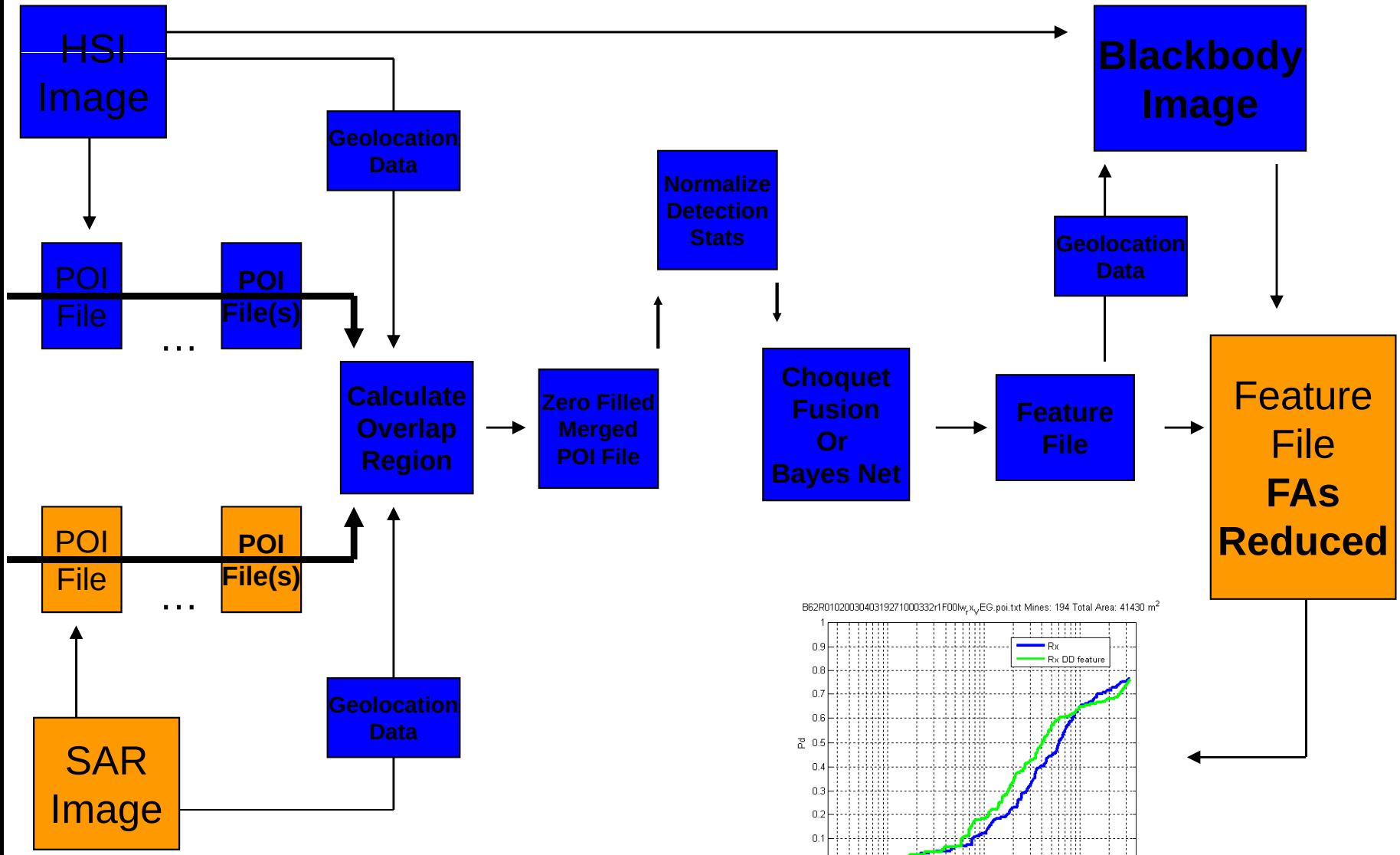
Decision Statistics / Features Used in Fusion Experiments

- Hyperspectral Imagery (AHI)
  - UFL
    - Reststrahlen Feature
    - Blackbody Feature
  - TRA
    - Rx
    - Mtd
- Radar Imagery (UltraSar / Lynx)
  - Duke
    - LGRF
    - RVM
  - ARL / Raytheon
    - Prescreener
  - UFL (Jian Li)
    - SAL BN

University of Florida

Fusion Infrastructure

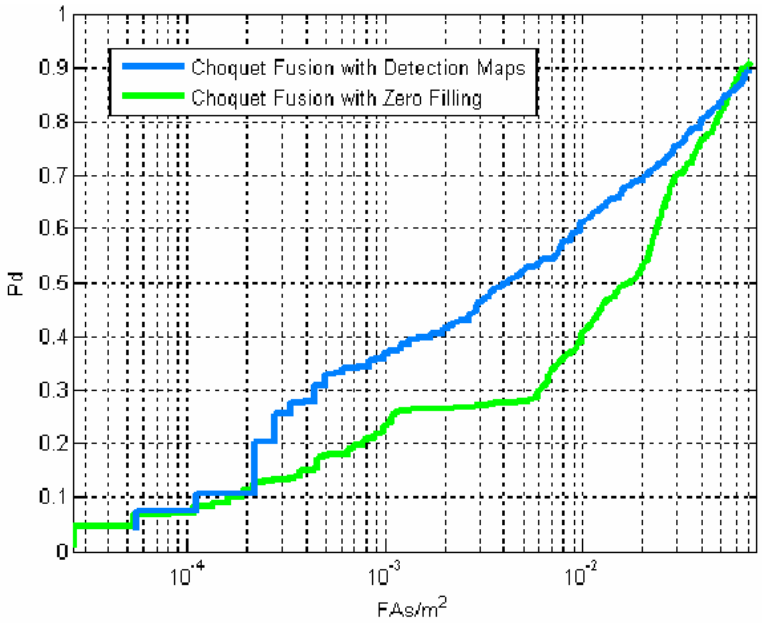
UF Fusion Diagram



University of Florida

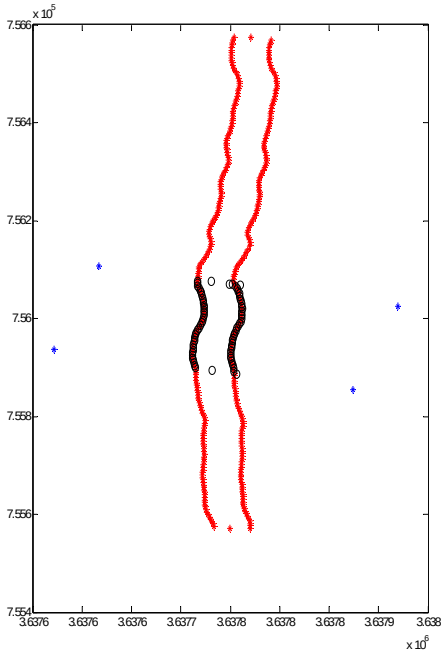
POI Merging Under Uncertainty

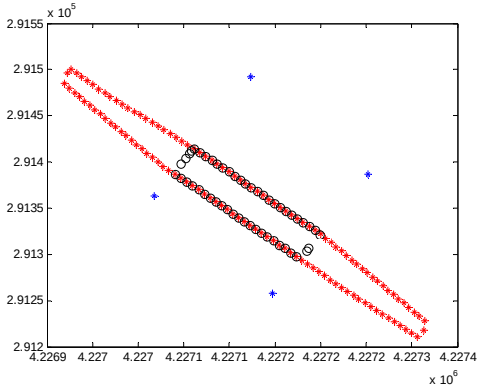
- Fusion of detector values
  - Fusion: some combination of detector outputs
  - How do we treat the case when we don't have all detector outputs
    - Pessimistic – fill with zeros
      - Cannot differentiate between Ignorance and Zero confidence
    - Optimistic – fill with a numbe below the detectors threshol
  - Previous experiments at Counterr have shown a significant decrease fusion performance when we treat ignorance and zero confidence th
  - Investigating new ways to deal wi lack of detection maps

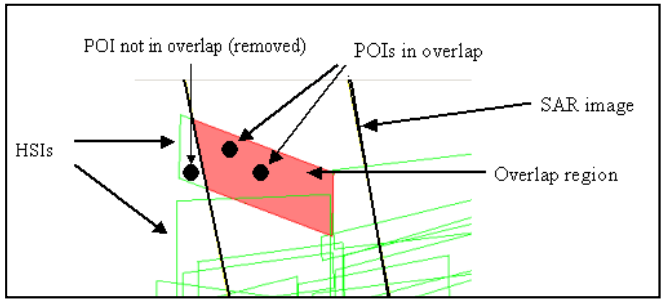


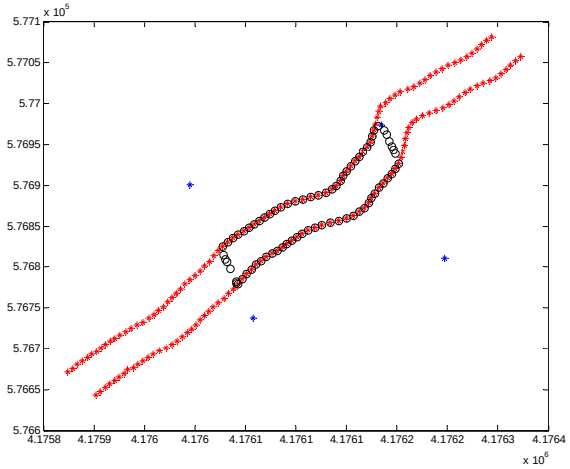
Overlap Areas

- Overlap Regions
  - Regions found using Geolocation Data and built-in polygon tools (Raytheon)
  - Area calculated using similar built-in polygon area tools (Raytheon)
  - Collect POIs and target points inside the overlap for testing and scoring







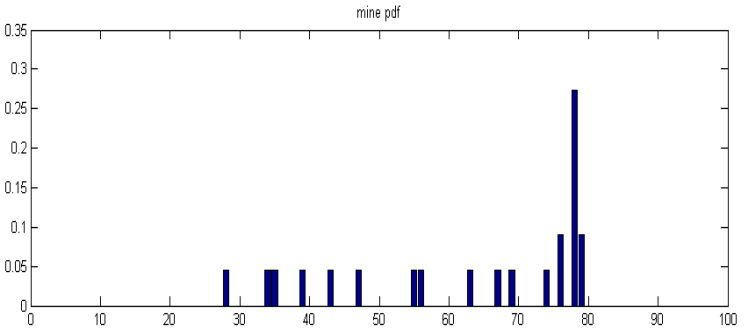


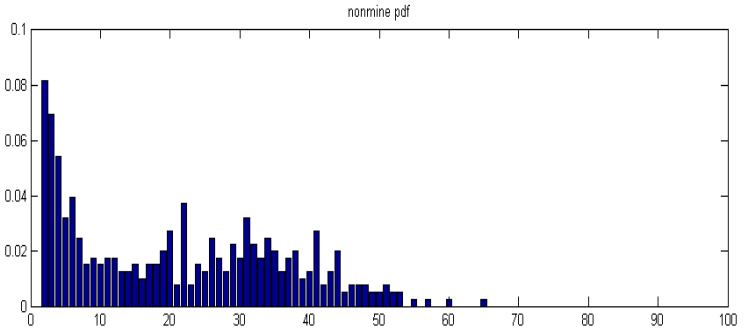
University of Florida

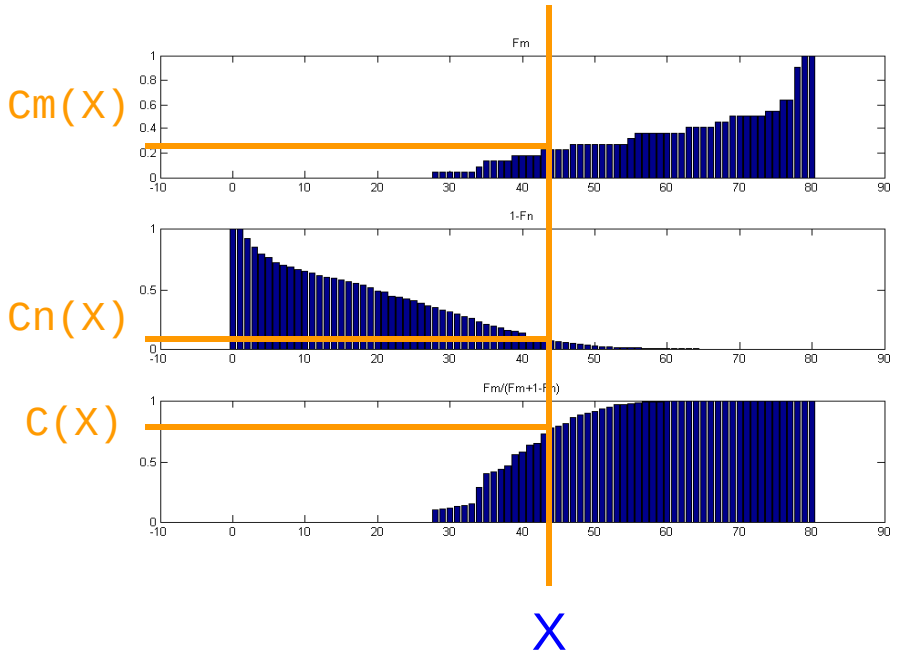
Fusion - Normalization

Suppose detector produces confidence values between 0 and 80 and has the following pdfs for mines and non-mines

```
Fm     = cumulative mines
Fn     = cumulative nonmines
X      = detector confidence
Cm(X) = Fm(X)
Cn(X) = 1-Fn(X)
C(X)   = Fm(X)/(Fm(X)+1-Fn(X))
```







University of Florida

Choquet Interpretation

- Let $X = \{x_1, x_2, \ldots, x_n\}$ be a finite set (algorithms)
- Let μ be a fuzzy measure on 2^X (nonlinear weights)
- Let $f : X \rightarrow [0,1]$ (confidence returned by algorithm)
- $f(x_{(1)}) \leq f(x_{(2)}) \leq \ldots \leq f(x_{(n)})$ (sorted confidences)
- Let $A_{(i)}=\{x_{(i)},\ldots,x_{(n)}\}$ and $\mu(A_{(n+1)})=0$
- The discrete Choquet integral of f is the sorted confidence times a weight indicating the value of adding the algorithm x_j to algorithms $x_{j+1}, x_{j+2}, \ldots, x_n$

$$C_\mu(f) = \sum_{j=1}^{n} f(x_{(j)})(\mu(A_{(j)}) - \mu(A_{(j+1)}))$$

University of Florida

Sugeno Choquet Measures

- λ-measures, or Sugeno measures

$$\text{For all } A, B \subseteq X \text{ with } A \cap B = \phi$$

$$\mu_\lambda(A \cup B) = \mu_\lambda(A) + \mu_\lambda(B) + \lambda\mu_\lambda(A)\mu_\lambda(B) \quad \text{for some } \lambda > -1$$

- Note that for λ = 0, the λ-measure is a probability measure
- Measure of a set more or less than the sum of the parts
- A λ-measure determined by *densities*: $\mu^i \equiv \mu(\{x_i\})$ since

$$X = \bigcup_{i=1}^{n} \{x_i\} \text{ and } \mu_\lambda(X) = 1 \quad \Rightarrow \quad (1+\lambda) = \prod_{i=1}^{n} (1+\lambda\mu^i)$$

Sugeno Choquet Fusion Optimization

Find Optimal Sugeno Measure via ***Gradient Descent***

- Let $C = \sum_{i=1}^{n} h(x_{(i)})[g(A_{(i)}) - g(A_{(i+1)})]$ where g is a Sugeno measure
- Minimize criterion $E = E\left[\frac{1}{2}(C(p) - O(p))^2\right]$ or Discriminative
- Differentiation yields

1 $$\Delta g_j = \alpha \frac{\partial E}{\partial g_j} = \alpha (C(p) - O(p)) \frac{\partial C}{\partial g_j}$$

2 $$\frac{\partial C}{\partial g_j} = \sum_{i=1}^{n} h(x_{(i)}) \left(\frac{\partial g(A_{(i)})}{\partial g_j} - \frac{\partial g(A_{(i+1)})}{\partial g_j} \right)$$

3 $$\frac{\partial g(A_{(i)})}{\partial g_j} = \begin{cases} 1 + \lambda g(A_{(i+1)}) + g_{(i)} g(A_{(i+1)}) \frac{\partial \lambda}{\partial g_j} + (1 + \lambda g_{(i)}) \frac{\partial g(A_{(i+1)})}{\partial g_j}, \\ \quad if\ (i) \neq n, (i) = j \\ (1 + \lambda g_{(i)}) \frac{\partial g(A_{(i+1)})}{\partial g_j} + g_{(i)} g(A_{(i+1)}) \frac{\partial \lambda}{\partial g_j}, \ if\ (i) \neq n, (i) \neq j \\ 1, \qquad if\ (i) = n,\ j = n \\ 0, \qquad if\ (i) = n,\ j \neq n \end{cases}$$

4 $$\frac{\partial \lambda}{\partial g_j} = \frac{\lambda^2 + \lambda}{(1 + g_j)\left[1 - (\lambda + 1)\sum_{i=1}^{n}\left(\frac{g_i}{1 + g_i \lambda}\right)\right]}, \lambda \neq 0$$

University of Florida

University of Florida

Minimum Error Classification

- **Cost functions do not use target outputs.**
- **Cost functions depend on difference between confidences of classes, e.g..**

$$d_i(x) = -C_{\mu_i}\left(f_i(x)\right) + \max_{j, j \neq i}\left\{C_{\mu_j}\left(f_j(x)\right)\right\}$$

- **Loss is assigned to misclassifications, e.g.**

$$l_i(x) = \begin{cases} \dfrac{1}{1+\exp\{-\alpha d_i(x)\}}, & d_i(x) > 0 \\ 0 & d_i(x) \leq 0 \end{cases}$$

- **Correctly classified samples have zero loss**

University of Florida

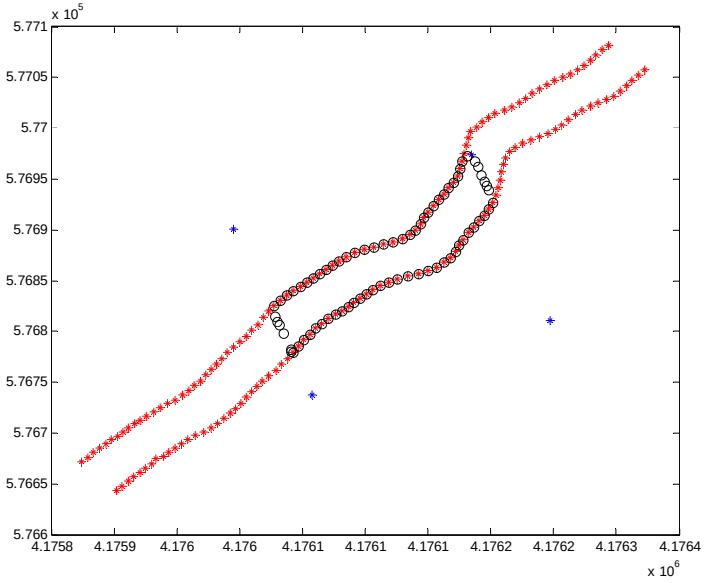
Experimental Procedure

- Each of the overlap regions (AHI / SAR) were tested and scored
  - Training was done using the other n-1 overlap regions from the same site. (Cross Validation)
  - POIs that were detections of fiducials and holes were removed for training.
- Choquet Fusion Results
  - Scoring
    - Stop light criteria
    - ROC curves
      - Fiducial, hole, and IR panel detections were counted as FAs

University of Florida

Scoring from an Entire Site

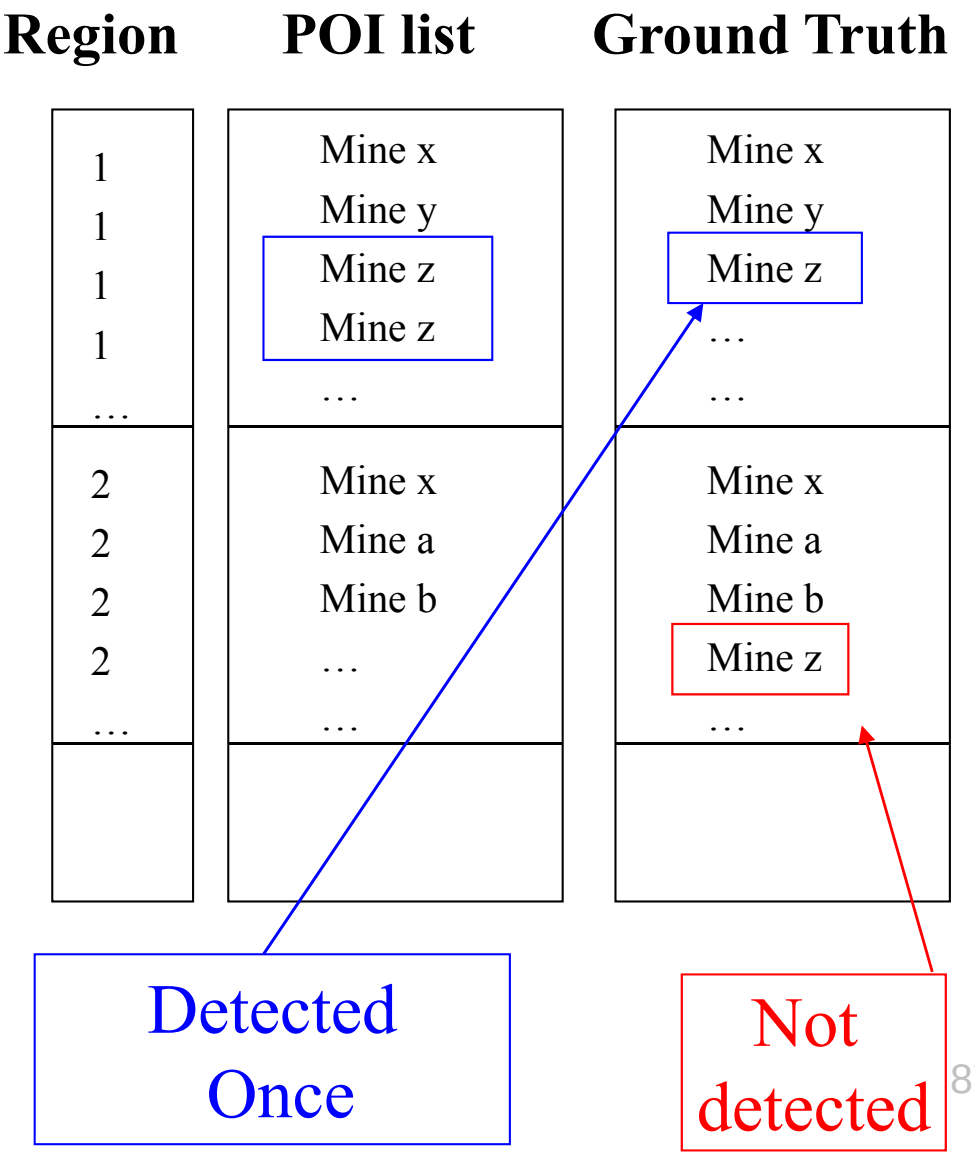
- Many Hyperspectral and SAR images overlapped with each other
  - Result:
    - Many encounters of the same mine were found in different overlap regions
    - How do we score the situation where one overlap region finds the mine and another does not?
      - Treat the multiple encounters of the same mine as completely different mines



How We Scored Results

- Concatenate POI and Ground Truth lists (Not Union)
  - Assists in keeping separate mine encounters, of the same mine, separate
  - Since all mines from each region are distinct, the total area is the sum of the individual regions' areas
- Example:

Two Mine z detections from Region 1 can only detect Mine z from the ground truth corresponding to Region 1. → Pd = 1/2



University of Florida

University of Florida

WAAMD Level 1 Evaluation

- Fixed Set of Images/Sensors Defined by WAAMD
- Multiple
  - Altitudes
  - Sites
  - Backgrounds
- Fusion Results with AHI LWIR and Lynx SAR
  - AHI RX (Ed Winter)
  - AHI MTD (Ed Winter)
  - AHI Blackbody
  - Lynx Random Set (UF)
  - LGRF (Duke)

University of Florida

Motivation for Bayesian Networks

- Previous Fusion Issues
  - Normalization issues
    - Fusion may require comparable feature values
  - Large amount of missing data
    - Features vectors are delivered with missing data
      - Some algorithms detect targets will others do not
- Bayesian Networks
  - Previous Issues
    - The outputs of random variables do not need to be comparable
      - Inference (probability of target) can be performed using incomparable feature values
    - Well-known EM algorithms can be used to compensate for missing data
  - New Issues
    - The Network Structure must be known to accurately perform inference
      - Unique method for structure estimation

University of Florida

Bayesian Networks: Parameter Learning

- The goal of a Bayesian Network is to learn the joint probability density function given an observed set of co-occurrences.
- Parameters of Bayesian Network: Discrete Random Variable Distributions
  - Given a network structure, learn the joint probability density function given an observed set of co-occurrences.
    - Used a MATLAB Bayesian Networks Package
  - The features used in fusion are not discrete so we cluster them using k-means clustering algorithm
  - Learning process was performed using training datasets.
- Testing data consisted of
    - Selecting testing data vector from testing dataset
    - Updating/instantiating values in the learned Bayesian Network
    - Performing inference to calculate the probability of mine.

University of Florida

Bayesian Networks Structure Learning

- Given a set of random variables: $V = \{X_1, X_2, ..., X_n\}$
- Goal: Find the *best* graph, G, that represents the relationships between the random variables.
- How do we define the *best* graph?

  There are several method, we use the Bayesian Scoring Criterion
  - Maximize the posterior

$$p(\mathbf{G} \mid \mathbf{d}) = \frac{p(\mathbf{d} \mid \mathbf{G})\, p(\mathbf{G})}{p(\mathbf{d})}$$

  - Using Bayes Rule, we can perform the above maximization by maximizing

$$p(\mathbf{d} \mid G)$$

  - with the assumption that p(**G**) is uniform. Note that p(**d**) is constant.

University of Florida

Structure Learning Methods

- MCMC Methods with missing data
  - Using a MCMC sampling method, such as Gibbs, sample data vector values (d=[x_1, x_2, …, x_n]) for a specific graph.
  - Using the collected set of data vectors, compute p(G|d).
  - Do this for all graphs being considered and select the graph with the largest p(G|d).
  - This is done because the initial data set is missing values. This will generate data samples without any missing data.

- Model Averaging
  - Instead of choosing a single graph, inference can be done using all graphs. Using the law of total probability, perform a weighted sum over all graphs to perform inference.
  - Model Averaging is often done when a single structure is not overwhelmingly most probable.
  - This is can be used with any structure learning method.

$$P(X_1 \mid X_2, \mathbf{d}) = \sum_i P(X_1 \mid X_2, \mathbf{d}, G_i) P(G_i \mid X_2, \mathbf{d})$$

University of Florida

Heuristic State Space Reduction / Computational Complexity

- There are approximately $2^{\frac{n^2}{2}-n}$ possible graphs given *n* nodes
- Also, the calculation of $P(\mathbf{d}\,|\,\mathbf{G})$ requires the calculation of all possible co-occurrences. This calculation can be exponential for each **G**.
  - We are looping through every graph. There are an exponential number of graphs.
  - For each graph, we are performing the calculation of P(**d**|**G**) which is an exponential calculation.
- If we limit the graph structure to trees
  - The number of possible graphs will be reduced.
  - We can calculate $P(\mathbf{d}\,|\,\mathbf{G})$ using only pairwise co-occurrences
  - This will significantly reduce the amount of computation needed.

University of Florida

Structure Sampling -- Gibbs Solution

- We have chosen to implement a Gibbs Sampler.
  - Create graph samples as opposed to data samples
  - Requires a convenient graph structure representation
- Represent graphs as vectors.
  - Elements in the vector represent possible connections in the graph.
  - If a vector element is 1, then a connection between the corresponding nodes in the Bayesian network exists. If the vector element is 0, there is no link between the two corresponding nodes in the graph.
- Create Graph
  - Visit each index in the graph
  - Calculate the posteriors of the two possible cases (0 , 1).
  - Toss a coin and set the value.
- Pick a graph based on the samples created.

Basic Gibbs Sampler

- Graph sample with *n* – nodes
- Algorithm
  - Compute all pairwise co-occurrences
  - Initialize a graph sample **G**
  - For each index 1 - $\frac{n^2}{2}-n$, calculate $P(g_i \mid g_{\forall j \neq i})$

$$\mathbf{G} = \begin{bmatrix} g_1 \\ g_2 \\ . \\ . \\ . \\ g_{\frac{n^2}{2}-n} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ . \\ . \\ . \\ 1 \end{bmatrix}$$

Since each g_i is binary, this will create one of two graphs $\mathbf{G}_1$, $\mathbf{G}_2$. Using

$$P(\mathbf{G}_i \mid \mathbf{d}) = \frac{P(\mathbf{d} \mid \mathbf{G}_i) P(\mathbf{G}_i)}{P(\mathbf{d})}$$

we can calculate the probability of the two different graphs given the data.

- Set the value of g_i using the ratio: $\frac{P(\mathbf{d} \mid \mathbf{G}_1)}{P(\mathbf{d} \mid \mathbf{G}_1) + P(\mathbf{d} \mid \mathbf{G}_2)}$
- After this is completed for each index, collect the sample.
- Repeat *M* times for a sample size of *M*.

University of Florida

University of Florida

Algorithm Details

- As previously mentioned, the state space of this problem is very large.
  - Reduce the space of possible graphs to trees.
    - Reduces posterior calculation
    - Reduces time needed to create samples.
    - May reduce accuracy of the hierarchal representation.
  - During the Gibbs sampling, if the addition of an edge creates a cycle.
    - Do not allow the addition of the edge.
    - Move to next vector index and proceed.
- Multiple restarts
  - Gibbs sampler to get stuck in local minima
  - Perform multiple restarts and consider multiple *best* solutions.

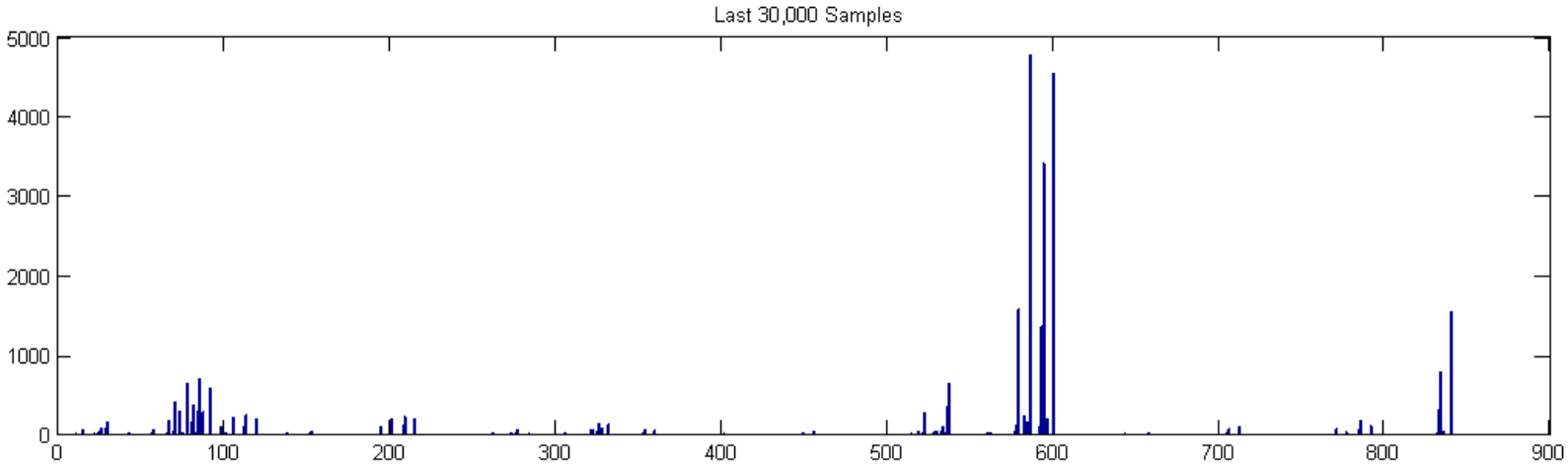
University of Florida

Structure Learning Experiments

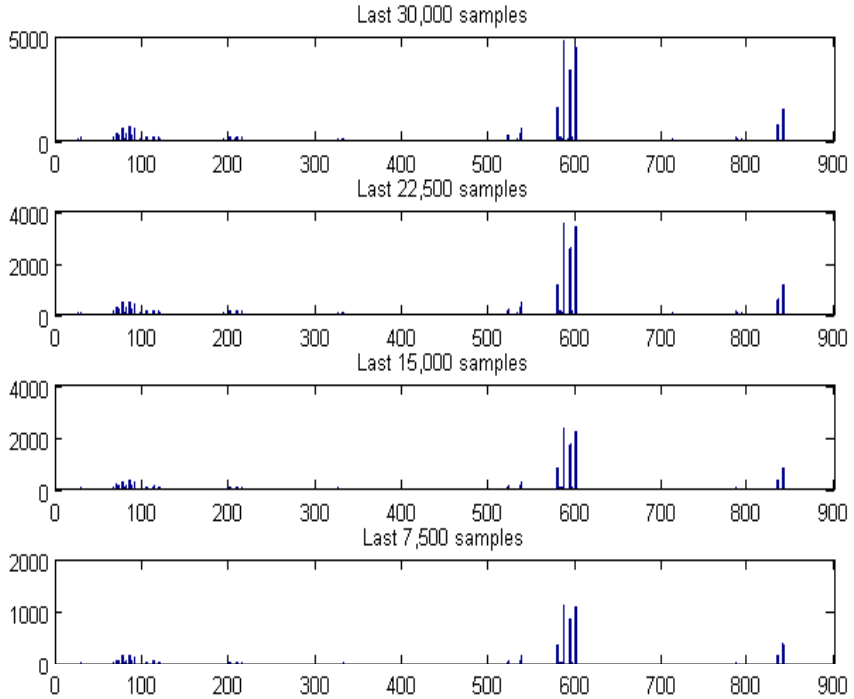
- Gibbs Sample implemented
  - Used 4 features (+1 mine label)
    - Rx
    - MTD
    - Reststrahlen
    - RS
  - Sampler was run for 30,000 iterations and 150,000 iterations (samples collected)
- Outputs
  - Histograms were created
    - Each number on the x-axis represents a unique graph structure
    - Binary graph vector is treated as a (binary) number for histogram representation
    - Histograms were saved after every ¼ of the samples were selected.
      - Check for convergence
    - Graph representations with high frequencies should best represent correlations between the features.

University of Florida

Structure Learning Results



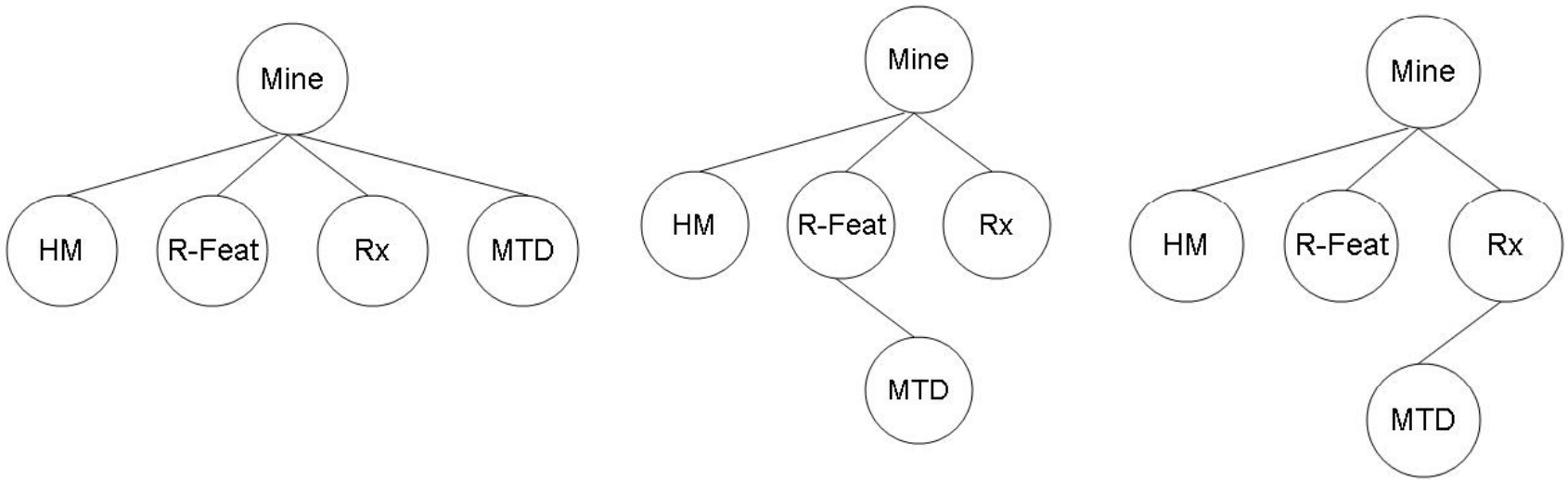
- Sampler experiment with 150,000 iterations
  - Allowed a minimum “burn-in” period of 120,000 iterations
  - Four histograms representing the samples collected with 4 different burn-in periods
  - Histogram shape seems to have stabilized



University of Florida

Structure Learning Results

- The histogram shown previously had 3 major peaks representing the following 3 graph configurations.
  - Other than uses in inference calculation, results can be useful by expressing
    - Importance of each feature
    - Correlations among features
    - Which features are most correlated with the presence of a mine

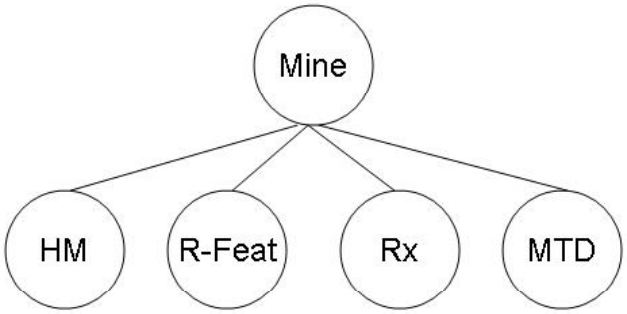


University of Florida

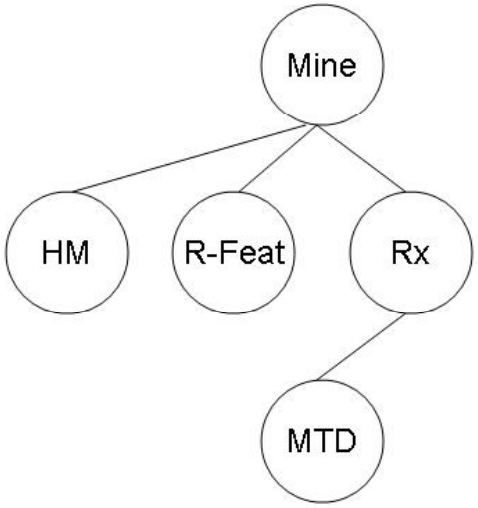
Learned Structures

- Fusion was performed using each of the 3 Bayesian network Configurations
  - The Results are compared to our existing fusion methods
  - The results from the best configuration are scored used the WAAMD “stop-light” criterion

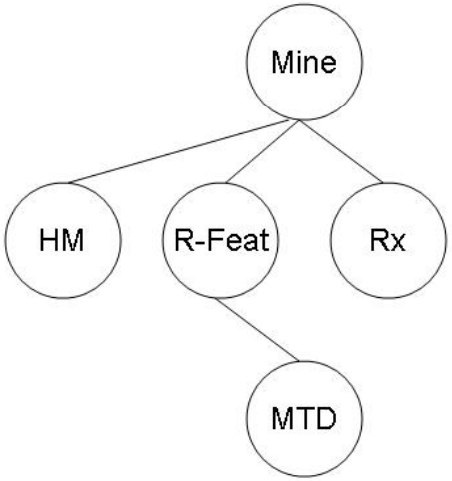
C1



C2



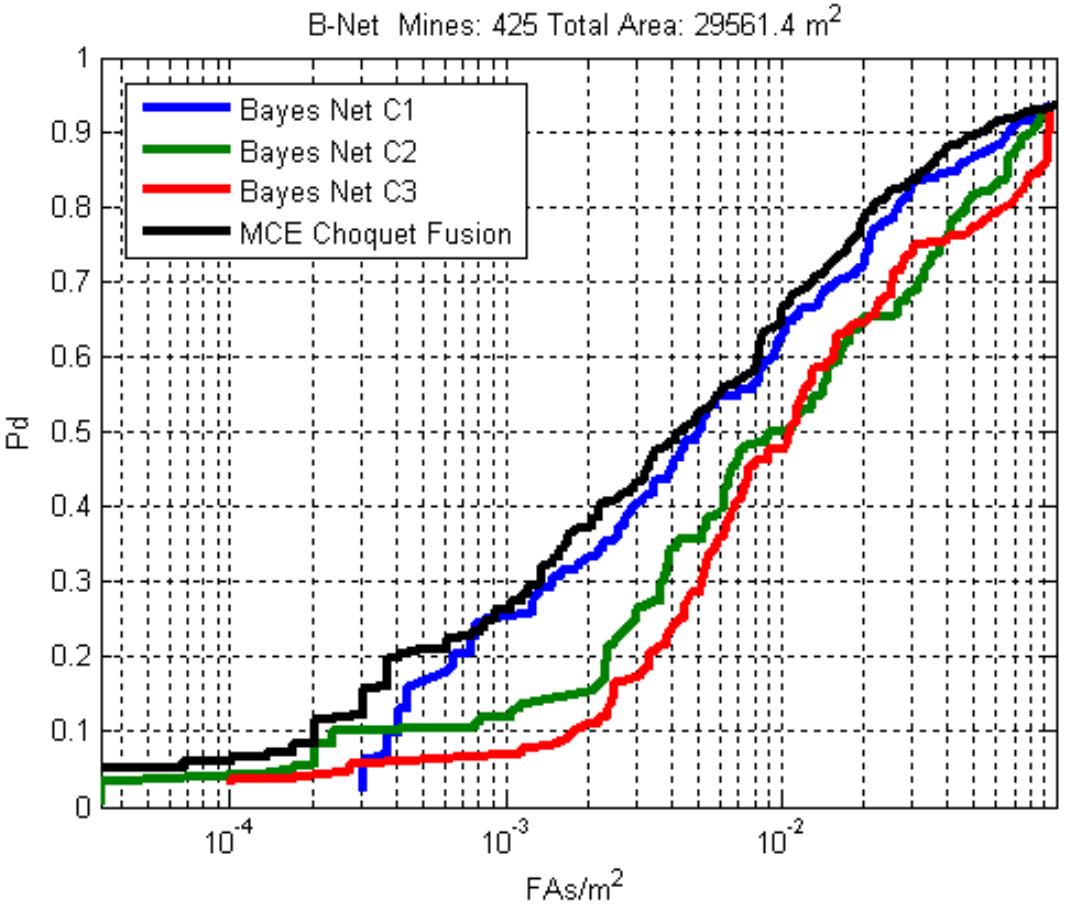
C3



University of Florida

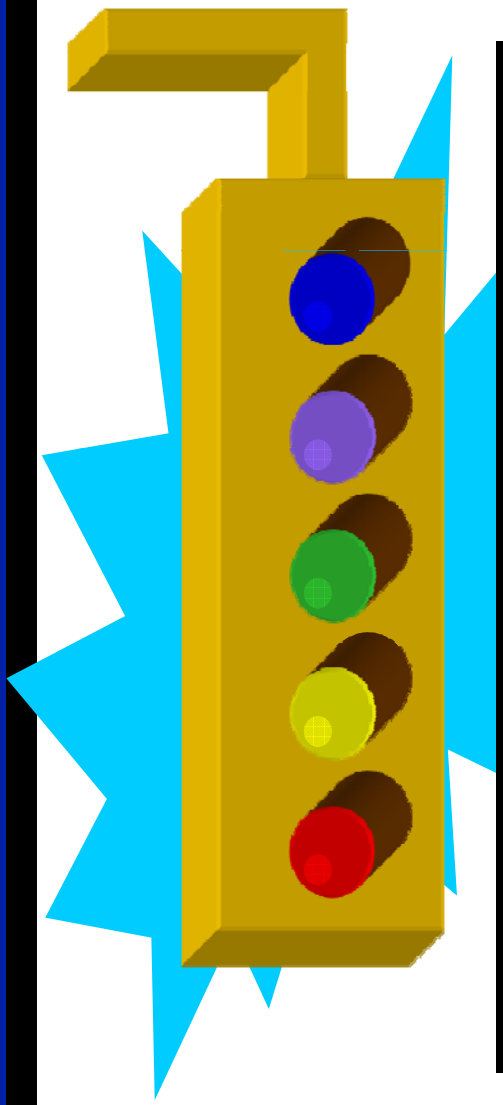
Overall Results of Bayesian Network Fusion

- **Countermine Site at YUMA**
- **Images (4 Overlap regions)**
  - 4 AHI images
    - B62R010200304031927100032r
    - B62R010200304040852000001r
    - B62R010200304040915200001r
    - B62R040200304041209060001r
  - 1 Lynx image
    - B35R20030429221958r1F00vv



University of Florida

WAAMD Level 1
Detection Performance Stoplight Criteria

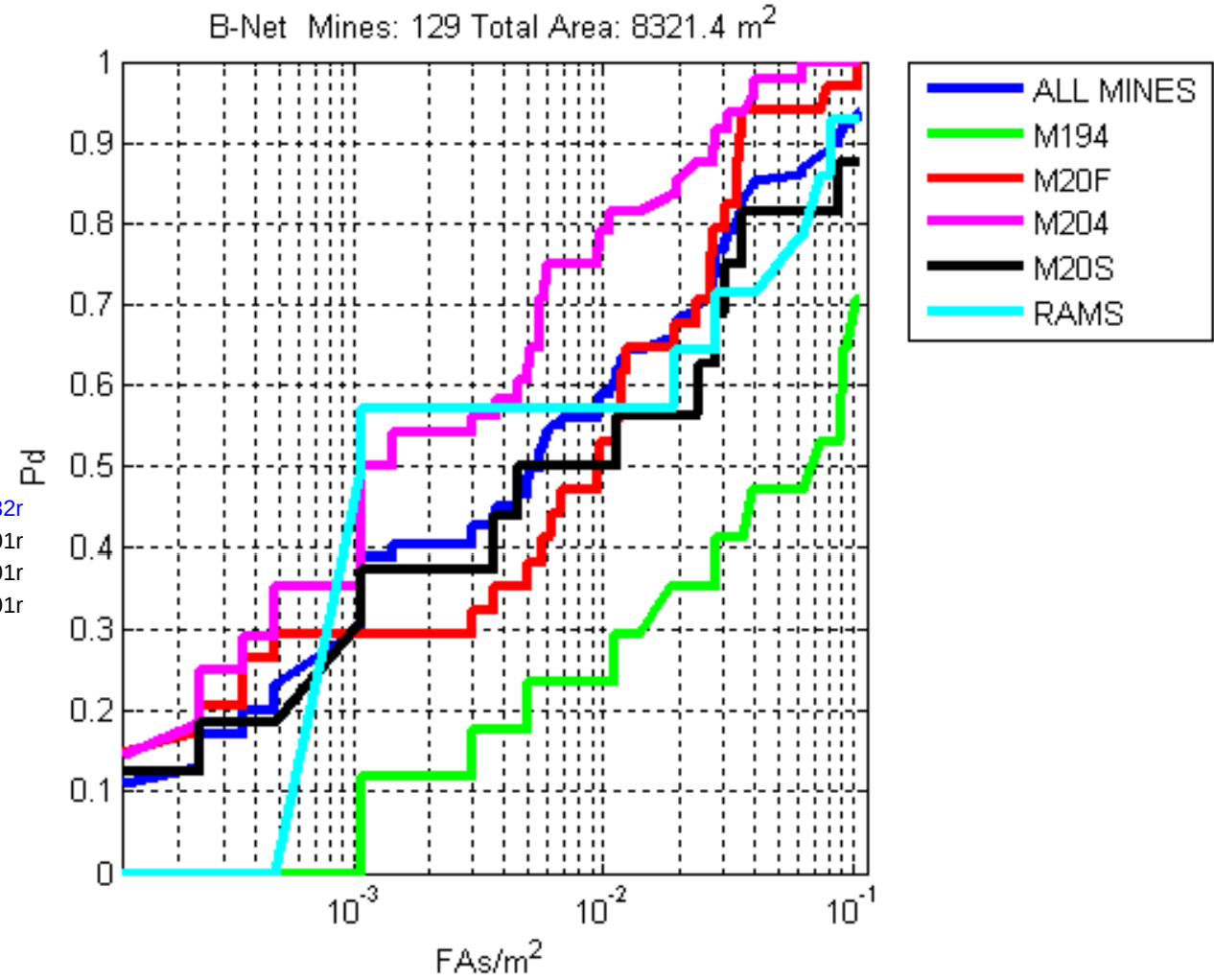


Color	Pd	FAR $1/m^2$	Detection Performance	Basis for Criteria
Blue	50-60%	10^{-4}	"Blue Ribbon"	Detect scatterable surface MF using density
Purple	50-60%	10^{-3}	Very Good	Typical pattern MF detection need
Green	50-60%	10^{-2}	Satisfactory	Notional limit for pattern MF detect & FBC scatterables
Yellow	30-50%	10^{-2}	Marginal	Needs improvement for MF detect
Red	< 30%	> 10^{-2}	Unsatisfactory	

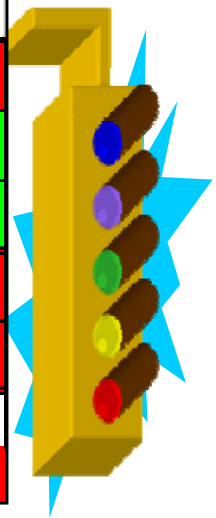
University of Florida

WAAMD Level 1 Eval
CM: Overlap 1

- **Countermine Site at YUMA**
- **Images (4 Overlap regions)**
  - 4 AHI images
    - B62R01020030403192710003 32r
    - B62R01020030404085200000 01r
    - B62R01020030404091520000 01r
    - B62R04020030404120906000 01r
  - 1 Lynx image
    - B35R20030429221958r1F00vv



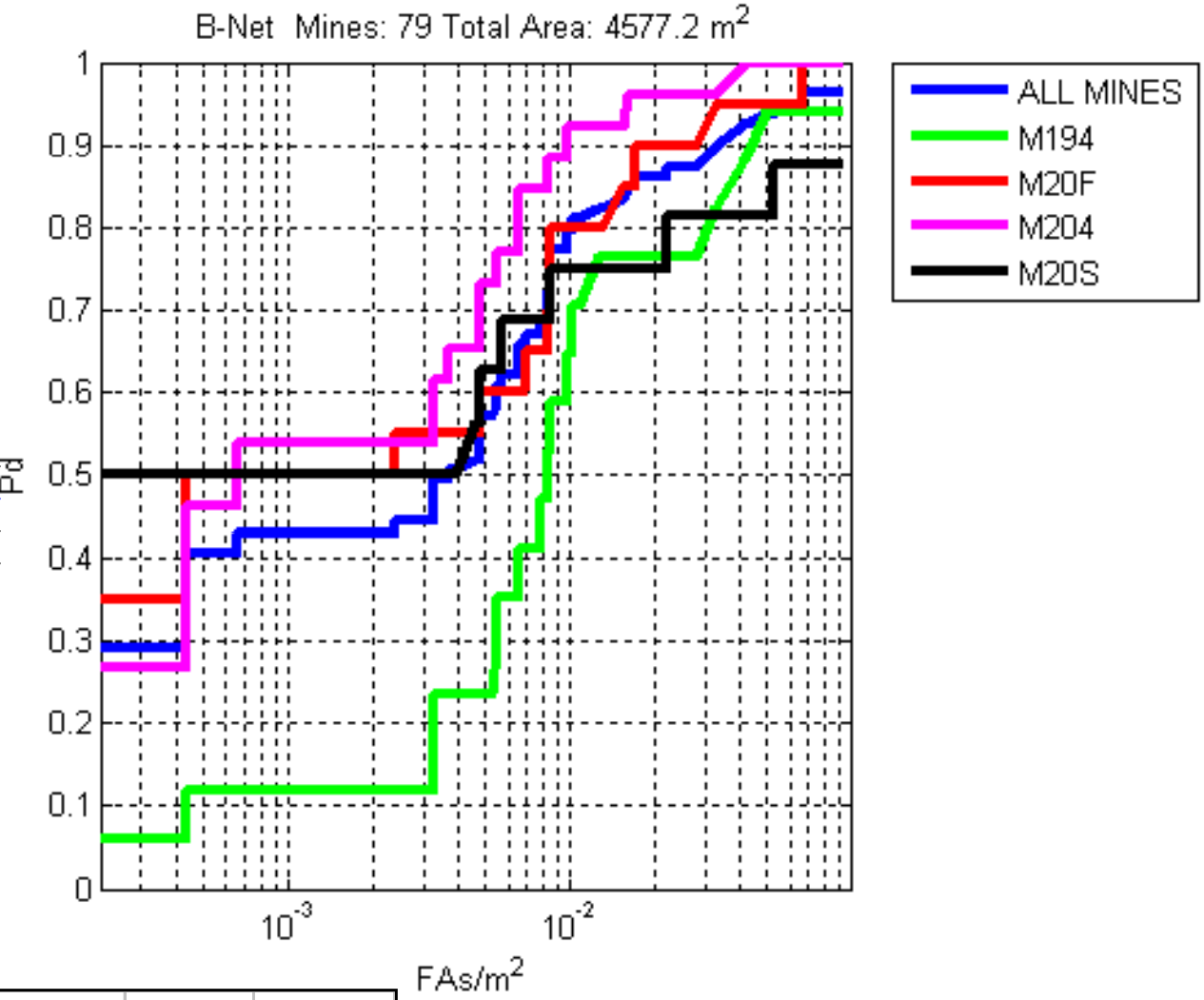
Mine Type	Bayes	Rx	MTD	RS	R-Feat
M19 4"	17	17	17	17	17
M20 4"	48	48	48	48	48
M20 F	34	34	34	34	34
M20 S	16	16	16	16	16
RAAM S	14	14	14	14	14
VS 1.6 S	0	0	0	0	0
ALL MINES	129	129	129	129	129



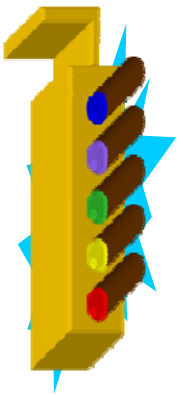
University of Florida

WAAMD Level 1 Eval
CM Overlap 2

- **Countermine Site at YUMA**
- **Images (4 Overlap regions)**
  - 4 AHI images
    - B62R01020030403192710003332r
    - B62R01020030404085200000001r
    - B62R01020030404091520000001r
    - B62R04020030404120906000001r
  - 1 Lynx image
    - B35R20030429221958r1F00vv



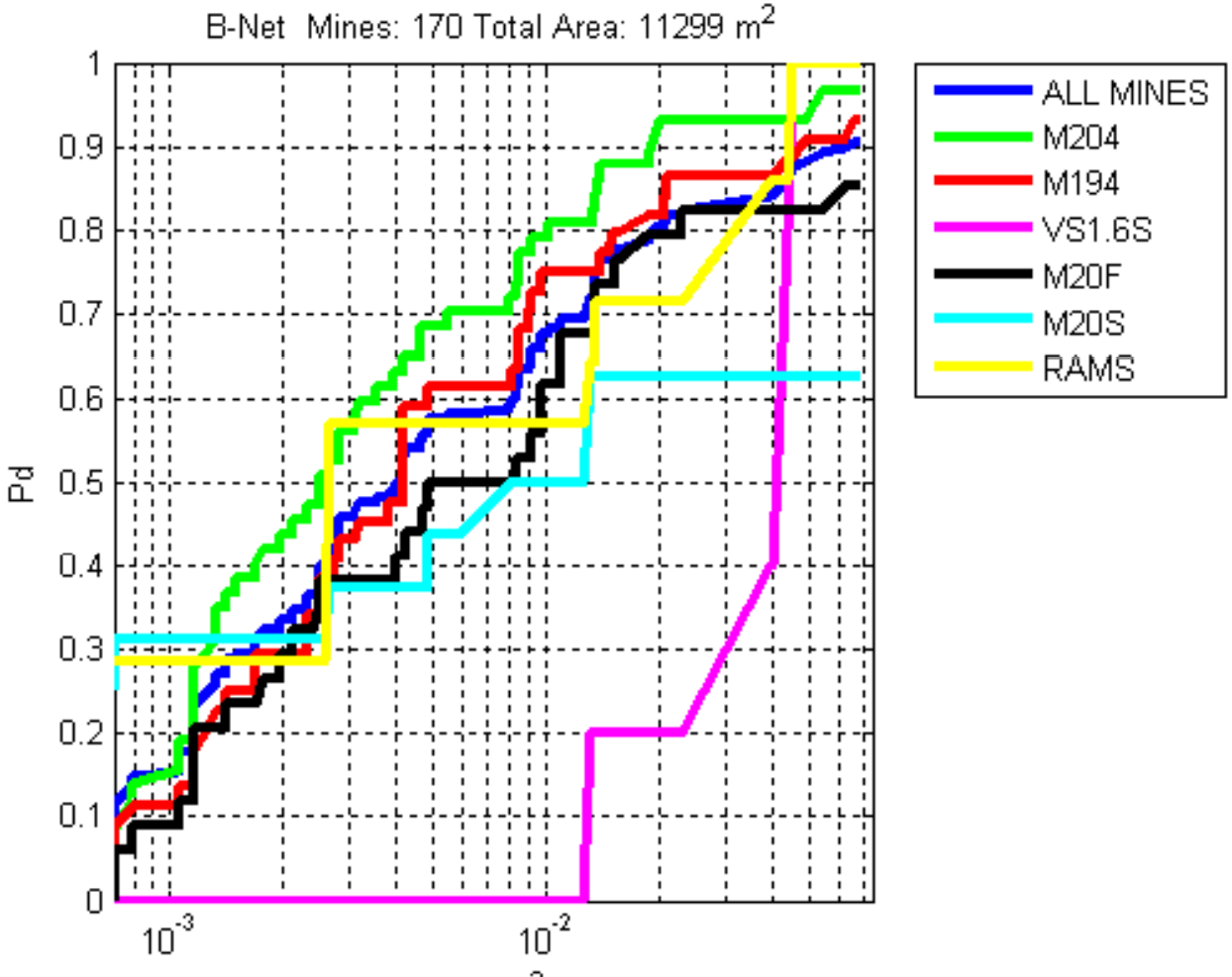
Mine Type	Bayes	Rx	MTD	RS	R-Feat
M19 4"	17	17	17	17	17
M20 4"	26	26	26	26	26
M20 F	20	20	20	20	20
M20 S	16	16	16	16	16
RAAM S	0	0	0	0	0
VS 1.6 S	0	0	0	0	0
ALL MINES	79	79	79	79	79



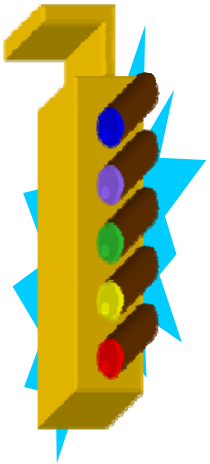
University of Florida

WAAMD Level 1 Eval
CM Overlap 3

- **Countermine Site at YUMA**
- **Images (4 Overlap regions)**
  - 4 AHI images
    - B62R01020030403192710003 32r
    - B62R01020030404085200000 01r
    - B62R01020030404091520000 01r
    - B62R04020030404120906000 01r
  - 1 Lynx image
    - B35R20030429221958r1F00vv



Mine Type	Bayes	Rx	MTD	RS	R-Feat
M19 4"	44	44	44	44	44
M20 4"	57	57	57	57	57
M20 F	34	34	34	34	34
M20 S	16	16	16	16	16
RAAM S	14	14	14	14	14
VS 1.6 S	5	5	5	5	5
ALL MINES	170	170	170	170	170



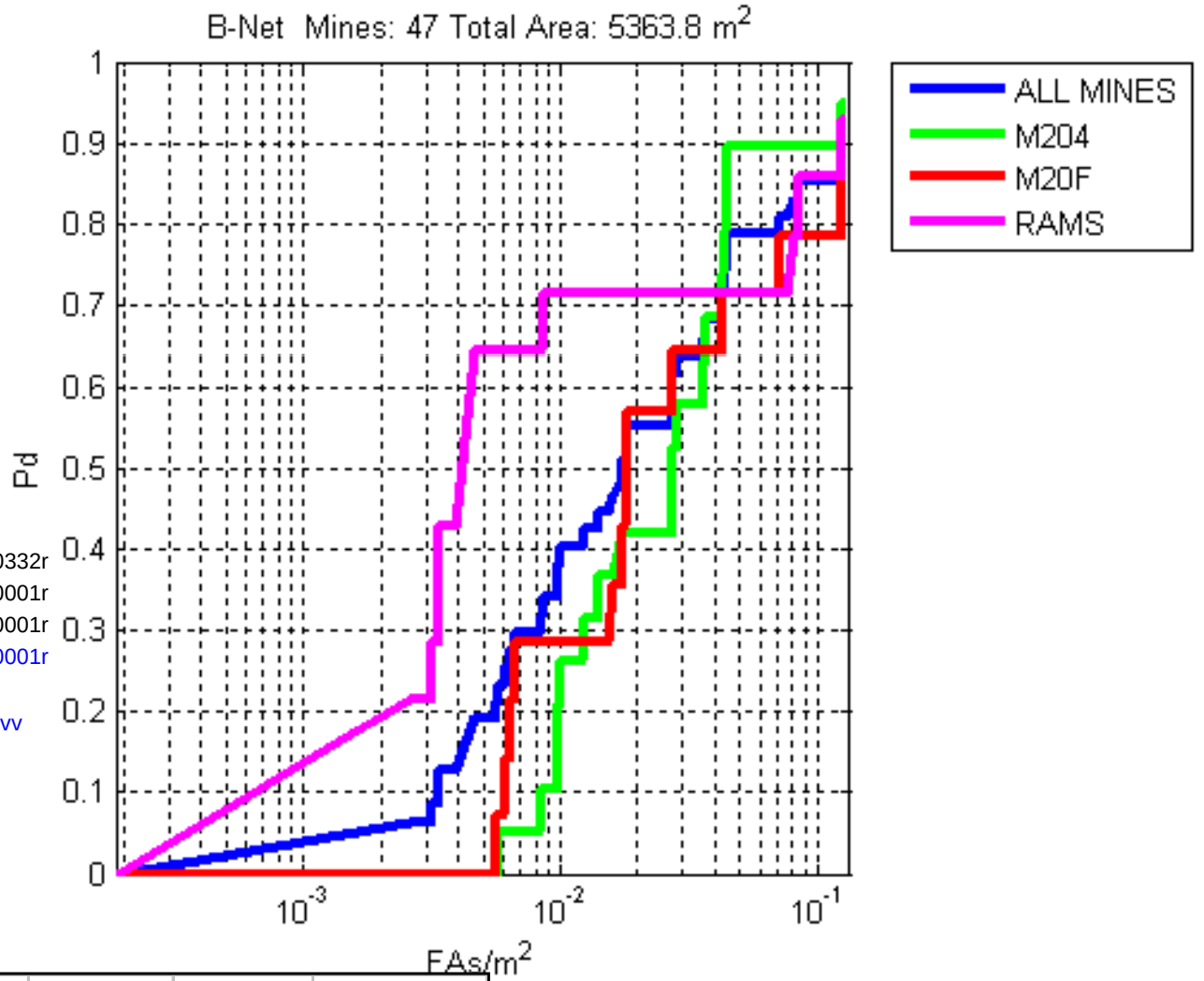
University of Florida

WAAMD Level 1 Eval
CM Overlap 4

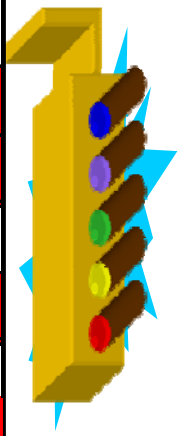
Countermine Site at YUMA

Images (4 Overlap regions)

- 4 AHI images
  - B62R01020030403192710003 32r
  - B62R01020030404085200000 01r
  - B62R01020030404091520000 01r
  - B62R04020030404120906000 01r
- 1 Lynx image
  - B35R20030429221958r1F00vv



Mine Type	Bayes	Rx	MTD	RS	R-Feat
M19 4"	0	0	0	0	0
M20 4"	19	19	19	19	19
M20 F	14	14	14	14	14
M20 S	0	0	0	0	0
RAAM S	14	14	14	14	14
VS 1.6 S	0	0	0	0	0
ALL MINES	47	47	47	47	47



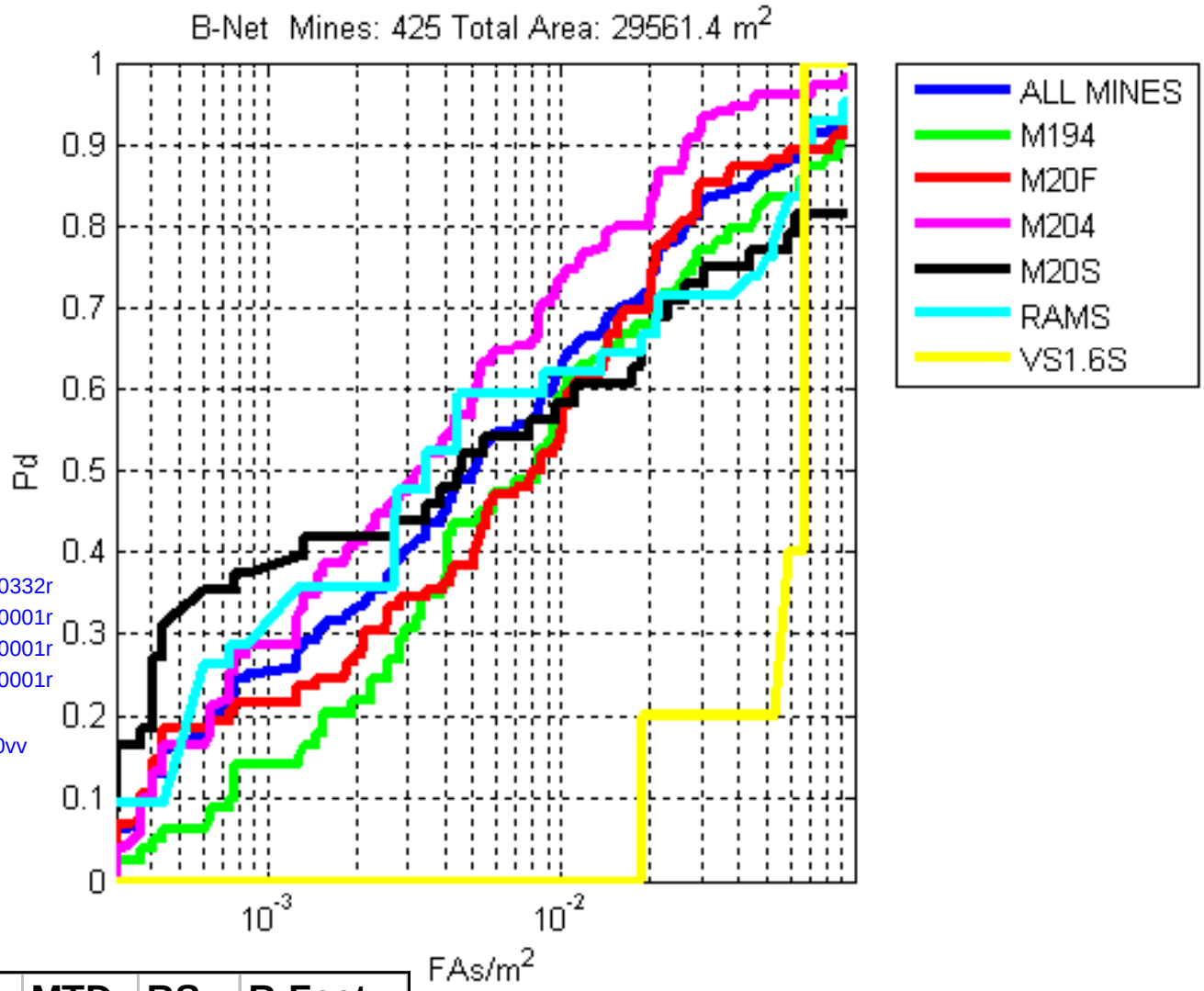
University of Florida

WAAMD Level 1 Eval ALL CM

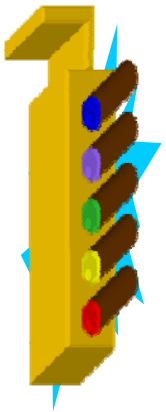
Countermine Site at YUMA

Images (4 Overlap regions)

- 4 AHI images
  - B62R01020030403192710003 32r
  - B62R01020030404085200000 01r
  - B62R01020030404091520000 01r
  - B62R04020030404120906000 01r
- 1 Lynx image
  - B35R20030429221958r1F00vv



Mine Type	Bayes	Rx	MTD	RS	R-Feat
M19 4"	78	78	78	78	78
M20 4"	150	150	150	150	150
M20 F	102	102	102	102	102
M20 S	48	48	48	48	48
RAAM S	42	42	42	42	42
VS 1.6 S	5	5	5	5	5
ALL MINES	425	425	425	425	425



University of Florida

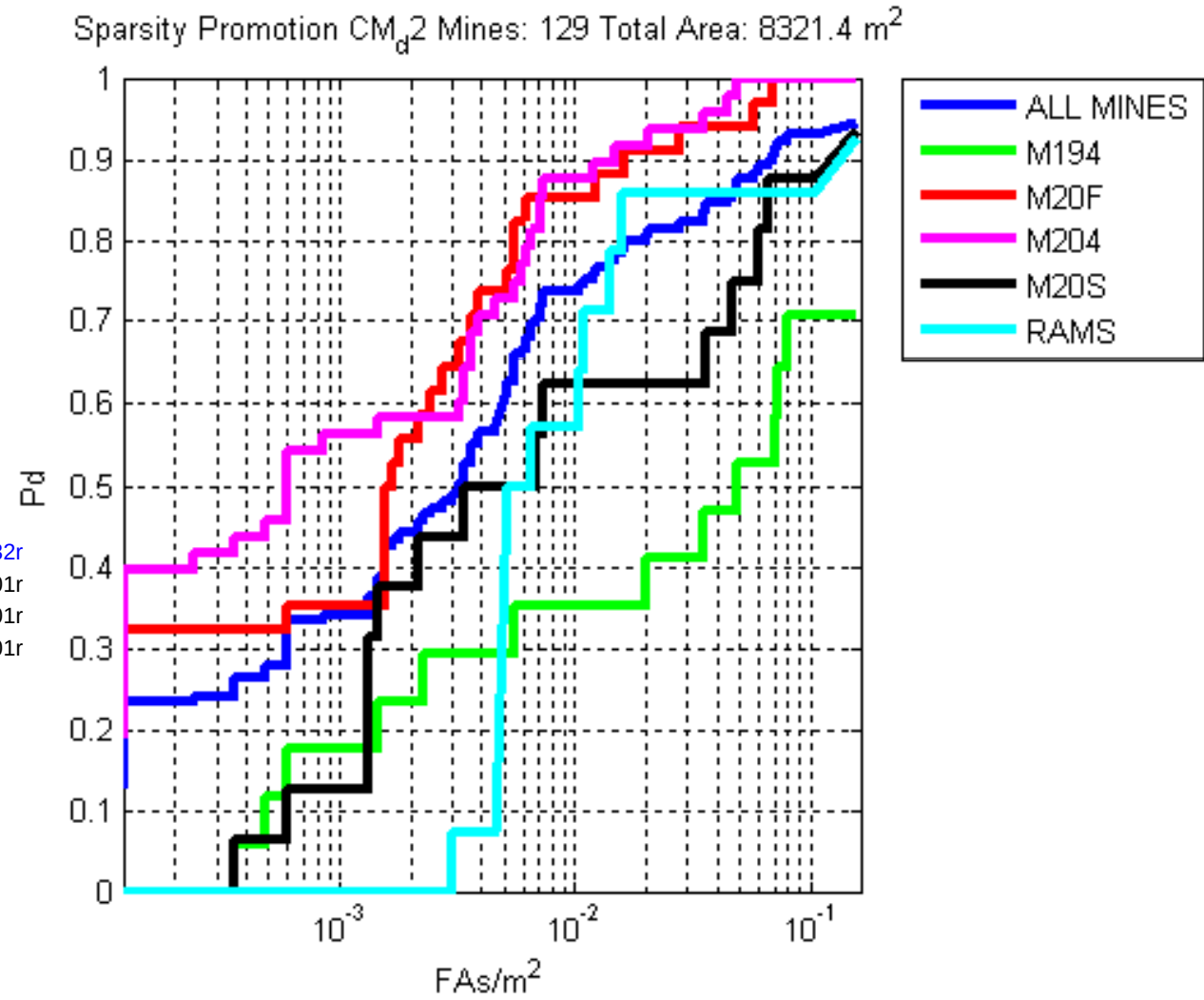
Sparsity Promotion Choquet Fusion Results

- Data was tested using Choquet Fusion with Sparsity Promotion
  - Results were scored using the stop-light criterion
  - Results were compared to previous fusion results

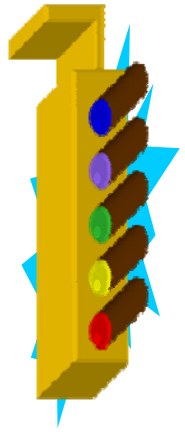
University of Florida

WAAMD Level 1 Eval
CM: Overlap 1

- **Countermine Site at YUMA**
- **Images (4 Overlap regions)**
  - 4 AHI images
    - B62R0102003040319271000332r
    - B62R0102003040408520000001r
    - B62R0102003040409152000001r
    - B62R0402003040412090600001r
  - 1 Lynx image
    - B35R20030429221958r1F00vv

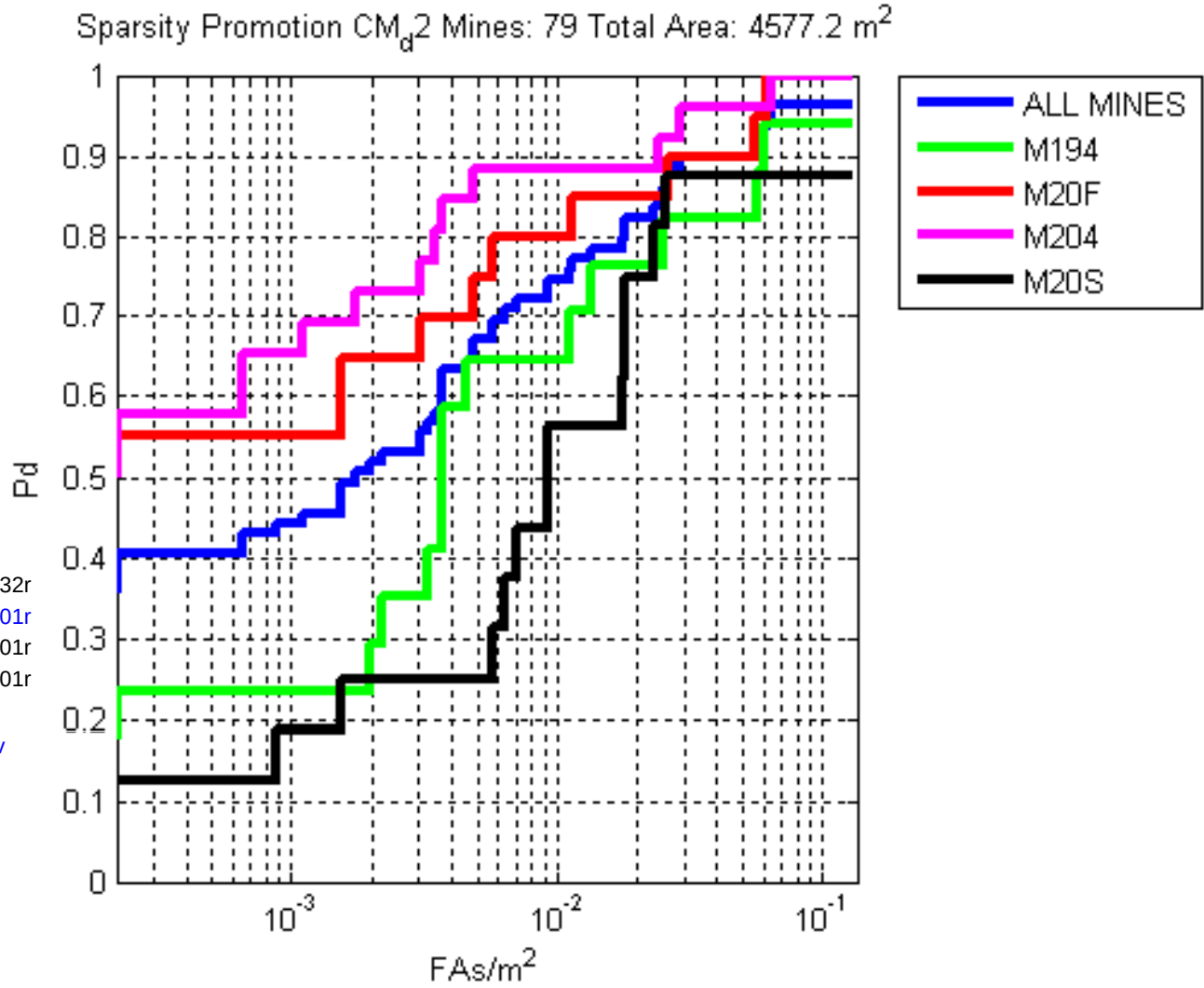


Mine Type	Sp. Choq	Rx	MTD	RS	LGRF Sup	AND
M19 4"	17	17	17	17	17	17
M20 4"	48	48	48	48	48	48
M20 F	34	34	34	34	34	34
M20 S	16	16	16	16	16	16
RAAM S	14	14	14	14	14	14
VS 1.6 S	0	0	0	0	0	0
ALL MINES	129	129	129	129	129	129



University of Florida

WAAMD Level 1 Eval
CM Overlap 2

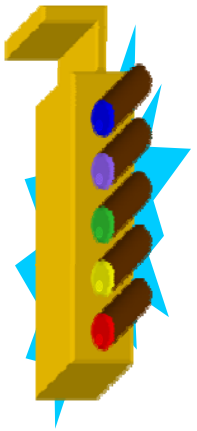


Countermine Site at YUMA

Images (4 Overlap regions)

- 4 AHI images
  - B62R01020030403192710003 32r
  - B62R01020030404085200000 01r
  - B62R01020030404091520000 01r
  - B62R04020030404120906000 01r
- 1 Lynx image
  - B35R20030429221958r1F00vv

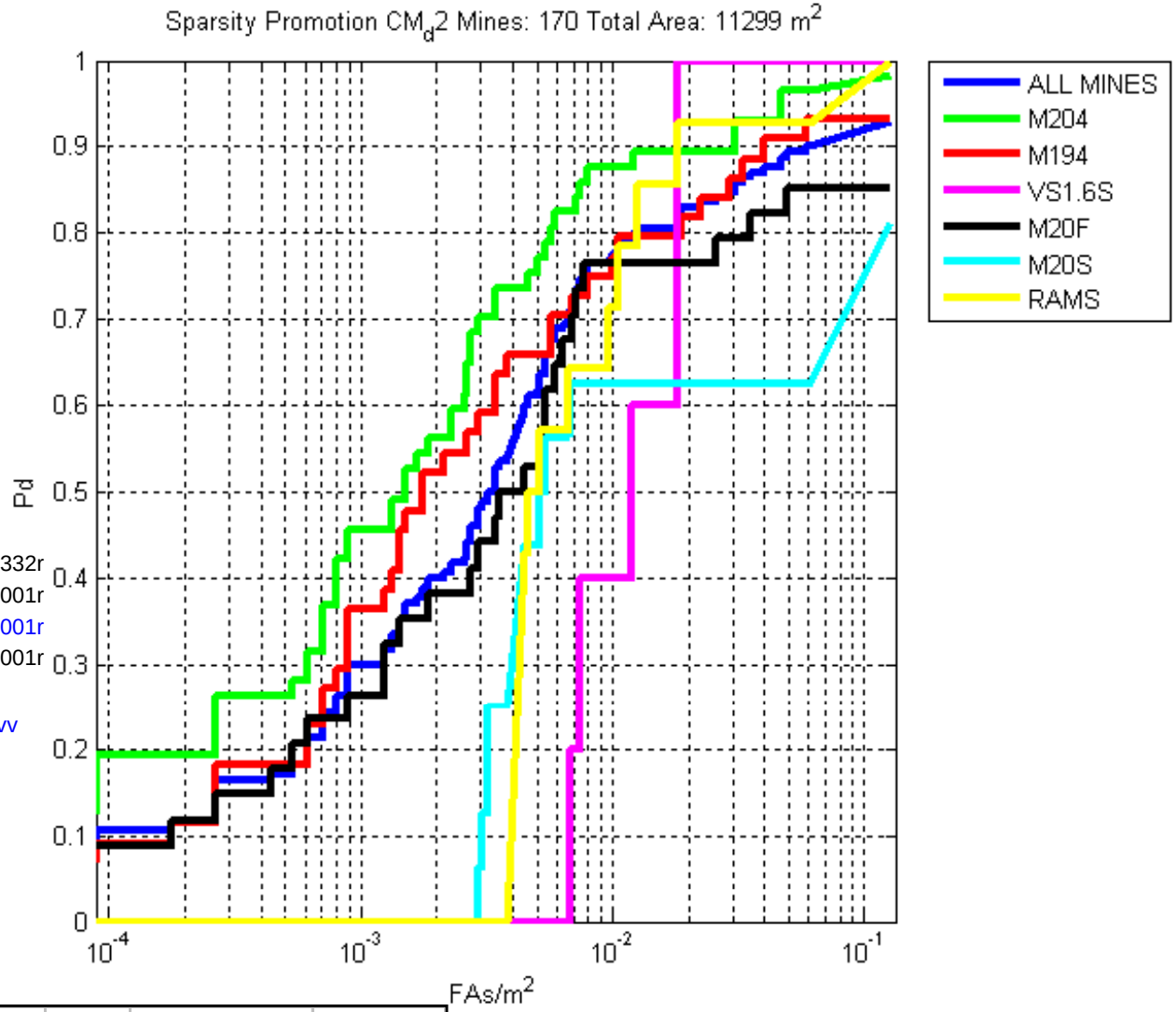
Mine Type	Sp. Choq	Rx	MTD	RS	LGRF Sup	AND
M19 4"	17	17	17	17	17	17
M20 4"	26	26	26	26	26	26
M20 F	20	20	20	20	20	20
M20 S	16	16	16	16	16	16
RAAM S	0	0	0	0	0	0
VS 1.6 S	0	0	0	0	0	0
ALL MINES	79	79	79	79	79	79



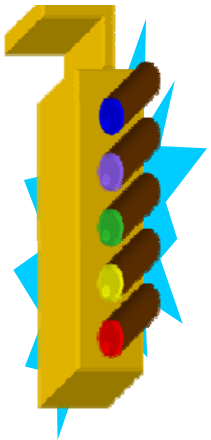
University of Florida

WAAMD Level 1 Eval
CM Overlap 3

- **Countermine Site at YUMA**
- **Images (4 Overlap regions)**
  - 4 AHI images
    - B62R01020030403192710003 32r
    - B62R01020030404085200000 1r
    - B62R01020030404091520000 01r
    - B62R04020030404120906000 01r
  - 1 Lynx image
    - B35R20030429221958r1F00vv

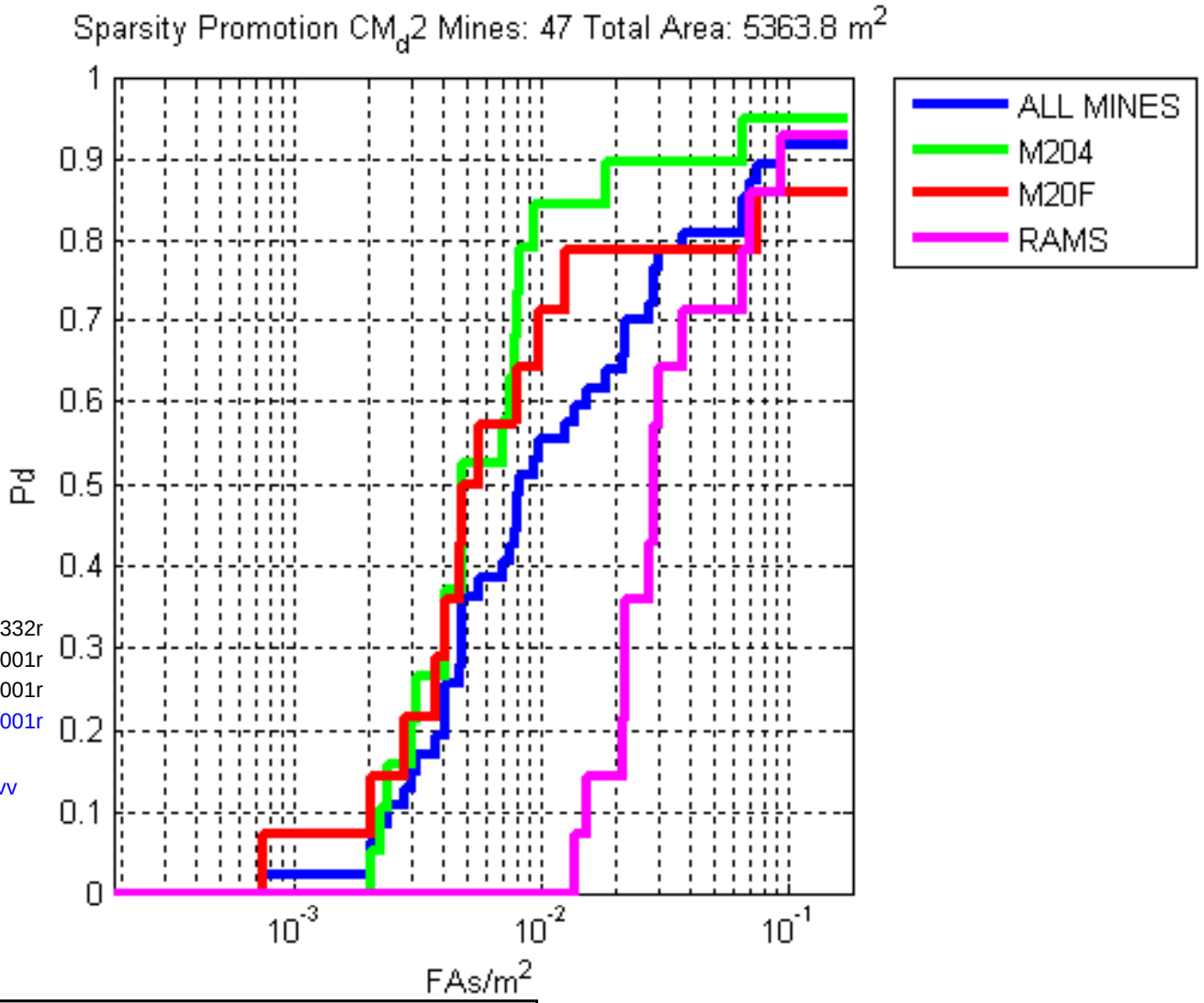


Mine Type	Sp. Choq	Rx	MTD	RS	LGRF Sup	AND
M19 4"	44	44	44	44	44	44
M20 4"	57	57	57	57	57	57
M20 F	34	34	34	34	34	34
M20 S	16	16	16	16	16	16
RAAM S	14	14	14	14	14	14
VS 1.6 S	5	5	5	5	5	5
ALL MINES	170	170	170	170	170	170



University of Florida

WAAMD Level 1 Eval
CM Overlap 4

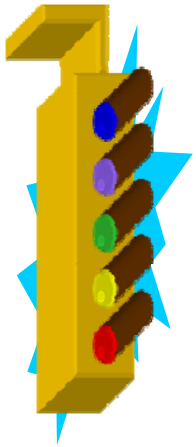


Countermine Site at YUMA

Images (4 Overlap regions)

- 4 AHI images
  - B62R01020030403192710003 32r
  - B62R01020030404085200000 01r
  - B62R01020030404091520000 01r
  - B62R04020030404120906000 01r
- 1 Lynx image
  - B35R20030429221958r1F00vv

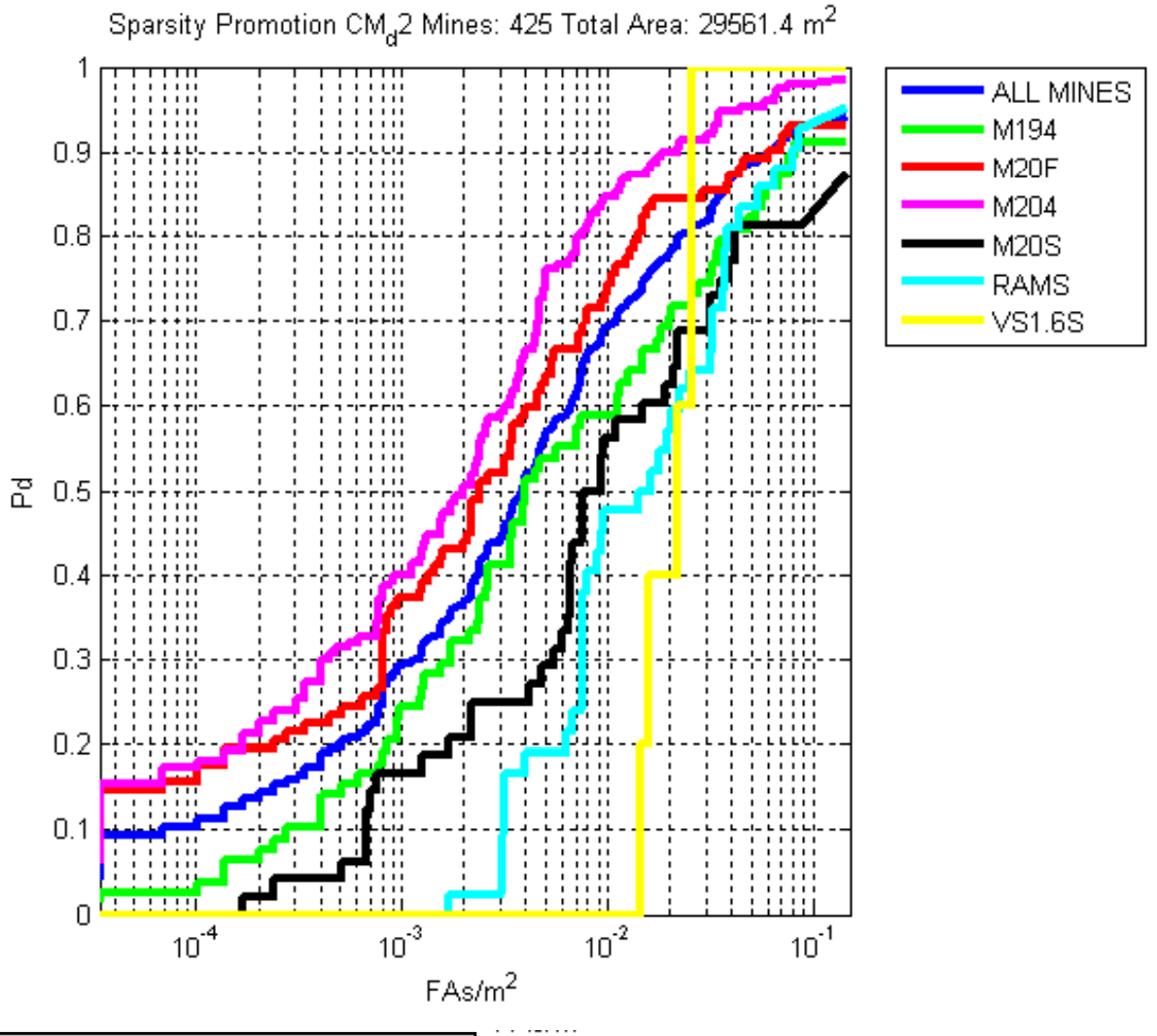
Mine Type	Sp. Choq	Rx	MTD	RS	LGRF Sup	AND
M19 4"	0	0	0	0	0	0
M20 4"	19	19	19	19	19	19
M20 F	14	14	14	14	14	14
M20 S	0	0	0	0	0	0
RAAM S	14	14	14	14	14	14
VS 1.6 S	0	0	0	0	0	0
ALL MINES	47	47	47	47	47	47



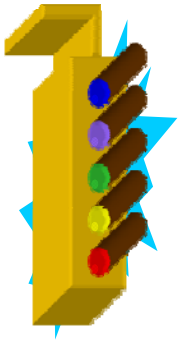
University of Florida

WAAMD Level 1 Eval
ALL CM

- **Countermine Site at YUMA**
- **Images (4 Overlap regions)**
  - 4 AHI images
    - B62R01020030403192710000332r
    - B62R01020030404085200000001r
    - B62R01020030404091520000001r
    - B62R04020030404120906000001r
  - 1 Lynx image
    - B35R20030429221958r1F00vv



Mine Type	Sp. Choq	Rx	MTD	RVM	LGRF Sup	AND
M19 4"	78	78	78	78	78	78
M20 4"	150	150	150	150	150	150
M20 F	102	102	102	102	102	102
M20 S	48	48	48	48	48	48
RAAM S	42	42	42	42	42	42
VS 1.6 S	5	5	5	5	5	5
ALL MINES	425	425	425	425	425	425



University of Florida

Fusion Experiments – Overlap Regions

Mine Distribution

- Mine Types used in Experiments
  - 96 surface
  - 84 flush
  - 206 buried

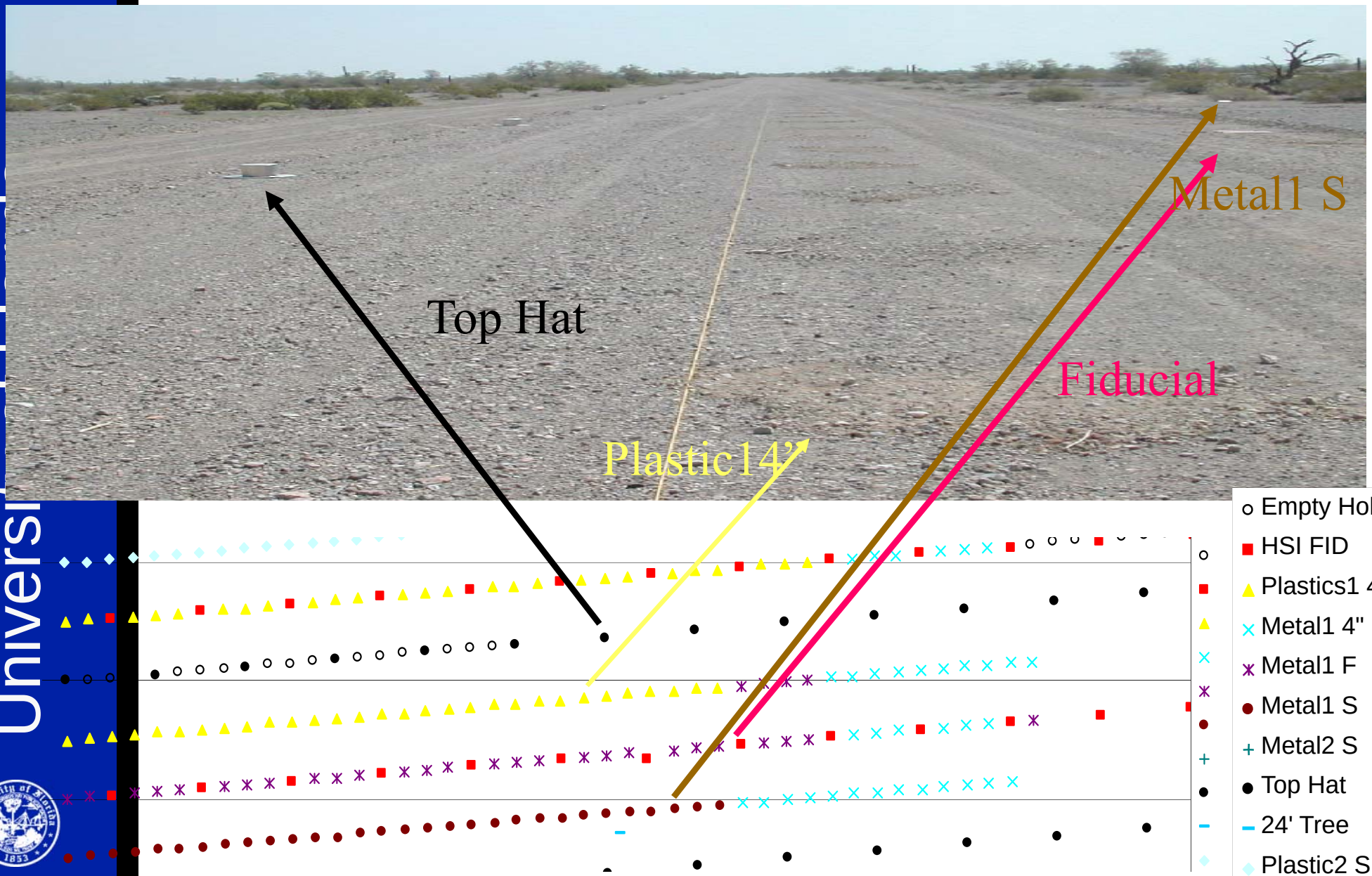
Mines	M19	M20	M20	M20	RAM	VS1.6	**Total**
Burial Depth	4"	4"	Flush	Surface	Surface	Surface	
Overlap Regions:							
1946	44	56	32	16	14	5	167
2535	17	42	24	16	14	0	113
769	0	8	13	1	14	0	36
2349	0	19	15	16	0	0	50
1495	11	9	0	0	0	0	20
All Overlap Regions	72	134	84	49	42	5	386

University of Florida

Experimental Procedure

- Each of the five overlap regions were tested
  - Training was done using the other four overlap regions
  - POIs that were detections of fiducials and holes were removed for training
- Shapley Index was calculated for each detector
  - Calculated using measures optimized from training
  - Is a measure of each detectors “importance”
- Choquet Fusion Results
  - Compared with OR operator
  - Scoring - ROC curves
    - Fiducial, hole, and IR panel detections were counted as FAs

Countermine Image 1616

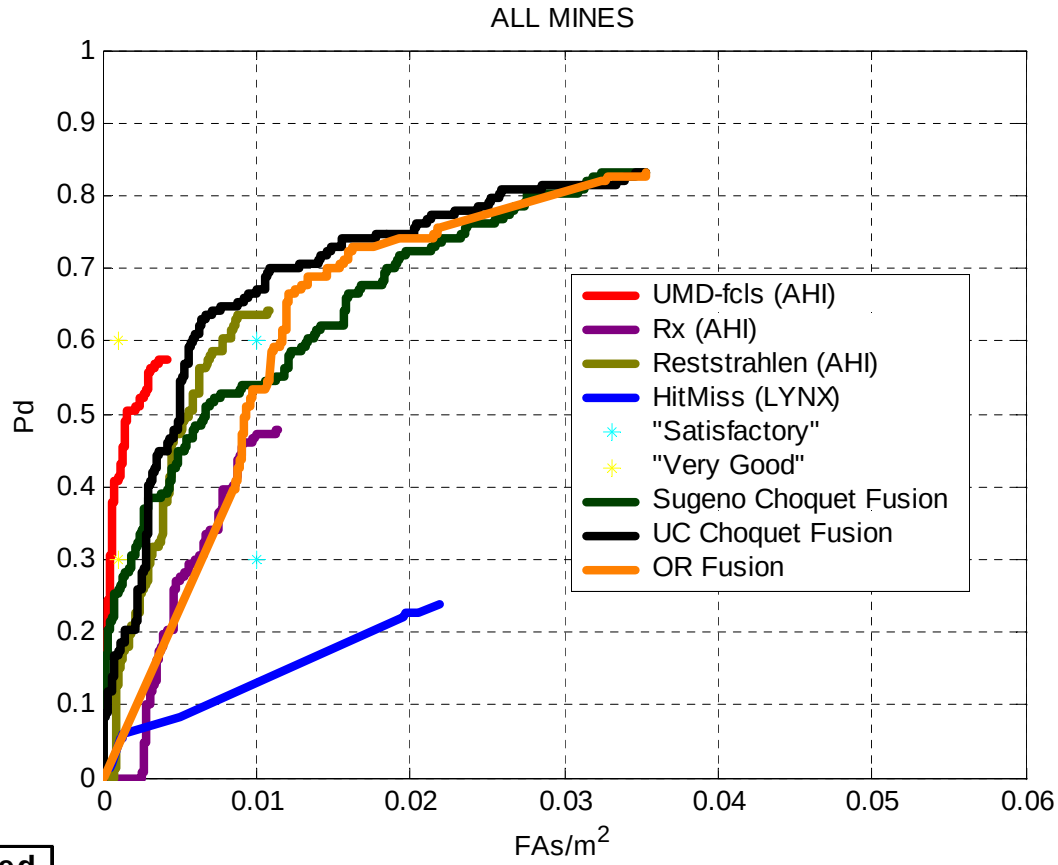


Results 1946

- Lynx image
  - B35R20030429221958r1F00vv
- AHI image
  - B62R01020030404091520019 46r1F00lw
- Info
  - Mine Distribution

Mine	Depth	Quantity
M19	4"	44
M20	4"	56
M20	Flush	32
M20	Surface	16
RAM	Surface	14
VS1.6	Surface	5
Total		167

  - Overlap Area: 6749.6 m^2
  - Shapley Indices



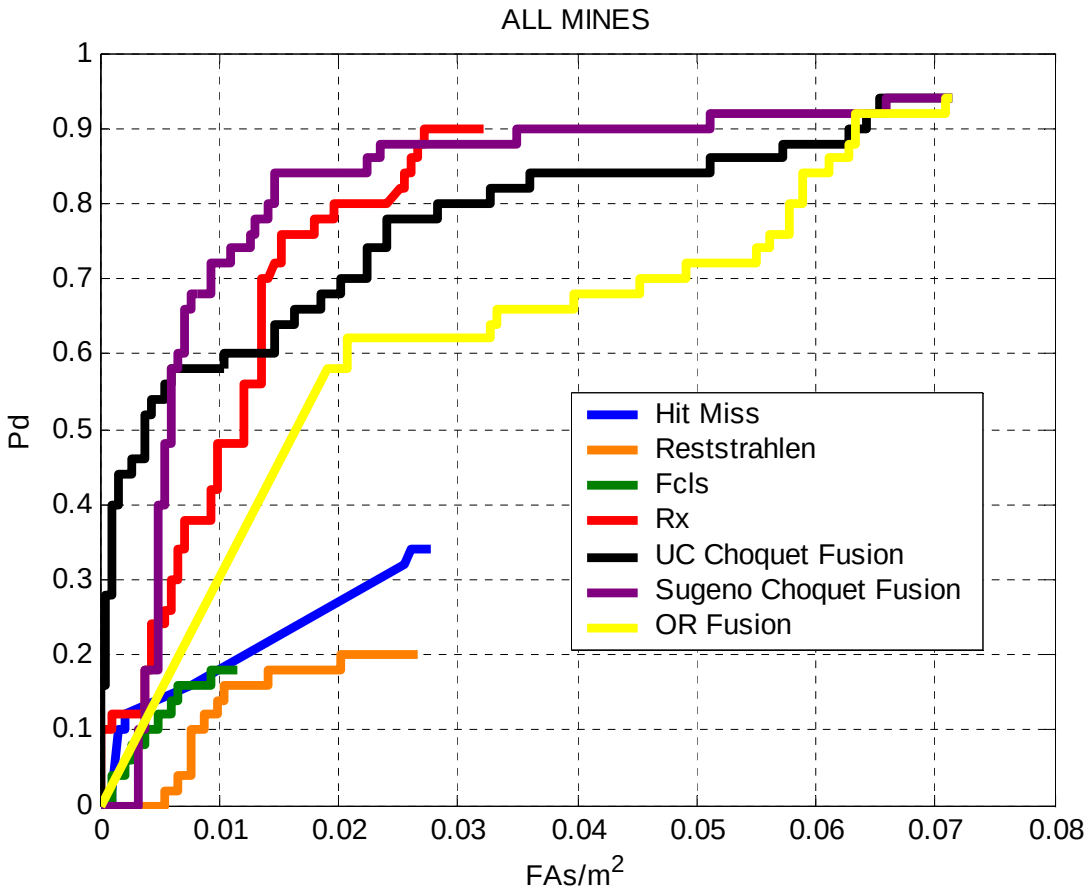
	Sugeno	Unconstrained
1946	**Shapley Index**	**Shapley Index**
Rx	0.57	0.19
Reststrahlen	0.04	0.20
Fcls	0.39	0.13
Hit Miss	0.00	0.16
BB Feature	0.00	0.33

Results 2349

- Lynx image
  - B35R20030429221958r1F00vv
- AHI image
  - B62R01020030404085200023 49r1F00lw
- Info
  - Mine Distribution

Mine	Depth	Quantity
M19	4"	0
M20	4"	19
M20	Flush	15
M20	Surface	16
RAM	Surface	0
VS1.6	Surface	0
Total		50

  - Overlap Area: 1834.5 m^2
  - Shapley Indices



	Sugeno	Unconstrained
2349	**Shapley Index**	**Shapley Index**
Rx	0.33	0.18
Reststrahlen	0.10	0.13
Fcls	0.57	0.17
Hit Miss	0.00	0.18
BB Feature	0.00	0.34

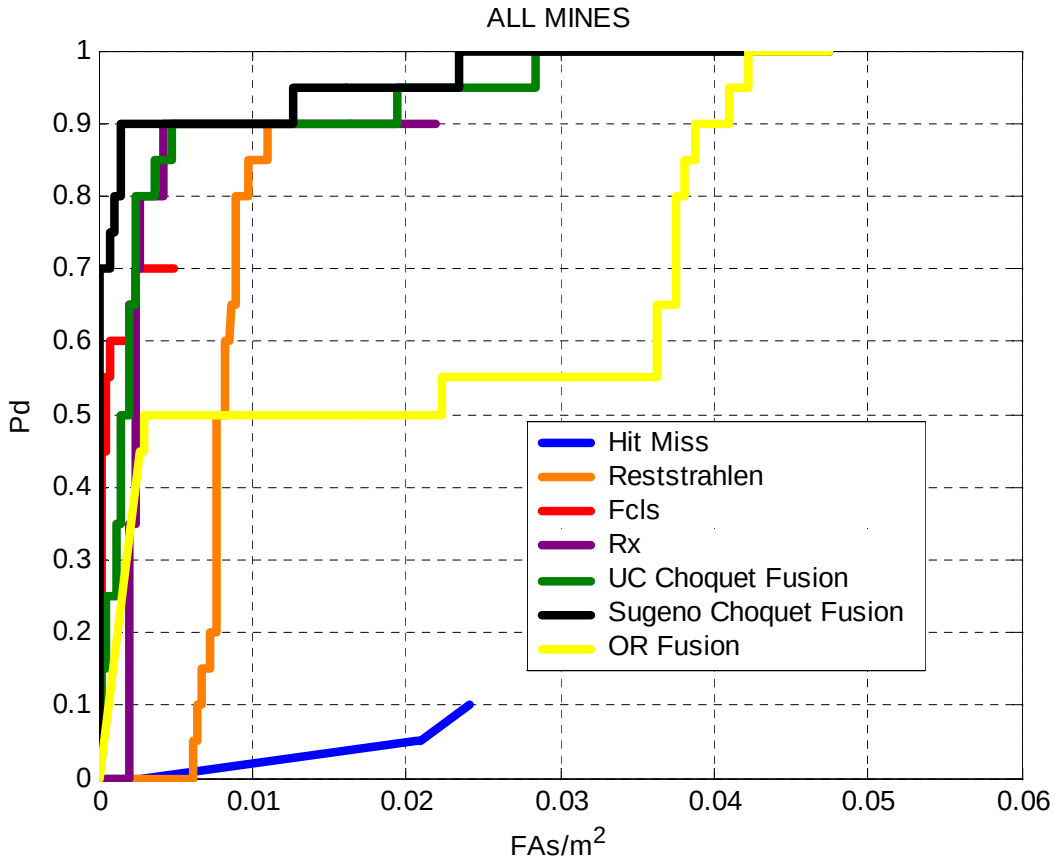
Results 1495

- Lynx image
  - B35R20030429221958r1F00vv
- AHI image
  - B62R020200304031931540 1495r1F00lw
- Info
  - Mine Distribution

Mine	Depth	Quantity
M19	4"	11
M20	4"	9
M20	Flush	0
M20	Surface	0
RAM	Surface	0
VS1.6	Surface	0
Total		20

  - Overlap Area: 4019.4 m^2
  - Shapley Indices

	Sugeno	Unconstrained
1495	**Shapley Index**	**Shapley Index**
Rx	0.61	0.21
Reststrahlen	0.00	0.15
Fcls	0.31	0.14
Hit Miss	0.08	0.18
BB Feature	0.00	0.31

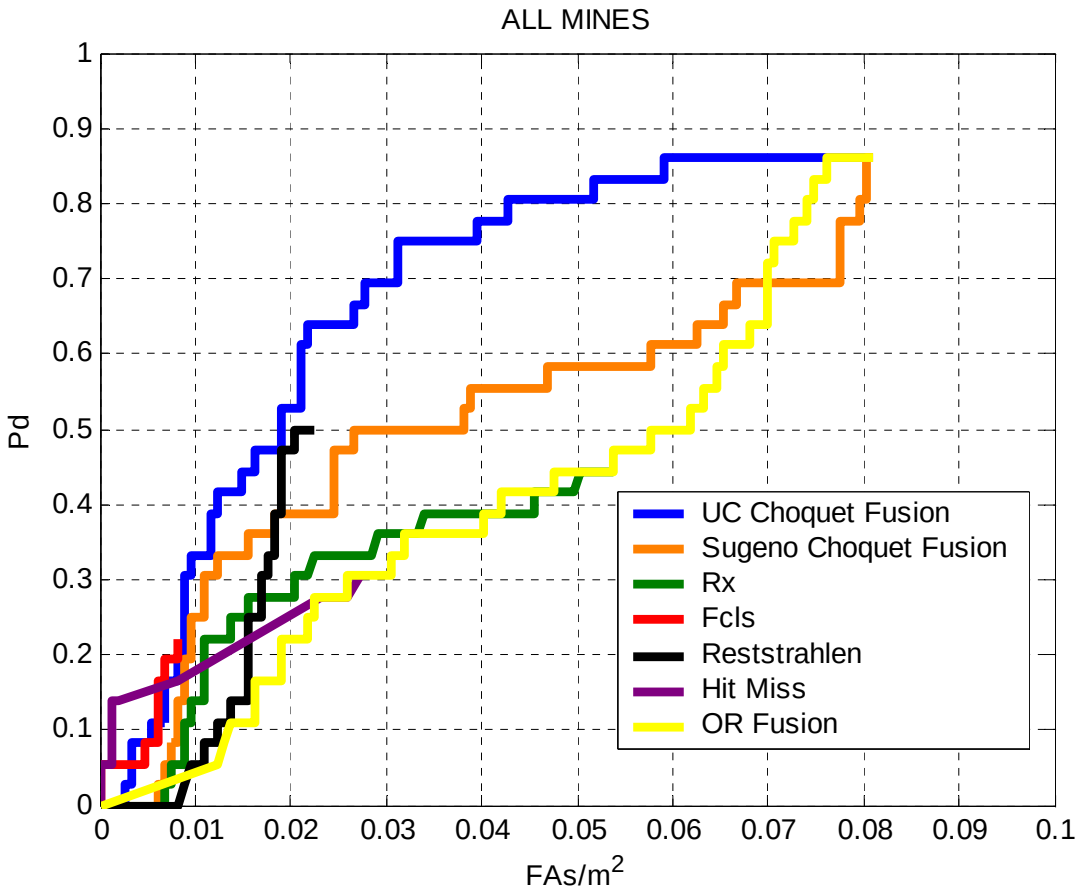


Results 0769

- Lynx image
  - B35R20030429221958r1F00vv
- AHI image
  - B62R0102003040408095500769r1F00lw
- Info
  - Mine Distribution

Mine	Depth	Quantity
M19	4"	0
M20	4"	8
M20	Flush	13
M20	Surface	1
RAM	Surface	14
VS1.6	Surface	0
Total		36

  - Overlap Area: 1470.9 m^2
  - Shapley Indices



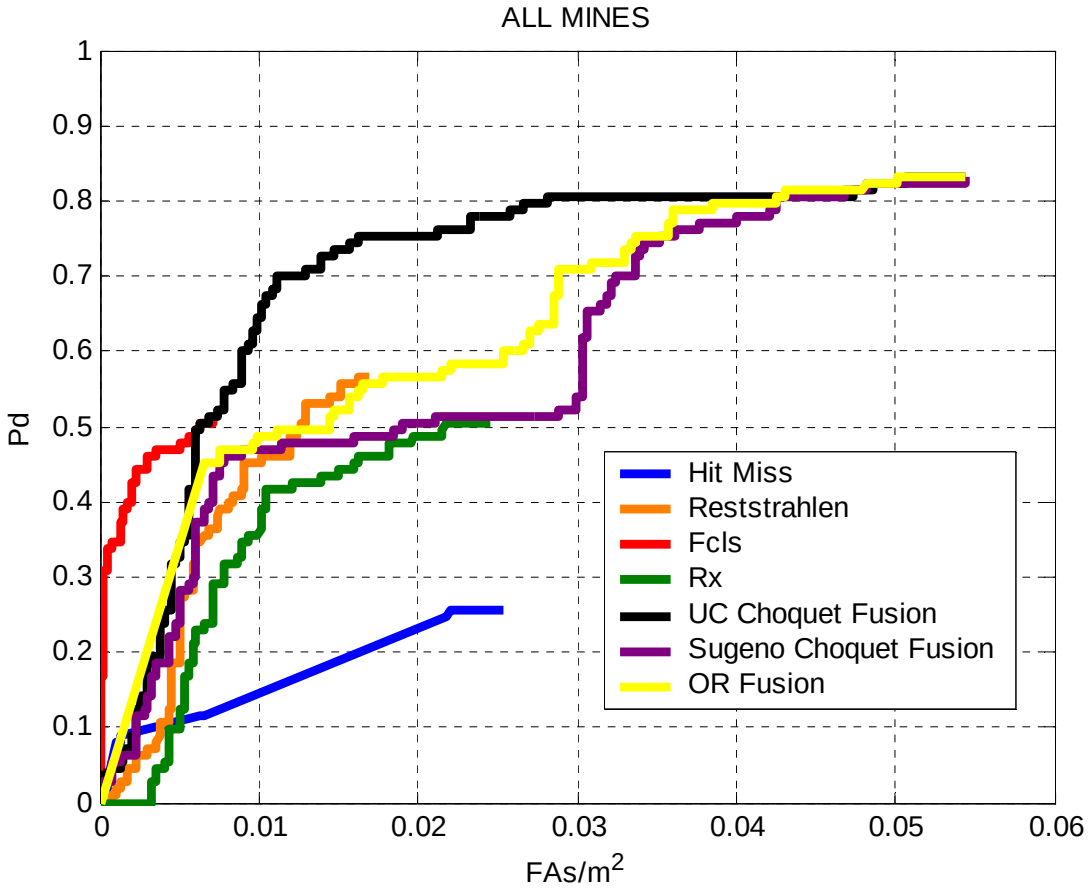
	Sugeno	Unconstrained
769	**Shapley Index**	**Shapley Index**
Rx	0.58	0.11
Reststrahlen	0.01	0.31
Fcls	0.41	0.10
Hit Miss	0.00	0.12
BB Feature	0.00	0.36

Results 2535

- Lynx image
  - B35R20030429221958r1F00vv
- AHI image
  - B62R010200304031927100253 5r1F00lw
- Info
  - Mine Distribution

Mine	Depth	Quantity
M19	4"	17
M20	4"	42
M20	Flush	24
M20	Surface	16
RAM	Surface	14
VS1.6	Surface	0
Total		113

  - Overlap Area: 3948.9 m^2
  - Shapley Indices



	Sugeno	Unconstrained
2535	**Shapley Index**	**Shapley Index**
Rx	0.70	0.16
Reststrahlen	0.02	0.20
Fcls	0.29	0.11
Hit Miss	0.00	0.16
BB Feature	0.00	0.37

University of Florida

Fusion Experiments Multiple Overlap Regions

- **Countermine Site at YUMA**
- **Image Overlap Regions**
  - 5 AHI images
  - 1 Lynx image
  - 2 Mirage images
- **Detectors and Features Used**
  - Rx (HSI)
  - Reststrahlen Ratio (HSI)
  - FCLS (HSI)
  - Hit Miss (SAR)
  - Blackbody Feature (HSI)
- **Choquet Fusion**
  - Unconstrained Measure
  - Sugeno Measure
  - OWA Measure

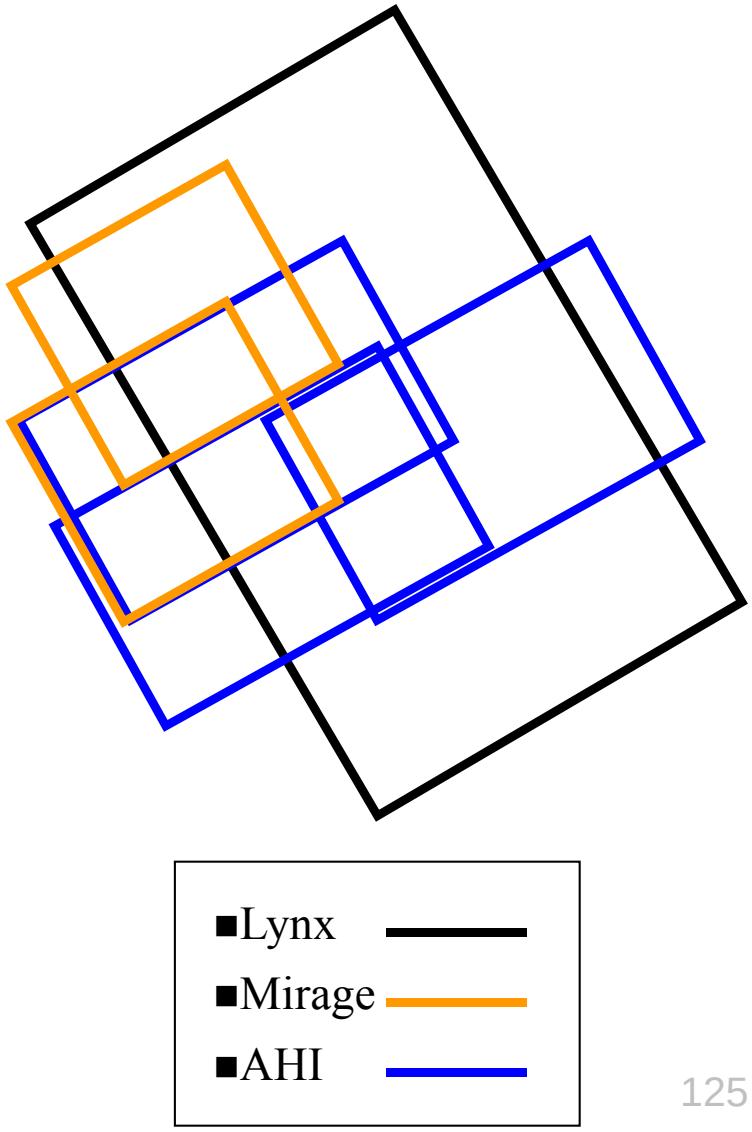
University of Florida

Experiments – Multiple Overlap Regions

- Each of the overlap regions were tested
  - Training was done using the other overlap regions consisting of the same SAR image
  - POIs that were detections of fiducials and holes were removed for training
- Shapley Index was calculated for each detector
  - Is a measure of each detectors “importance”
  - Calculated using measures optimized from training
- Choquet Fusion Results
  - Compared with OR and MEAN operators
  - Scoring - ROC curves
    - Fiducial, hole, and IR panel detections were counted as FAs

How We Combine Results From Multiple Overlap Regions for Scoring

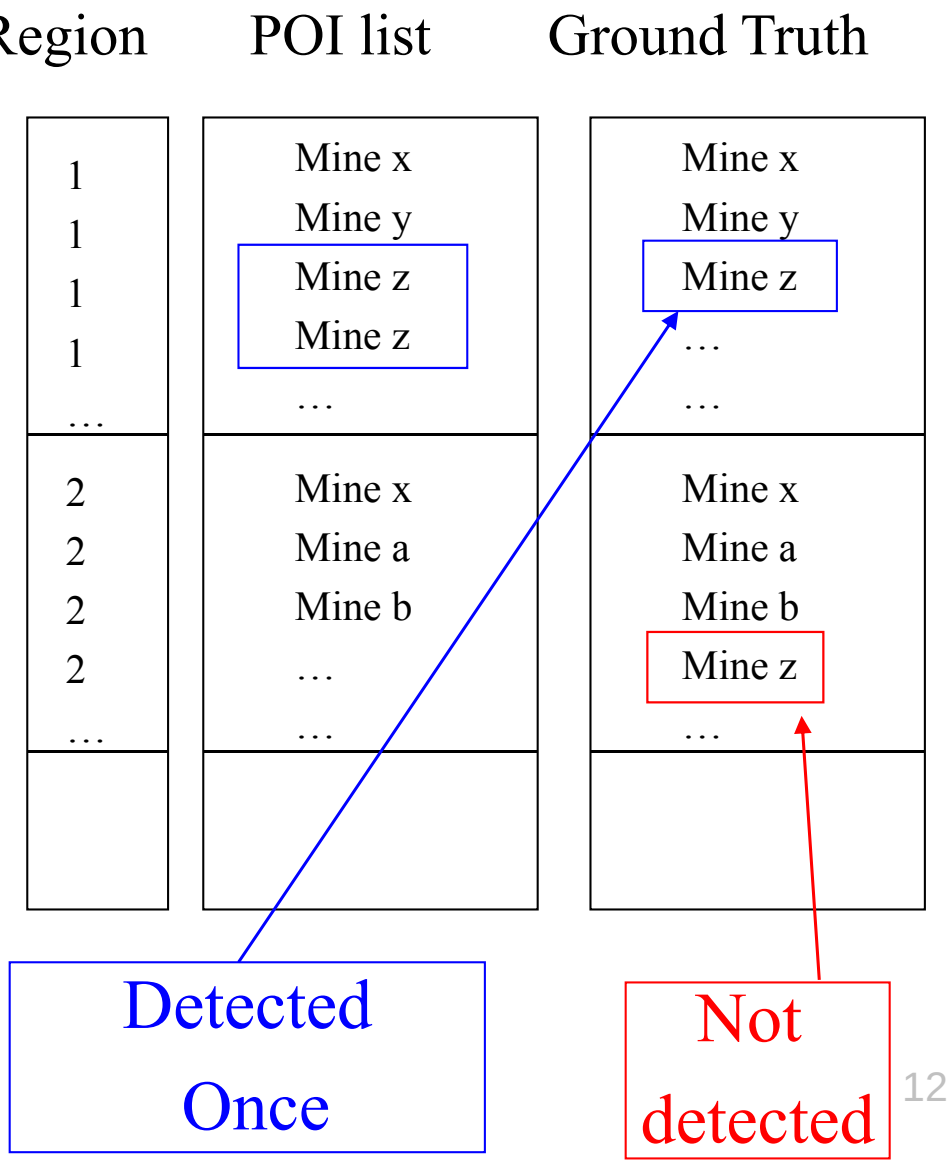
- Many Hyperspectral and SAR images overlapped with each other
  - Result:
    - Many encounters of the same mine were found in different overlap regions
    - How do we score the situation where one overlap region finds the mine and another does not?
      - Treat the multiple encounters of the same mine as completely different mines



University of Florida

How We Scored Results

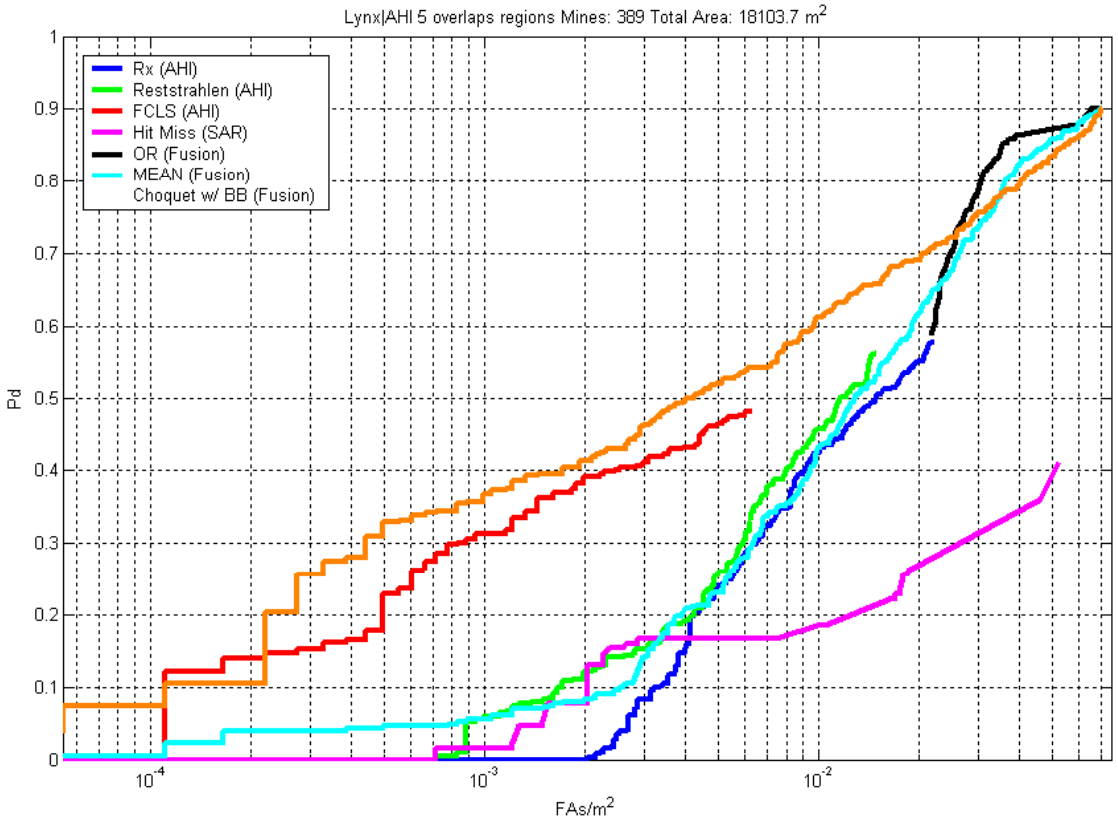
- Concatenate POI and Ground Truth lists (Not Union)
  - Assists in keeping separate mine encounters, of the same mine, separate
  - Since all mines from each region are distinct, the total area is the sum of the individual regions' areas
- Example:



Two Mine z detections from Region 1 can only detect Mine z from the ground truth corresponding to Region 1.
→ Pd = 1/2

Lynx / AHI Overlap Regions

- Countermine Site at YUMA
- Images (5 Overlap regions)
  - 5 AHI images
    - ☑ B62R01020030403192710025
    - ☑ B62R01020030404080955007
    - ☑ B62R01020030404085200023
    - ☑ B62R01020030404091520019
    - ☑ B62R02020030403193154014
  - 1 Lynx image
    - ☑ B35R20030429221958r



Mine	Depth	Quantity
M19	4"	72
M20	4"	134
M20	Flush	87
M20	Surface	49
RAM	Surface	42
VS1.6	Surface	5
Total		389

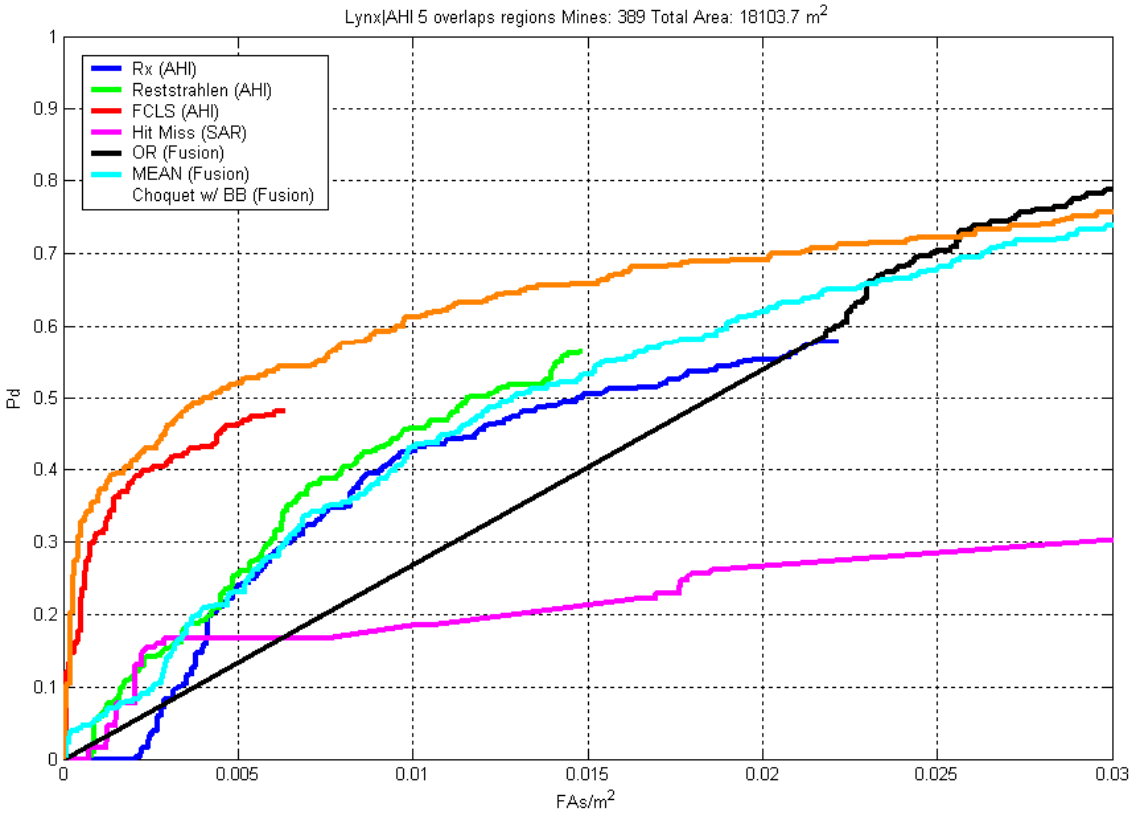
University of Florida

University of Florida

Lynx / AHI Overlap Regions

- Countermine Site at YUMA
- Images (5 Overlap regions)
  - 5 AHI images
    - ☑ B62R0102003040319271002535r
    - ☑ B62R010200304040809550076 9r
    - ☑ B62R010200304040852000234 9r
    - ☑ B62R010200304040915200194 6r
    - ☑ B62R020200304031931540149 5r
  - 1 Lynx image
    - ☑ B35R20030429221958r

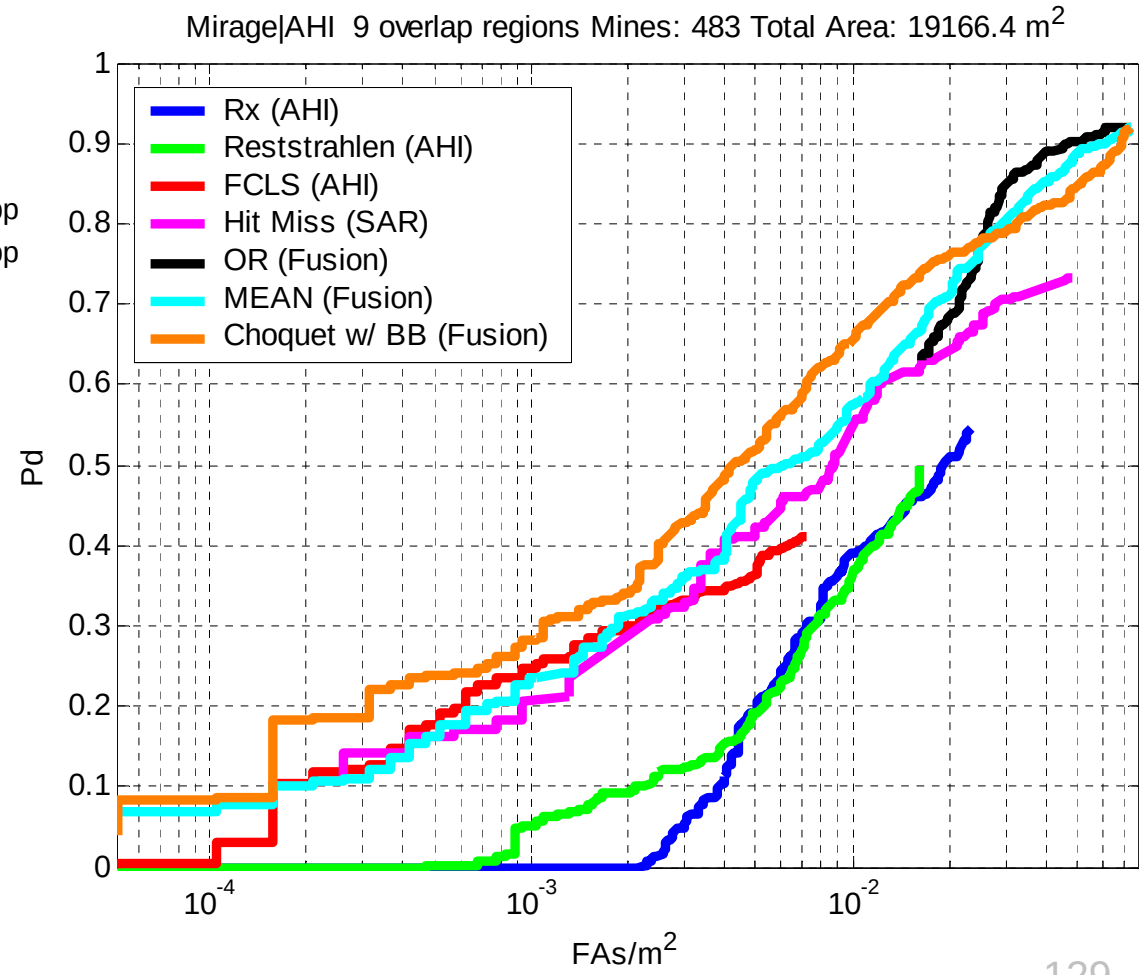
Mine	Depth	Quantity
M19	4"	72
M20	4"	134
M20	Flush	87
M20	Surface	49
RAM	Surface	42
VS1.6	Surface	5
Total		389



Mirage / AHI Overlap Regions

- Countermine Site at YUMA
- Images (5 Overlap regions)
  - 5 AHI images
    - ☑ B62R0102003040319271002535r
    - ☑ B62R0102003040408095500769r
    - ☑ B62R0102003040408520002349r
    - ☑ B62R0102003040409152001946r
    - ☑ B62R0202003040319315401495r
  - 2 Mirage images
    - ☑ B50R2003040406380004r000F00pp
    - ☑ B50R2003040406380002r000F00pp

Mine	Depth	Quantity
M19	4"	90
M20	4"	124
M20	Flush	119
M20	Surface	73
RAM	Surface	72
VS1.6	Surface	5
Total		483



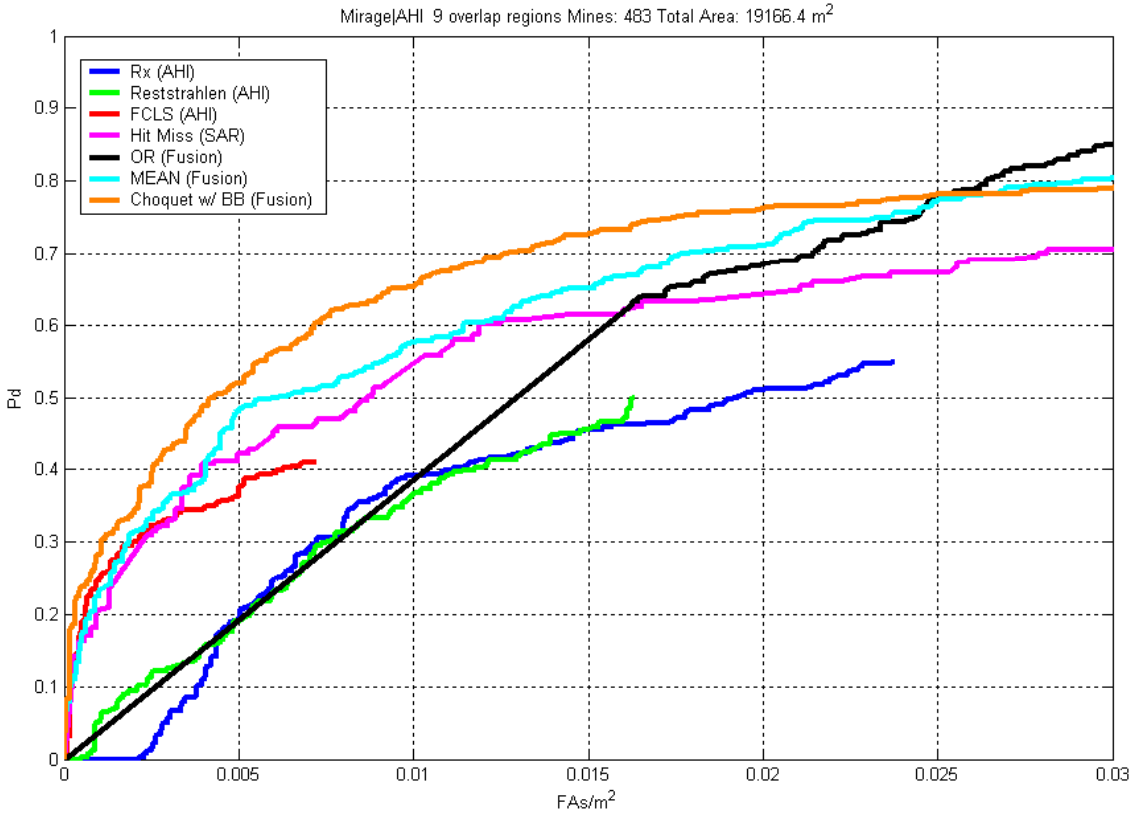
University of Florida

University of Florida

Mirage / AHI Overlap Regions

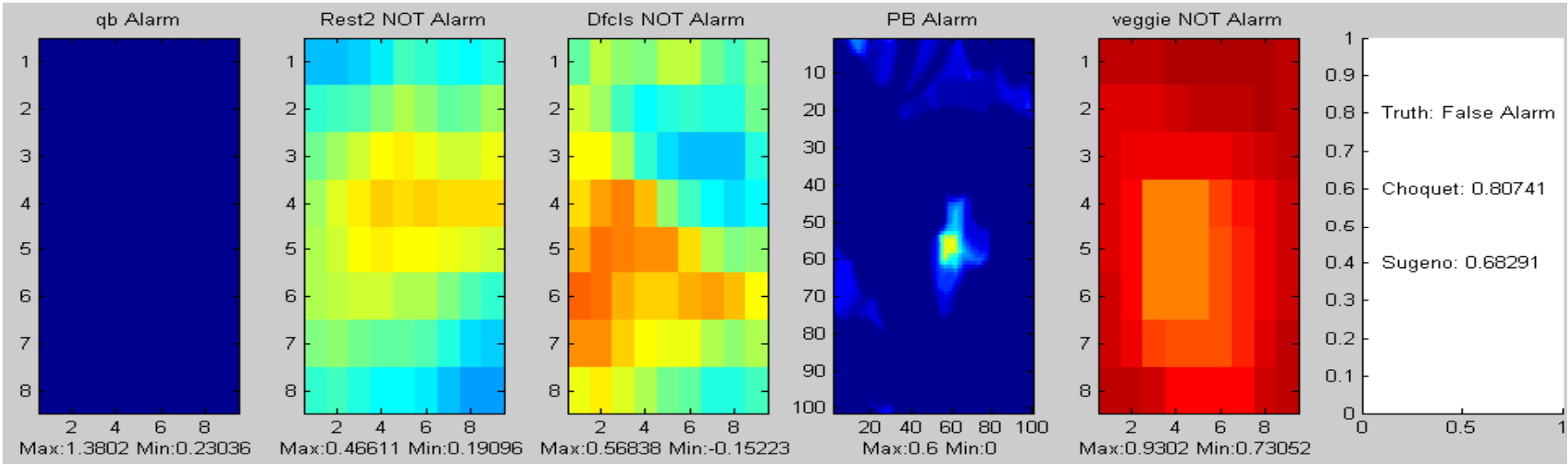
- Countermine Site at YUMA
- Images (5 Overlap regions)
  - 5 AHI images
    - ☑ B62R0102003040319271002535r
    - ☑ B62R010200304040809550076 9r
    - ☑ B62R010200304040852000234 9r
    - ☑ B62R010200304040915200194 6r
    - ☑ B62R020200304031931540149 5r
  - 2 Mirage images
    - ☑ B50R2003040406380004r000F00pp
    - ☑ B50R2003040406380002r000F00pp

Mine	Depth	Quantity
M19	4"	90
M20	4"	124
M20	Flush	119
M20	Surface	73
RAM	Surface	72
VS1.6	Surface	5
Total		483

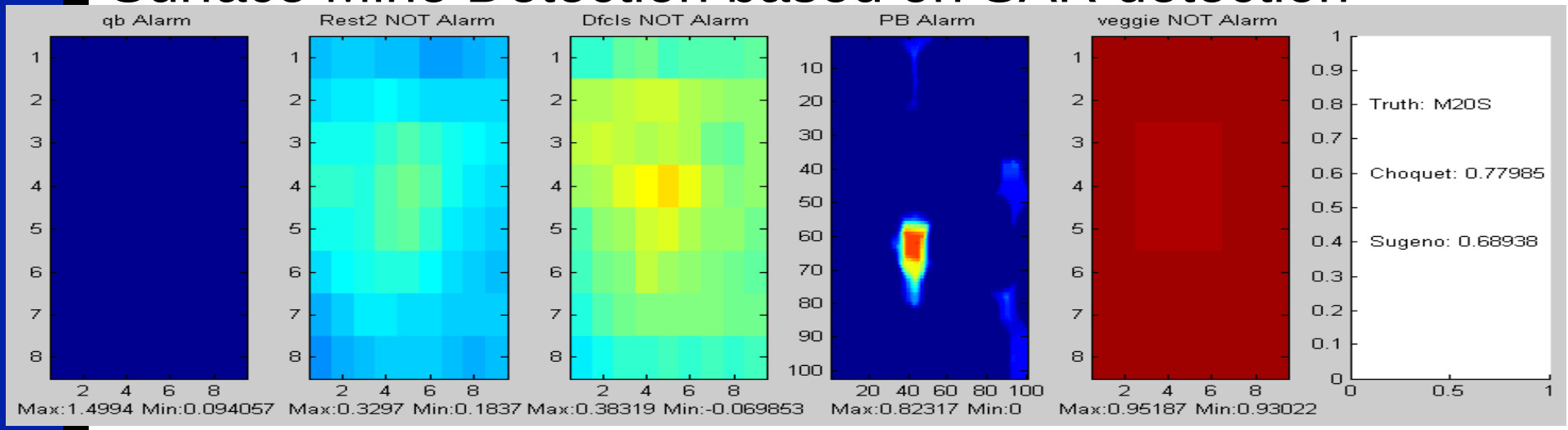


Examples

False Alarm Reduced Confidence via BB Feature



Surface Mine Detection based on SAR detection

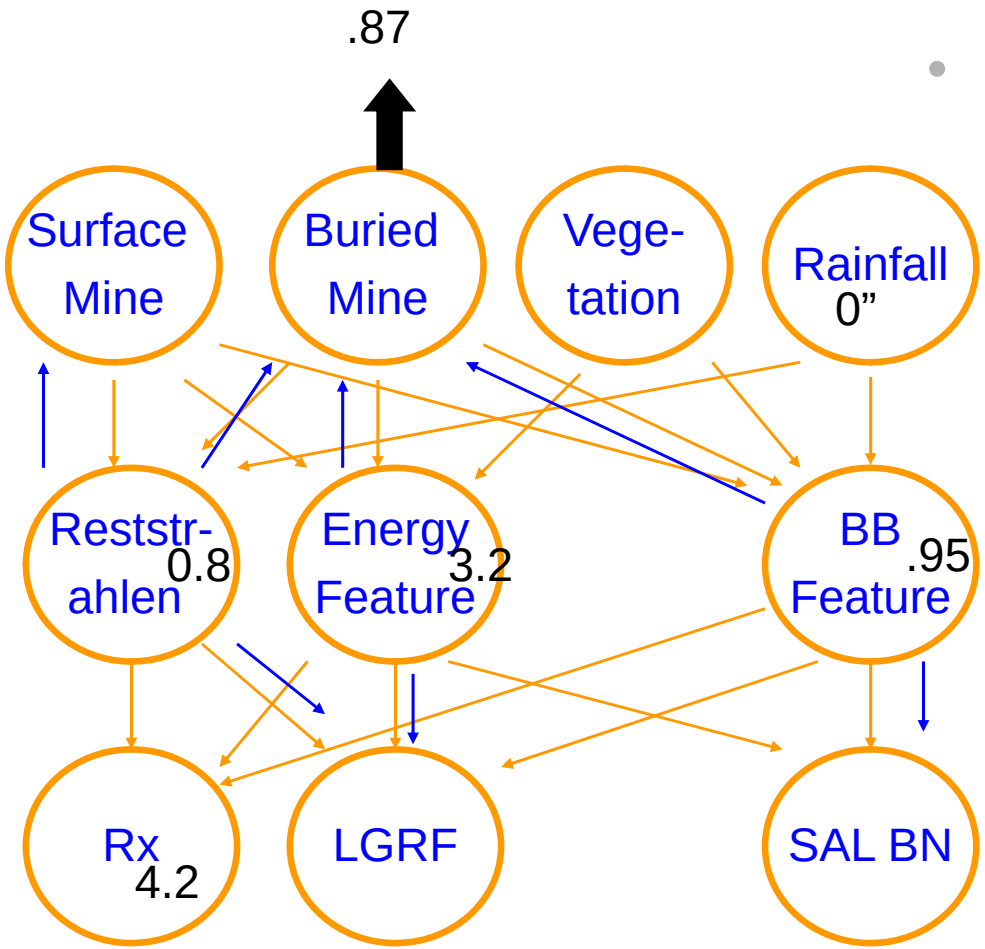


University of Florida

University of Florida

ROC Based Optimization

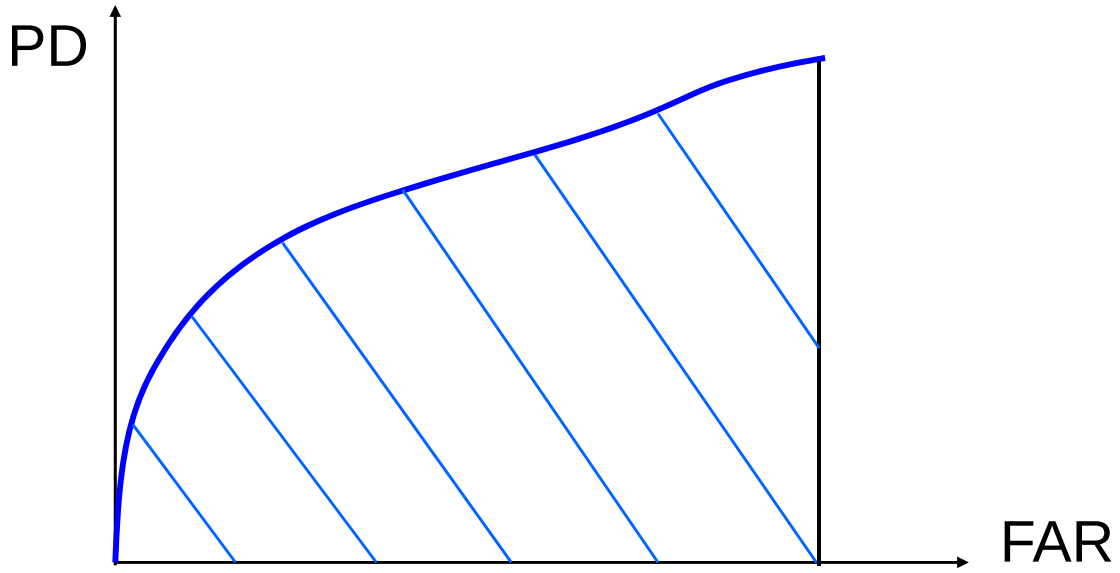
Fusion Method Under Investigation



- Bayesian Networks
  - Statistical model consisting of DAG and conditional densities
  - Network’s structure is based on *Causal* relationship of the nodes
  - Performs inference through message passing algorithm
  - Do not need all or certain features or decision statistics (just whatever we have for each POI)

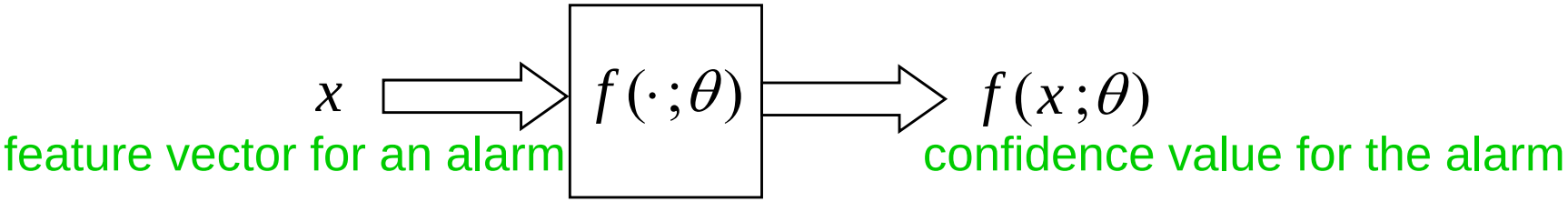
ROC Based Optimization

- Performance is evaluated by using ROC curves – generally the larger the area below the ROC, the better
- Therefore, we developed an algorithm for learning parameters that maximize the area under the ROC



University of Florida

ROC Based Optimization (cont.)



training data

$x^i, i = 1,2,\cdots,M$

M: number of mines

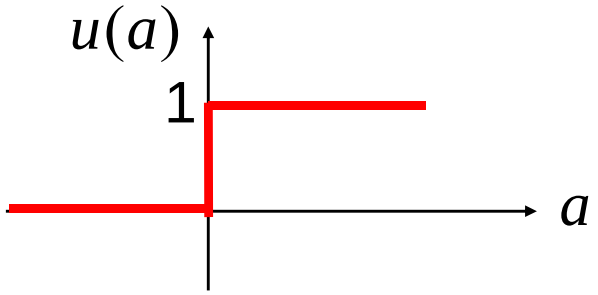
$y^j, j = 1,2,\cdots,N$

N: number of non-mines

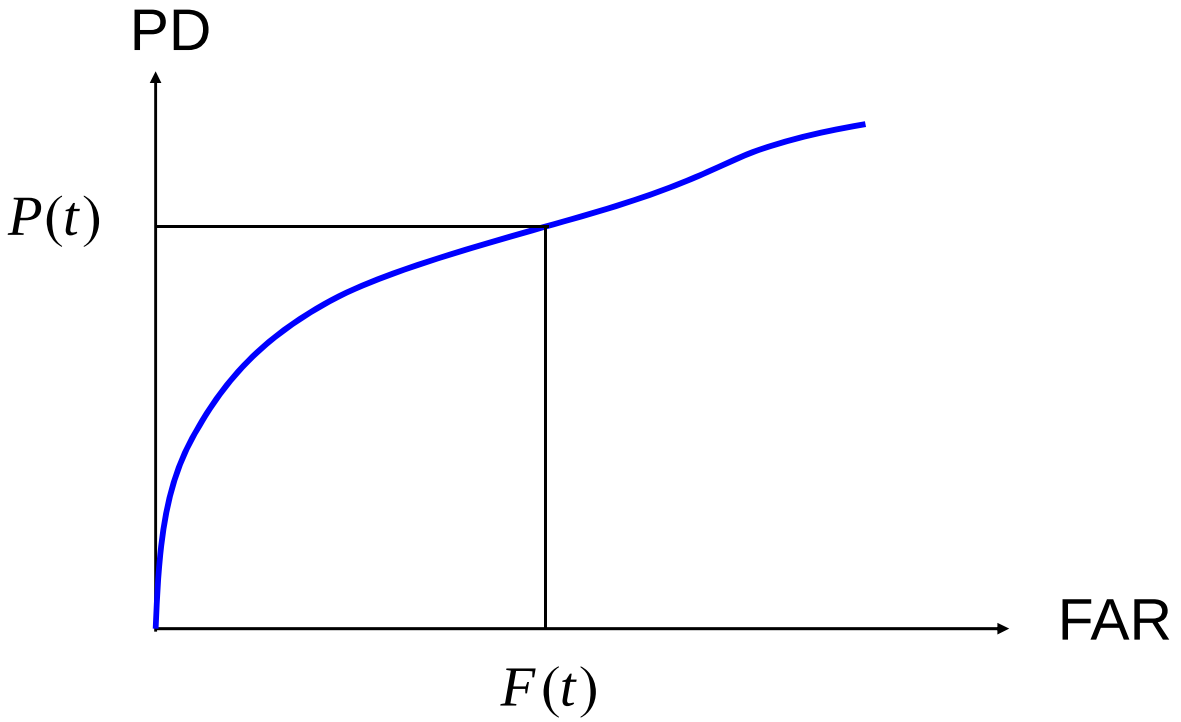
At threshold t,

$$\text{PD} = P(t) = \frac{1}{M}\sum_{i=1}^{M} u(f(x^i;\theta) - t)$$

$$\text{FAR} = F(t) = \frac{1}{\text{total area}}\sum_{j=1}^{N} u(f(y^j;\theta) - t)$$



ROC Based Optimization (cont.)



Area below the curve $= J = \int_0^\infty \mathrm{PD}\,|_{\mathrm{FAR}=F(t)}\, dF(t) \approx \int_0^\infty P(t)F'(t)dt$

$$\approx \int_0^\infty P(t)\frac{F(t-\Delta t)-F(t+\Delta t)}{2\Delta t}dt$$

University of Florida

ROC Based Optimization (cont.)

Area below the curve $= J \propto \int_0^{\infty} P(t)[F(t-\Delta t) - F(t+\Delta t)]dt$

$$= \int_0^{\infty} [\frac{1}{M}\sum_{i=1}^{M} u(f(x^i;\theta)-t)][\frac{1}{\text{total area}}\sum_{j=1}^{N} u(f(y^j;\theta)-(t-\Delta t)) - \frac{1}{\text{total area}}\sum_{j=1}^{N} u(f(y^j;\theta)-(t+\Delta t))]$$

$$\propto \int_0^{\infty} [\sum_{i=1}^{M} u(f(x^i;\theta)-t)]\{\sum_{j=1}^{N} [u(f(y^j;\theta)-(t-\Delta t)) - u(f(y^j;\theta)-(t+\Delta t))]\}dt$$

$$= \sum_{i=1}^{M}\sum_{j=1}^{N} \int_0^{\infty} \underbrace{u(f(x^i;\theta)-t)}_{\text{term 1}}[\underbrace{u(f(y^j;\theta)-(t-\Delta t)) - u(f(y^j;\theta)-(t+\Delta t))}_{\text{term 2}}]dt$$

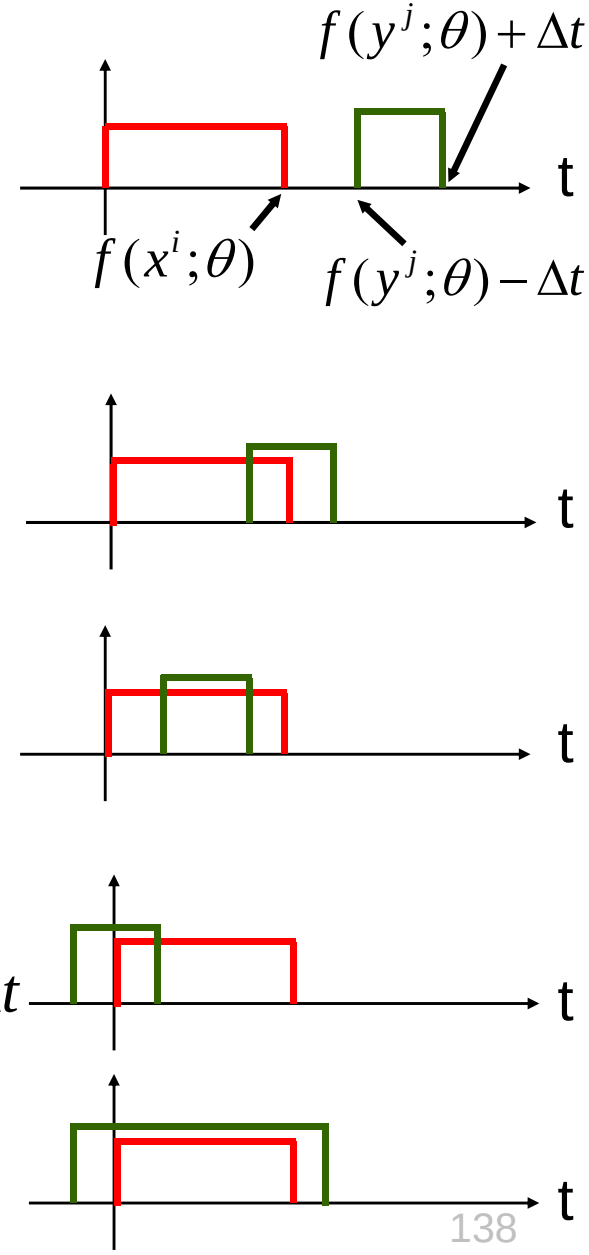
$$\text{term 1} = \begin{cases} 1 & \text{if } t \le f(x^i;\theta) \\ 0 & \text{otherwise} \end{cases} \qquad \text{term 2} = \begin{cases} 1 & \text{if } f(y^j;\theta) - \Delta t \le t \le f(y^j;\theta) + \Delta t \\ 0 & \text{otherwise} \end{cases}$$

Let $G(f(x^i;\theta), f(y^j;\theta)) \equiv \text{term1} \times \text{term2}$

ROC Based Optimization (cont.)

$$J = \sum_{i=1}^{M}\sum_{j=1}^{N}\int_0^{\infty} G(f(x^i;\theta), f(y^j;\theta))dt$$

$$\int_0^{\infty} G(f(x^i;\theta), f(y^j;\theta))dt = \begin{cases} 0 \\ \int_{f(y^j;\theta)-\Delta t}^{f(x^i;\theta)} 1\cdot dt \\ = f(x^i;\theta) - f(y^j;\theta) + \Delta t \\ \int_{f(y^j;\theta)-\Delta t}^{f(y^j;\theta)+\Delta t} 1\cdot dt = 2\Delta t \\ \int_0^{f(y^j;\theta)+\Delta t} 1\cdot dt = f(y^j;\theta) + \Delta t \\ \int_0^{f(x^i;\theta)} 1\cdot dt = f(x^i;\theta) \end{cases}$$



University of Florida

ROC Based Optimization Algorithm

1. J is expressed as a function of $f(x^i;\theta)$ and $f(y^j;\theta)$

2. Compare $f(x^i;\theta)$ and $f(y^j;\theta)$ for each pair of i and j

3. Compute update term for θ for each pair of i and j

   -- applicable to any $f(\cdot;\theta)$ as long as derivative of $f(\cdot;\theta)$ with respect to θ can be determined

   -- gradient-descent like methods can be used

University of Florida

WAAMD and SAR Processing

University of Florida

WAAMD Data

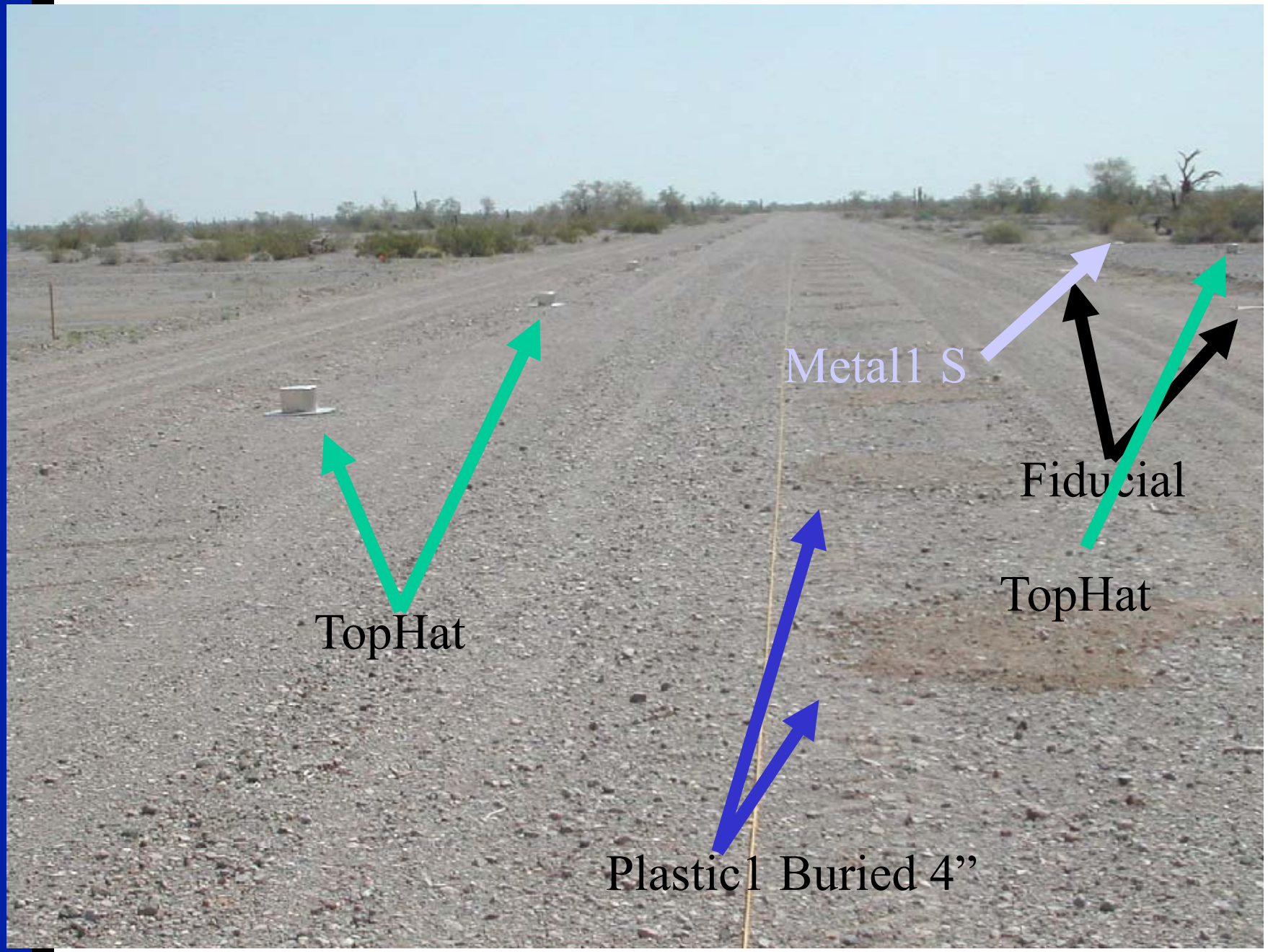
- Goal is to processing co-registered, WAAMD SAR and Hyperspectral data.
- SAR is geometrically warped and ground truthed
- Small amount of AHI is geometrically warped, none ground truthed yet.
- Initial looks at endmember, illustrate feedback to physics, understand state of the art
- RX processing and problems with geo-registration

University of Florida

WAAMD Data
Endmember Calculation

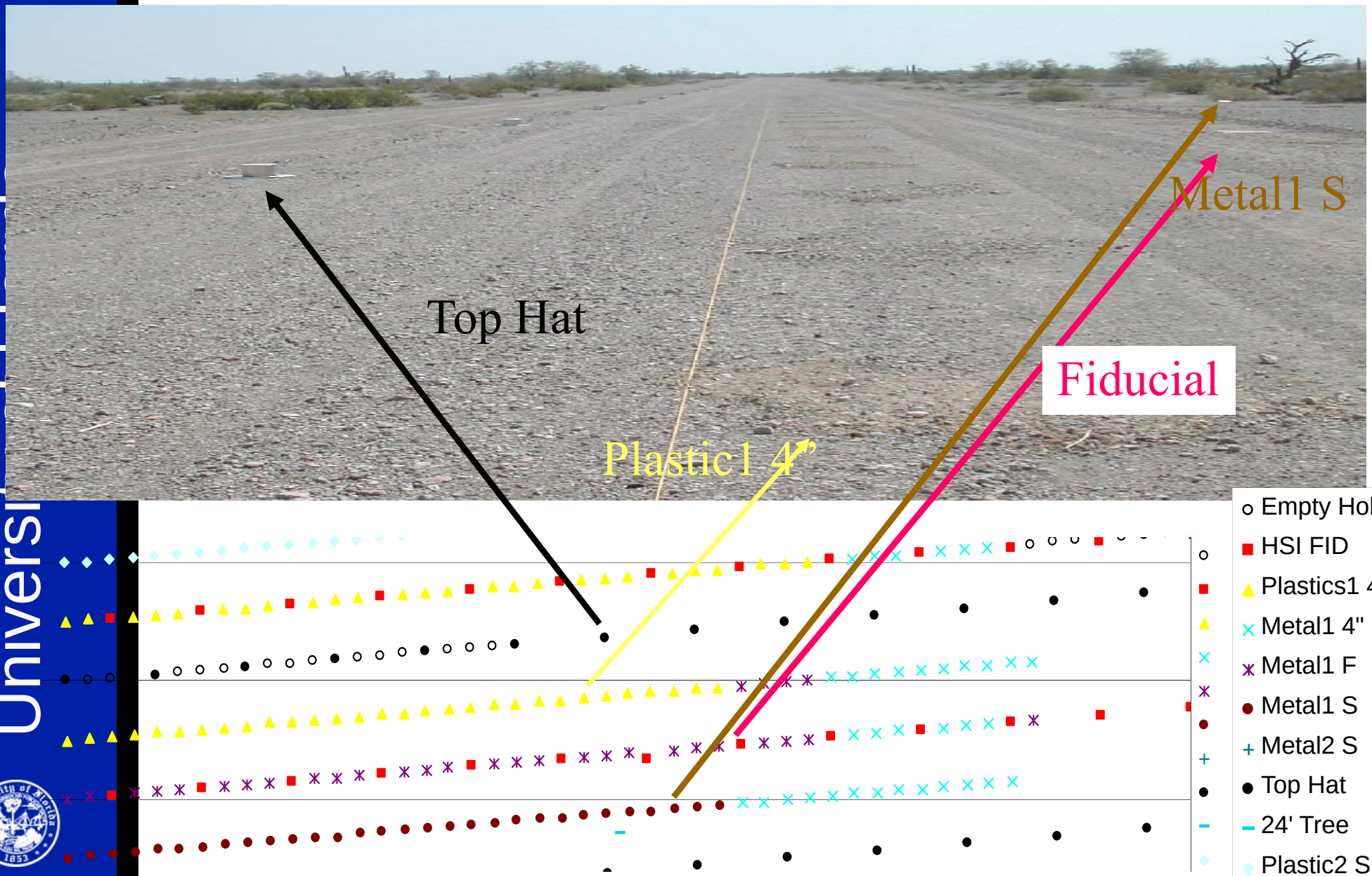
- AHI image from Yuma 2003
- Automatic Identification of Endmembers using Boardman's Pixel Purity Index, ENVI
- Scene Segmentation via Spectral Angle Mapper, ENVI

Countermine Site Image 1615

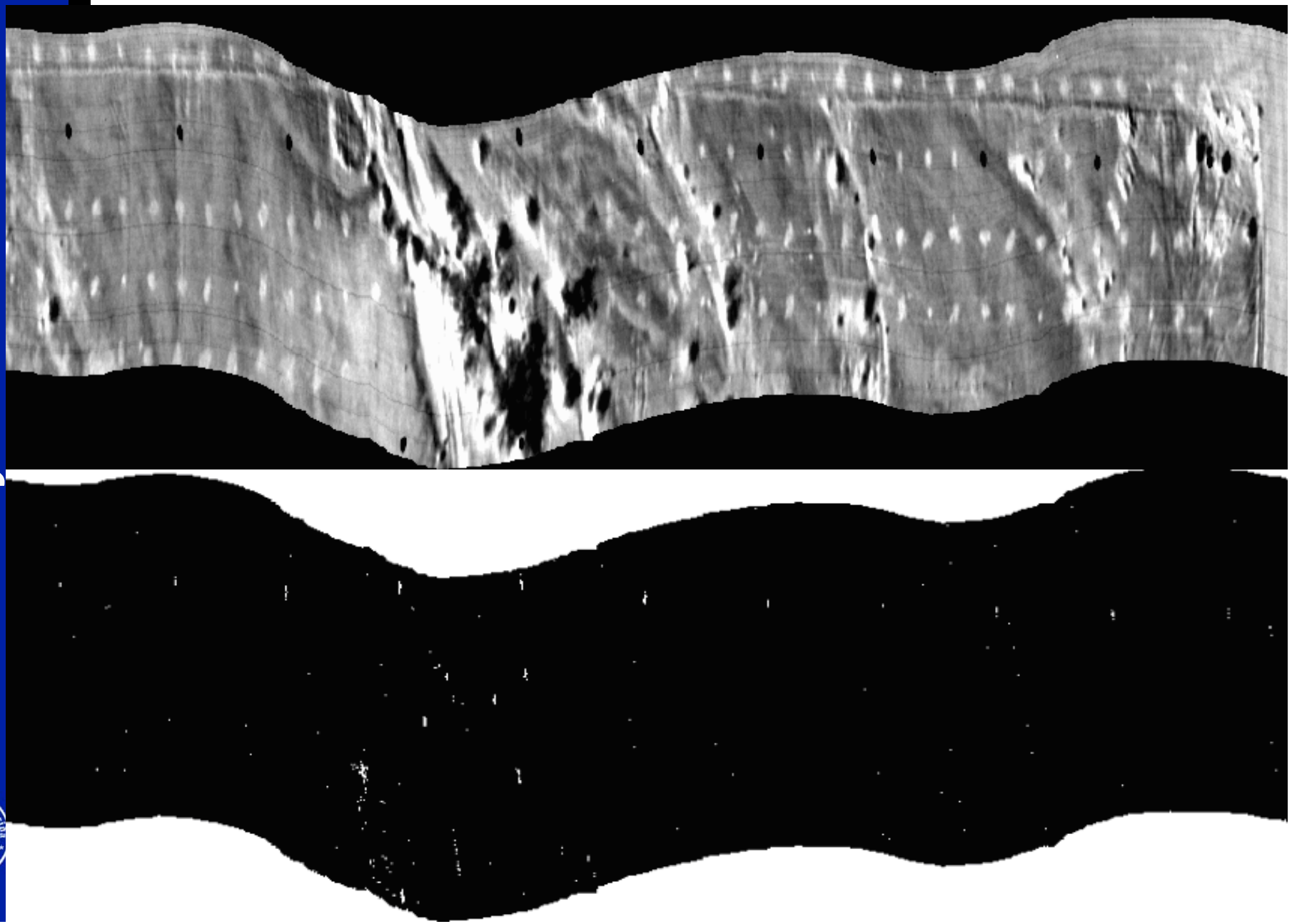


University of Florida

Countermine Image 1616

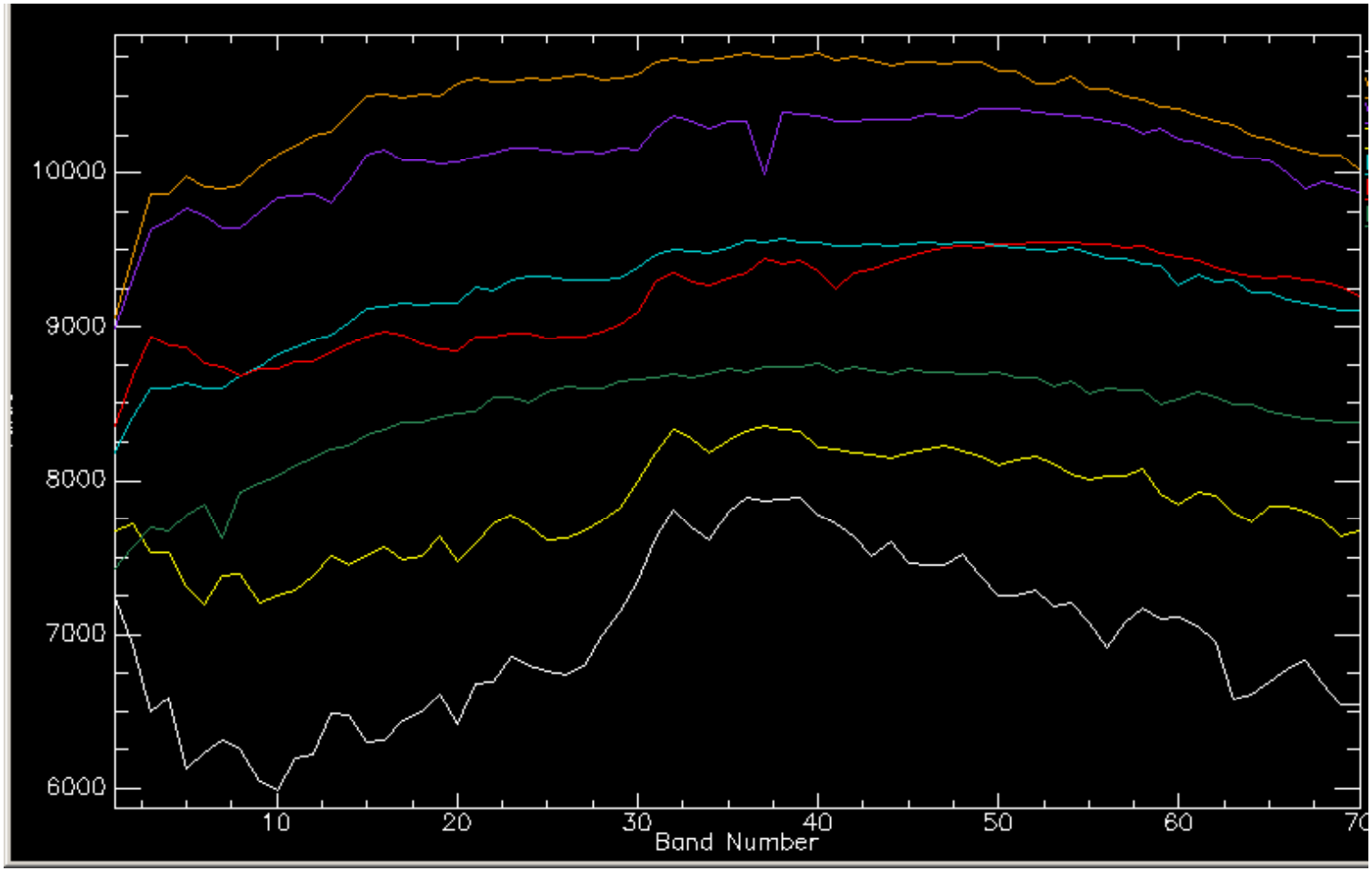


Original AHI + Spectrally Pure Pixels



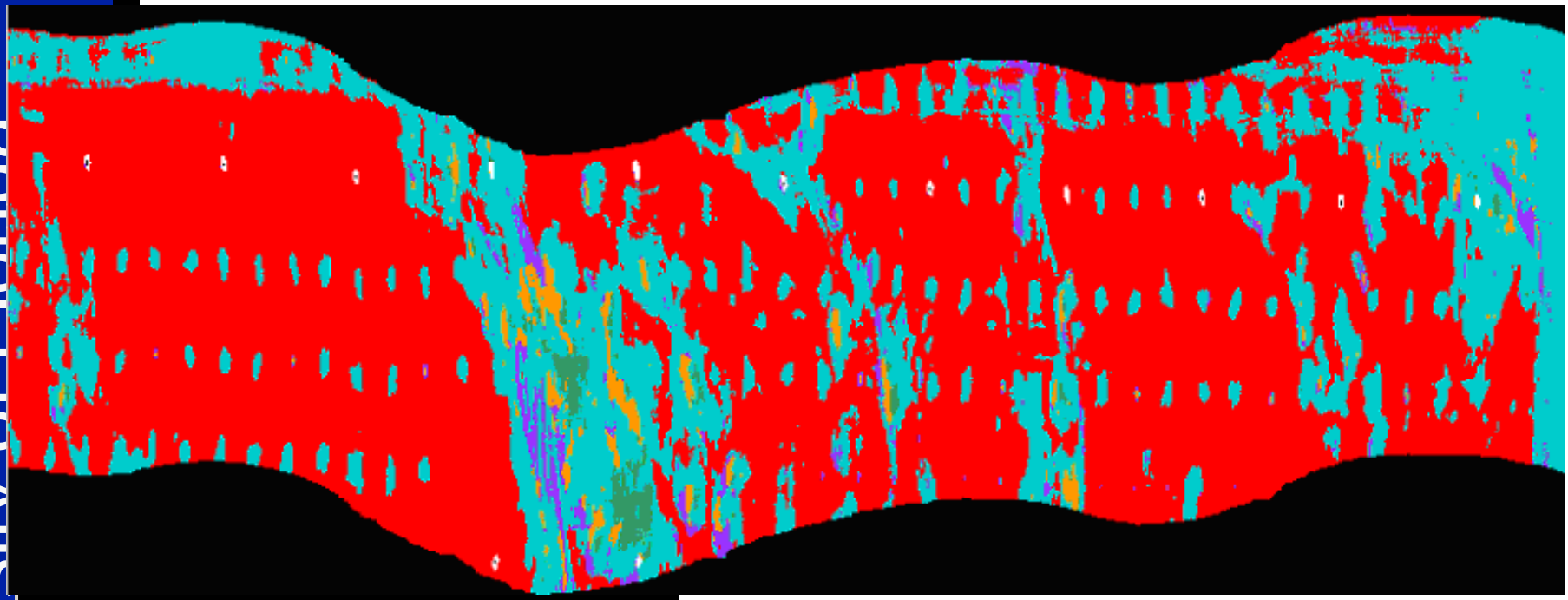
University of Florida

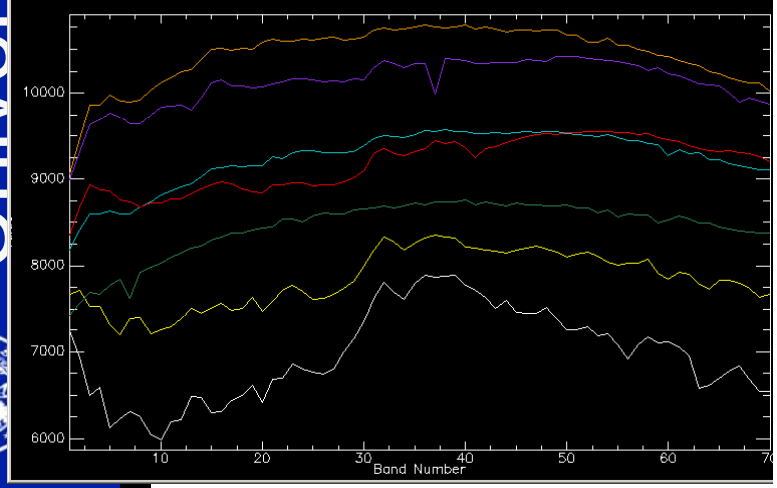
Pure Spectra + Plastic Mine Spectrum

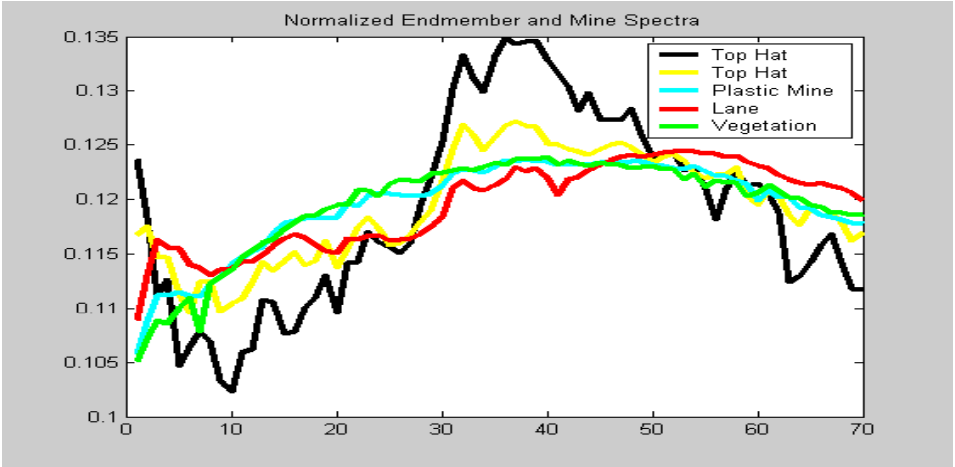


University of Florida

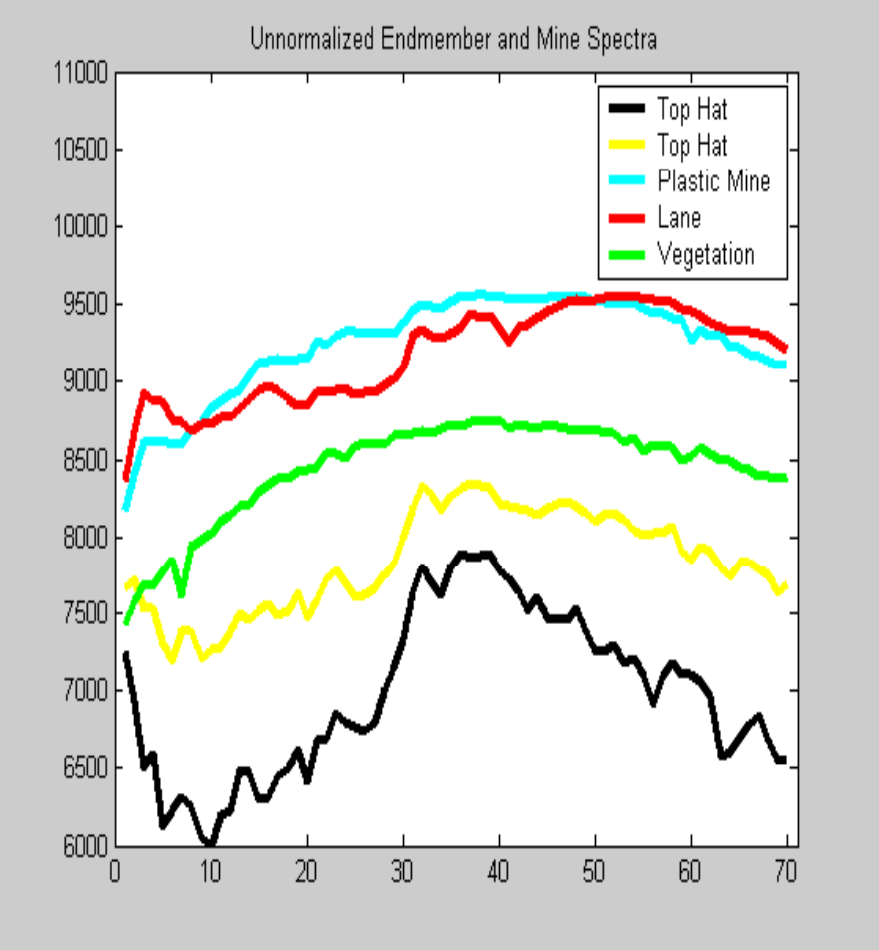
Scene Segmentation via Spectral Angle Mapping with Mine

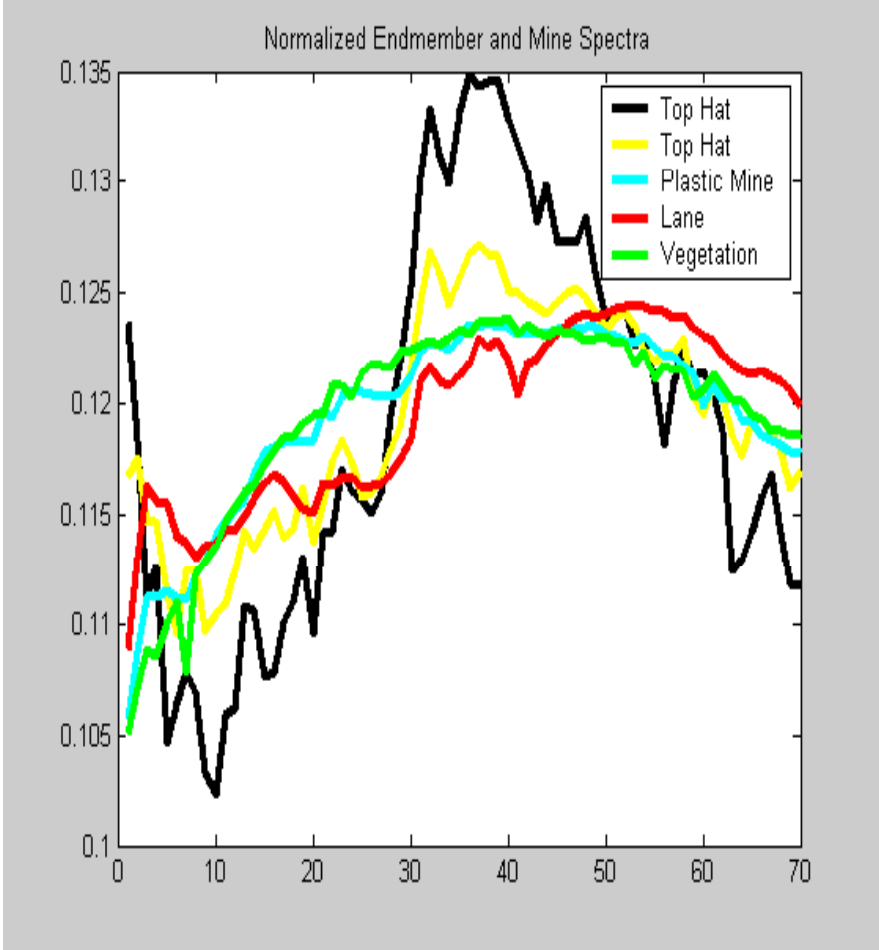






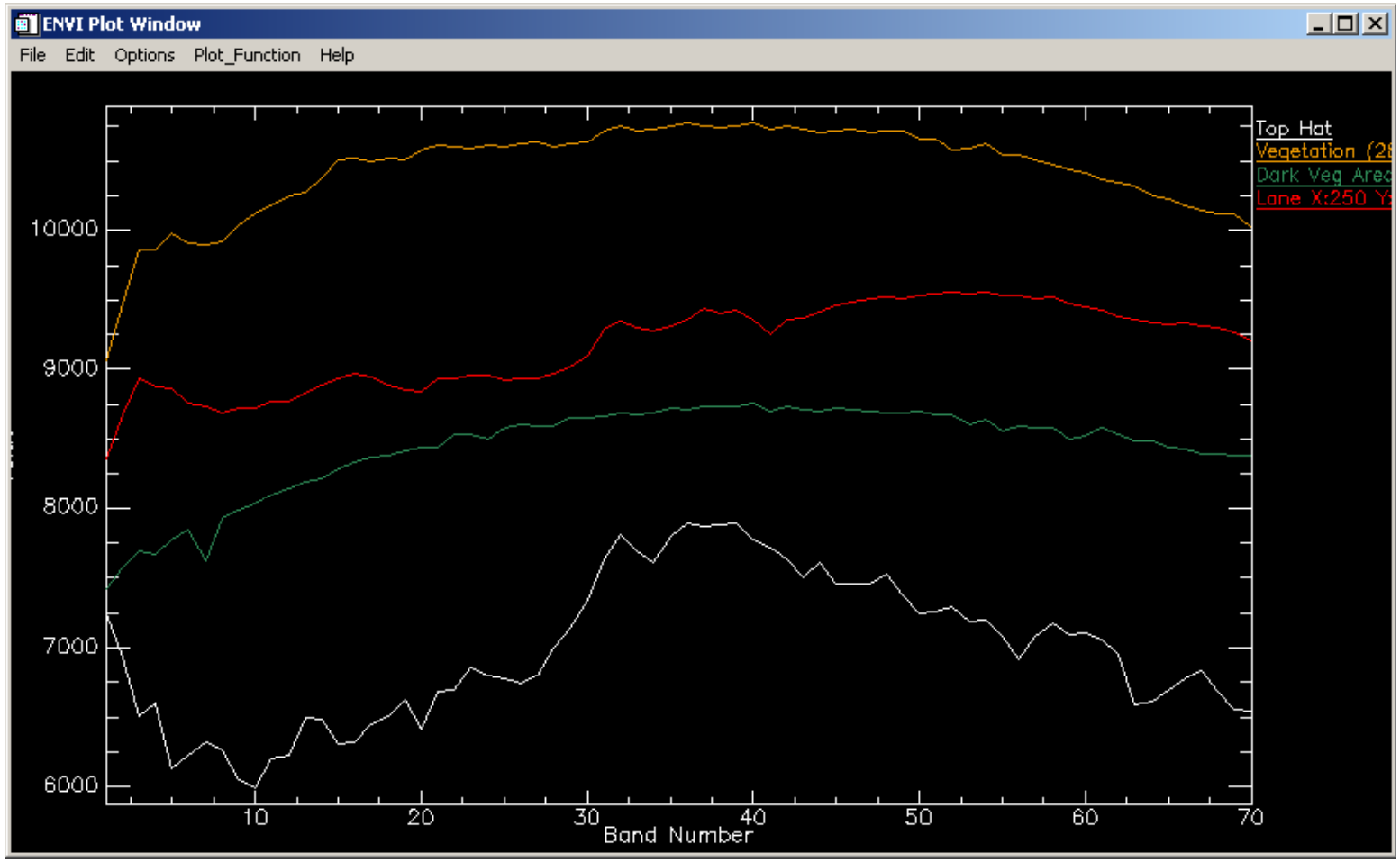
Unnormalized vs. Normalized Spectra



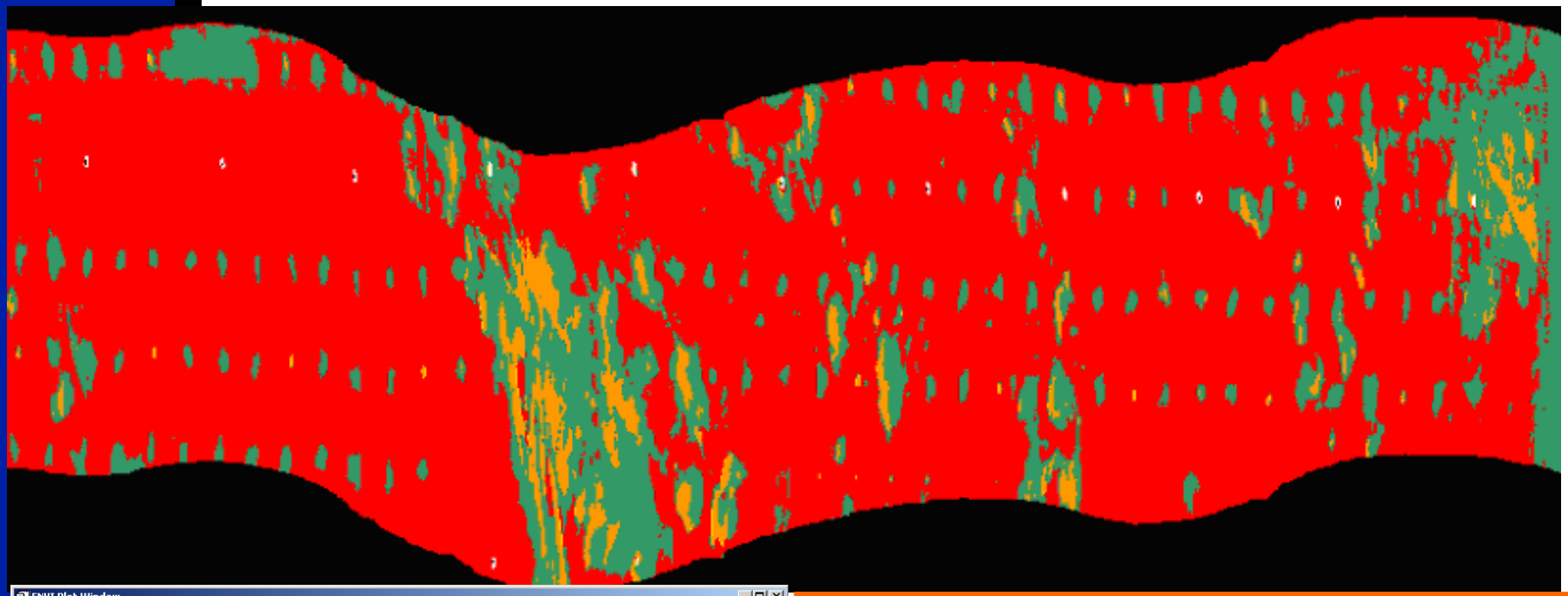


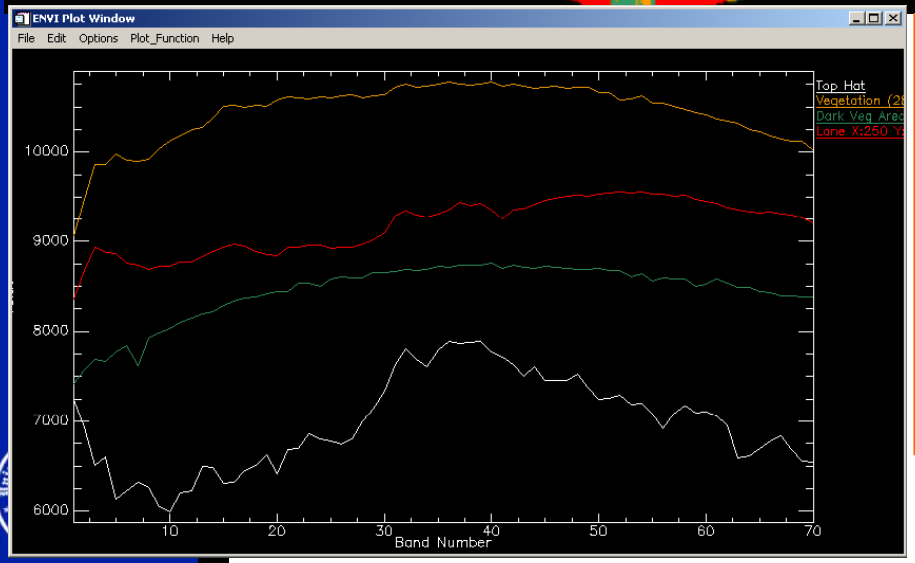
University of Florida

Found Endmembers with No Duplication



Scene Segmentation via Spectral Angle Mapping: no Mine



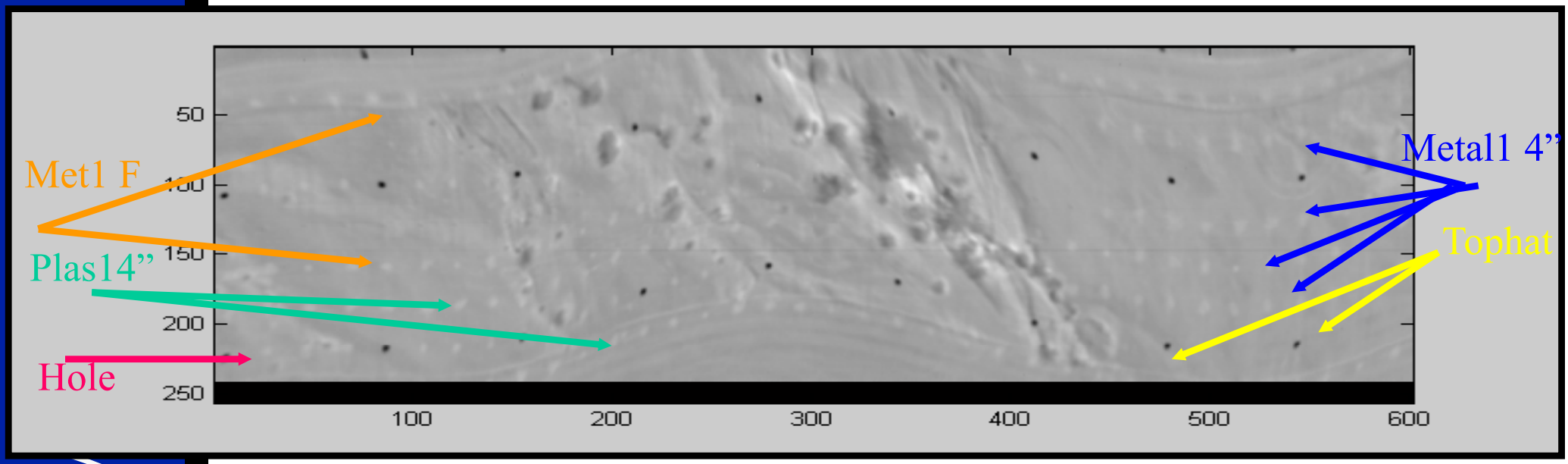


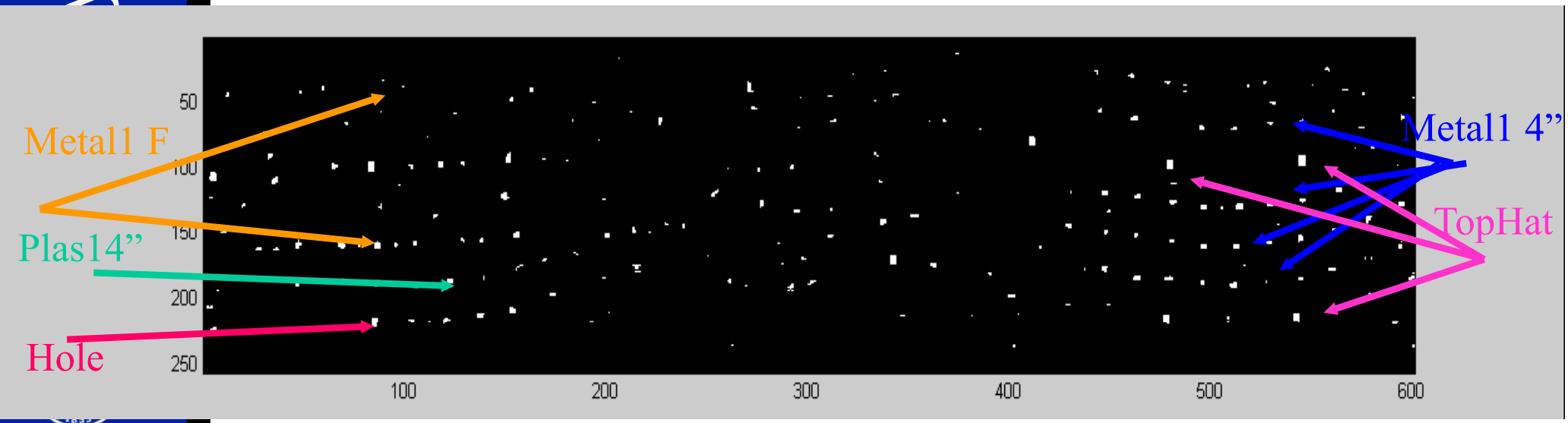
Q1: Why do mines (disturbed earth) look like vegetation? What is the physical basis?

Q2: Is this behavior consistent with different endmember calculation algorithms?

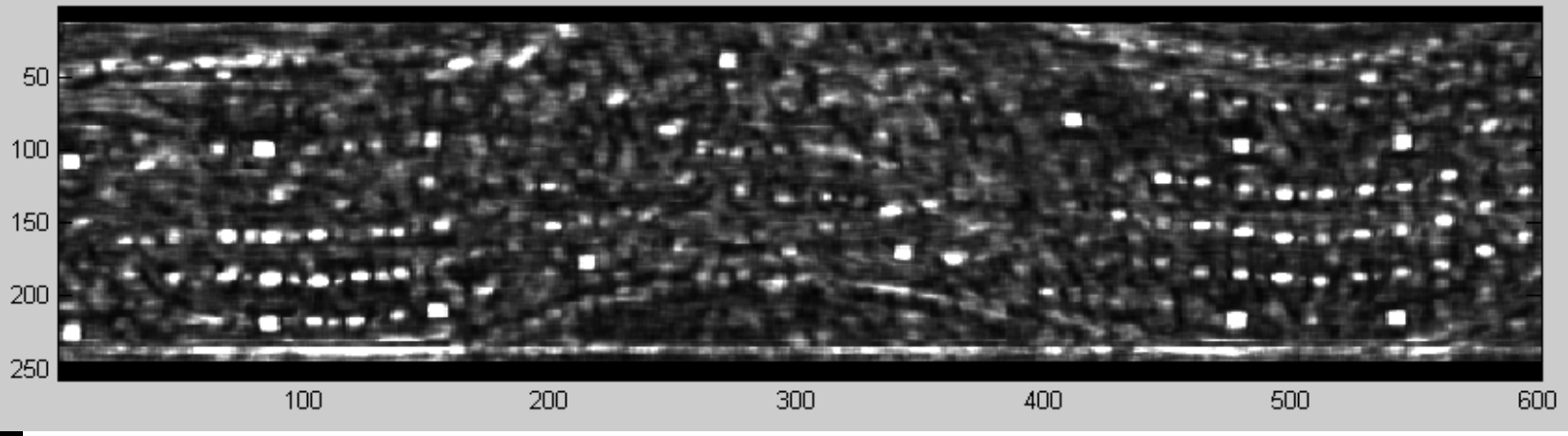
Q3: Can improved endmember and unmixing algorithms separate them, especially using better estimates of the spectral distributions?

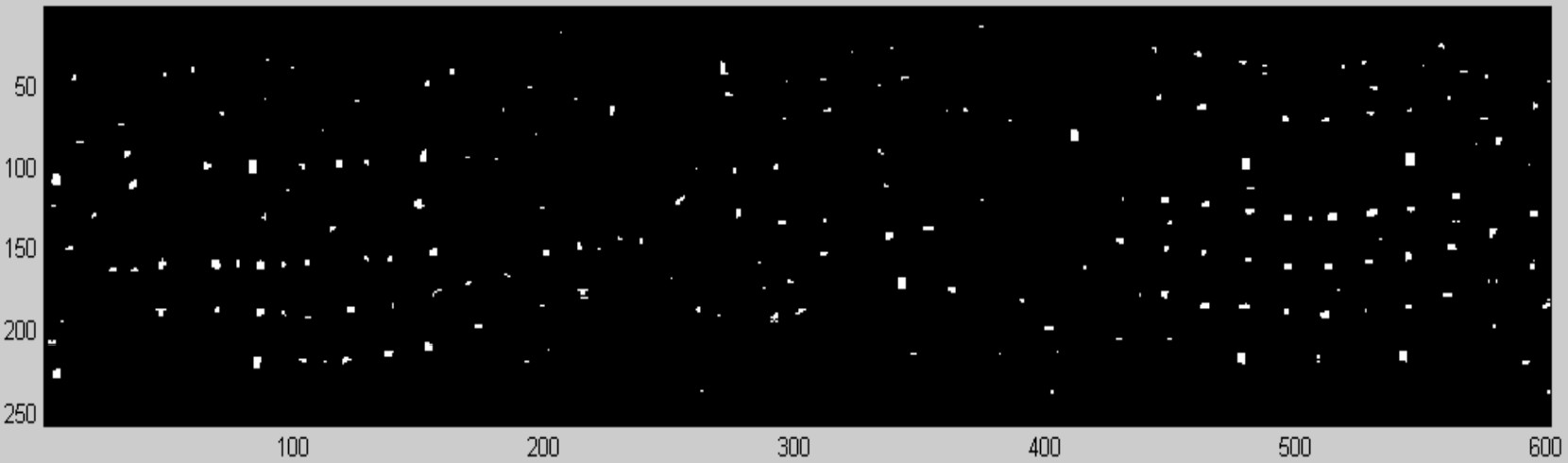
Results of RX





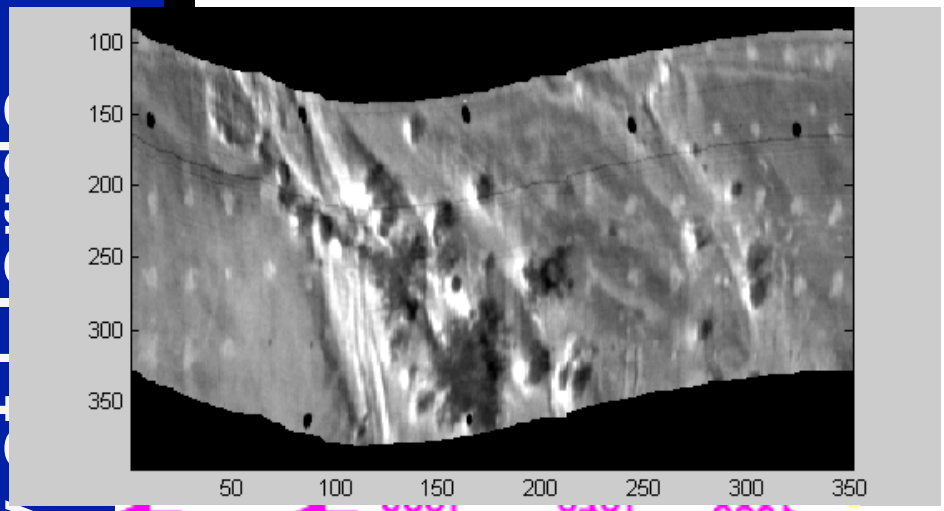
Background Removal Gaussian and Hit-Miss

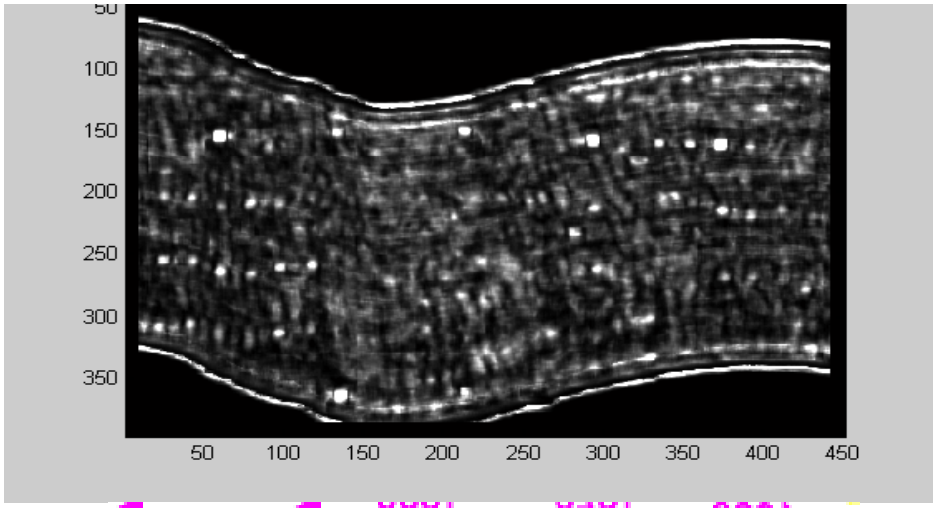


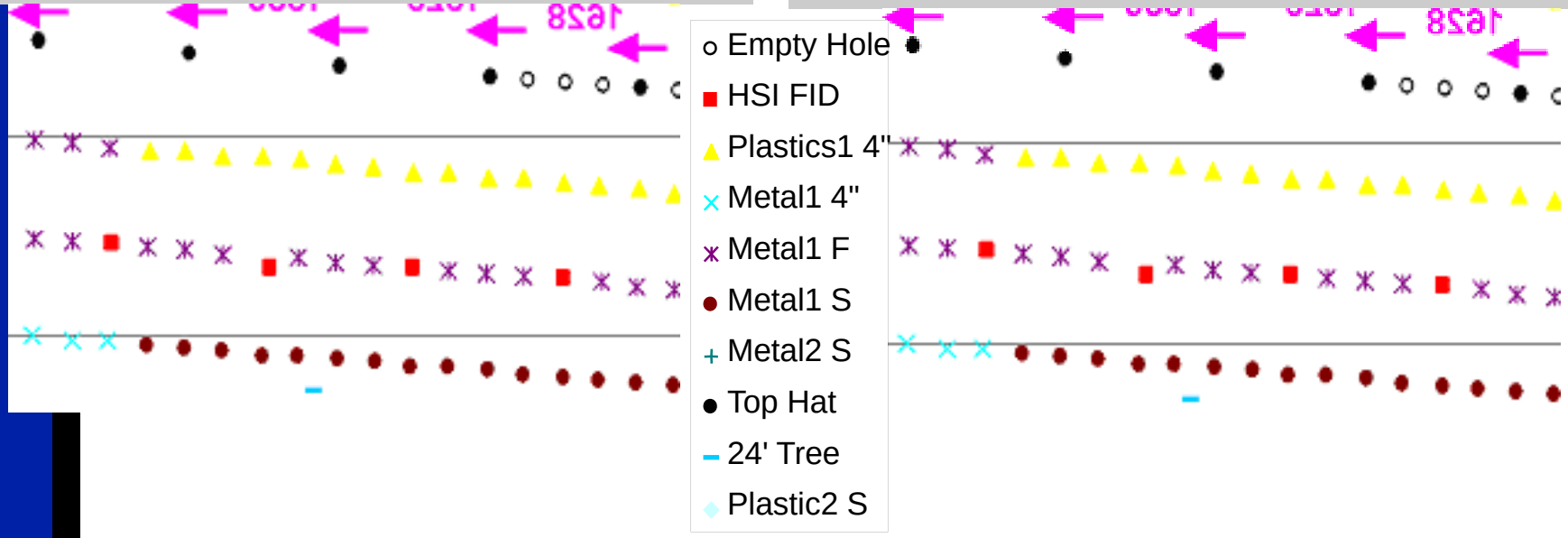


- Image: rest_cm2_am_alt1000_s300_r4_20030401_104600_r_bin3_s10e220.img

Portion of AHI Image and RX on AHI with Ground Truth – WAAMD Yuma 2003

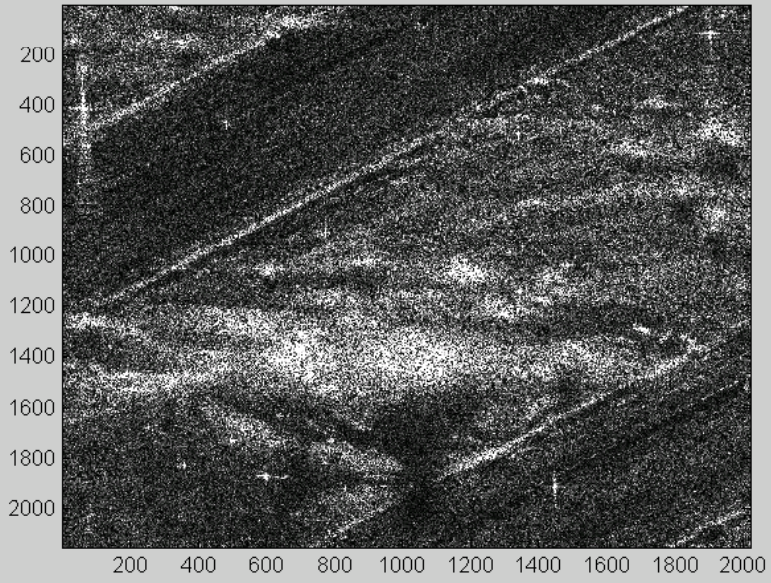


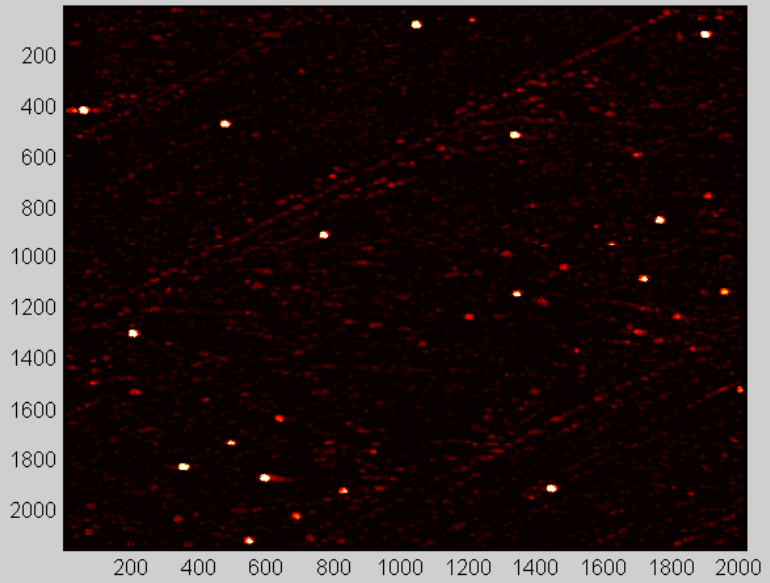




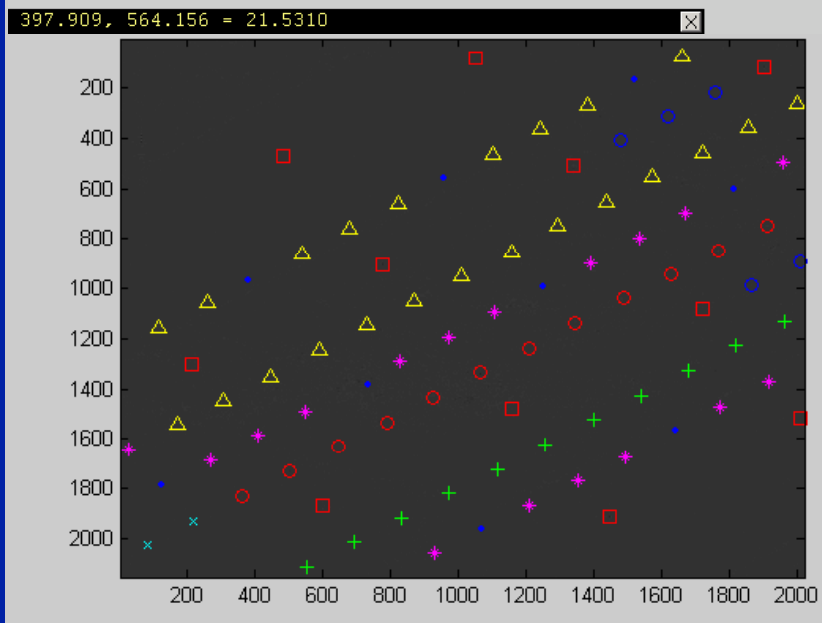
AHI Rotated with Ground Truth Overlay from SAR Image

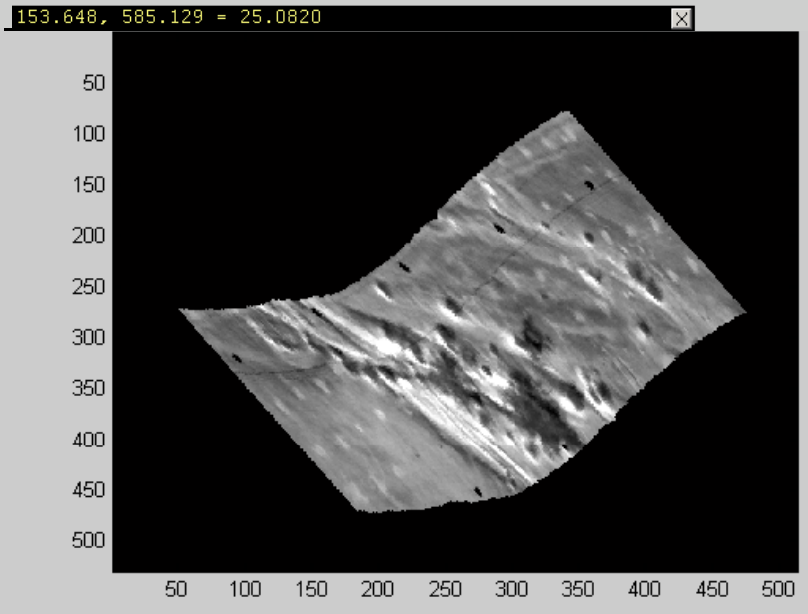
Ground Truth Veridian SAR





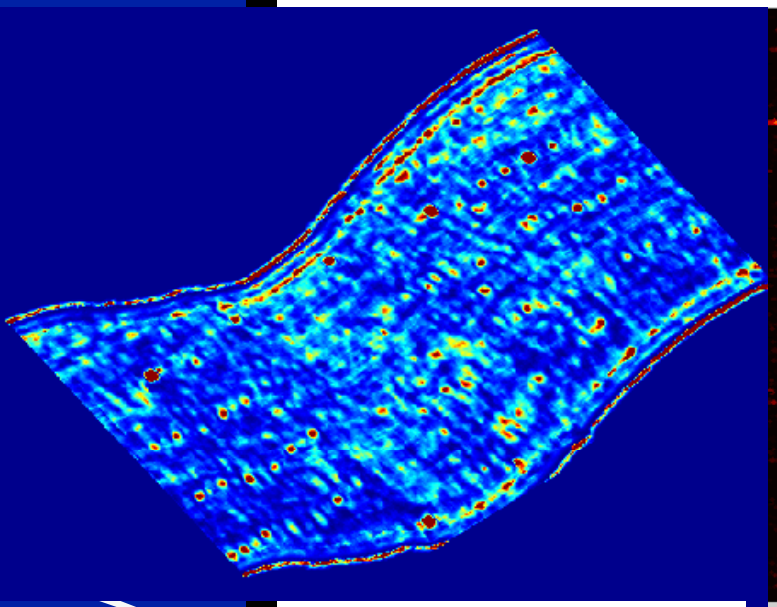
RX on SAR

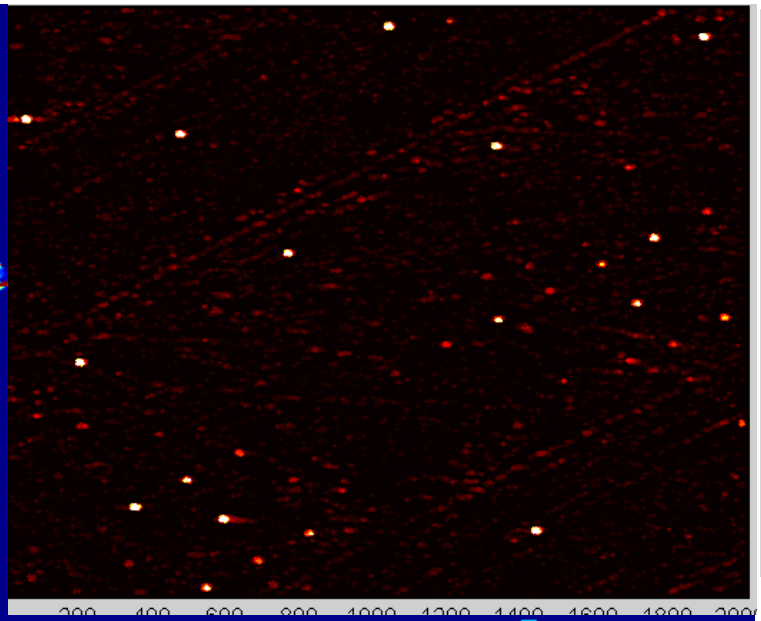


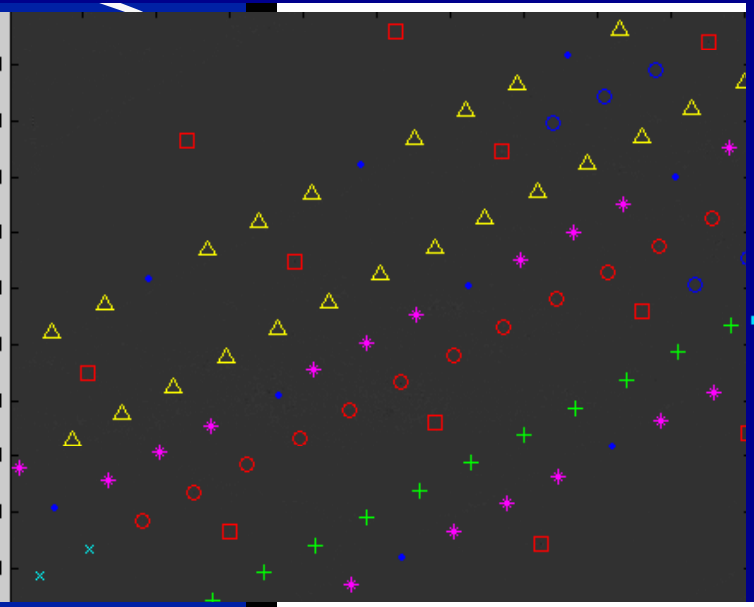


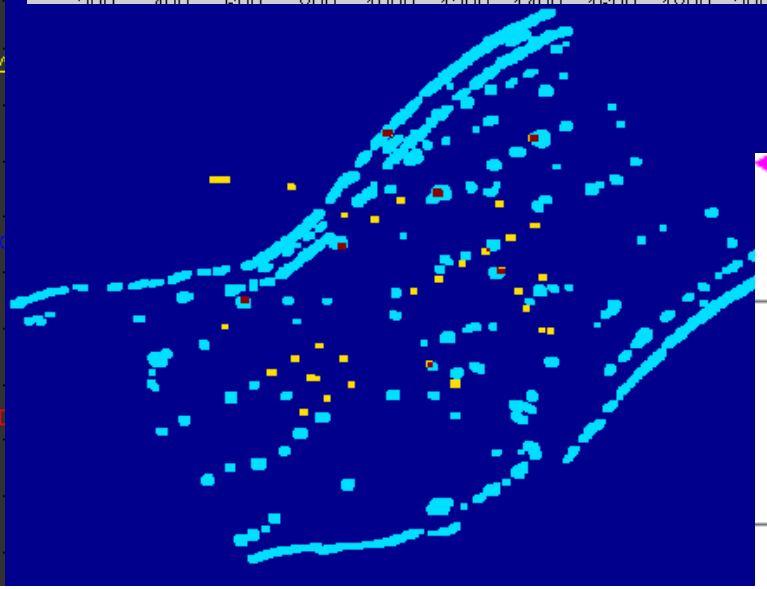
Rotate AHI

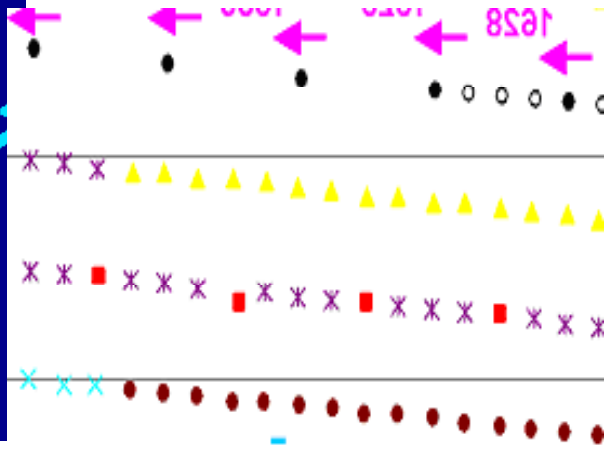
AHI Rotated with SAR Image











- Empty Hole
- HSI FID
- Plastic 4"
- Metal1 4"
- Metal1 F
- Metal1 S
- Metal2 S
- Top Hat
- 24' Tree
- Plastic2 S

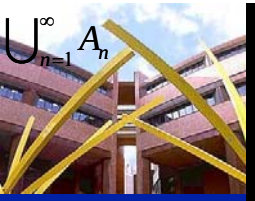
University of Florida

Choquet Integral
– Initial Research –

University of Florida

Choquet Integrals wrt Random Sets

- GOAL: Develop rigorous mathematical basis for multi-sensor fusion that incorporates most previous methodologies and apply this mathematical framework to the fusion of SAR and HSI.
- Developed and Investigated Continuous Choquet Integral Theory
  - Literature Review of Measure Theoretic Approach to Probability
- Established Relationship Between Choquet Integrals and Random Shapes
- Discovered Relationship Between Choquet Integrals, Capacity Functionals, Dempster-Shafer
- Implemented and Tested Capacity Optimization



University of Florida

Choquet Integrals wrt Random Sets σ-algebras

A *σ-algebra* on a set Ω is a collection of subsets of Ω, Σ(Ω), with the properties that:

(i) $\Omega \in \Sigma(\Omega)$,

(ii) If $A \in \Sigma(\Omega)$, then $A^c \in \Sigma(\Omega)$.

(iii) If $A = \bigcup_{n=1}^{\infty} A_n$ and if $A_n \in \Sigma(\Omega)$ for n = 1, 2, 3… then $A \in \Sigma(\Omega)$

These are the *Sample Spaces* in Probability Classes

Choquet Integrals wrt Random Sets Positive Measures

A *positive measure* is a countably additive function

$$\mu: \Sigma(\Omega) \rightarrow [0, \infty]$$

Countable additivity ≡

$$\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i) .$$

A probability measure is a measure with the property that $\mu(\Omega) = 1$.

In this case, the triple $(\Omega, \Sigma(\Omega), \mu)$ is called a probability space.

University of Florida

Choquet Integrals wrt Random Sets Fuzzy Measures (Capacities)

Let $\Sigma(\Omega)$ be a σ-algebra on Ω. A fuzzy measure on Ω is a function

$$\mu: \Sigma(\Omega) \rightarrow [0, 1]$$

with the properties that

(i) $\mu(\phi) = 0$ and $\mu(\Omega) = 1$,

(ii) If $A \subseteq B$, then $\mu(A) \leq \mu(B)$,

(iii) If $F_1 \subseteq F_2 \subseteq \ldots \subseteq F_n \subseteq \ldots$ is a montonically non-decreasing sequence of elements of $\Sigma(\Omega)$,

then $\lim_{n\rightarrow\infty} \mu(F_n) = \mu(\bigcup_{n=1}^{\infty} F_n)$.

University of Florida

Choquet Integrals wrt Random Sets
Random Variables and Random Sets
(Measurable Functions Set-Valued Random Variables)

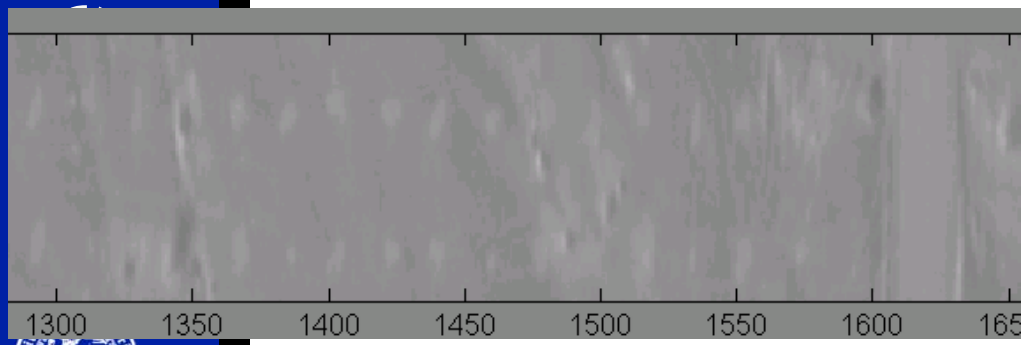
- A typical definition of a random variable is

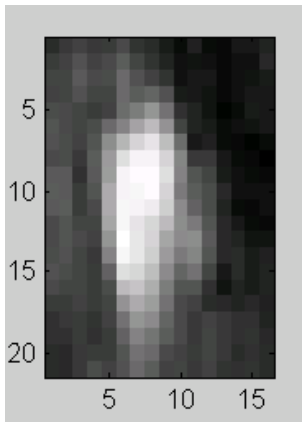
  $X: \Omega \to \Re$ (Ω is a s-algebra or Sample Space, $\Re$ denotes the real numbers)

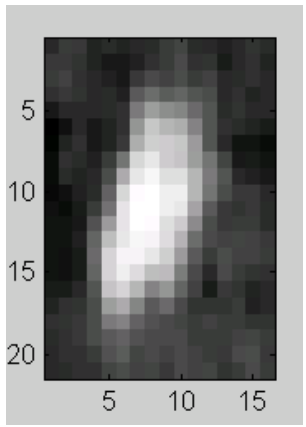
  X is a measurable function Experiment whose outcome is a number

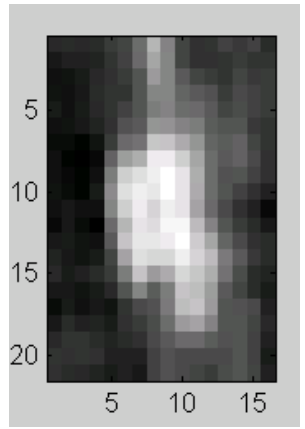
- Analogously a definition of a random set is
  - Let S be a set of subsets of a space.

  $X: \Omega \to S$ Experiment whose outcome is a set









University of Florida

Choquet Integrals wrt Random Sets
Capacity Functionals of Random Sets

- A random variable is completely determined by it's Probability Distribution

- A random set is completely determined by it's Capacity Functional

$$T_X(A) = P(X \cap A \neq \phi)$$

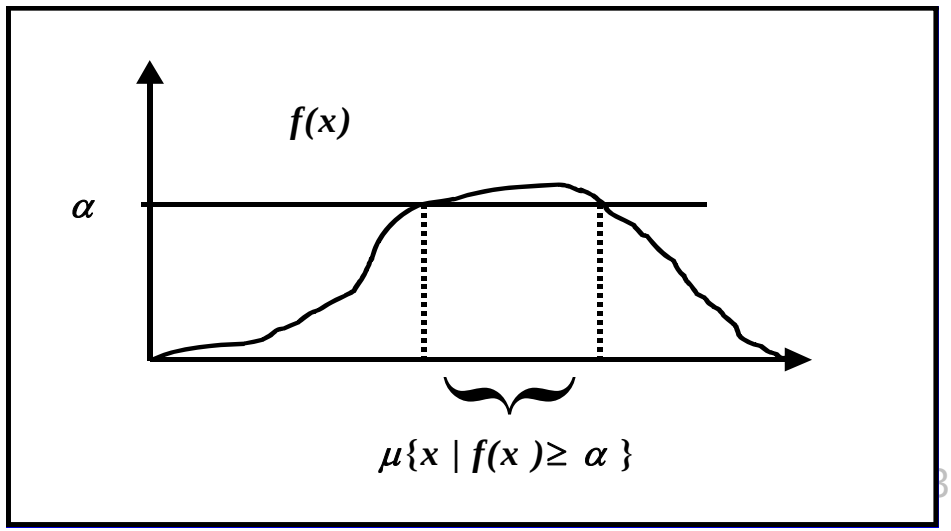
$$= P(\{\omega \mid X(\omega) \cap A \neq \phi\})$$

Continuous Choquet Integral

Let $f : \Omega \rightarrow [0, 1]$ be a measureable function with respect to the σ-algebra $\Sigma(\Omega)$.

The Choquet integral of f with respect to the fuzzy measure μ on $\Sigma(\Omega)$ is defined by

$$C_\mu(f) = \int_0^1 \mu(\{x \mid f(x) > \alpha\}) d\alpha \ .$$



University of Florida

University of Florida

Discrete

Represent Voting, MAX, MIN, Median, alpha-trimmed means, arithmetic and geometric means, and many others

- Let X = $\{x_1, x_2, \ldots, x_n\}$ be a discrete set
- Let μ be a fuzzy measure on 2^X
- Let $f : X \rightarrow [0,1]$
- Re-number X so that $f(x_{(1)}) \leq f(x_{(2)}) \leq \ldots \leq f(x_{(n)})$
- Let $A_{(i)}=\{x_{(i)},\ldots,x_{(n)}\}$ and $\mu(A_{(n+1)})=0$
- The discrete Choquet integral of f is

$$C_\mu(f) = \sum_{j=1}^{n} \mu(A_{(j)})(f(x_{(j)}) - f(x_{(j-1)})) \qquad f(x_{(0)}) \equiv 0$$

$$= \sum_{j=1}^{n} f(x_{(j)})(\mu(A_{(j)}) - \mu(A_{(j+1)})) \qquad \mu(A_{(n+1)}) \equiv 0$$

Where are we?

Fuzzy Measures

-Non-additive measures
-Generalization of Probability measures

Choquet Integrals

-Integrals wrt fuzzy measures
-Different measures → different operators
-Feature extraction
-Fusion

Random Sets

-Model Random Shapes
- ≡ Capacity Functionals
⊄ Fuzzy Measures

Choquet Integrals

Wrt

Random Sets

University of Florida

Choquet Dilation of Image by a Random Shape

Let f denote a measurable function $f: \Re^2 \rightarrow [0,1]$ and let

$$f_\alpha \equiv \{(x,y) \mid f((x,y)) > \alpha\} .$$

If **X** represents a random shape and f a subimage of an image, then the Choquet integral,

$$C_{T_X}(f) = \int_0^1 T_X(f_\alpha) d\alpha = \int_0^1 P[X \cap f_\alpha \neq \phi] d\alpha .$$

can be interpreted as the average probability that an α-cut of f intersects with the shape **X**.

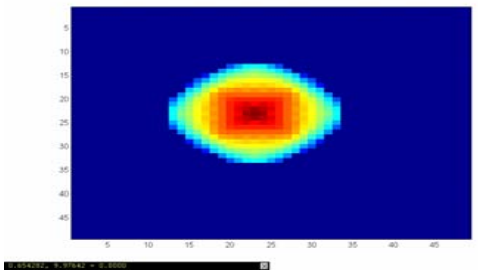
This is a generalization of morphological dilation.

University of Florida

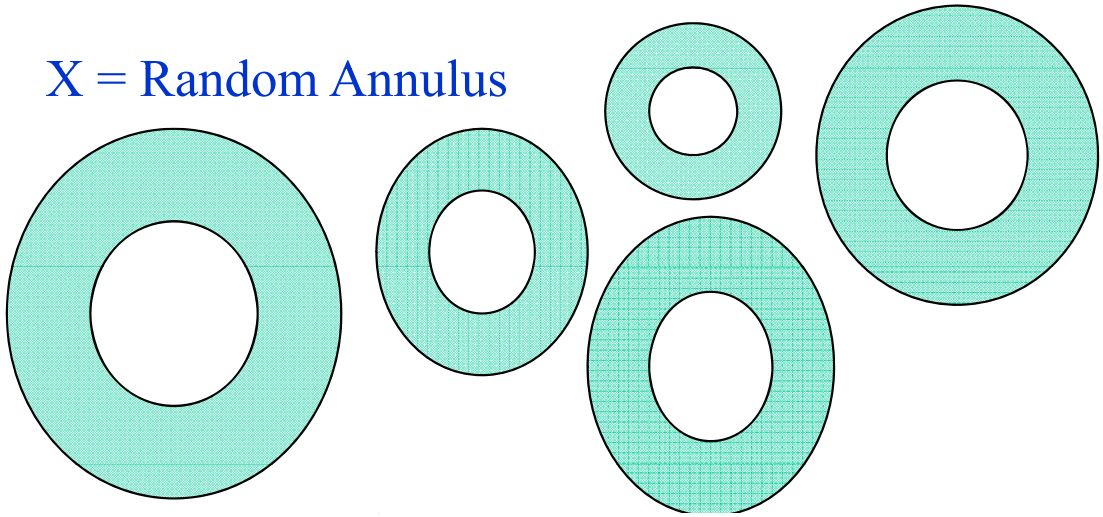
Choquet Dilation of Image by a Random Shape

$$C_{T_X}(f) = \int_0^1 T_X(f_\alpha)d\alpha = \int_0^1 P[X \cap f_\alpha \neq \phi]d\alpha\ .$$

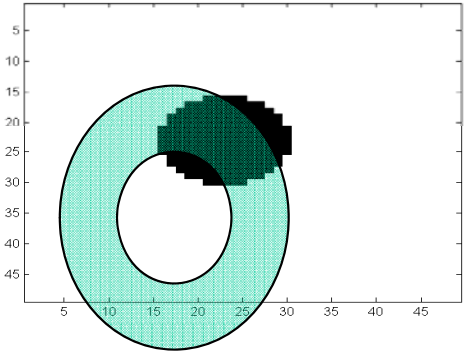
Image

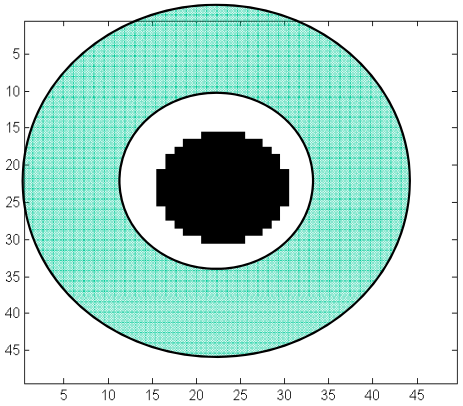


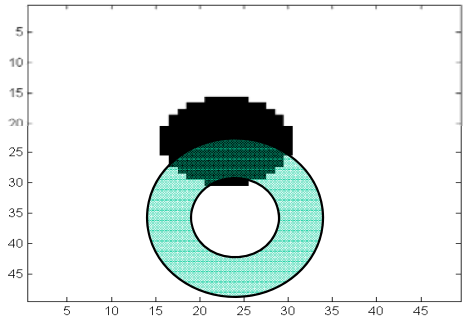
X = Random Annulus



α-cut







University of Florida

Choquet Erosion of Image by a Random Shape

- Note that

$$C_{T_X}(1-f) = \int_0^1 P[X \cap (1-f)_\alpha \neq \phi]\, d\alpha$$

- For any realization $\tilde{X}$ of X, $\tilde{X} \cap (1-f)_\alpha = \phi$ if and only if $\tilde{X} \subset f_{1-\alpha}$.

- Hence, by changing variables of integration

or $$C_{T_X}(1-f) = 1 - \int_0^1 P[X \subset f_\alpha]\, d\alpha$$

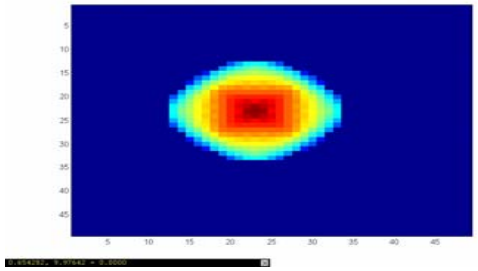
$$\boxed{1 - C_{T_X}(1-f) = \int_0^1 P[X \subset f_\alpha]\, d\alpha}$$

University of Florida

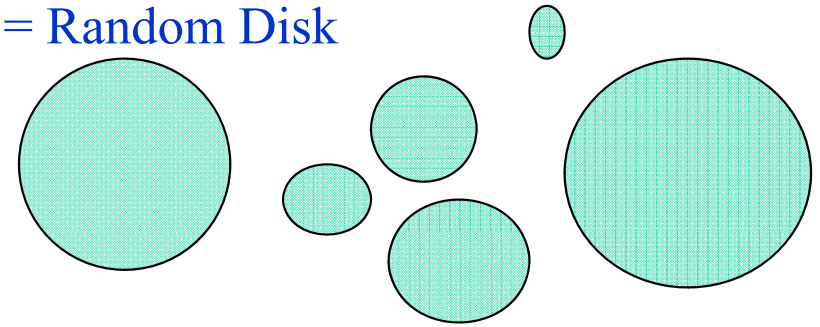
Choquet Erosion of Image by a Random Shape

$$1 - C_{T_X}(1 - f) = \int_0^1 P[X \subset f_\alpha]\, d\alpha$$

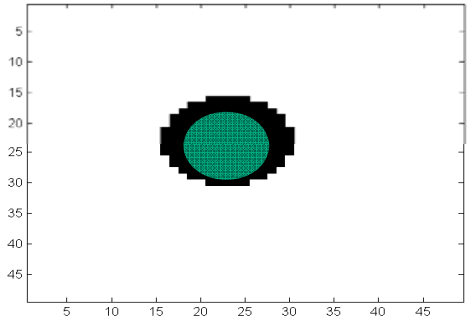
Image

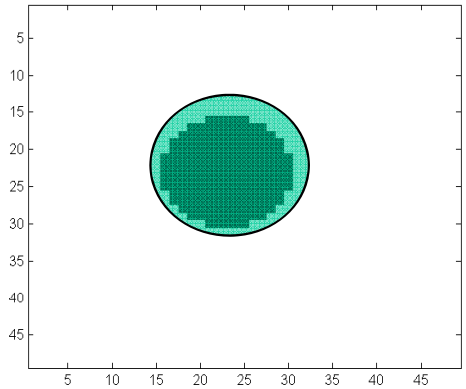


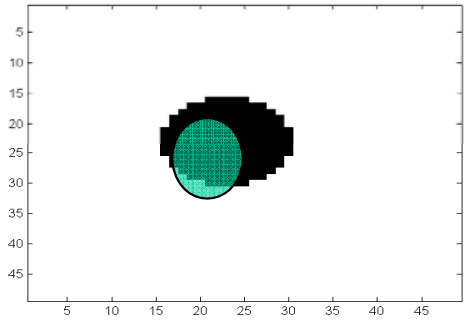
X = Random Disk



α-cut



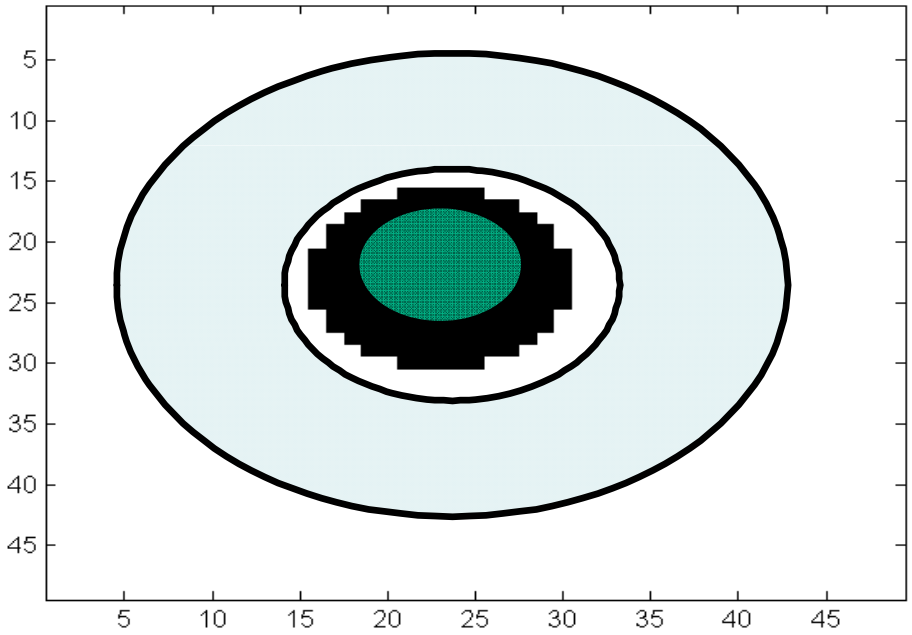




University of Florida

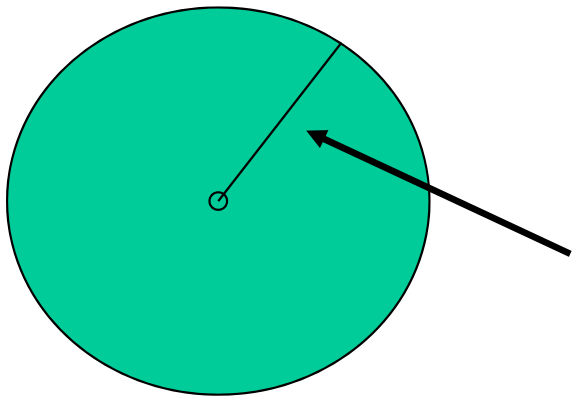
Random Hit-Miss transform

- Erode Foreground by Random Shape
- Erode Background by Complementary Random Shape
- Derive Analytical Forms for Random Disk/Annulus
- Will extend to more complex shapes, e.g. detector outputs
- Same mathematics applies to decision-level fusion



Example – Erosion by Random Disk

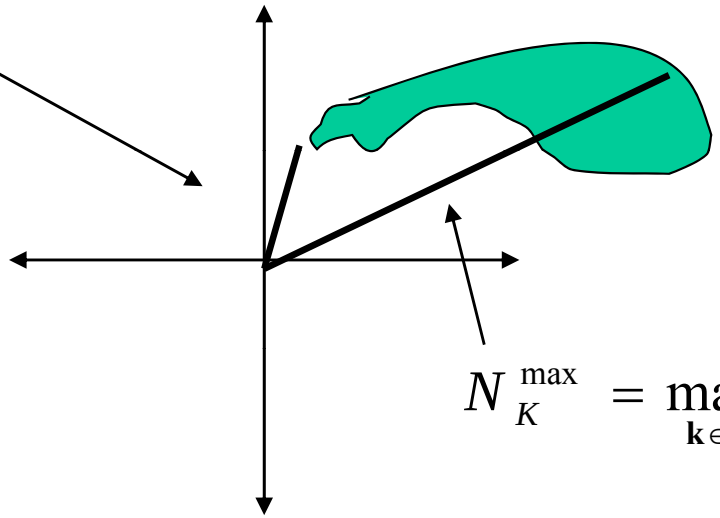
- Random Disk X
  - Random radius given by $N(r_d, \sigma_d)$.
  - If $r_d >> \sigma_d$ (e.g. $r_d > 3\sigma_d$), then the probability of a negative radius is negligible.



University of Florida

Reduction of Capacity Functional to Radial Distribution for Random Disk

$$N_K^{\min} = \min_{\mathbf{k} \in K} \{d(\mathbf{0}, \mathbf{k})\} = \min_{\mathbf{k} \in K} \{\|\mathbf{k}\|\}$$

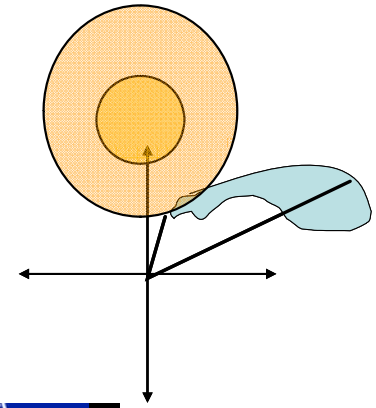


$$N_K^{\max} = \max_{\mathbf{k} \in K} \{d(\mathbf{0}, \mathbf{k})\} = \max_{\mathbf{k} \in K} \{\|\mathbf{k}\|\}$$

Reduction of Capacity Functional to Radial Distribution for Random Disk

- Let D_r = disk of radius r centered at origin
- Note that $D_r \cap K \neq \phi$ iff $N_k^{min} \leq r$. Hence
- $T_X(K) = P_X[X \cap K \neq \phi] = P(N_k^{min} \leq r)$.
- Complementary Error Function
- Probabilistic Dilation

$$erfc(x) = \int_x^{\infty} e^{-t^2} dt$$



$$C_\mu(f) = \int_0^1 P(r \geq N_{f_\alpha}^{\min}) d\alpha = \frac{1}{2} \int_0^1 erfc\left(\frac{N_{f_\alpha}^{\min} - r_d}{\sqrt{2}\sigma_d}\right) d\alpha$$

Probabilistic Morphology for Random Disk

Random Dilation

$$C_{\mu}(f) = \frac{1}{2}\int_0^1 erfc\left(\frac{N_{f_\alpha}^{\min} - r_d}{\sqrt{2}\sigma_d}\right)d\alpha$$

Random Erosion

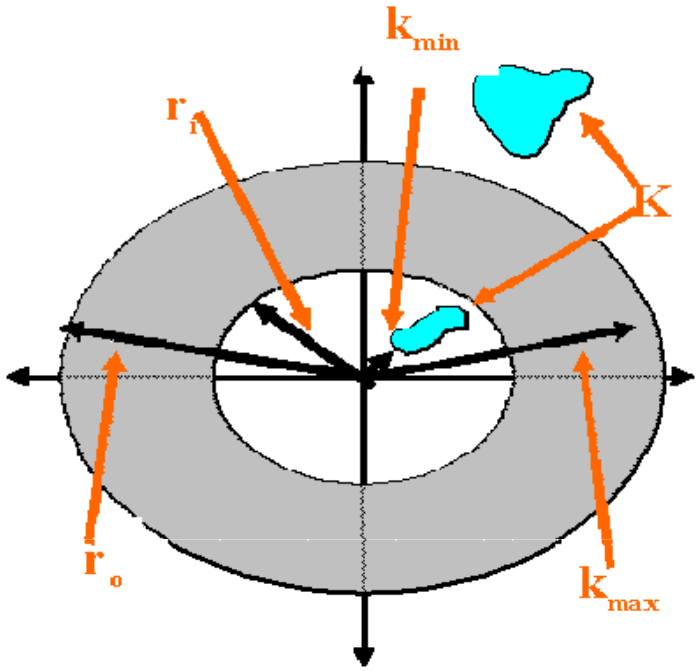
$$C_{\mu}(1-f) = 1 - \frac{1}{2}\int_0^1 erfc\left(\frac{N_{(1-f)_\alpha} - r_d}{\sqrt{2}\sigma_d}\right)d\alpha$$

University of Florida

Characterization of Annulus Hit Probability

Doesn't work if not path connected

Consider each path connected component



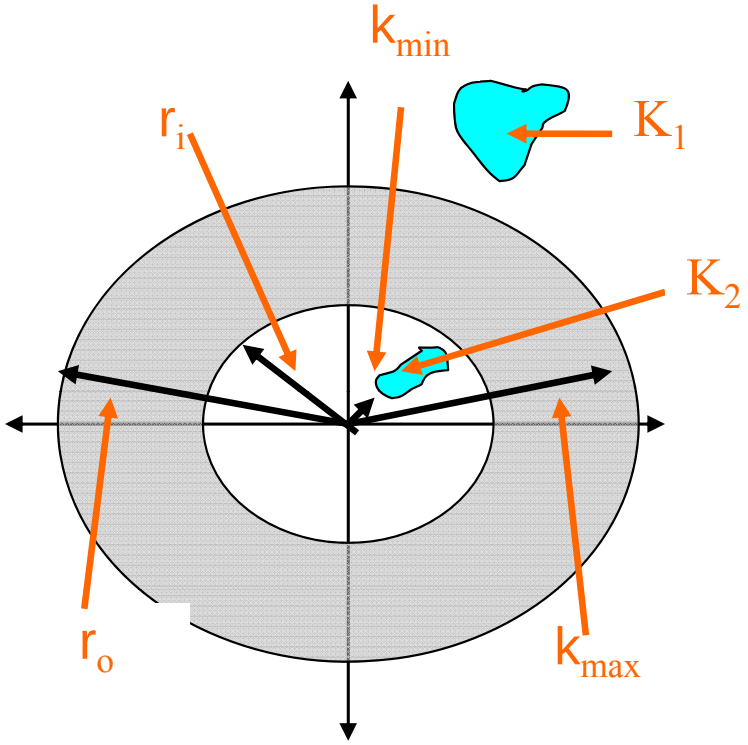
Characterization of Annulus Hit Probability Reduction to Path-Connected Case

- $K = K_1 \cup K_2 \cup \ldots \cup K_m$

$$P[A \cap K = \phi] = 1 - P[A \cap K \neq \phi]$$

$$P[A \cap K = \phi] = \prod_{i=1}^{m} P[A \cap K_i = \phi]$$

$$= \prod_{i=1}^{m} (1 - P[A \cap K_i \neq \phi]).$$



Reduction to path connected case

computable via complementary error function

University of Florida

University of Florida

The Identical Mathematical Framework is Applicable to Fusion, e.g. Dempster-Shafer

- Previously shown Choquet can represent voting, geometric, arithmetic, and robust means (asymptotically) and variety of other decision-level fusion methods
- Dempster-Shafer Theory is built on the notion of Belief and Plausibility functions.
- Belief and Plausibility functions can be defined as Capacity Functionals of Random Sets
- Choquet integrals represent expected values of Belief and Plausibility functions. Thus we should be able to use Expectation-Maximization to find optimal Belief/Plausibility functions for fusion.

Dempster-Shafer Representation

- Assume set of states S, unobservable
- Observations are random variables, X
- Define

$$\rho(s,x)=\begin{cases}1 & \text{if state } s \text{ is compatible with observation } x\\ 0 & \text{if state } s \text{ is not compatible with observation } x\end{cases}$$

- And

$$U_\rho(\omega)=\{s\in S \mid \rho(s,X(\omega))=1\}$$

- Then $U_\rho(\omega)$ is a random set

University of Florida

Dempster-Shafer Representation

- The capacity functional of $U_\rho(\omega)$ is the plausibility

$$Pl_\rho(A) = P\lfloor U_\rho(\omega) \cap A \neq \phi \rfloor$$
$$= P[\{\omega \in \Omega \,|\, U_\rho(\omega) \cap A \neq \phi \quad \text{and} \, U_\rho(\omega) \neq \phi\}]$$

- The belief is the dual

$$Bel_\rho(A) = P\lfloor U_\rho(\omega) \subset A \rfloor$$

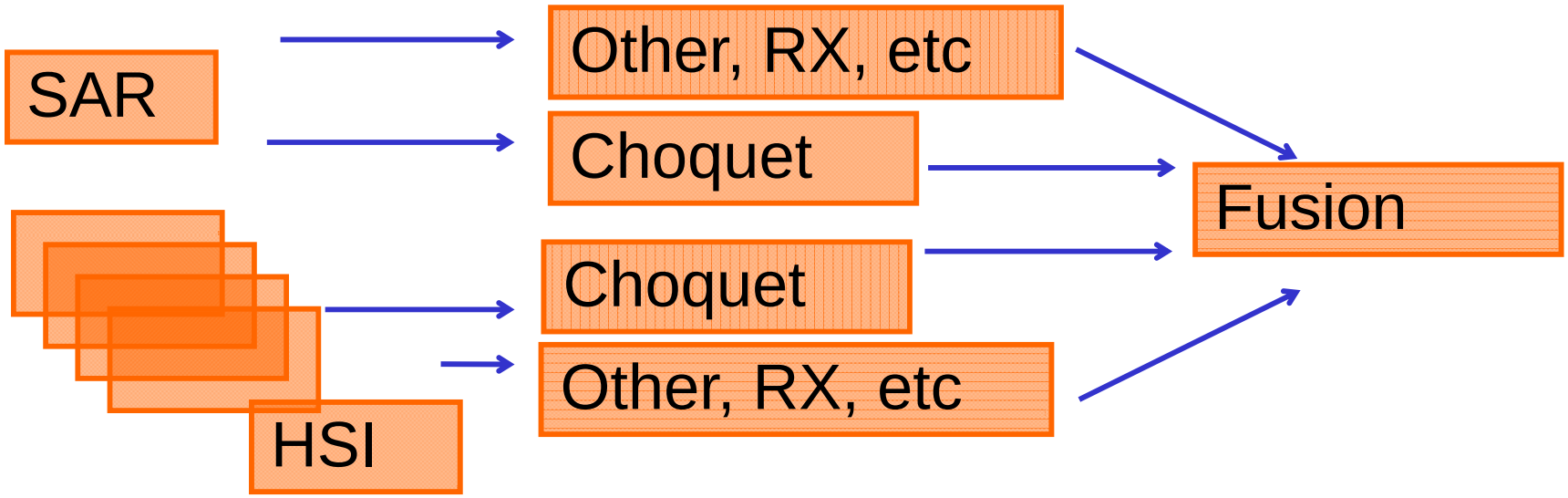
- The Choquet integral is the expected value of the Belief/Plausibility functions (EM)

$$C_{Bel_\rho}(f) = 1 - \int_0^1 Bel_\rho(1 - f_\alpha) d\alpha \quad \text{and} \quad \mathrm{C_{Pl}}(f) = \int_0^1 Pl_\rho(f_\alpha) d\alpha$$

University of Florida

Choquet Integral Fusion

- Goal: Develop Optimization Methods for Fusion Operators Based on Generic Choquet Integral Representation using EM, Quadratic Programming, and other methodologies



Alternating Quadratic Optimization

$$\mathbf{u} = [\mu(\{x_1\}) \quad \mu(\{x_2\}) \quad \ldots \quad \mu(\{x_n\}) \quad \mu(\{x_1, x_2\}) \quad \ldots \quad \mu(\{x_1, x_2, \ldots, x_n\})]^T$$

Minimize $$\frac{1}{2}\mathbf{u}^T\left(\sum_{p=1}^{P}\mathbf{v}^p\mathbf{v}^{pT}\right)\mathbf{u} + \mathbf{u}^T\sum_{p=1}^{P}\left(f_g^p(x_{(1)}) - o_p\right)\mathbf{v}^p$$

Subject to $Qu \leq d$

The matrix Q represents the constraints $-\mu(A_{(i)})+\mu(A_{(i+1)})\leq 0$ and $\mu(A_{(n)})\leq 1$

University of Florida

Kernel Matched Signal Detectors for Hyperspectral Target Detection

Heesung Kwon

Nasser M. Nasrabadi

U.S. Army Research Laboratory, Attn: AMSRL-SE-SE

2800 Powder Mill Road, Adelphi, MD 20783, USA

Outline

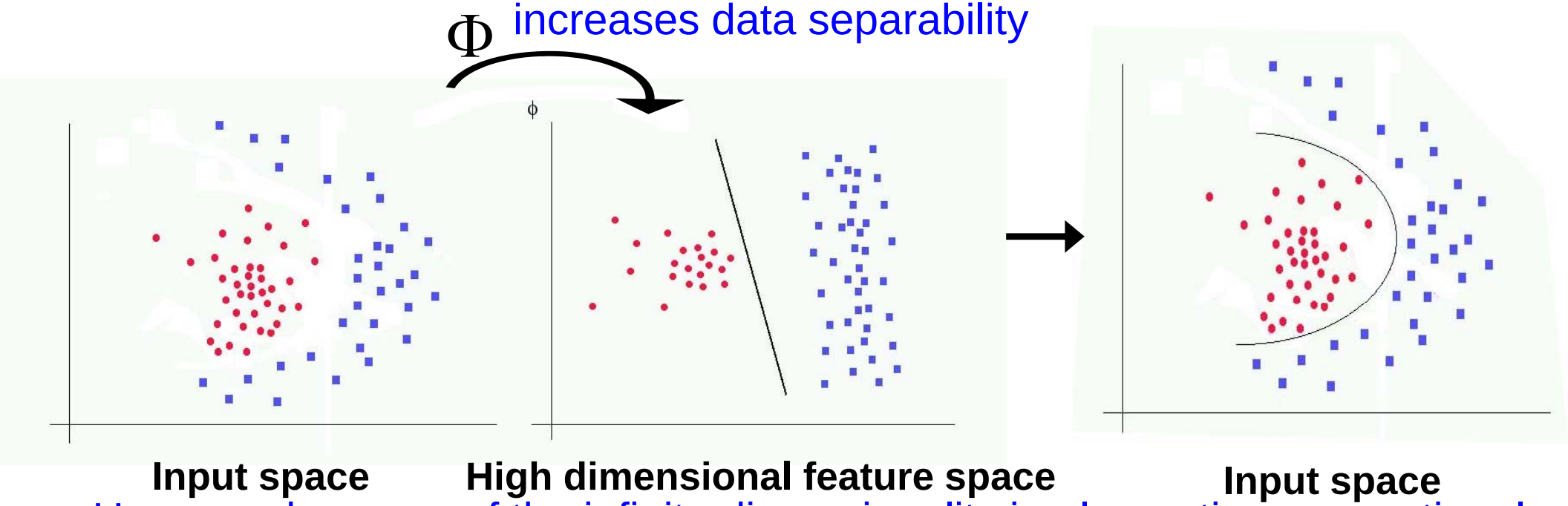
- Exploitation of Nonlinear Correlations Using Matched Filters
- Why Kernels
- Kernel Trick
- Conventional matched filters
- Kernel matched filters
- Detection results

Nonlinear Mapping of Data Exploitation of Nonlinear Correlations

- Nonlinear mapping Φ

$$\Phi : X \to F \qquad x \mapsto \Phi(x) = (\sqrt{\lambda_1}\psi_1(x), \sqrt{\lambda_2}\psi_2(x), \cdots)$$

- Statistical learning (VC): Mapping into a higher dimensional space increases data separability



Input space **High dimensional feature space** **Input space**

- However, because of the infinite dimensionality implementing conventional detectors in the feature space is not feasible using conventional methods
- Kernel trick : $k(x,y) = <\Phi(x), \Phi(y)>$
- Convert the detector expression into dot product forms →

Kernel-based nonlinear version of the conventional detector

Kernel Trick

$$k(\mathrm{x}, \mathrm{y}) = < \Phi(\mathrm{x}), \Phi(\mathrm{y}) >$$

- Consider 2-D input patterns $\mathrm{x} = (x_1, x_2)$, where $\mathrm{x} = (x_1, x_2) \in R^2$
- If a 2nd order monomial is used as the nonlinear mapping

$$\Phi : R^2 \rightarrow R^3, \Phi(\mathrm{x}) = \mathrm{x}^2$$

$$(x_1, x_2) \mapsto (z_1, z_2, z_3) := (x_1^2, \sqrt{2} x_1 x_2, x_2^2)$$

- Example of the kernel trick

$$< \Phi(\mathbf{x}), \Phi(\mathbf{y}) >= (x_1^2, \sqrt{2} x_1 x_2, x_2^2)(y_1^2, \sqrt{2} y_1 y_2, y_2^2)^T = x_1^2 y_1^2 + 2 x_1 x_2 y_1 y_2 + x_2^2 y_2^2$$

$$= ((x_1, x_2)(y_1, y_2)^T)^2 =< \mathbf{x}, \mathbf{y} >^2 := k(\mathbf{x}, \mathbf{y})$$

$$k(\mathrm{x}, \mathrm{y}) = < \Phi(\mathrm{x}), \Phi(\mathrm{y}) >, \; k : \text{kernel function}$$

- This property generalizes for $\mathbf{x}, \mathbf{y} \in R^N$ and $d \in R$

$$k(\mathrm{x}, \mathrm{y}) =< \mathrm{x}, \mathrm{y} >^d$$

Examples of Kernels

1. Gaussian RBF kernel: $k(\mathbf{x},\mathbf{y}) = \exp(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}) = \Phi(\mathrm{x})\Phi(\mathrm{y})$

Possible realization of

$$\Phi(\mathrm{x}) = (\sqrt{\lambda_1}\psi_1(\mathrm{x}), \sqrt{\lambda_2}\psi_2(\mathrm{x}), \cdots)$$

2. Inverse multiquadric kernel: $k(\mathbf{x},\mathbf{y}) = \frac{1}{\sqrt{\|\mathrm{x}-\mathrm{y}\|^2 + c^2}}$

3. Spectral angle-based kernel: $k(\mathbf{x},\mathbf{y}) = \frac{\mathrm{xy}}{\|\mathrm{x}\|\|\mathrm{y}\|}$

4. Polynomial kernel: $k(\mathbf{x},\mathbf{y}) = ((\mathbf{x}\cdot\mathbf{y}) + \theta)^d$

Matched Subspace Detection (MSD)

- Consider a linear mixed model:

$$H_0 : \mathbf{y} = \mathbf{B}\boldsymbol{\zeta} + \mathbf{n}, \qquad \text{Target absent} \approx N(\mathbf{B}\boldsymbol{\zeta}, \sigma^2\mathbf{I})$$

$$H_1 : \mathbf{y} = \mathbf{T}\boldsymbol{\theta} + \mathbf{B}\boldsymbol{\zeta} + \mathbf{n} \qquad \text{Target present} \approx N(\mathbf{T}\boldsymbol{\theta} + \mathbf{B}\boldsymbol{\zeta}, \sigma^2\mathbf{I})$$

- where $\mathbf{T}$ and $\mathbf{B}$ represent matrices whose column vectors span the target and the background subspaces
  $\boldsymbol{\zeta}$ and $\boldsymbol{\theta}$ are unknown vectors of coefficients, $\mathbf{n}$ is a Gaussian random noise distributed as $N(0, \sigma^2\mathbf{I})$
- The log Generalized likelihood ratio test (GLRT) is given by

$$L_2(\mathbf{y}) = \frac{\mathrm{p}(\mathrm{y} \mid \text{signal present})}{\mathrm{p}(\mathrm{y} \mid \text{signal absent})} = \frac{\mathbf{y}^T(\mathbf{I} - \mathbf{P}_\mathbf{B})\mathbf{y}}{\mathbf{y}^T(\mathbf{I} - \mathbf{P}_\mathbf{BT})\mathbf{y}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta$$

- where $\mathbf{P}_\mathbf{B} = \mathbf{B}\mathbf{B}^T$, $\mathbf{P}_\mathbf{TB} = [\mathbf{T}\ \mathbf{B}]\{[\mathbf{T}\ \mathbf{B}]^T[\mathbf{T}\ \mathbf{B}]\}^{-1}[\mathbf{T}\ \mathbf{B}]$

Kernel Matched Subspace Detection

- Define the matched subspace detector in the feature space
- To kernelize we use the kernel PCA, and kernel function properties as shown below

$$L_2(\Phi(y)) = \frac{\Phi(y)^T(I_\Phi - P_{B_\Phi})\Phi(y)}{\Phi(y)^T(I_\Phi - P_{B_\Phi T_\Phi})\Phi(y)} = \frac{\Phi(y)^T(I_\Phi - B_\Phi B_\Phi^T)(y)}{\Phi(y)^T\begin{bmatrix}T_\Phi & B_\Phi\end{bmatrix}\begin{bmatrix}T_\Phi^T T_\Phi & T_\Phi^T B_\Phi \\ B_\Phi^T T_\Phi & B_\Phi^T B_\Phi\end{bmatrix}^{-1}\begin{bmatrix}T_\Phi^T \\ B_\Phi^T\end{bmatrix}\Phi(y)}$$

$$B_\Phi = Z_{B_\Phi}\beta, T_\Phi = Z_{T_\Phi}\tau, \quad B_\Phi^T\Phi(y) = \beta^T k(Z_B, y), \text{ and } T_\Phi^T\Phi(y) = \tau^T k(Z_T, y)$$

$$\therefore \Phi(y)^T B_\Phi B_\Phi \Phi(y) = k(Z_B, y)^T \beta\beta^T k(Z_B, y)$$

$$L_{2k} = \frac{k(y, y) - k(Z_B, y)^T \beta\beta^T k(Z_B, y)}{k(y, y) - \begin{bmatrix}\tau^T k(Z_B, y) & \beta^T k(Z_B, y)\end{bmatrix}^T \Lambda_1^{-1}\begin{bmatrix}\tau^T k(Z_B, y) \\ \beta^T k(Z_B, y)\end{bmatrix}}$$

$$\Lambda_1 = \begin{bmatrix}\tau^T K(Z_T, Z_T)\tau & \tau^T K(Z_T, Z_B)\beta \\ \beta^T K(Z_B, Z_T)\tau & \beta^T K(Z_B, Z_B)\beta\end{bmatrix}$$

MSD vs. Kernel MSD

- GLRT for the MSD:

$$L_2(\mathbf{y}) = \frac{\mathbf{y}^T(\mathbf{I} - \mathbf{P_B})\mathbf{y}}{\mathbf{y}^T(\mathbf{I} - \mathbf{P_{BT}})\mathbf{y}}$$

- Nonlinear GLRT for the MSD in feature space:

$$L_2(\Phi(\mathbf{y})) = \frac{\Phi(\mathbf{y})^T(\mathbf{I}_\Phi - \mathbf{P}_{\mathbf{B}_\Phi})\Phi(\mathbf{y})}{\Phi(\mathbf{y})^T(\mathbf{I}_\Phi - \mathbf{P}_{\mathbf{B}_\Phi\mathbf{T}_\Phi})\Phi(\mathbf{y})}$$

- Kernelized GLRT for the kernel MSD:

$$L_{2k} = \frac{k(y,y) - k(Z_B,y)^T \beta\beta^T k(Z_B,y)}{k(y,y) - \begin{bmatrix}\tau^T k(Z_B,y) & \beta^T k(Z_B,y)\end{bmatrix}^T \Lambda_1^{-1} \begin{bmatrix}\tau^T k(Z_B,y) \\ \beta^T k(Z_B,y)\end{bmatrix}}$$

Orthogonal Subspace Projector vs. Kernel OSP

- The model in the nonlinear feature space is

$$H_{0_\Phi} : \Phi(\mathbf{y}) = \mathbf{B}_\Phi \zeta_\Phi + \mathbf{n}_\Phi, \quad \text{Target absent}$$
$$H_{1_\Phi} : \Phi(\mathbf{y}) = \mathrm{s}_\Phi \mu_\Phi + \mathbf{B}_\Phi \zeta_\Phi + \mathbf{n}_\Phi \quad \text{Target present}$$

- The MLE for μ_ϕ in feature space is given as

$$\mu_\Phi = \frac{\Phi(\mathrm{s})^T (\mathbf{I}_\Phi - \mathbf{P}_{\mathbf{B}_\Phi}) \Phi(\mathbf{y})}{\Phi(\mathrm{s})^T (\mathbf{I}_\Phi - \mathbf{P}_{\mathbf{B}_\Phi}) \Phi(\mathrm{s})} \underset{H_{0\Phi}}{\overset{H_{1\Phi}}{\gtrless}} \eta$$

- The kernel version of μ_ϕ is given as

$$\mu_k = \frac{\mathrm{k}(\mathrm{s},\mathrm{y}) - \mathrm{k}(Z_B,\mathrm{s})^T \beta\beta^T \mathrm{k}(Z_B,\mathrm{y})}{\mathrm{k}(\mathrm{s},\mathrm{s}) - \mathrm{k}(Z_B,\mathrm{s})^T \beta\beta^T \mathrm{k}(Z_B,\mathrm{s})}$$

Linear Spectral Matched Filter & Nonlinear Spectral Matched Filter

- Spectral signal model

$$H_0 : \mathbf{x} = \mathbf{n}, \qquad a = 0: \text{ no target,}$$
$$H_1 : \mathbf{x} = a\mathbf{s} + \mathbf{n} \qquad a > 0: \text{ target present}$$

$\mathbf{n}$: background clutter noise
$\mathbf{s}$: target spectral signature,

- Linear matched filter is given as:

$$y(\mathrm{x}) = \mathrm{w}^{\mathrm{T}}\mathrm{x} = \frac{\mathrm{s}^{\mathrm{T}}\mathrm{C}^{-1}\mathrm{x}}{\mathrm{s}^{\mathrm{T}}\mathrm{C}^{-1}\mathrm{s}},$$

- In the feature space, the equivalent signal model

$$H_0 : \Phi(\mathrm{x}) = \mathrm{n}_\Phi, \qquad \text{No target}$$
$$H_1 : \Phi(\mathrm{x}) = a_\Phi \Phi(\mathrm{s}) + \mathrm{n}_\Phi \qquad \text{Target present}$$

- Output of the matched filter in the feature space

$$y(\Phi(\mathrm{x})) = \mathrm{w}_\Phi^{\mathrm{T}}\Phi(\mathrm{x}) = \frac{\Phi(\mathrm{s})\mathrm{C}_\Phi^{-1}\Phi(\mathrm{x})}{\Phi(\mathrm{s})\mathrm{C}_\Phi^{-1}\Phi(\mathrm{s})}$$

Kernelization of Spectral Matched Filter in Feature space

- Using the following properties of PCA and Kernel PCA

$$\mathbf{C}_{\Phi}^{-1} = \mathbf{V}_{\Phi}\Lambda^{-1}\mathbf{V}_{\Phi}^{\mathrm{T}}, \qquad \mathbf{V}_{\Phi} = [\mathbf{v}_{\Phi}^{1}, \mathbf{v}_{\Phi}^{2}, \cdots, \mathbf{v}_{\Phi}^{M}]$$

- Each eigenvector can be represented in terms of the input data

$$\mathbf{V}_{\Phi} = \mathbf{X}_{\Phi}\mathbf{B}, \qquad \mathbf{B} = [\mathbf{b}^{1}, \mathbf{b}^{2}, \cdots, \mathbf{b}^{M}]$$

- Inverse Covariance matrix is now

$$\mathbf{C}_{\Phi}^{-1} = \mathbf{X}_{\Phi}\mathbf{B}\Lambda^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{X}_{\Phi}^{\mathrm{T}}$$

- Kernel matrix, $\mathbf{K}$, spectral decomposition (kernel PCA)

$$\mathbf{K}^{-1} = \frac{1}{M}\mathbf{B}\Lambda^{-1}\mathbf{B}^{T}, \quad \text{where } \mathbf{K}(\mathbf{X},\mathbf{X}) = \mathbf{K}_{ij} = k(\mathbf{x}_i,\mathbf{x}_j) \quad \mathbf{x}_i,\mathbf{x}_j \in \mathbf{X}$$

$$y(\Phi(\mathrm{x})) = \frac{\Phi(\mathrm{s})^{\mathrm{T}}\mathrm{X}_{\Phi}\mathrm{B}\Lambda^{-1}\mathrm{B}^{\mathrm{T}}\mathrm{X}_{\Phi}^{\mathrm{T}}\Phi(\mathrm{x})}{\Phi(\mathrm{s})\mathrm{X}_{\Phi}\mathrm{B}\Lambda^{-1}\mathrm{B}^{\mathrm{T}}\mathrm{X}_{\Phi}^{\mathrm{T}}\Phi(\mathrm{s})},$$

- The kernelized version of matched filter

$$y(\mathrm{k}_{\mathrm{x}}) = \frac{\mathrm{k}(\mathrm{X},\mathrm{s})\mathrm{K}^{-1}\mathrm{k}(\mathrm{X},\mathrm{x})}{\mathrm{k}(\mathrm{X},\mathrm{s})\mathrm{K}^{-1}\mathrm{k}(\mathrm{X},\mathrm{s})},$$

$$\mathrm{k}(\mathrm{X},\mathrm{s}) = (k(\mathrm{x}_1,\mathrm{s}), k(\mathrm{x}_2,\mathrm{s}), \cdots, k(\mathrm{x}_{\mathrm{N}},\mathrm{s}))^{\mathrm{T}}$$
$$\mathrm{k}(\mathrm{X},\mathrm{x}) = (k(\mathrm{x}_1,\mathrm{x}), k(\mathrm{x}_2,\mathrm{x}), \cdots, k(\mathrm{x}_{\mathrm{N}},\mathrm{x}))^{\mathrm{T}}$$

Conventional MF vs. Kernel MF

- Conventional spectral matched filter

$$y(x) = w^T x = \frac{s^T C^{-1} x}{s^T C^{-1} s},$$

- Nonlinear matched filter

$$y(\Phi(x)) = w_\Phi^T \Phi(x) = \frac{\Phi(s) C_\Phi^{-1} \Phi(x)}{\Phi(s) C_\Phi^{-1} \Phi(s)}$$

- Kernel matched filter

$$y(k_x) = \frac{k(X, s) K^{-1} k(X, x)}{k(X, s) K^{-1} k(X, s)}$$

Adaptive Subspace Detection (ASD) & Nonlinear ASD

- Consider a linear mixed model:

$$H_0 : \mathbf{r} = \mathbf{n}, \qquad \text{Target absent} \approx (0, \mathbf{C})$$

$$H_1 : \mathbf{r} = \mathbf{U}\theta + \sigma\,\mathbf{n} \qquad \text{Target present} \approx (\mathbf{U}\theta, \sigma^2\mathbf{C})$$

where $\mathbf{U}$ represent the target subspace and $\mathbf{C}$ is the background covariance.

- The ASD is given by

$$D_{ASD}(\mathbf{r}) = \frac{\mathbf{r}^T \hat{\mathbf{C}}^{-1}\mathbf{U}(\mathbf{U}^T\hat{\mathbf{C}}^{-1}\mathbf{U})^{-1}\mathbf{U}^T\hat{\mathbf{C}}^{-1}\mathbf{r}}{\mathbf{r}^T\hat{\mathbf{C}}^{-1}\mathbf{r}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta$$

- The model in the nonlinear feature space is

$$H_{0_\Phi} : \Phi(\mathbf{r}) = \mathbf{n}_\Phi, \qquad \text{Target absent}$$

$$H_{1_\Phi} : \Phi(\mathbf{r}) = \mathbf{U}_\Phi\theta_\Phi + \sigma\,\mathbf{n}_\Phi \qquad \text{Target present}$$

- The ASD in feature space is given as

$$D_{ASD}(\Phi(\mathbf{r})) = \frac{\Phi(\mathbf{r})^T\hat{\mathbf{C}}_\Phi^{-1}\mathbf{U}_\Phi(\mathbf{U}_\Phi^T\hat{\mathbf{C}}_\Phi^{-1}\mathbf{U}_\Phi)^{-1}\mathbf{U}_\Phi^T\hat{\mathbf{C}}_\Phi^{-1}\Phi(\mathbf{r})}{\Phi(\mathbf{r})^T\hat{\mathbf{C}}^{-1}\Phi(\mathbf{r})} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{ASD}$$

ASD vs. Kernel ASD

- GLRT for the ASD:

$$D_{ASD}(\mathrm{r}) = \frac{\mathrm{r}^{\mathrm{T}}\,\hat{\mathbf{C}}^{-1}\mathbf{U}(\mathbf{U}^{\mathrm{T}}\hat{\mathbf{C}}^{-1}\mathbf{U})^{-1}\,\mathbf{U}^{\mathrm{T}}\hat{\mathbf{C}}^{-1}\,\mathrm{r}}{\mathrm{r}^{\mathrm{T}}\,\hat{\mathbf{C}}^{-1}\,\mathrm{r}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta$$

- Nonlinear GLRT for the ASD in feature space:

$$D_{ASD}(\Phi(\mathrm{r})) = \frac{\Phi(\mathrm{r})^{\mathrm{T}}\,\hat{\mathbf{C}}_{\Phi}^{-1}\mathbf{U}_{\Phi}(\mathbf{U}_{\Phi}^{\mathrm{T}}\hat{\mathbf{C}}_{\Phi}^{-1}\mathbf{U}_{\Phi})^{-1}\,\mathbf{U}_{\Phi}^{\mathrm{T}}\hat{\mathbf{C}}_{\Phi}^{-1}\,\Phi(\mathrm{r})}{\Phi(\mathrm{r})^{\mathrm{T}}\,\hat{\mathbf{C}}^{-1}\,\Phi(\mathrm{r})} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{ASD}$$

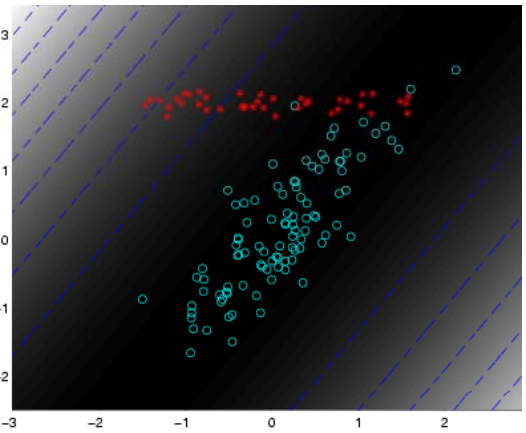
- Kernelized GLRT for the kernel ASD:

$$D_{KASD}(\mathrm{r}) = \frac{\mathrm{K_r}\,[\tau^{\mathrm{T}}\mathrm{K}(\mathrm{X},\mathrm{Y})^{\mathrm{T}}\mathrm{K_b}(\mathrm{X},\mathrm{X})^{-1}\mathrm{K}(\mathrm{X},\mathrm{Y})\tau\,]^{-1}\,\mathrm{K_r^T}}{\mathrm{k}(\mathrm{r},\mathrm{X})^{\mathrm{T}}\mathrm{K_b}(\mathrm{X},\mathrm{X})^{-1}\mathrm{k}(\mathrm{r},\mathrm{X})}$$

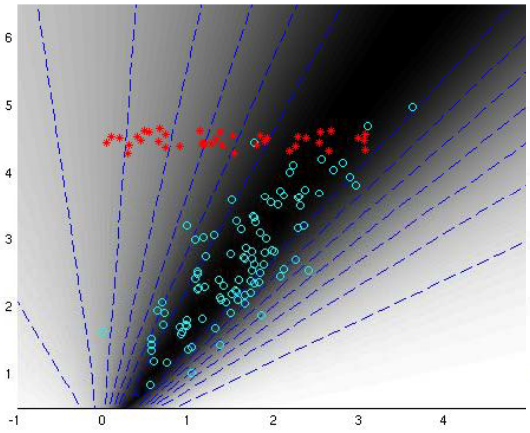
τ corresponds to eigenvectors of kernel matrix K(Y,Y)

A 2-D Gaussian Toy Example

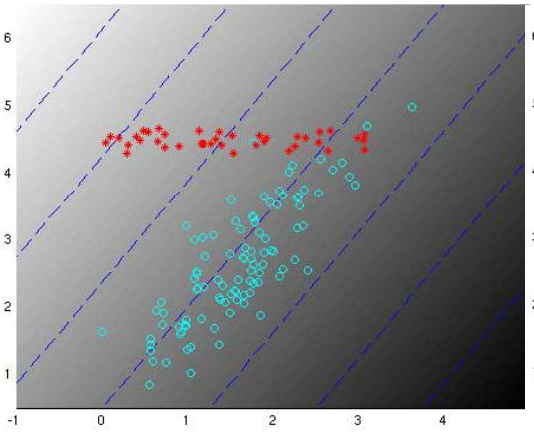
- Red dots belong to class H_1, blue dots belong to H_0



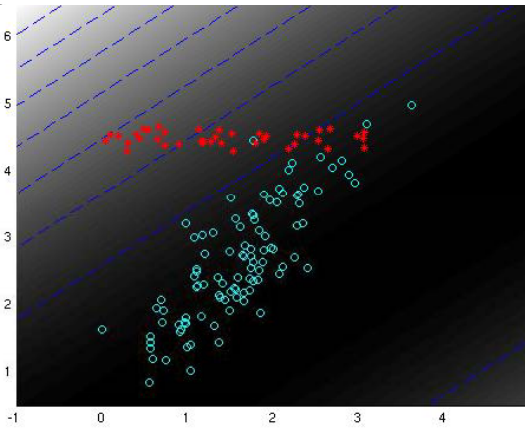
(a) MSD



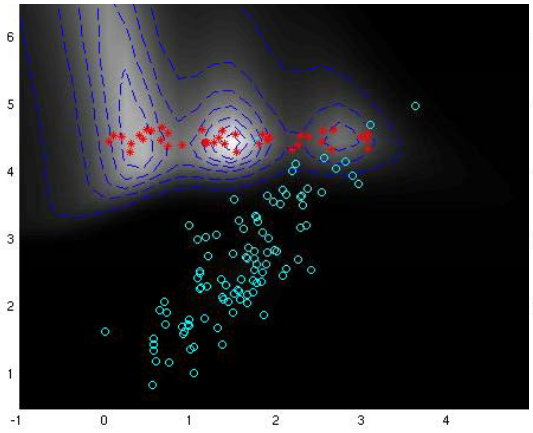
(c) ASD



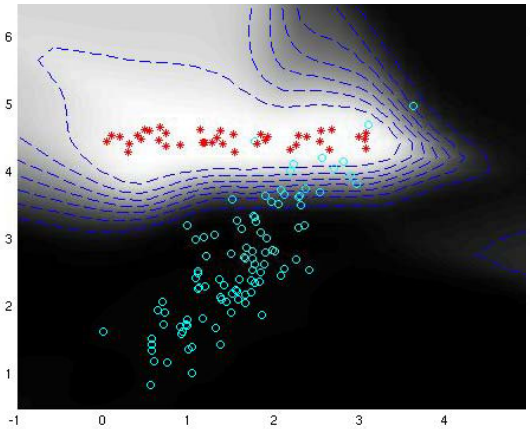
(e) OSP



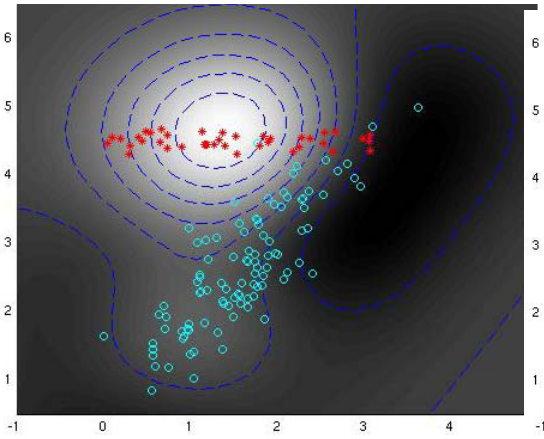
(g) SMF



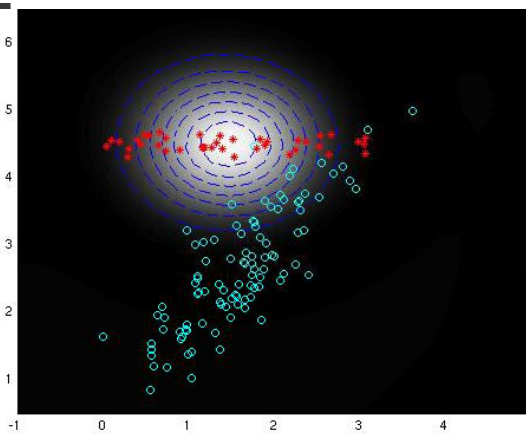
(b) KMSD



(d) KASD



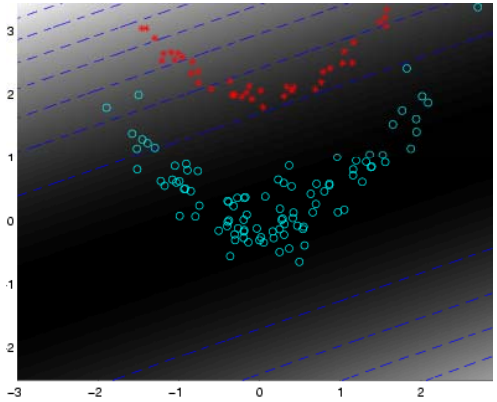
(f) KOSP



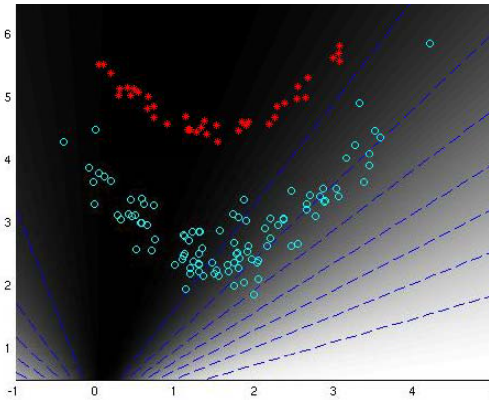
(h) KSMF

A 2-D Toy Example

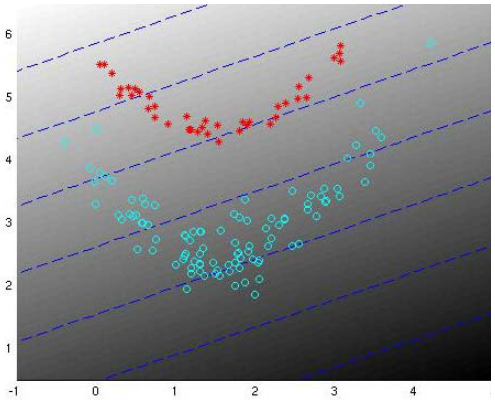
- Red dots belong to class H_1, blue dots belong to H_0



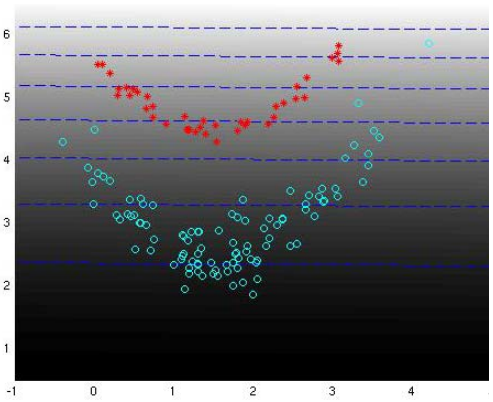
(a) MSD



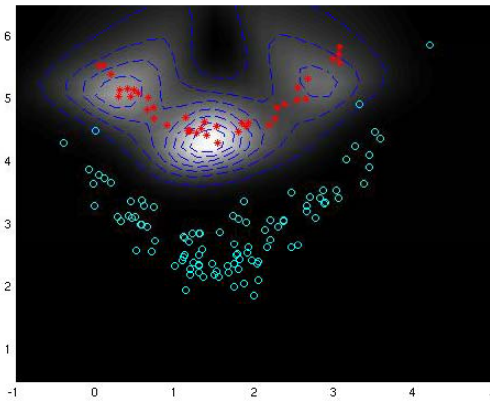
(c) ASD



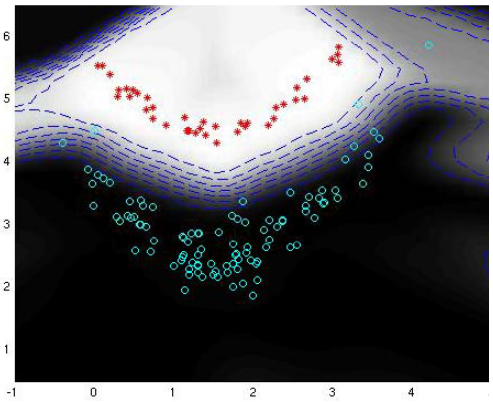
(e) OSP



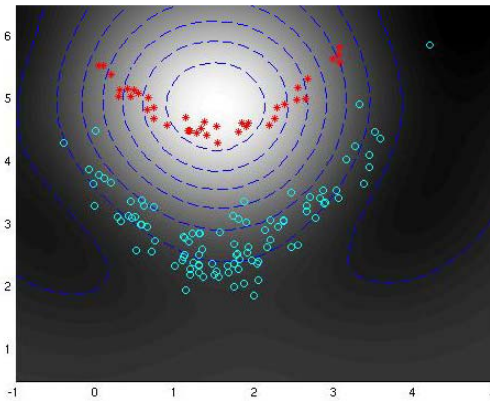
(g) SMF



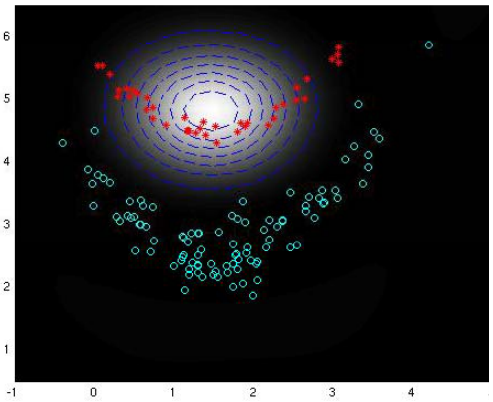
(b) KMSD



(d) KASD

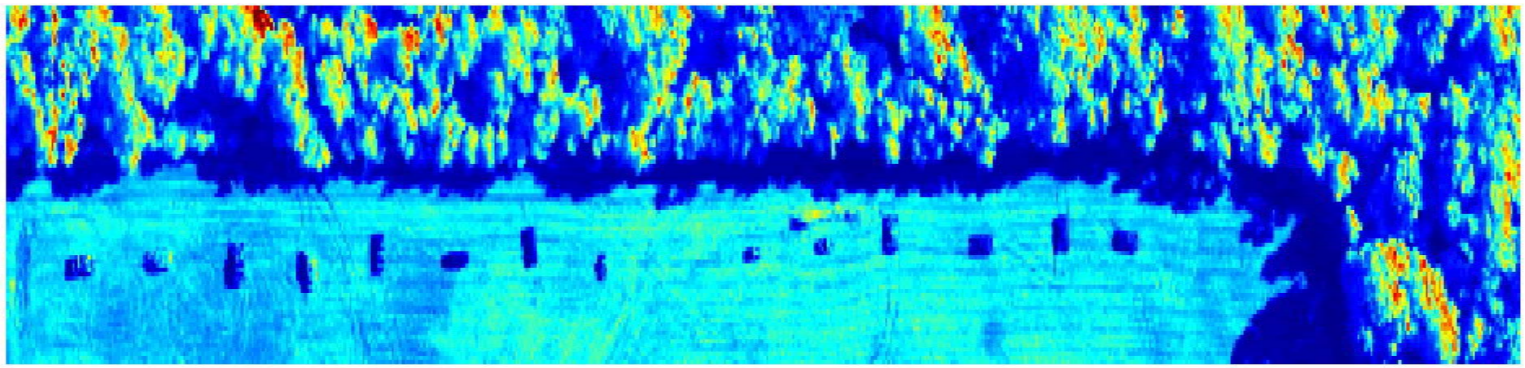


(f) KOSP

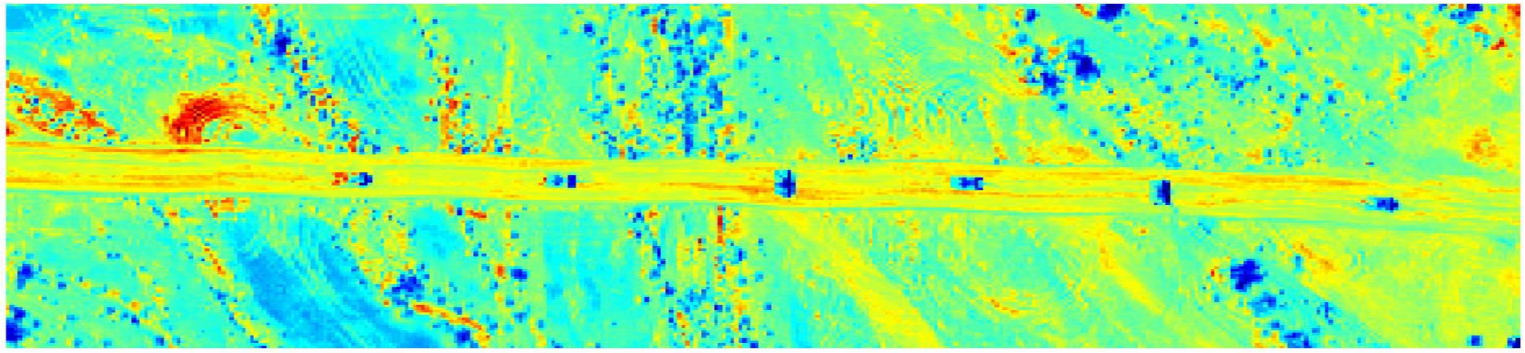


(h) KSMF

Test Images

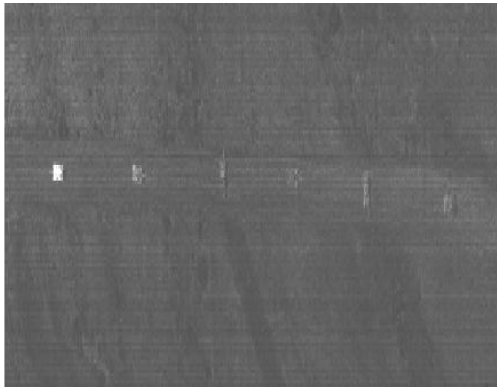


Forest Radiance I



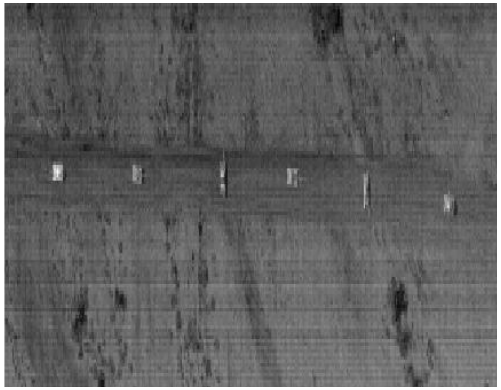
Desert Radiance II

Results for DR-II Image



(a) MSD

(c) ASD



(e) OSP

(g) SMF

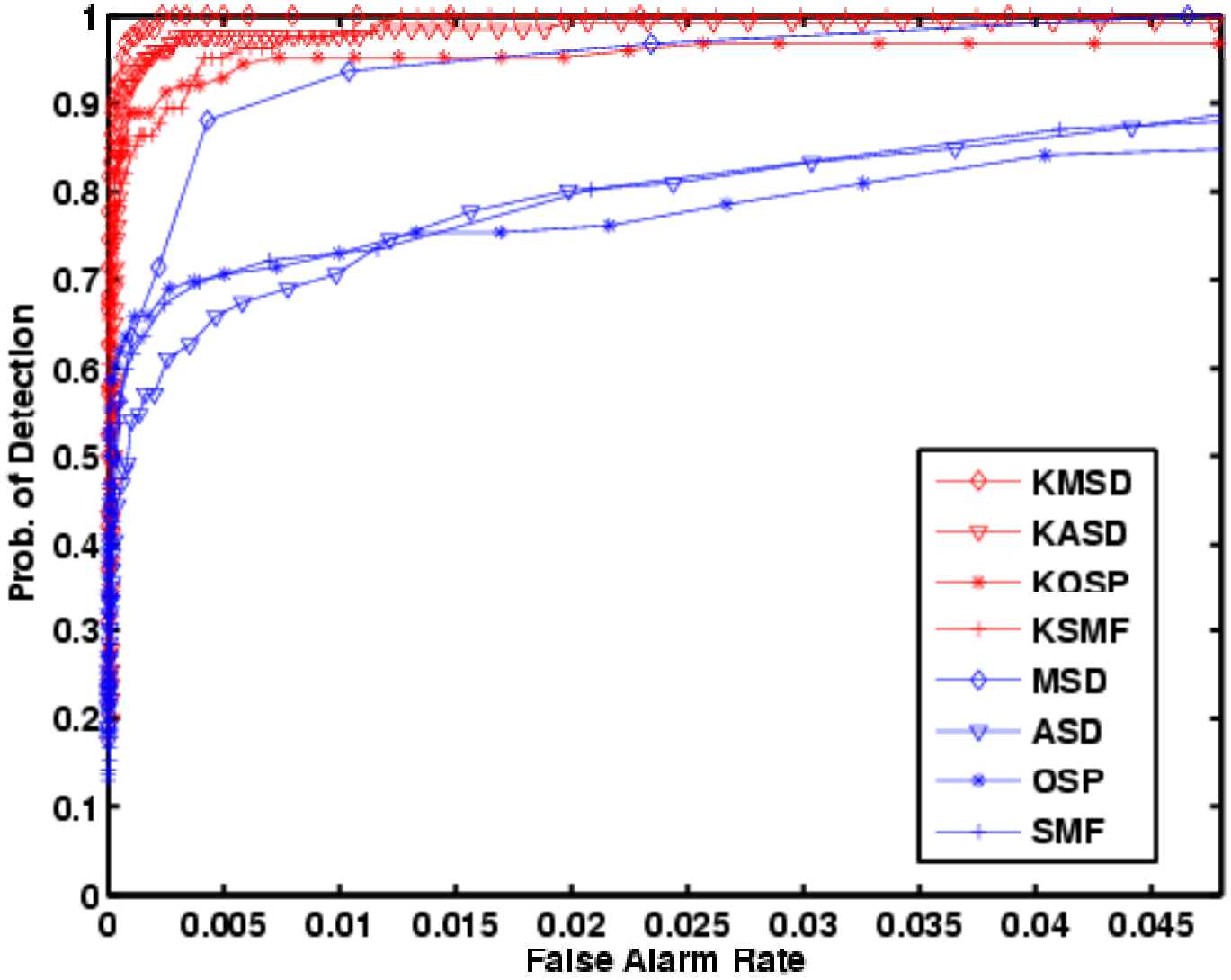
(b) KMSD

(d) KASD

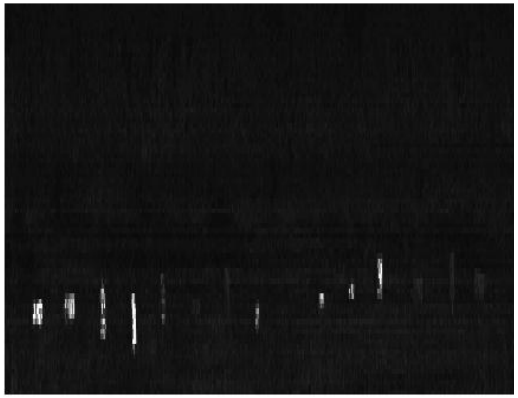
(f) KOSP

(h) KSMF

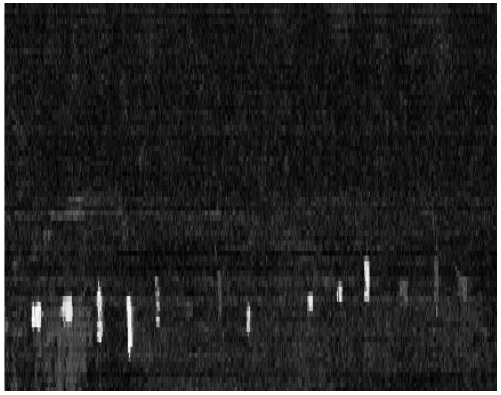
ROC Curves for DR-II Image



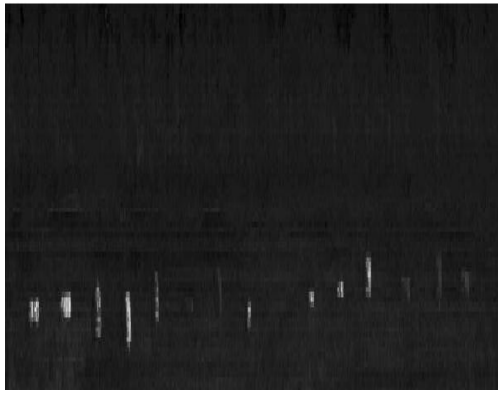
Results for FR-II Image



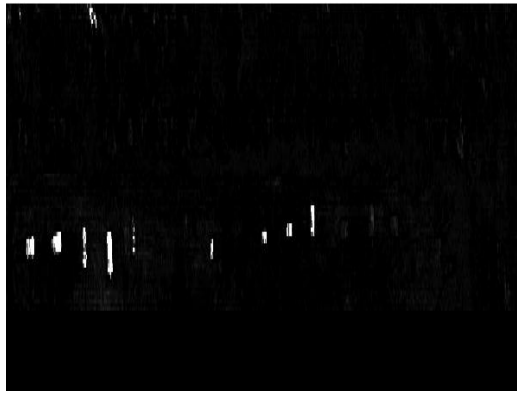
(a) MSD



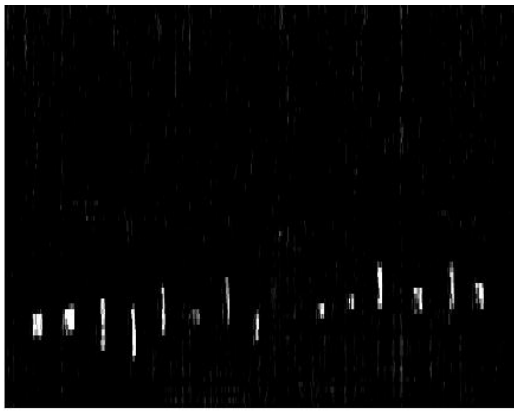
(c) ASD



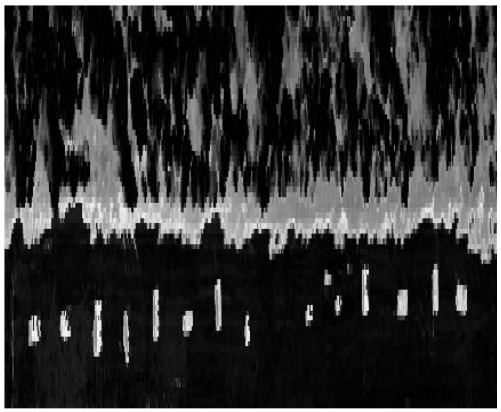
(e) OSP



(g) SMF

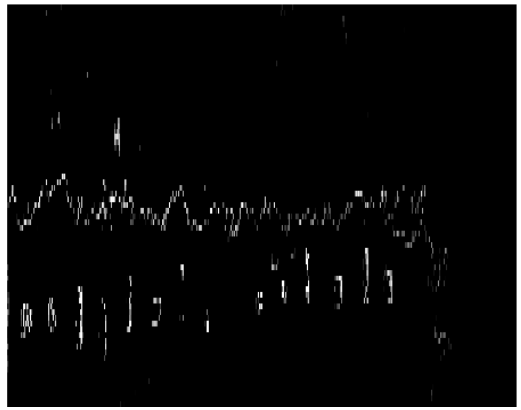


(b) KMSD



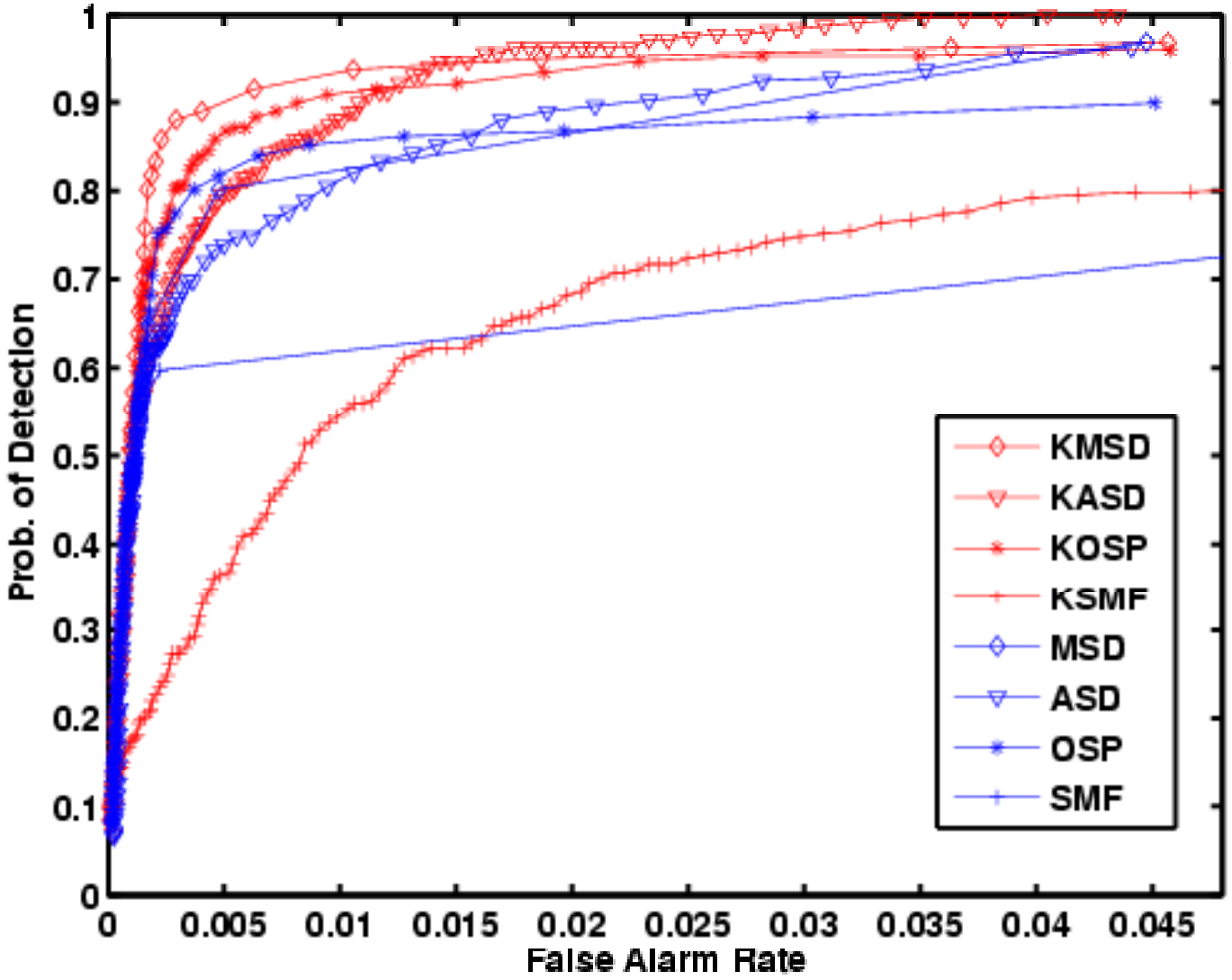
(d) KASD

(f) KOSP



(h) KSMF

ROC Curves for FR-II Image



Scene Anomaly Detection @ Ground Level Using HSI

Dalton Rosario
Army Research Laboratory

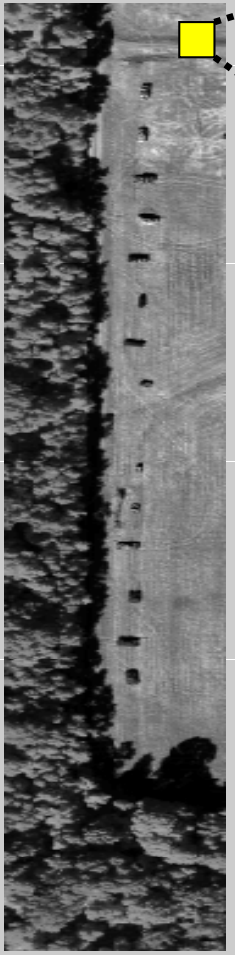
Rama Chellappa
University of Maryland

Outline

- Motivation/Idea
- New Family of Anomaly Detectors for HSI
  - Ground Vehicle Detection –Top View
  - Scene Anomaly Detection @ Ground Level
- Final Remarks

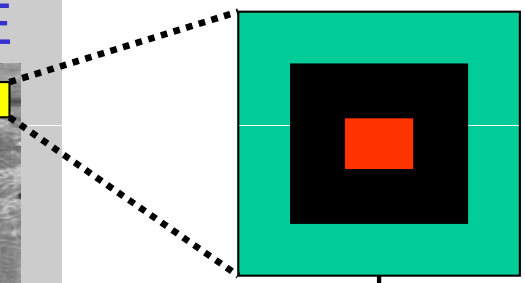
Physical Motivation

HYDICE



Aberdeen, MD

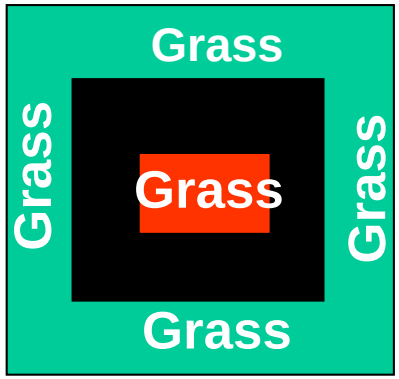
Visible to SWIR Hyperspectral Data



Inside-Outside Windows

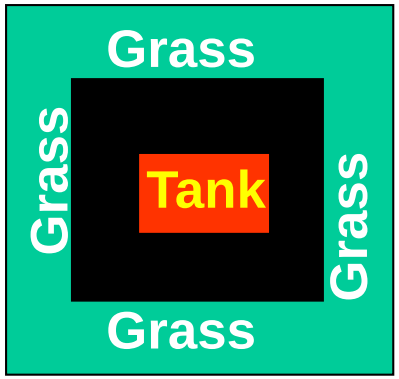
Local Anomaly Detector (e.g., FLD)

Case 1



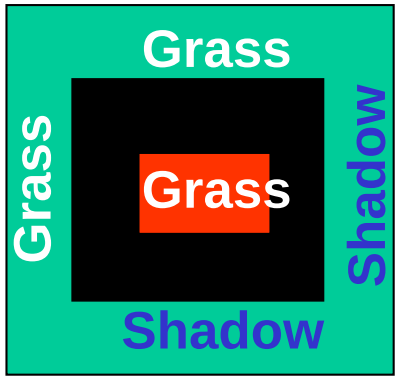
Non-Anomaly

Case 2

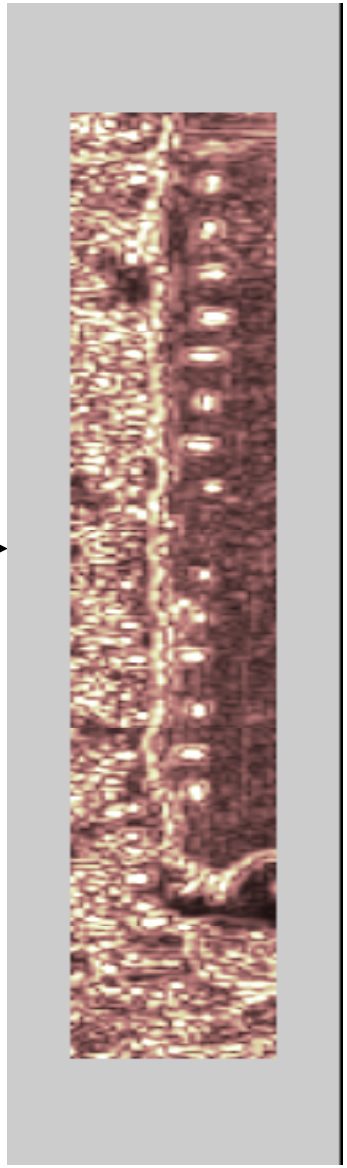


Anomaly

Case 3



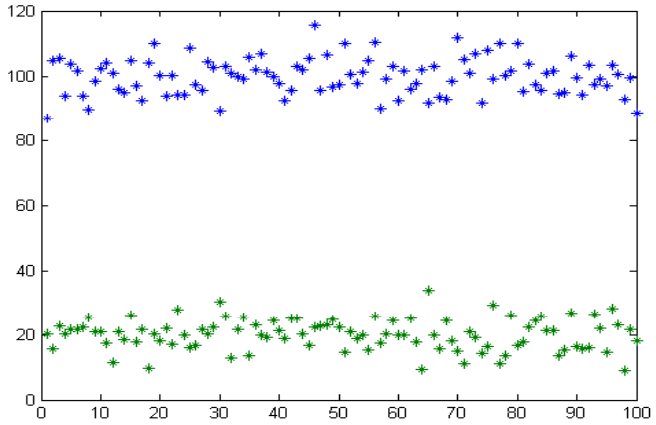
Anomaly

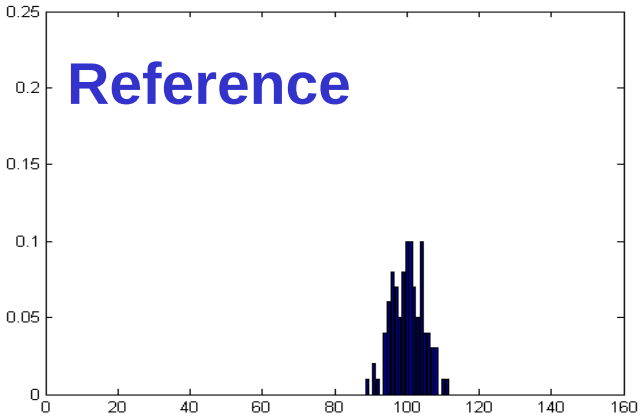


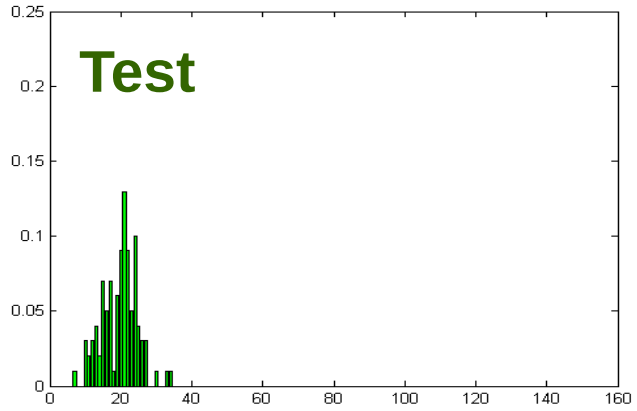
Output Surface

Statistical Motivation

Case 2

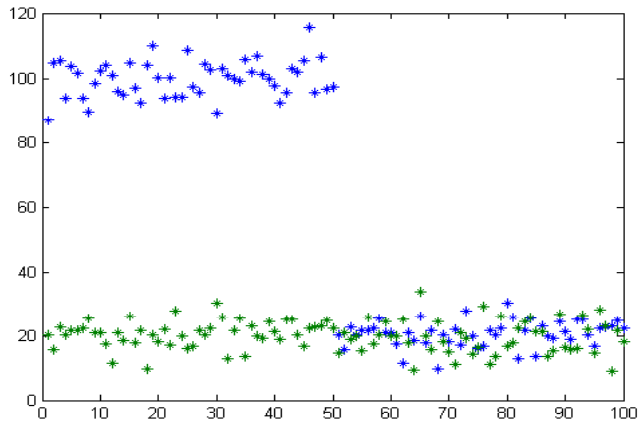


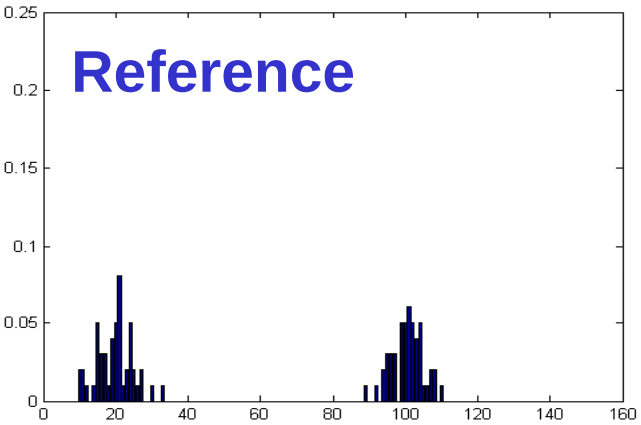


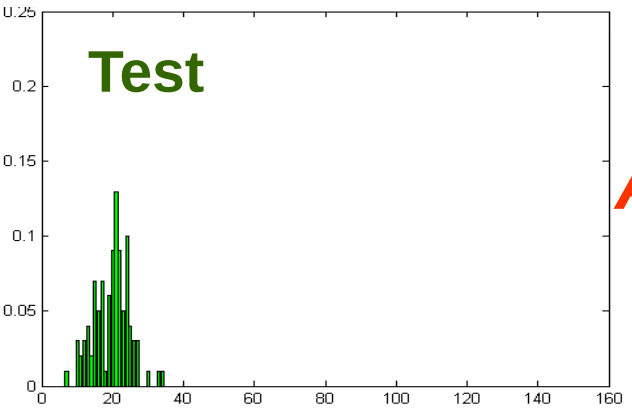


Anomaly

Case 3



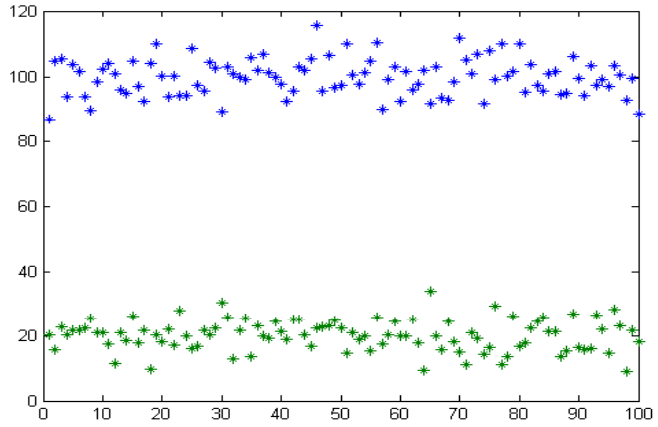


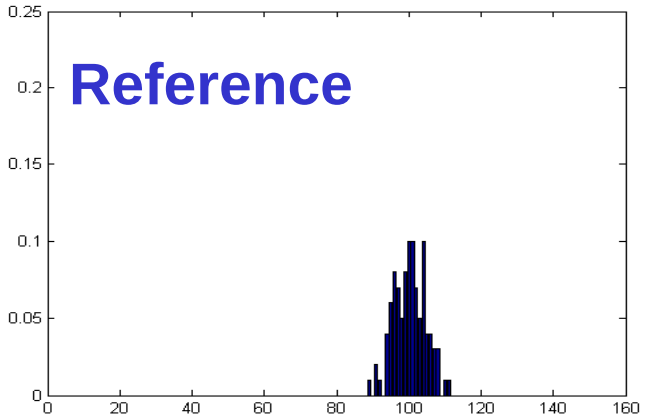


Anomaly

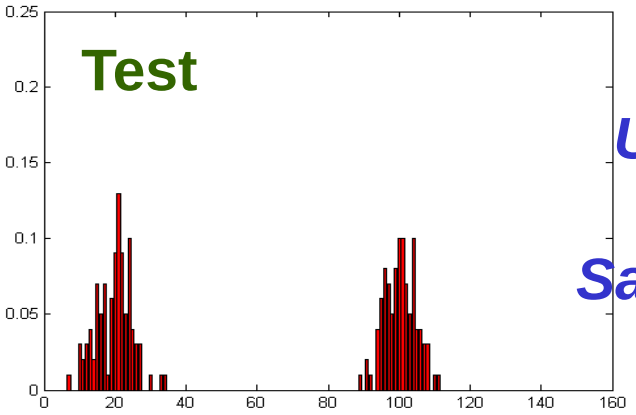
Indirect Comparison: *Combine & Compare*

Case 2



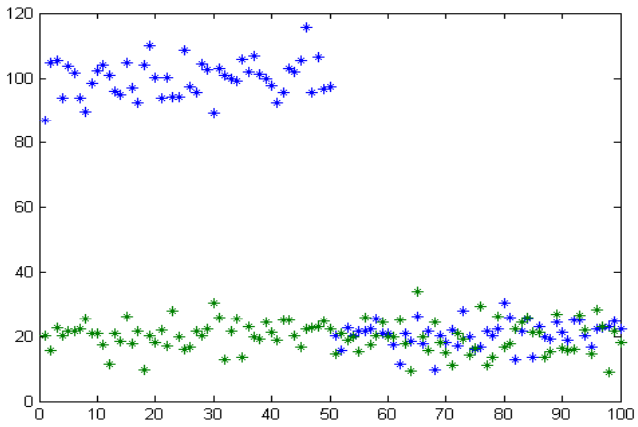


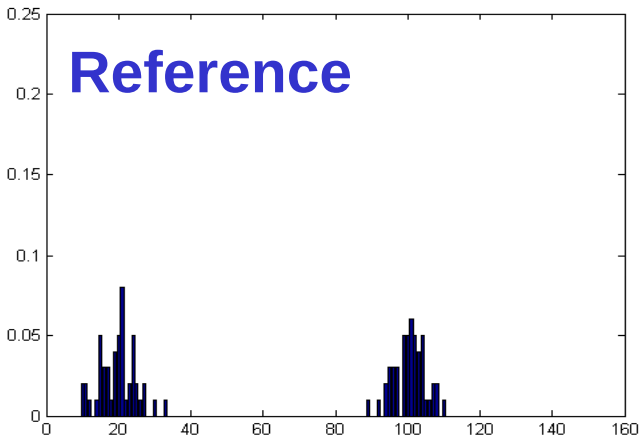
Anomaly

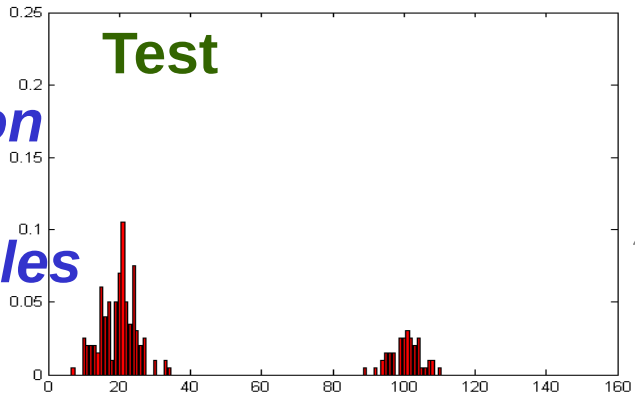


Union of Samples

Case 3



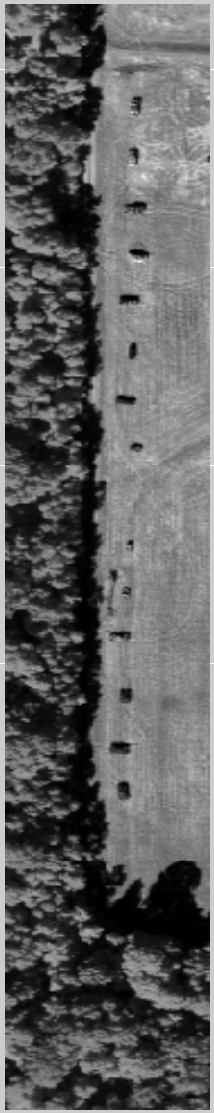




Softer Anomaly

New Family of Anomaly Detectors

Scene



V-SWIR

SemiP Detector

$$\eta_1 = \rho_1 \hat{\beta}^2 \left[\hat{E}(\mathbf{t}^2) - \hat{E}^2(\mathbf{t}) \right] \Rightarrow \chi_1^2$$

CFT Detector

$$\eta_3 = \rho_3 \left(\hat{\beta}_1^2 / S_1^2 \right) \Big/ \left(\hat{\beta}_2^2 / S_2^2 \right) \Rightarrow F_{1,1}$$

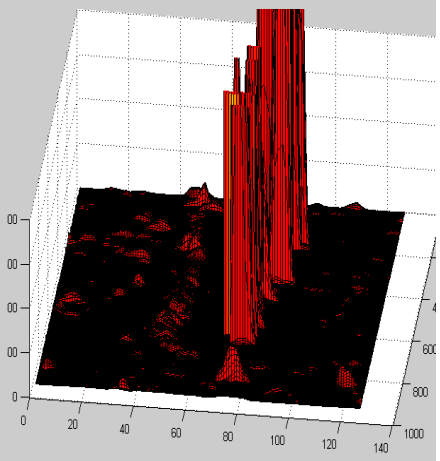
AsemiP Detector

$$\eta_2 = \frac{\hat{\tilde{\beta}}^2}{\left(\frac{1}{n_0} + \frac{1}{n_1} \right)} \frac{S_t^2}{S^4} \Rightarrow \chi_1^2$$

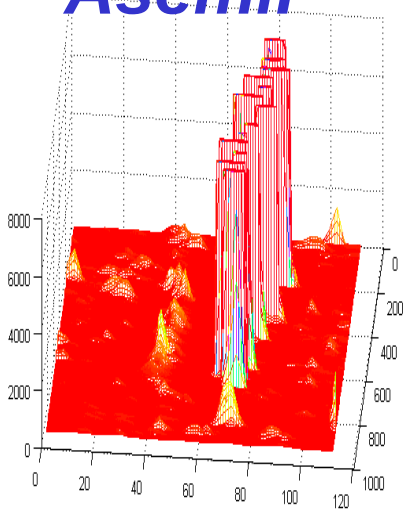
ANOVA Detector

$$\eta_4 = \frac{\sum_{i=1}^{2} n_i \left(\bar{x}_i - \bar{\bar{x}} \right)^2}{\frac{1}{N-2} \sum_{i=1}^{2} (n_i - 1) s_i^2} \sim F_{1,N-2}$$

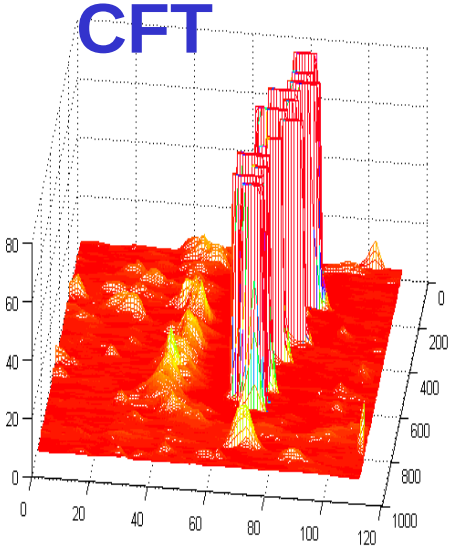
SemiP



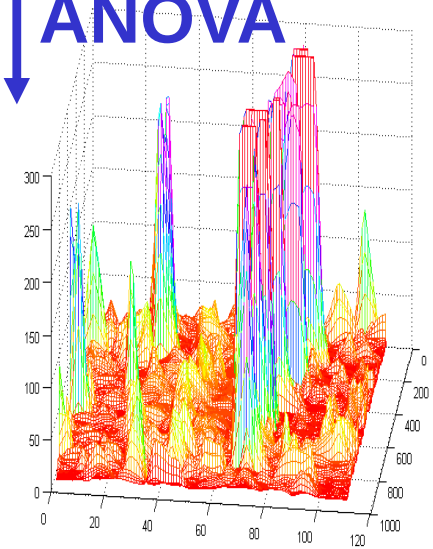
AsemiP

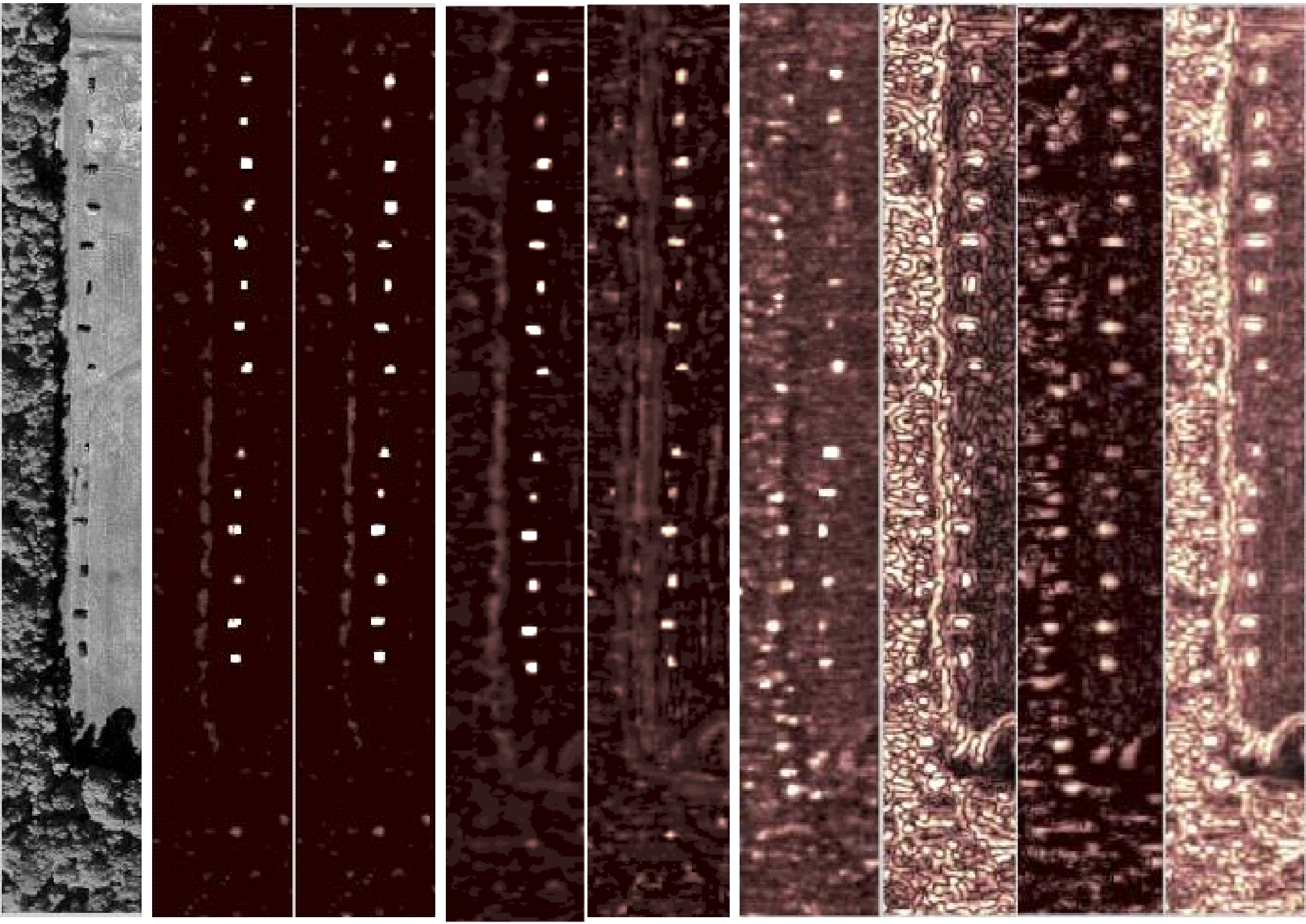


CFT



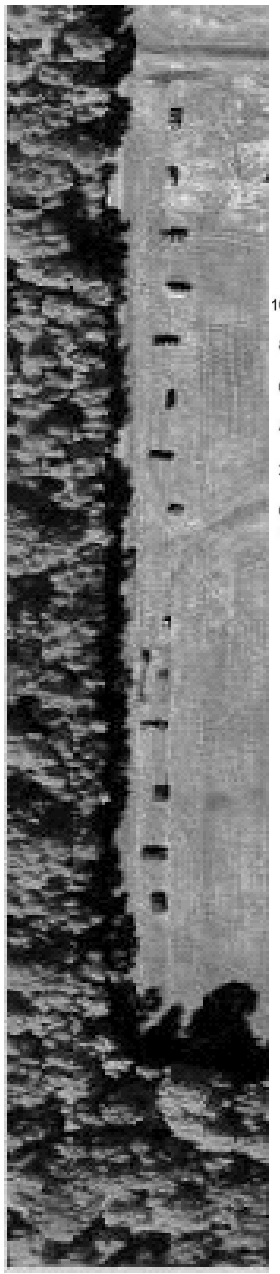
ANOVA



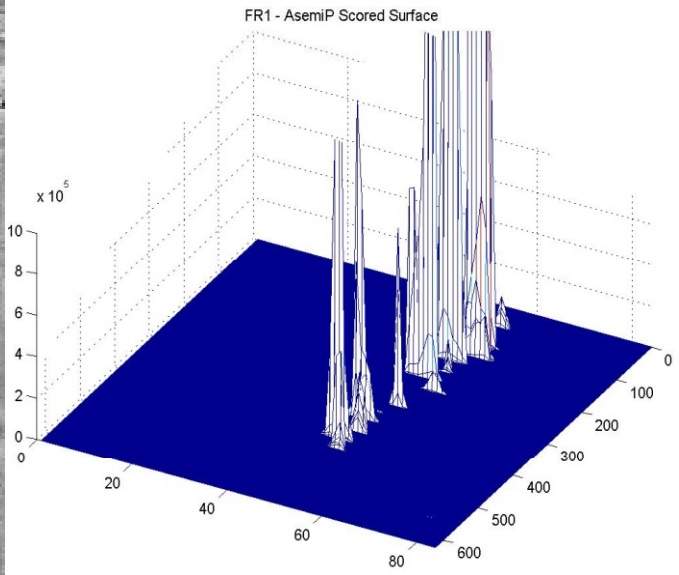


Scene *SemiP* *AsemiP* CFT ANOVA RX PCA EST FLD

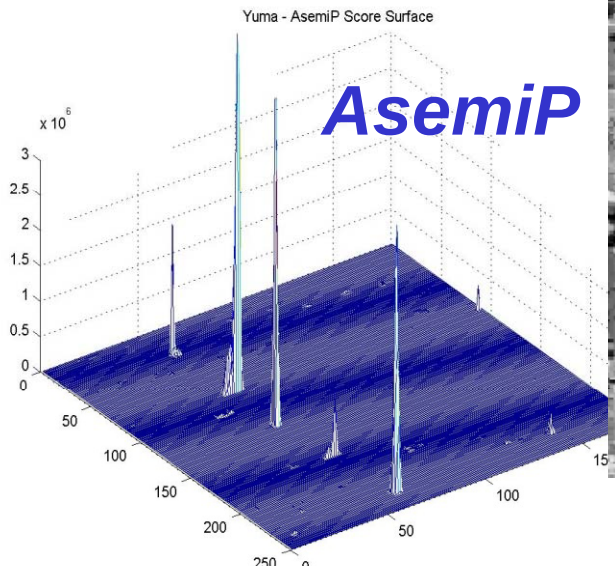
AsemiP Anomaly Detector



AsemiP



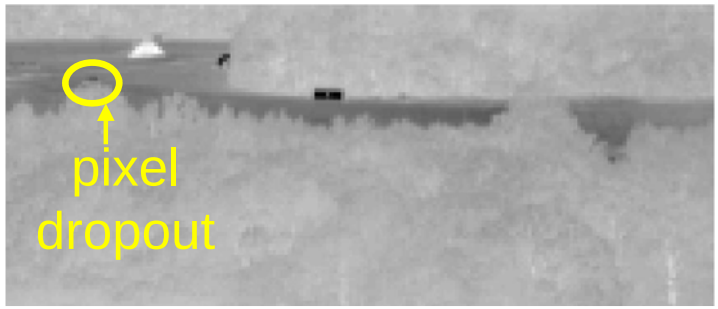
Desert Radiance



AsemiP

V-SWIR

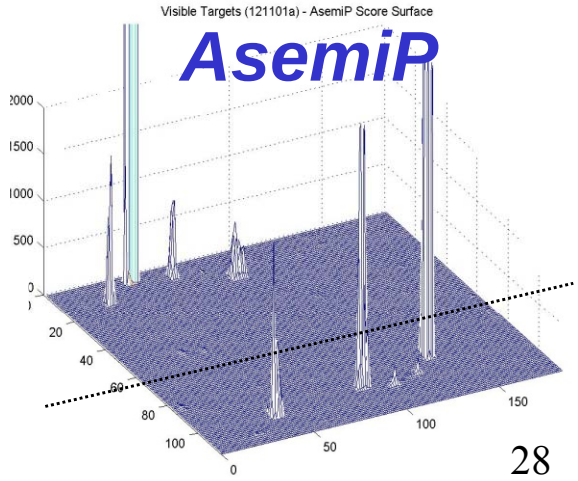
← **Forest Radiance** →



LWIR

No Targets

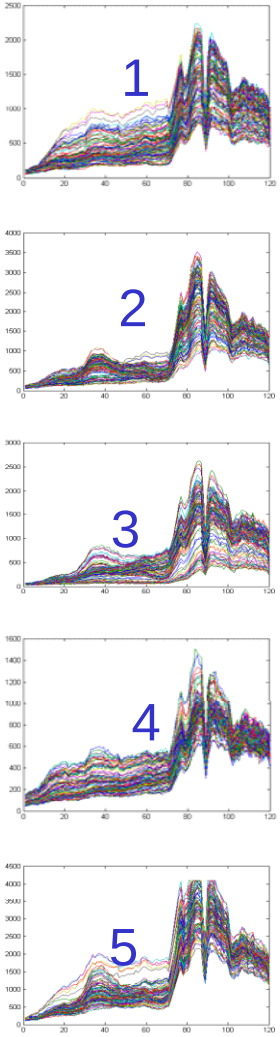
V-SWIR



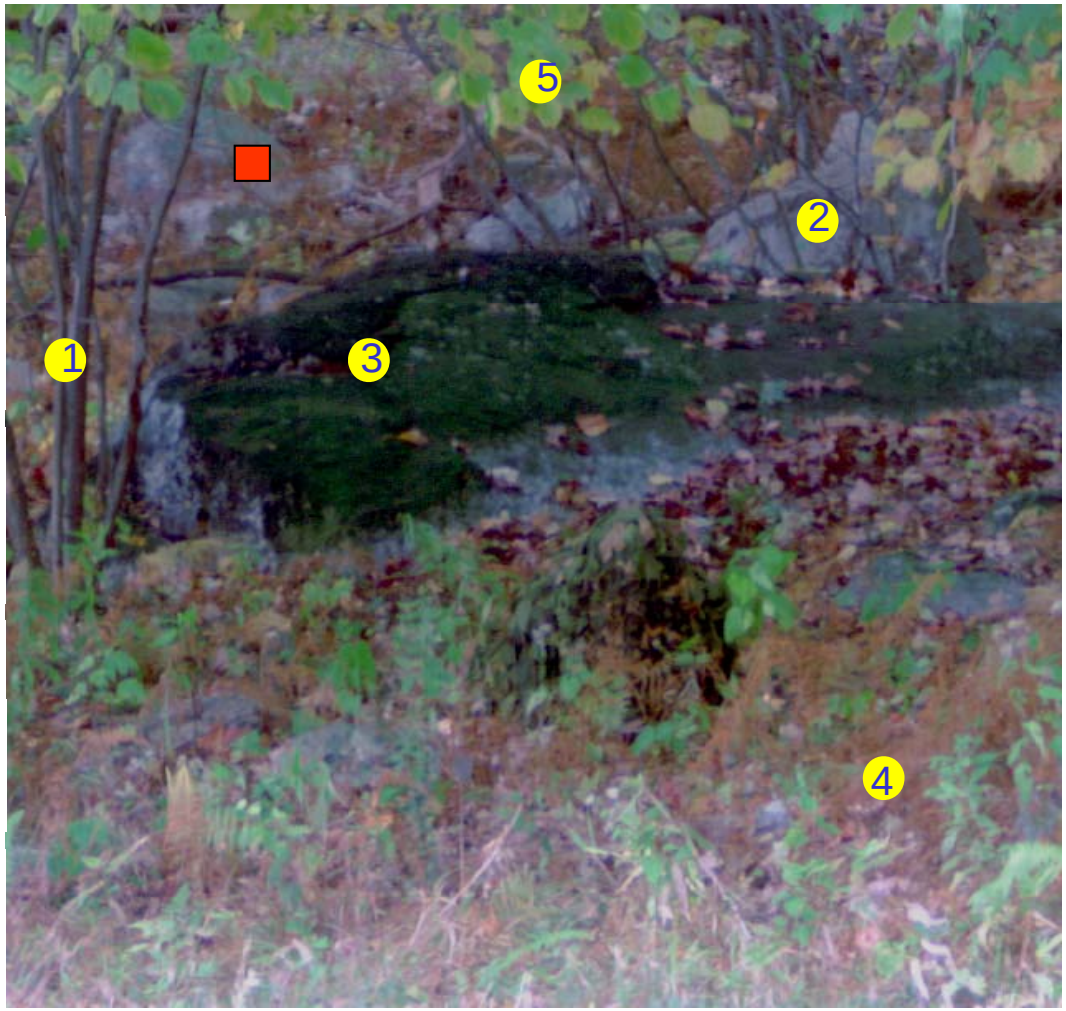
Scene Anomaly Detection @ Ground Level

***AsemiP* Algorithm**
Multi-Sample Extension

***10x10* area**

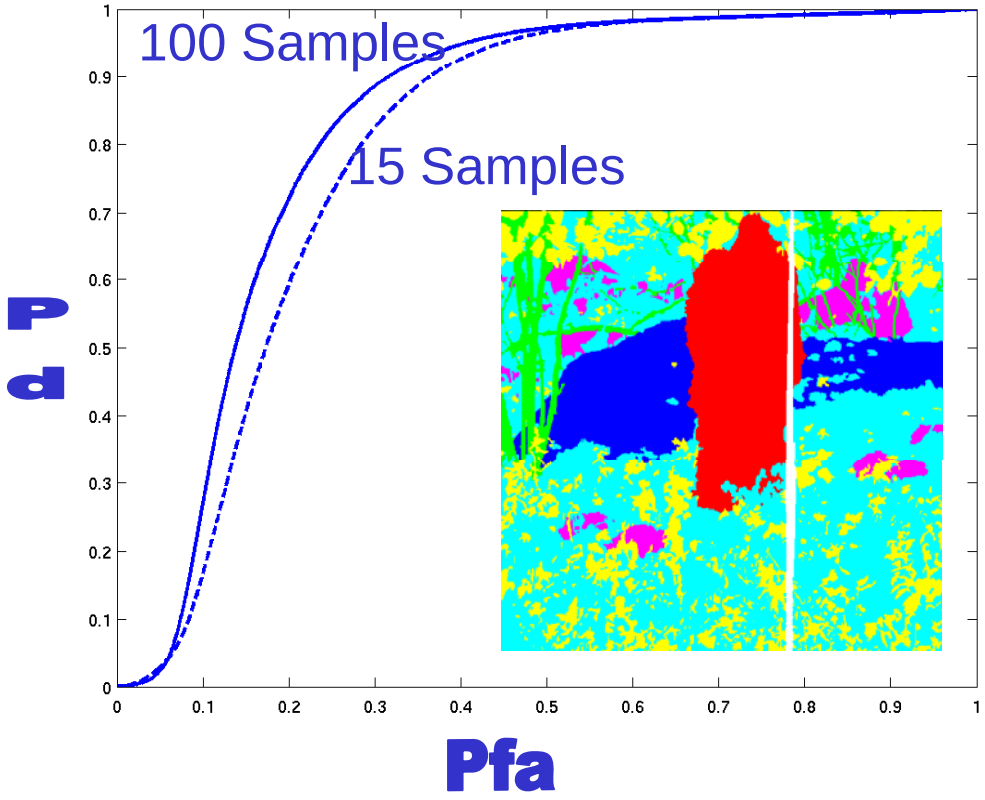


All Samples: 0.12% (500 / 373,321)

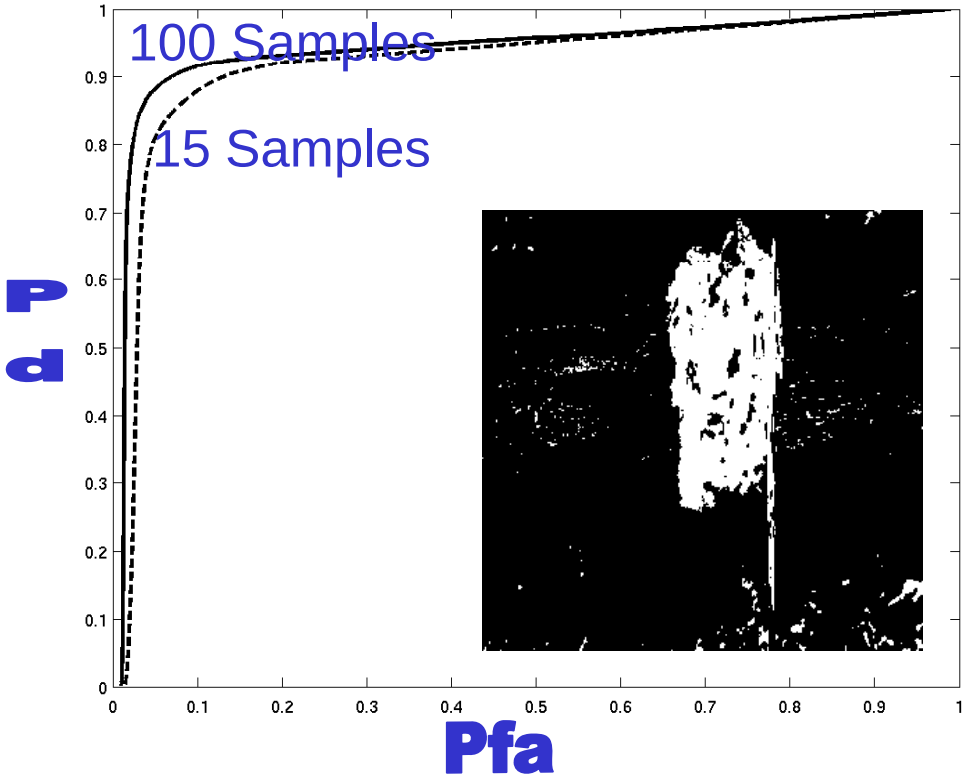


Six-Class Anomaly Detection

Supervised Learning Approach



Unsupervised Learning Approach



Artificial Neural Network

AsemiP Anomaly Detector

Final Remarks

- Statistical-Motivated Idea
- New Family of Anomaly Detectors
- Many Applications

Follow Up

- Auto Sampling
- Unsupervised Learning
  - Target Detection/Classification

Clutter Complexity Analysis of Hyperspectral Images

Clark Atlanta University

Clutter Complexity for Hyperspectral Imagery

Research Summary

Personnel

- Dr. Lance Kaplan
  - Moved to ARL halfway through program
- Dr. Péter Molnár
  - Took over for Dr. Kaplan
- Oladipo Fadiran

Research Focus

- Develop computationally simple clutter complexity measures.
- Extend clutter complexity measures developed for FLIR imagery to hyperspectral scenes.
- Clutter complexity measures will provide:
  - a priori information regarding the difficulty to detect a target in a scene.
  - fairer ATR comparisons over disparate databases.
  - validation of synthetic scenes.

Clutter Complexity Measure (CCM)

- **Goal:** to derive an objective measure of clutter complexity in hyperspectral images
  - as indication of the inherent difficulty of target recognition by ATRs,
  - for upper bound on the performance of ATR algorithms.
- **Approach:** find the aggregation of image metrics and statistics that correlate best with ATR baseline performance.
- Properties of the CCM:
  - independent of any particular ATR,
  - obtained with much lower computational complexity compared to a typical ATR.
- **Possible applications:**
  - objective basis for comparing different ATRs,
  - quick pre-assessment of image quality.

Definition of CCM

- The Clutter Complexity ***Measure*** is a function of selected image ***metrics*** that correlate best with baseline ATR performance.
- Metrics are:
  - Descriptive of scene parametric variation and significant for ATR performance
  - Computing them only requires a priori information on the order of spatial extent of the target in the scene at the most
  - Algorithmically uncomplicated, and easy to implement
- They fall into four main categories:
  - Single band metrics
  - Metrics based on band information content
  - Metrics based on anomaly detectors
  - Multi-band metrics
- Total of 129 metrics has been used.

Project Results

Clutter Complexity Measure

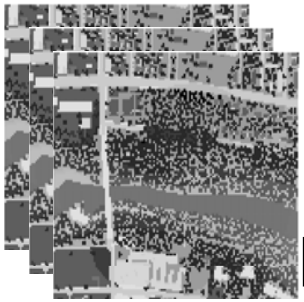
- Developed a method to compute a Clutter Complexity Measure (CCM) for HSI images that predicts ATR performance.
- The CCM is computed on a set of sample images that are representative to a particular scene or application.
- The resulting CCM estimates the expected ATR performance for similar images.

Adaptive Sampling by Histogram Equalization (ASHE)

- Developed a novel algorithm to efficiently produce synthesized HIS images for a range clutter complexity values.
- This algorithm can be used to adaptively sample multi-dimensional functions for which obtaining a sample point is (computationally) expensive.

Development and Uses of a Clutter Complexity Measure (CCM)

Training Set



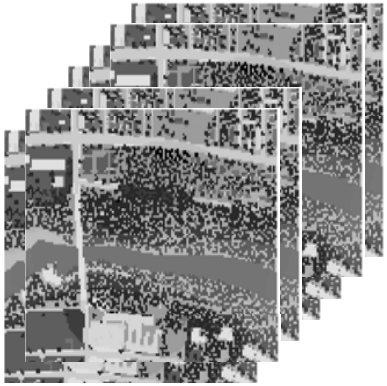
Complete Database

129 Statistical Image Clutter Features

Clutter Features Space Reduction through Novel Factor Analysis Method

Established Baseline ATR Performance

Derived CCM as Combination of Clutter Features that Correlates best with ATR Baseline Performance



Selection from Database

Clutter Complexity Measure

Predicts Baseline ATR Performance independent of any particular ATR

Measure of Image Quality

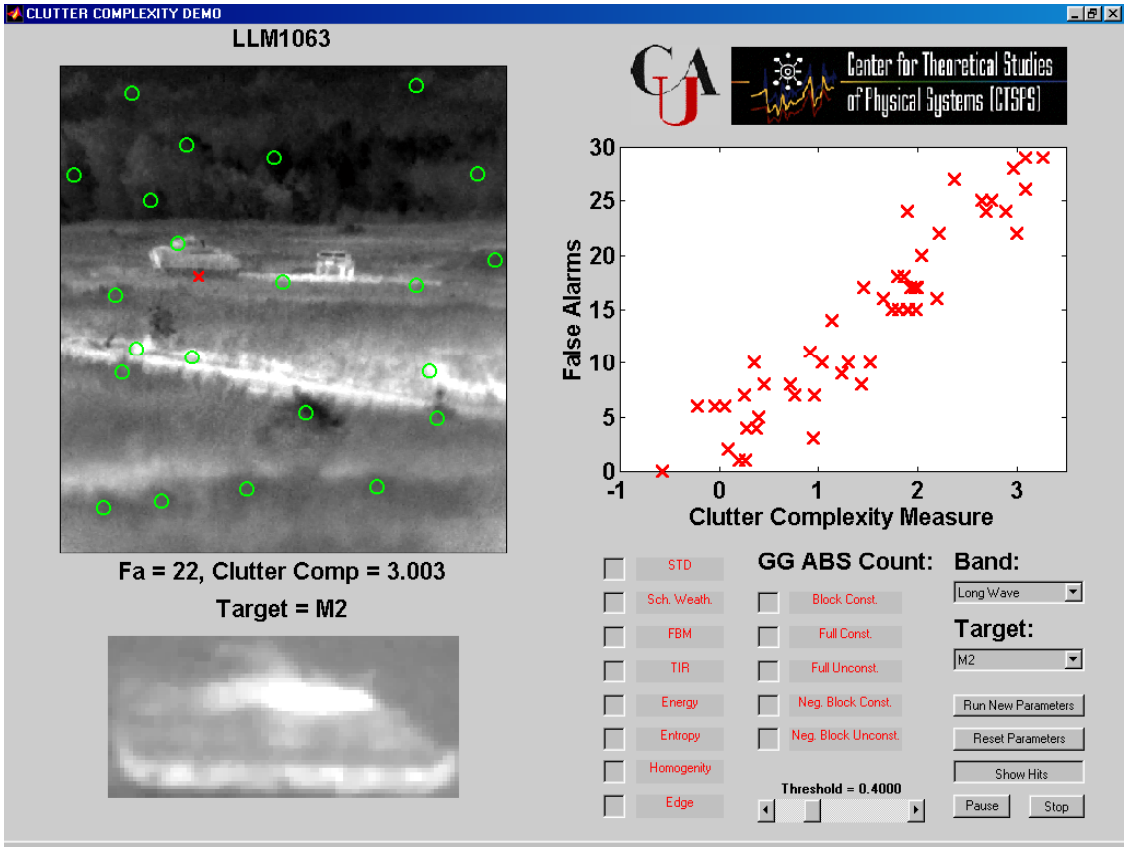
Clutter Complexity for Hyperspectral Imagery

2003 Research Efforts

Research Focus

- Develop computationally simple clutter complexity measures.
- Extend clutter complexity measures developed for FLIR imagery to hyperspectral scenes.
- Clutter complexity measures will provide:
  - a priori information regarding the difficulty to detect a target in a scene.
  - fairer ATR comparisons over disparate databases.
  - validation of synthetic scenes.

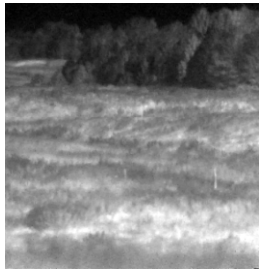
Clutter Complexity for IR Imagery



CAU Clutter Complexity Tool

Low

Medium

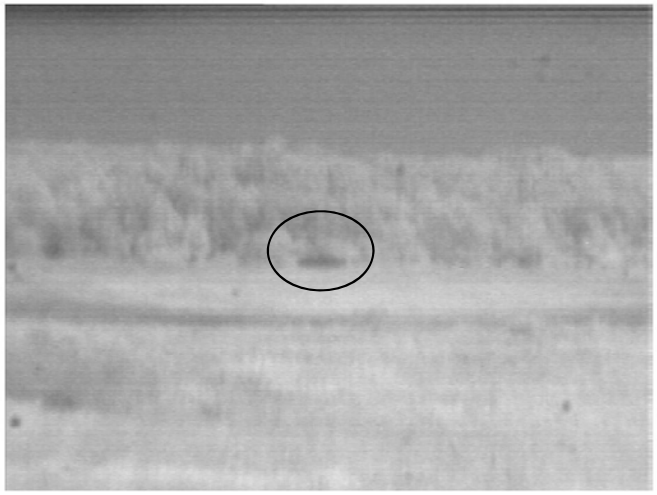


High

The Approach

- Use a database of hyperspectral scenes.
- Divide the database into smaller partitions where target characteristics are held constant.
- Evaluate ATR performance on each partition.
- Develop features that may measure clutter.
- Use the weighted sum of features that maximized correlation with ATR bounds as the complexity metric.
- Compare metric against real ATR performance over each partition using a disparate dataset.

AOTF Imagery

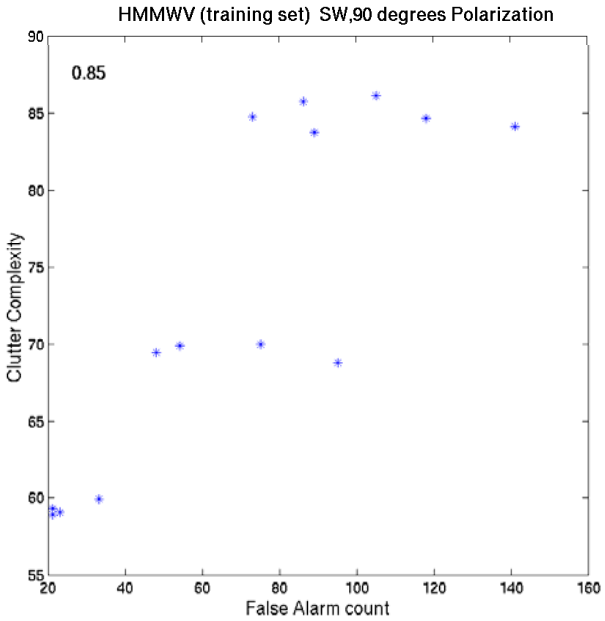


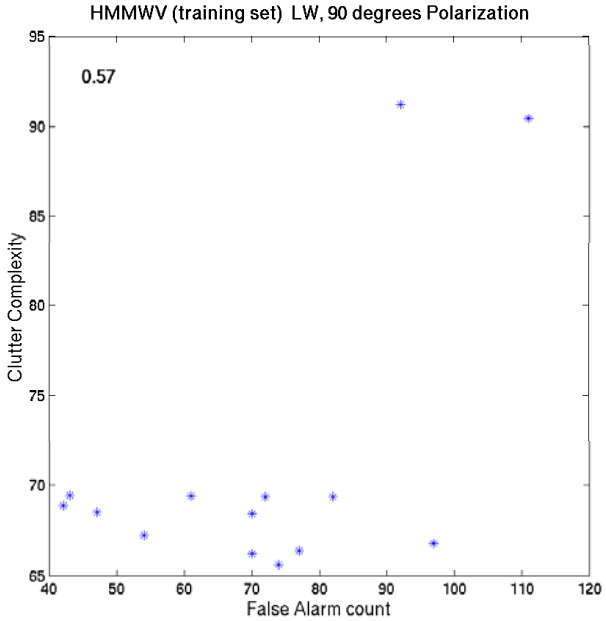
Personnel carrier
Distance = 2 km

HMMWV
Distance = 1.2 km

- Spectral Range: 460-1000nm in 20nm steps
- Polarizations: 0°, 45°, 90° and 135°
- 22 different scenes
- Limited background diversity

Clutter Complexity of Spectral Bands

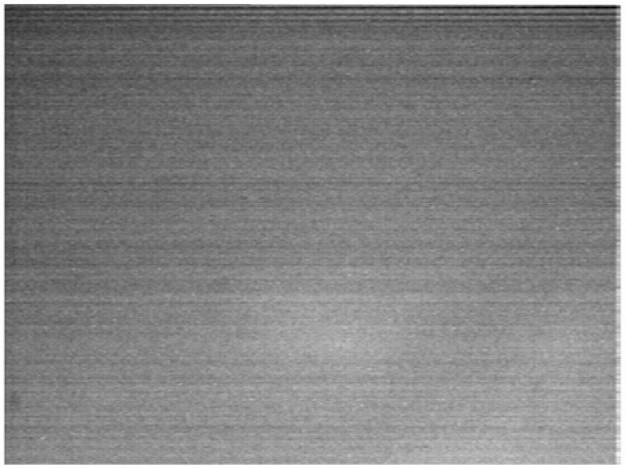




460-730nm ***730-1000nm***

HMMWV

Clutter Complexity Example

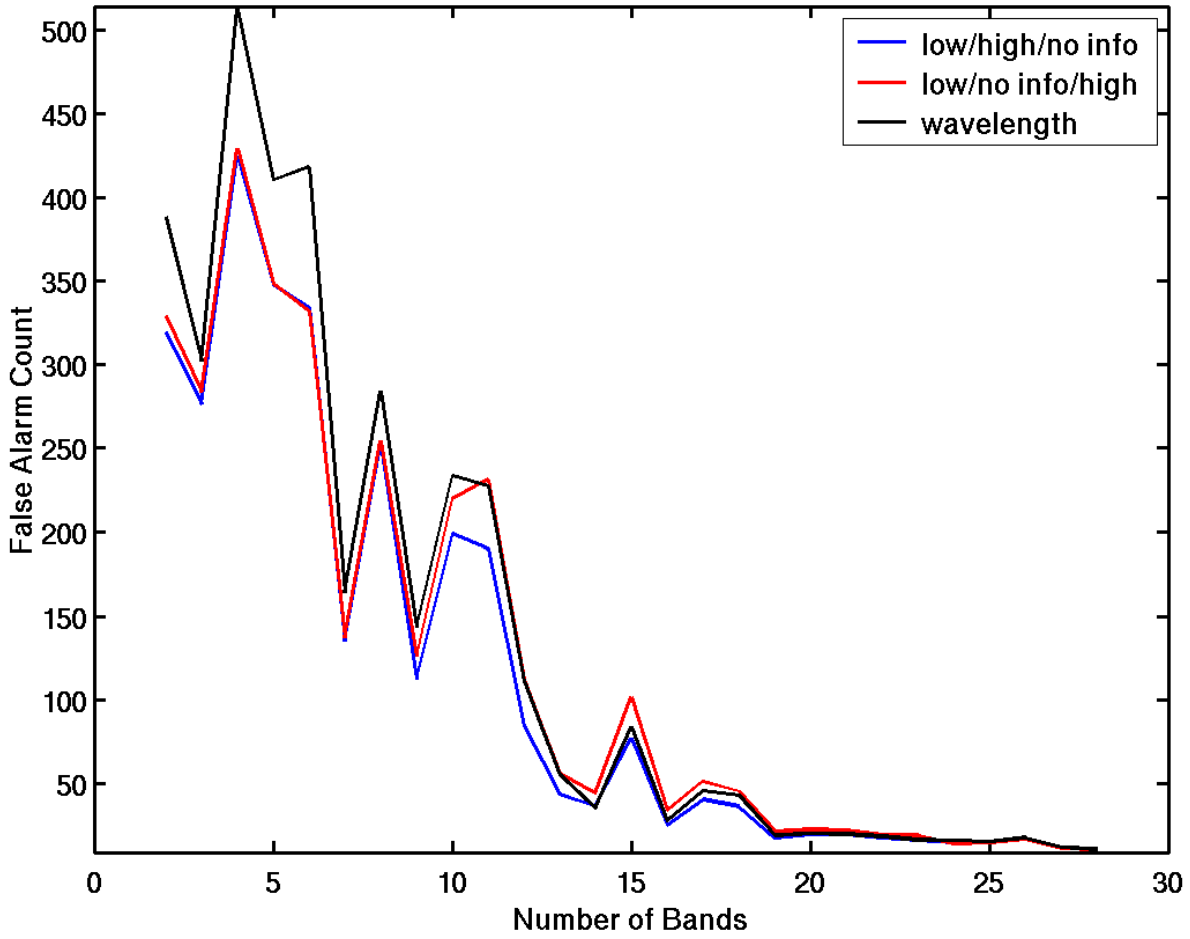


Low complexity image
Clutter complexity = 38.84

Medium complexity image
Clutter complexity = 73.56

Highly complex image
Clutter complexity = 112.91

Band Selection Results

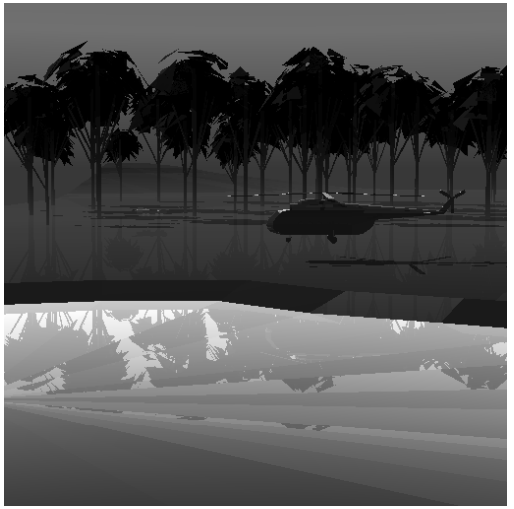


Data Generation

- Use DIRSIG to generate hyperspectral data of various “clutter complexity” for a multitude of targets.
- Progress to date:
  - Successful installation of latest release of DIRSIG and MODTRAN .
  - Can generate scenes with different target types.
  - Working on the addition of clutter objects to the scene, e.g. trees, rocks, etc.

Example Synthetic Data

380nm



640nm

960nm

1560nm

Future Work

- Further study of clutter complexity for different spectral bands.
- Hyperspectral clutter complexity:
  - 3-D Features
  - Matched-filter ATR
  - Aggregation of 3-D features into a clutter complexity measure
  - Requires lots of data!

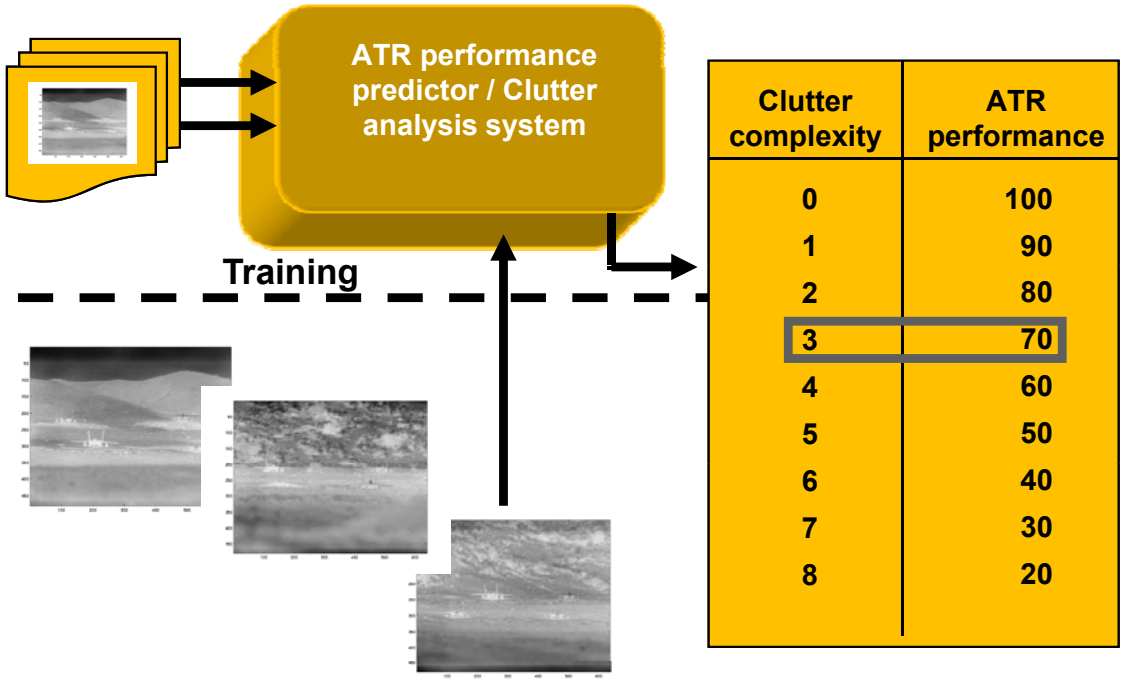
Clutter Complexity for Hyperspectral Imagery

2004 Research Efforts

Research Objective

- **Goal:** To develop a computationally efficient measure of clutter complexity for hyperspectral imagery.

Impact of Research

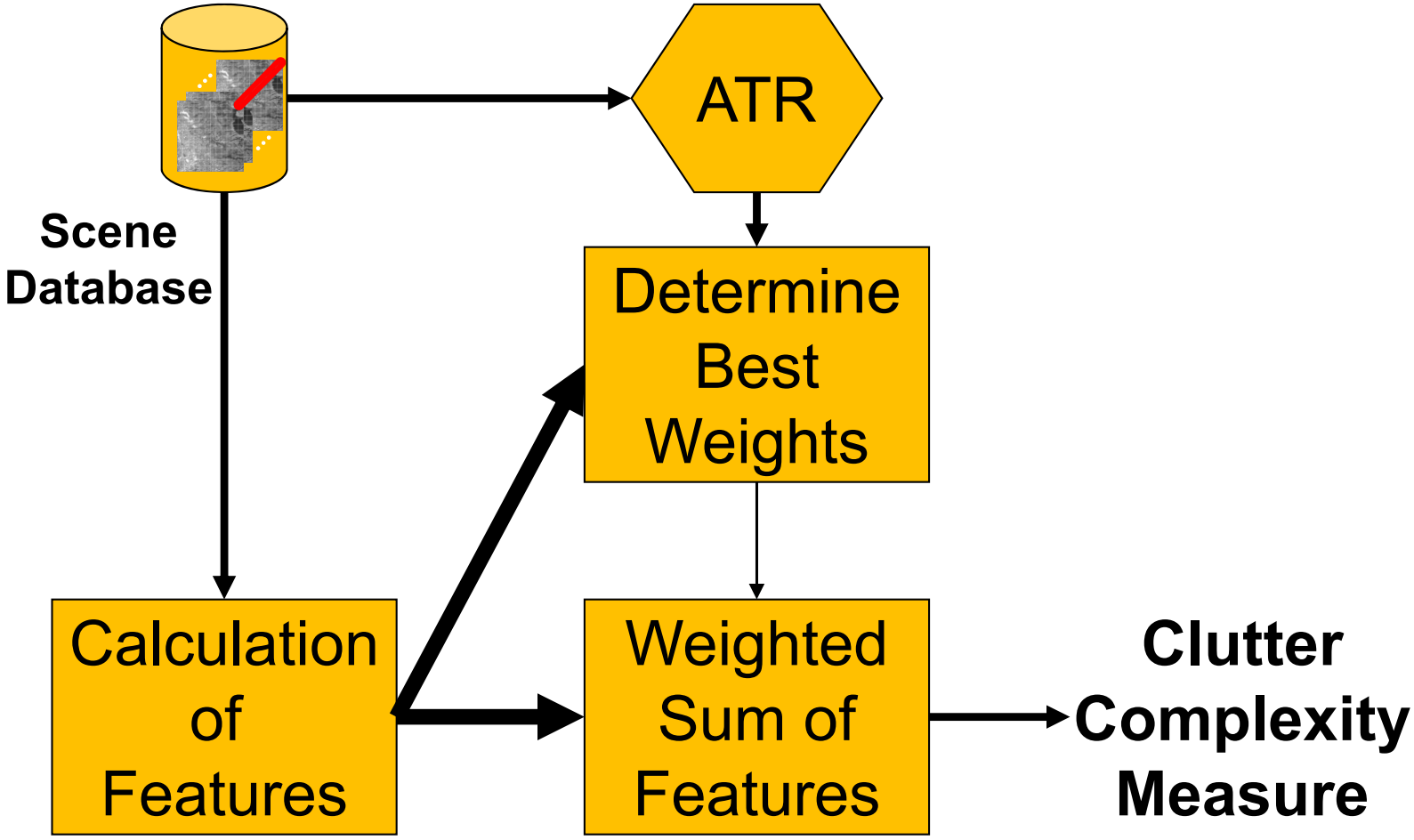


- Information to manage ATR resources
- Evaluation of “complexity” of databases
- Synthetic scene validation

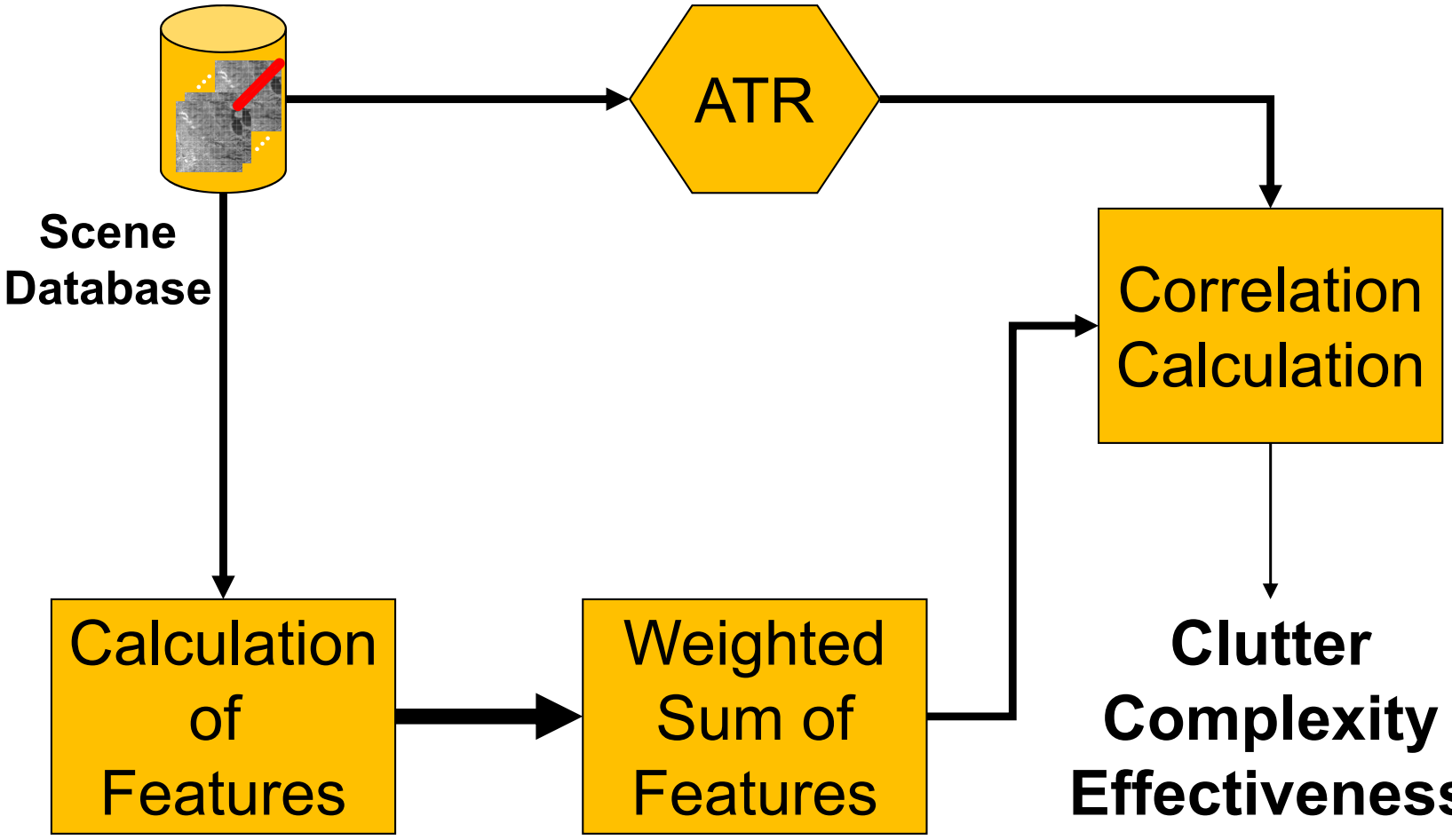
Collaborations

- **University of Maryland:** Anomaly detectors as scene features
- **Rochester Institute of Technology:** Synthetic scene generation
- **Georgia Tech Research Institute:** Evaluation of clutter complexity measures over real imagery

Training


ATR
Scene Database
Determine Best Weights
Calculation of Features
Weighted Sum of Features
Clutter Complexity Measure

Testing


Scene Database
ATR
Correlation Calculation
Calculation of Features
Weighted Sum of Features
Clutter Complexity Effectiveness

AOTF Imagery

Personnel carrier
Distance = 2 km

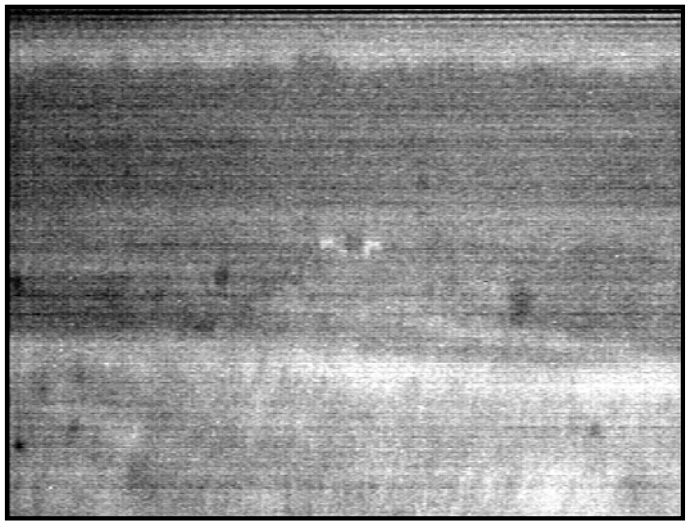
HMMWV
Distance = 1.2 km

- Spectral Range: 460-1000nm in 20nm steps
- Polarizations: 0°, 45°, 90° and 135°
- 22 different scenes
- Limited background diversity

Image Features

- **Standard Deviation:** Global standard deviation
- **Schmieder Weathersby:** Average standard deviation
- **FBM Hurst Parameter:** Quantification of texture roughness
- **Target Interference Ratio:** Average contrast
- **Energy:** Average histogram energy
- **Entropy:** Average histogram entropy
- **Homogeneity:** Average pixel variations
- **Outlier Ratio:** Average percentage of outlier pixels

Band Complexity

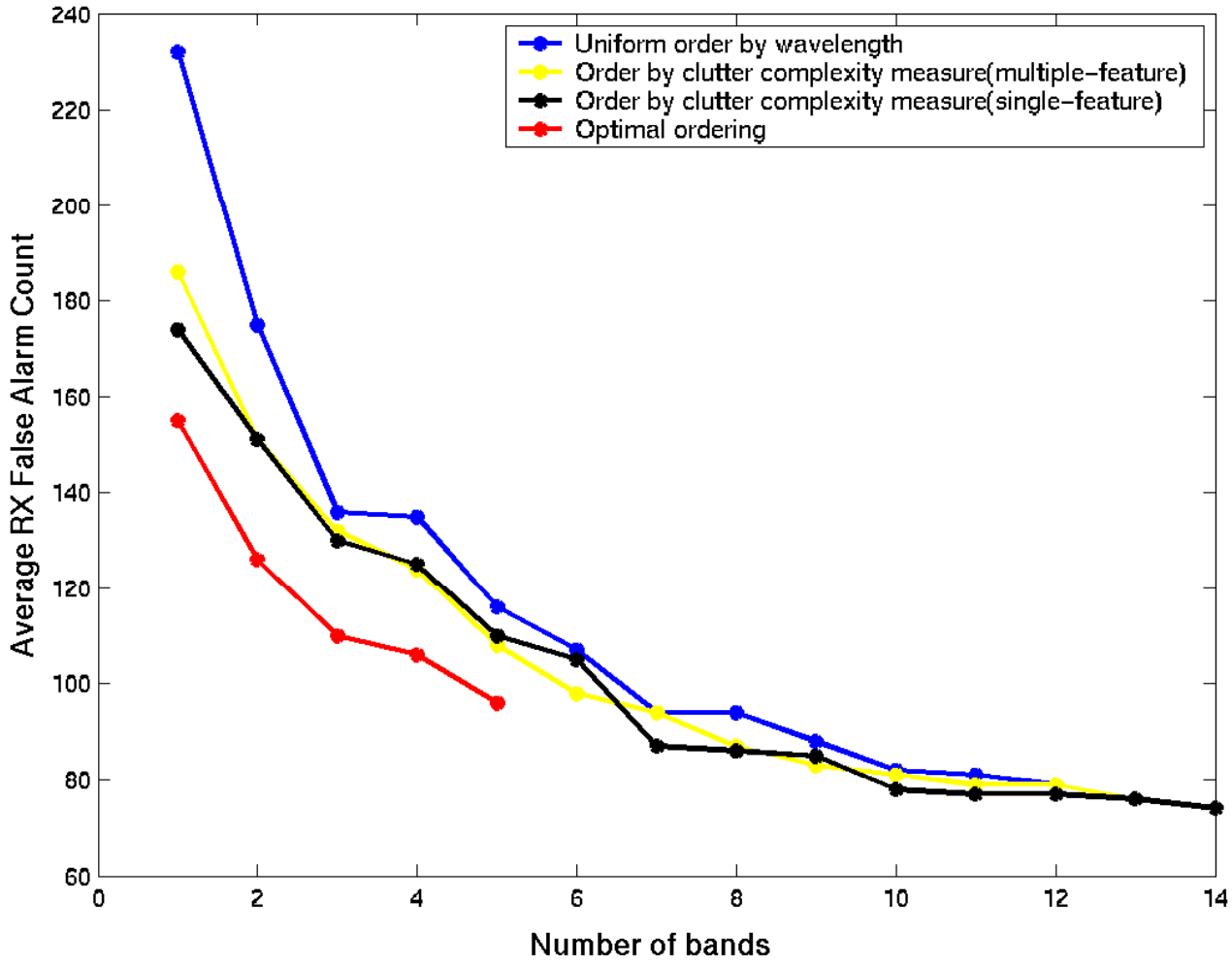


Low complexity band
Clutter complexity = 38.84

Medium complexity band
Clutter complexity = 73.56

Highly complex band
Clutter complexity = 112.91

Band Selection

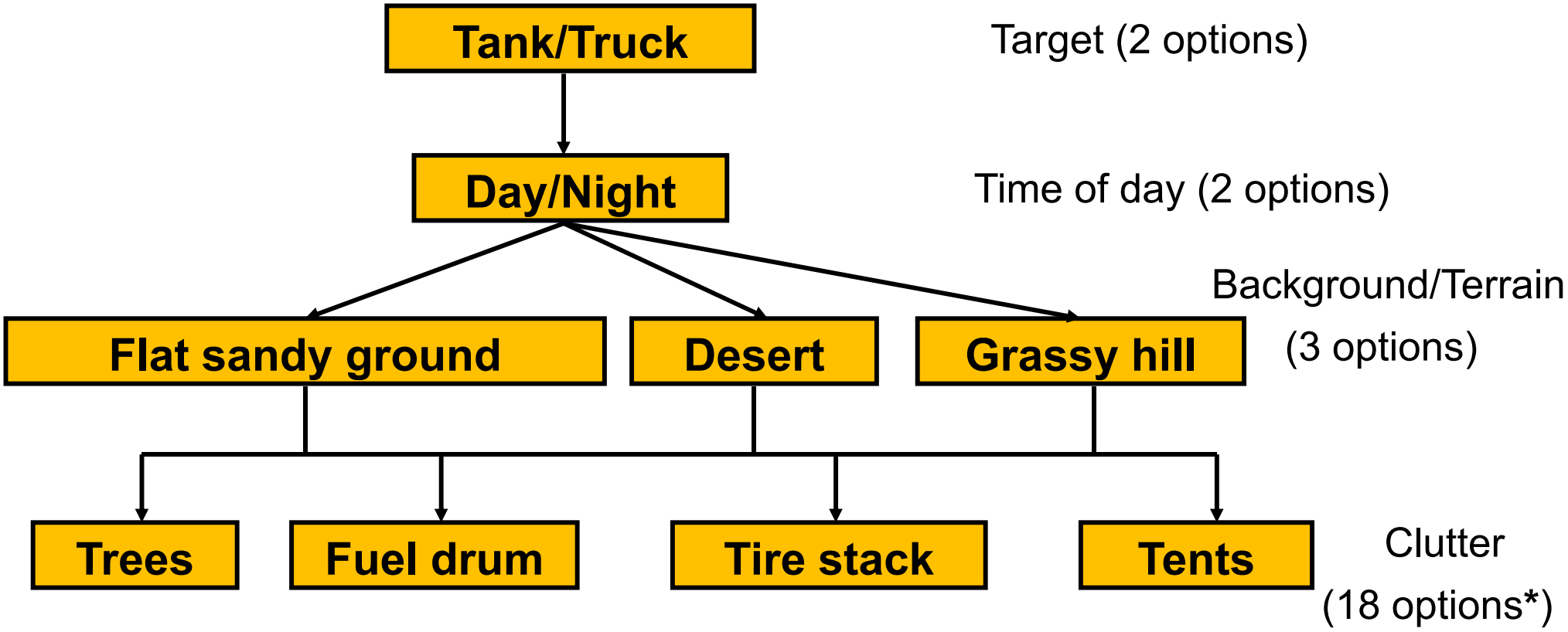


Synthetic Scene Generation

- Sensor: Framing array, single shot image acquisition, 50mm focal length.
- Geometry: stand-off of 2km in forward-looking arrangement
- Size and Resolution: 512×512 pixels, 1.93m.
- Wavelength: 8-13 microns with 40 nanometer (nm) steps resulting in 126 bands per hyperspectral image cube

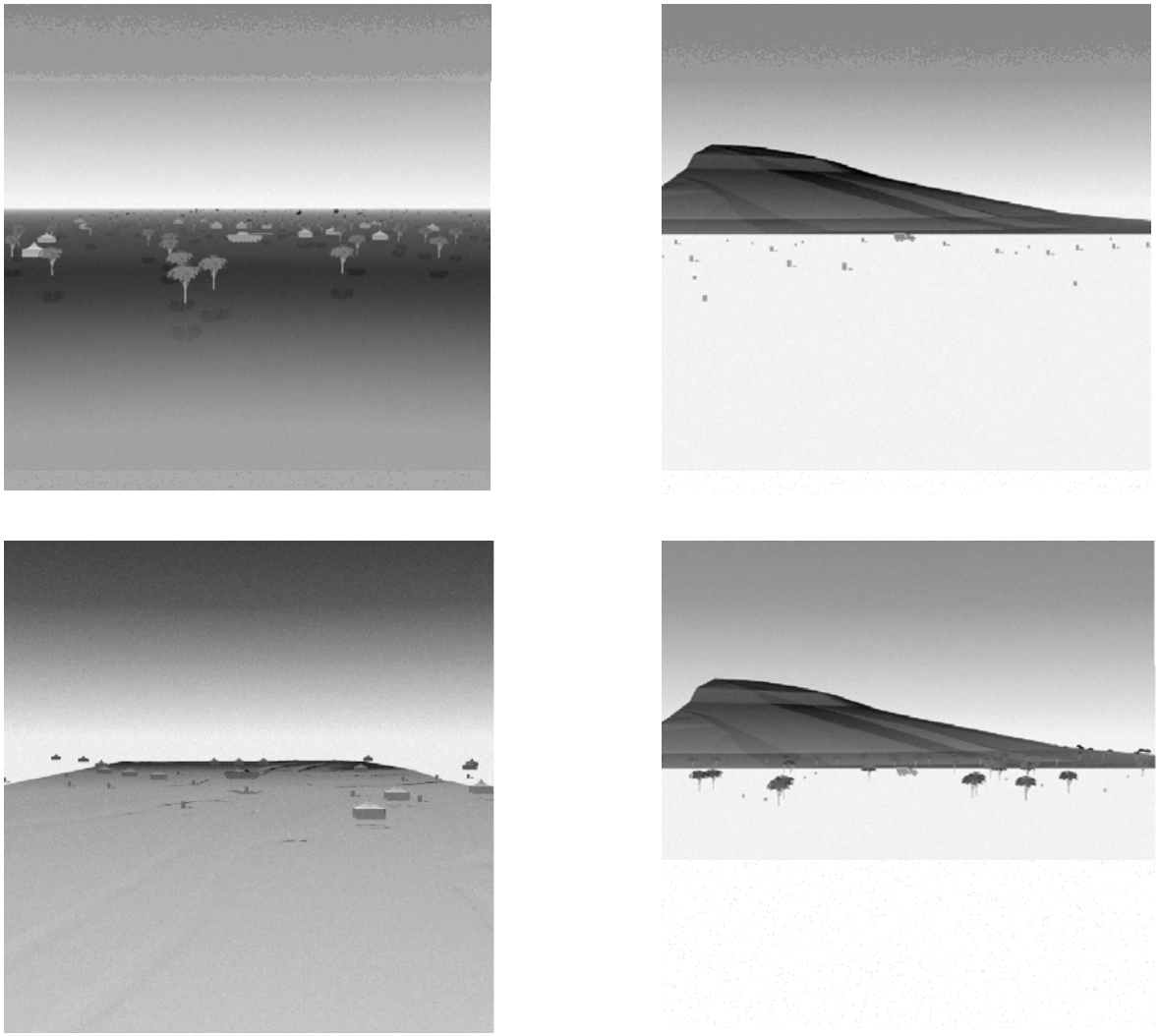
Synthetic Database

216 Hyperspectral Image Cubes



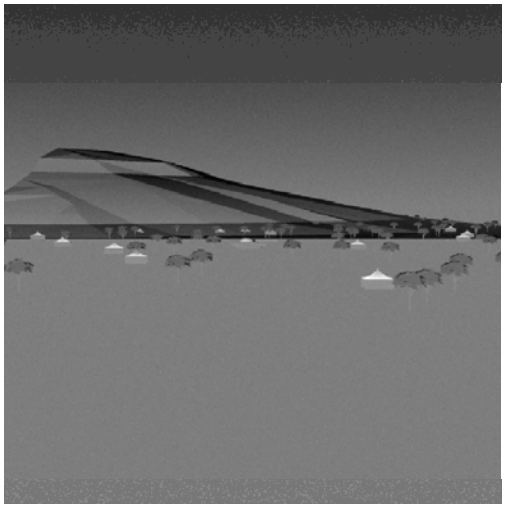
* Of the 4 types of cluttering objects to be used, all possible combinations of 2 are used resulting in 6 sets. Each of these sets also have 3 levels determined by the number of objects resulting in 18 scenarios.

Examples of Synthetic Scenes

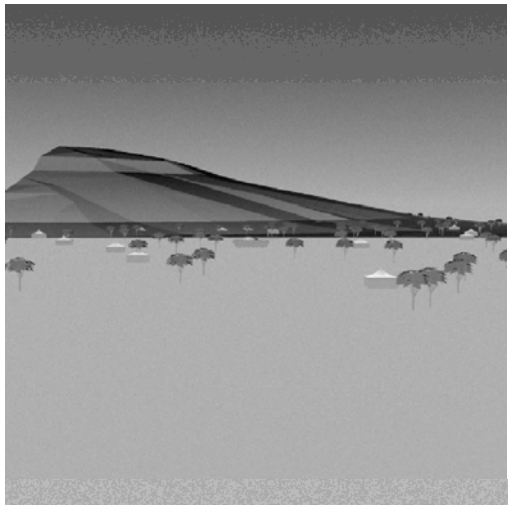


Single bands from hyperspectral cubes, $\lambda = 8.40$ microns for all images

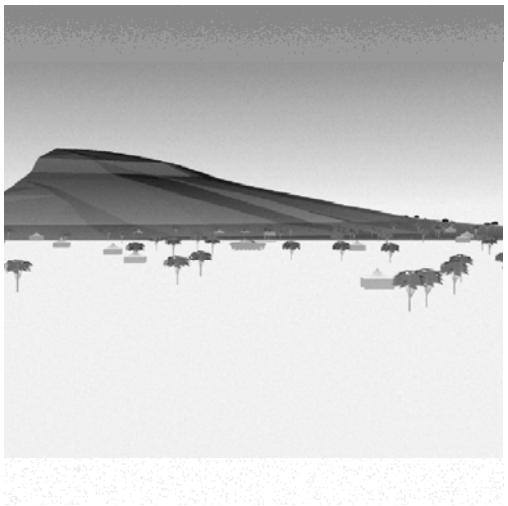
Example of Synthetic Bands



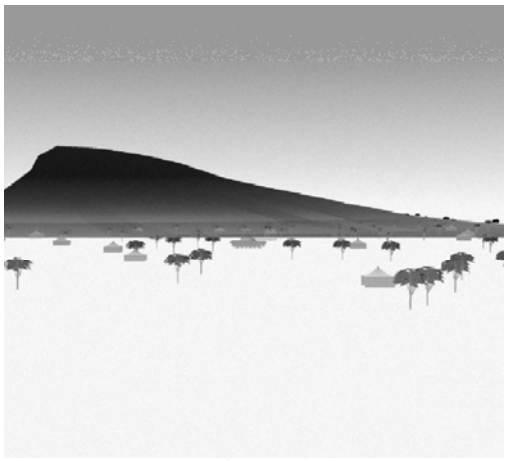
λ = 8.40 microns



λ = 10.00 microns



λ = 11.60 microns



λ = 12.80 microns

Hyperspectral Target Detection

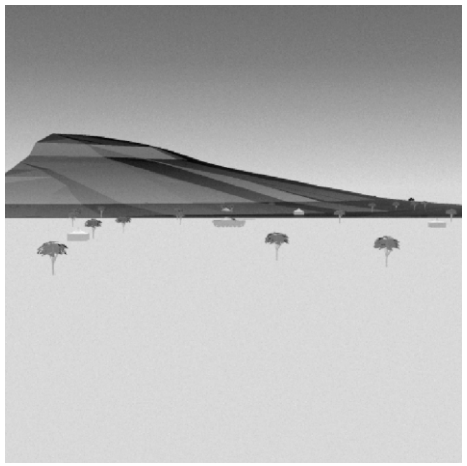
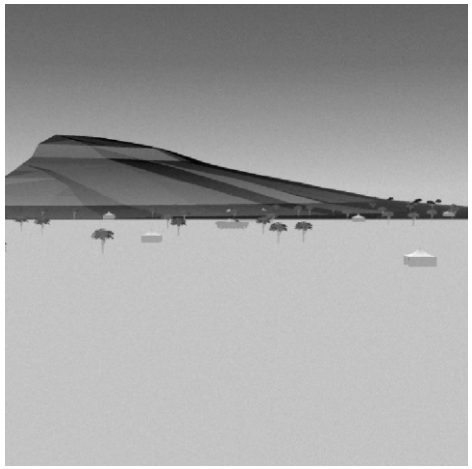
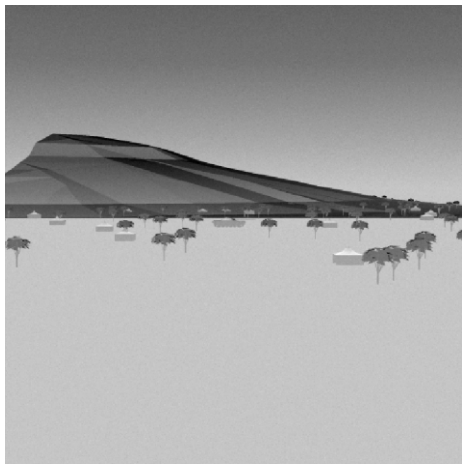
Single pixel ATR: Matched filtering via the adaptive coherence estimator (ACE):

$$\frac{\left| s^T \hat{R}_b^{-1} x \right|^2}{(s^T \hat{R}_b^{-1} s)(x^T \hat{R}_b^{-1} x)},$$

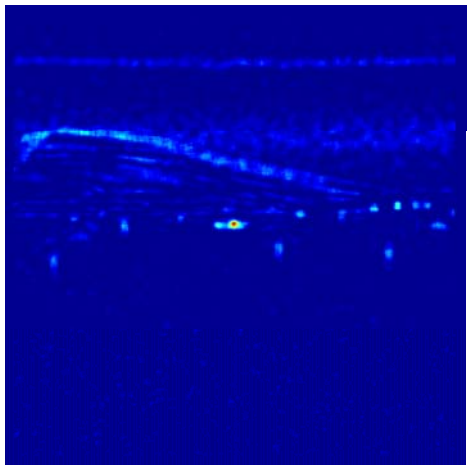
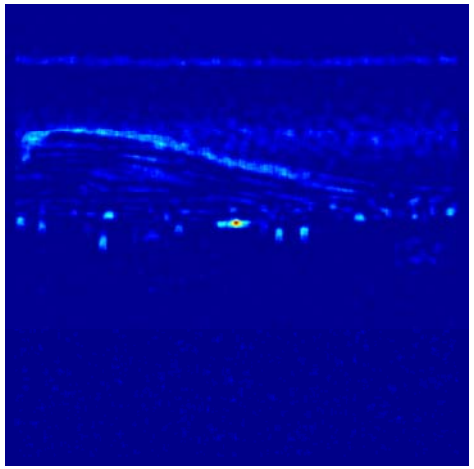
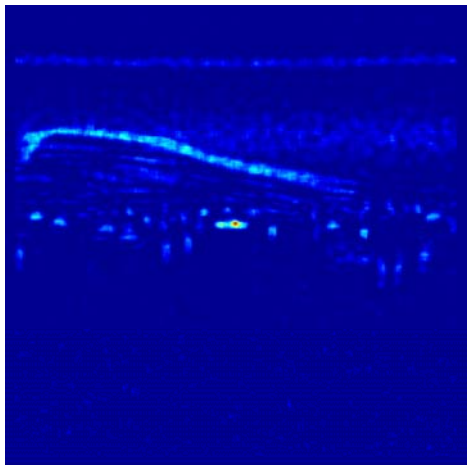
where s is the target template and x is the vector under test.

Target Detection Example

Scenes

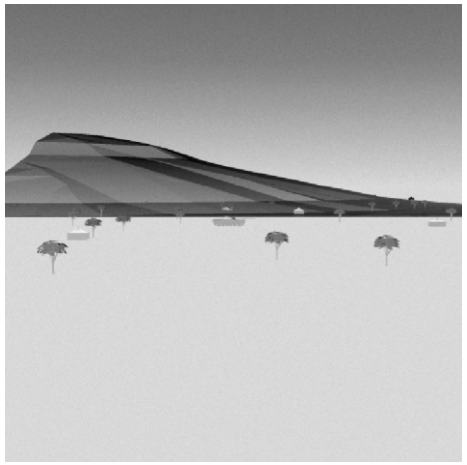
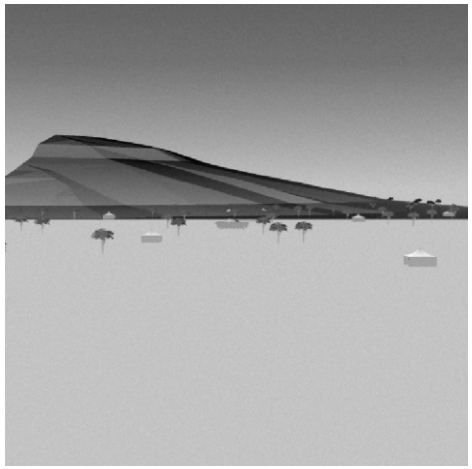
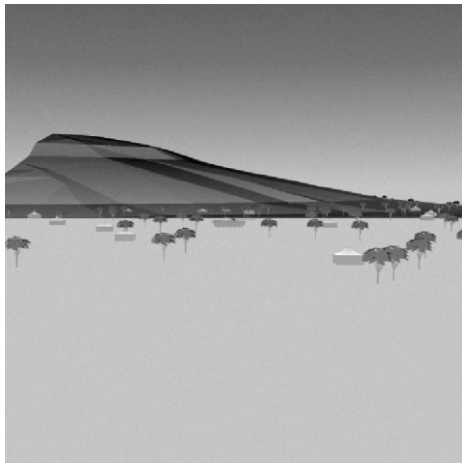




ACE Statistic

Target Detection Example (cont.)

Scenes

Detections

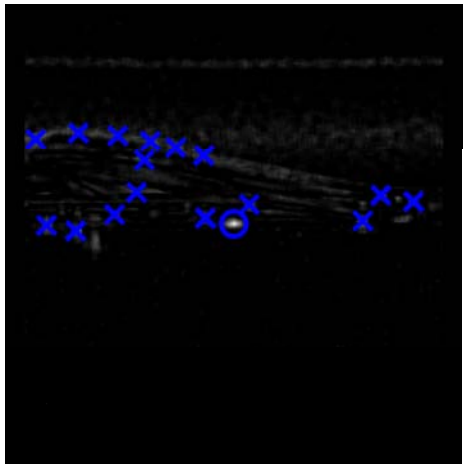
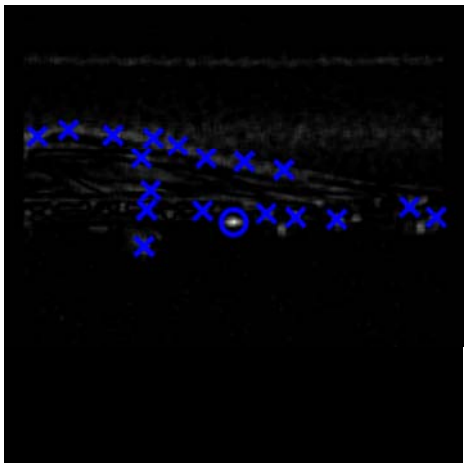
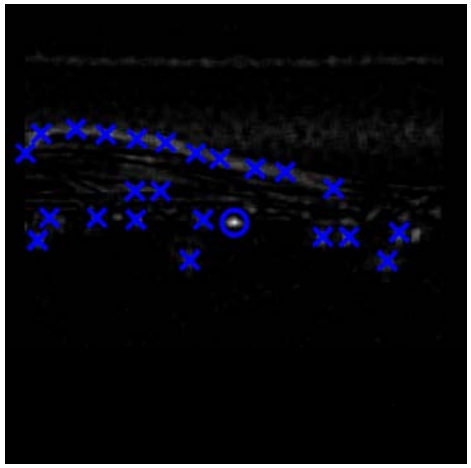




Image Features –Anomaly Detectors

- **Dot Product –**

$$(X / \|X\|) \bullet (Y / \|Y\|)$$

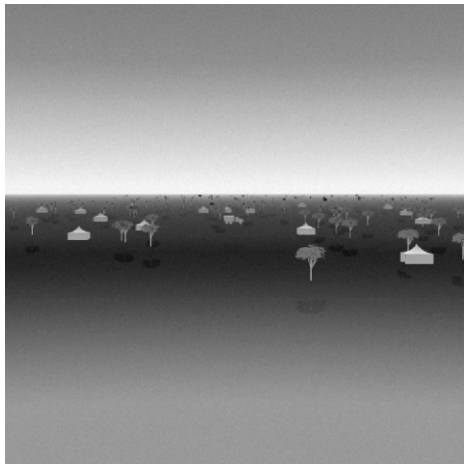
- **Relative Entropy or Kullback-Leibler Distance –**

$$\sum p(x) \times \log \frac{p(x)}{q(x)}$$

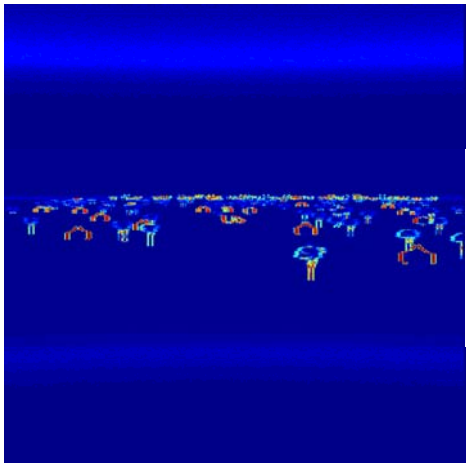
- **RX Algorithm –**

$$S = (x-u)' R^{-1} (x-u)$$

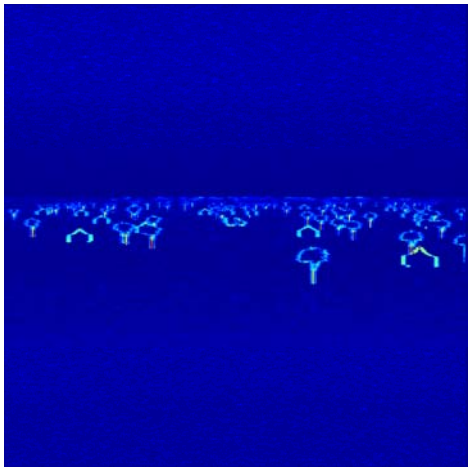
Anomaly Detection Examples



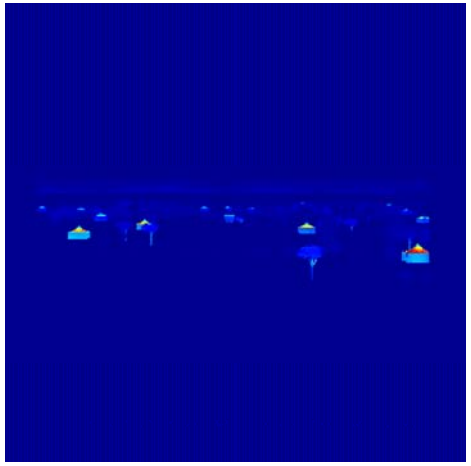
Scene



Dot Product



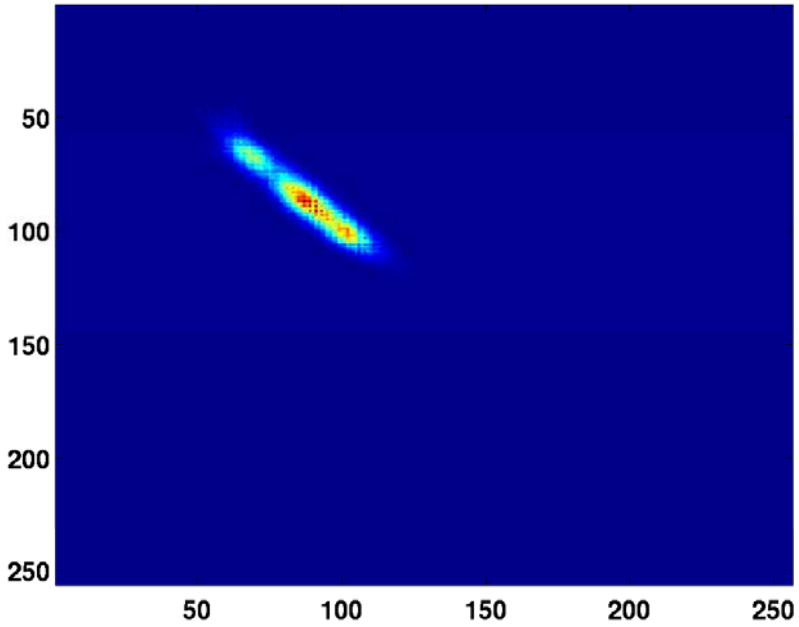
Relative Entropy



RX

Image Features – Gray-Level Co-occurrence Matrices (GLCM)

Single band of the scene image

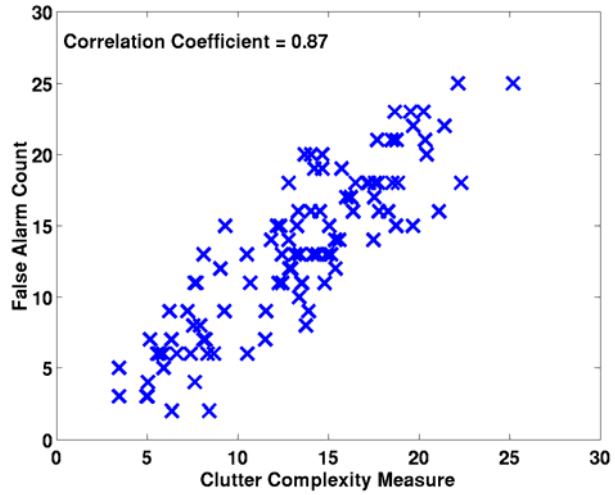
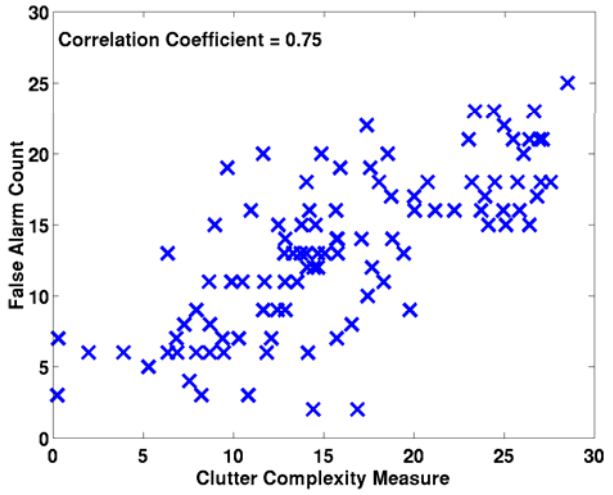
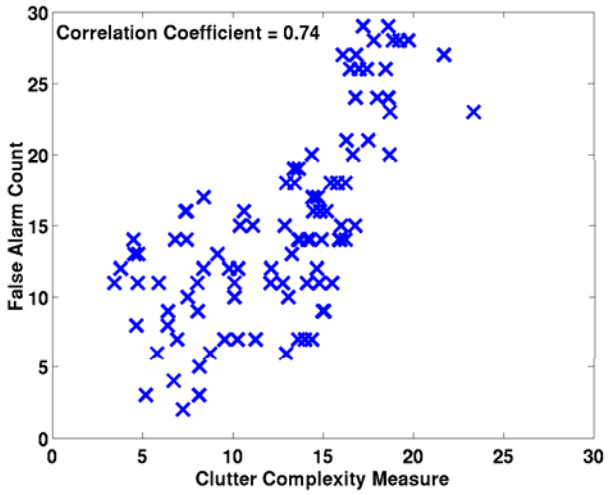
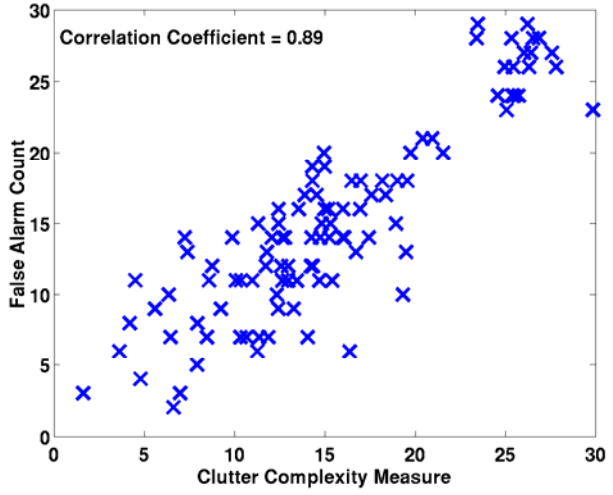


GLCM

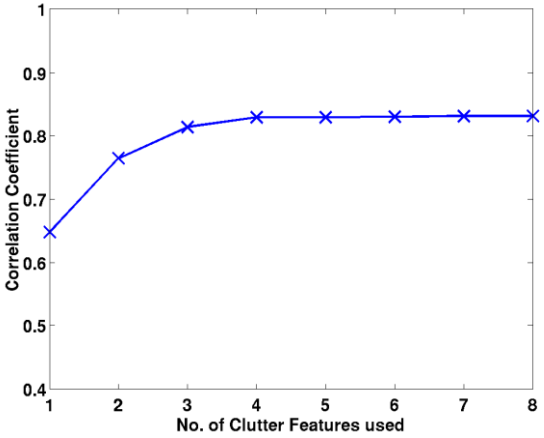
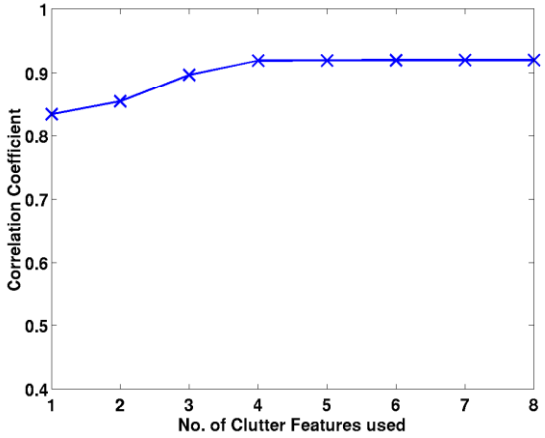
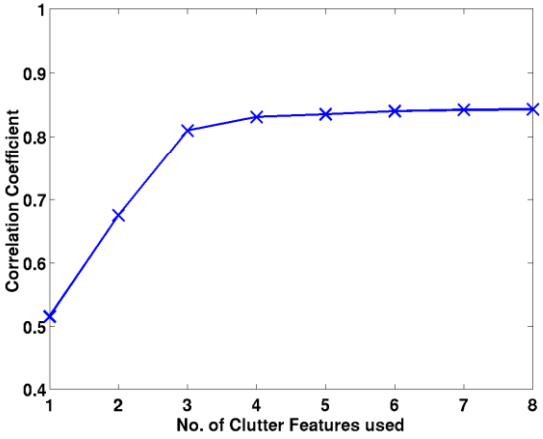
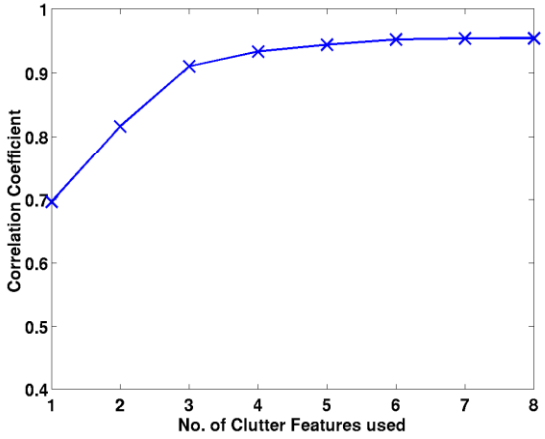
Image Features Derived from the GLCM

- Maximum value in GLCM – $\max_{i,j}(C_{ij})$
- Entropy – $\sum_i \sum_j C_{ij} \log(C_{ij})$
- Energy – $\sum_i \sum_j C_{ij}^2$
- Contrast – $\sum_i \sum_{j \neq i} C_{ij} |i-j|^2$
- Homogeneity – $\sum_i \sum_{j \neq i} C_{ij} / |i-j|$

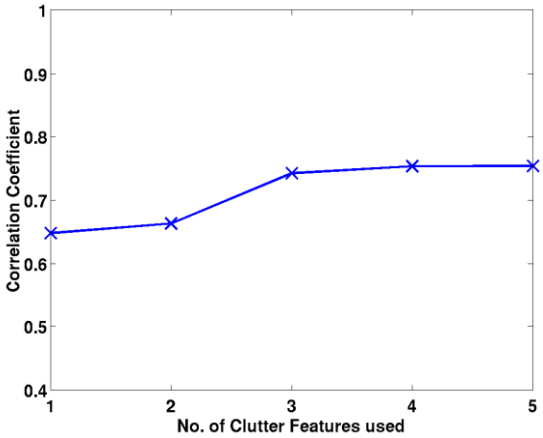
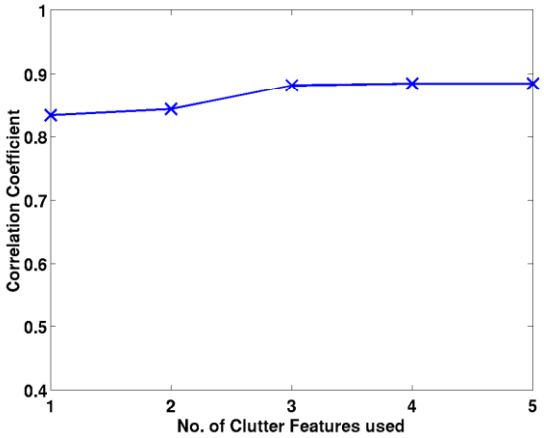
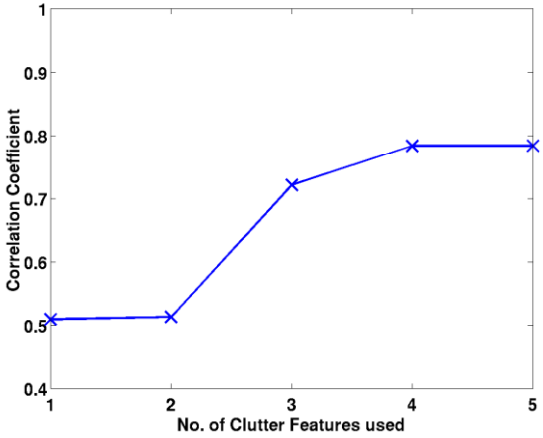
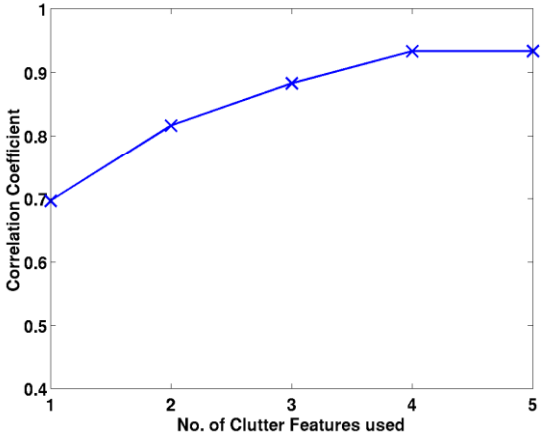
CCM vs. ATR Performance

Test \ Train	Tank	Truck
Tank		
Truck		

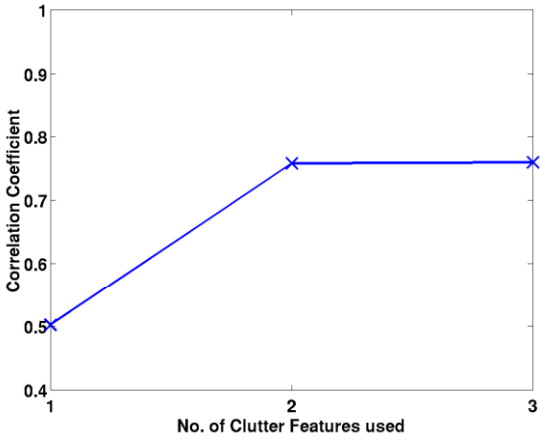
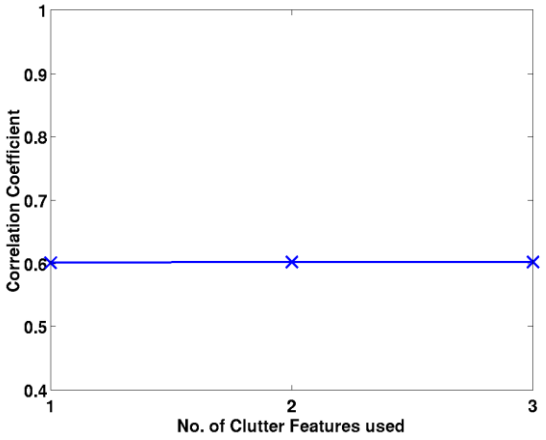
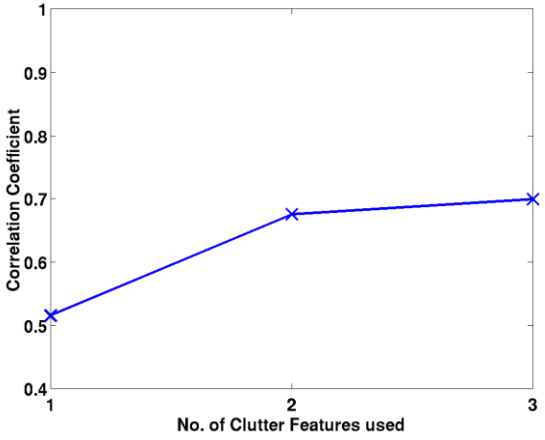
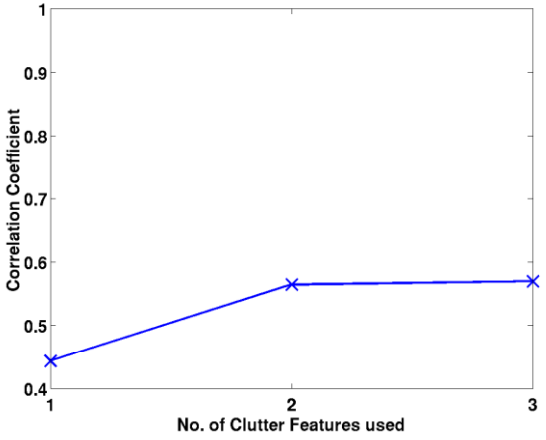
Number of Features for the CCM

Target \ Time	Day	Night
Tank		
Truck		

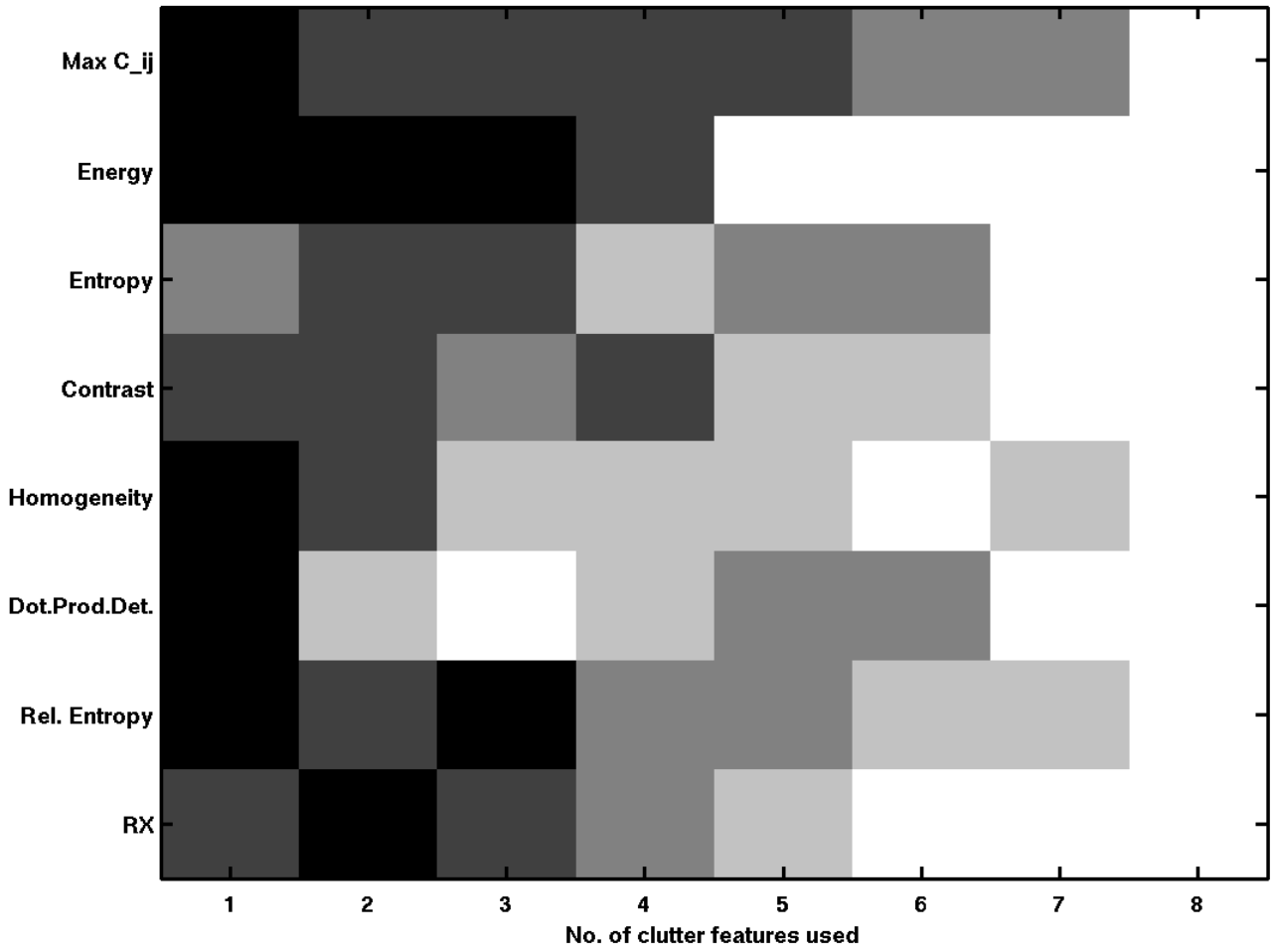
Number of Texture Features for the CCM

Target \ Time	Day	Night
Tank		
Truck		

Number of Anomaly Features for the CCM

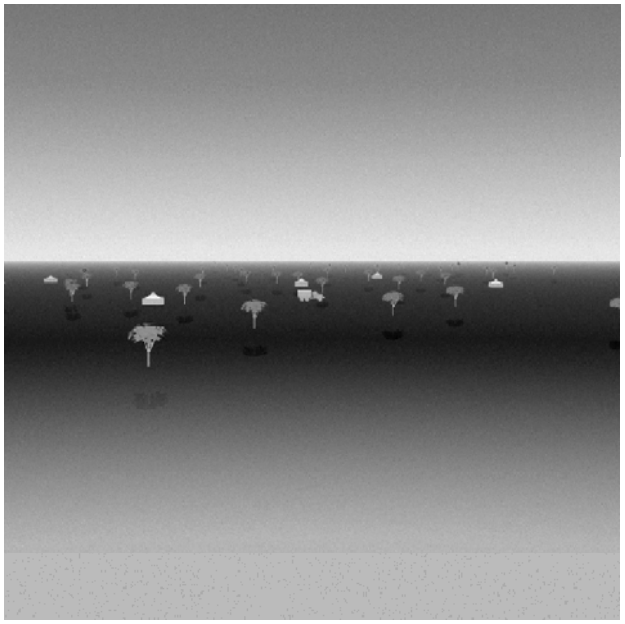
Target \ Time	Day	Night
Tank		
Truck		

Optimal Combination of Clutter Features

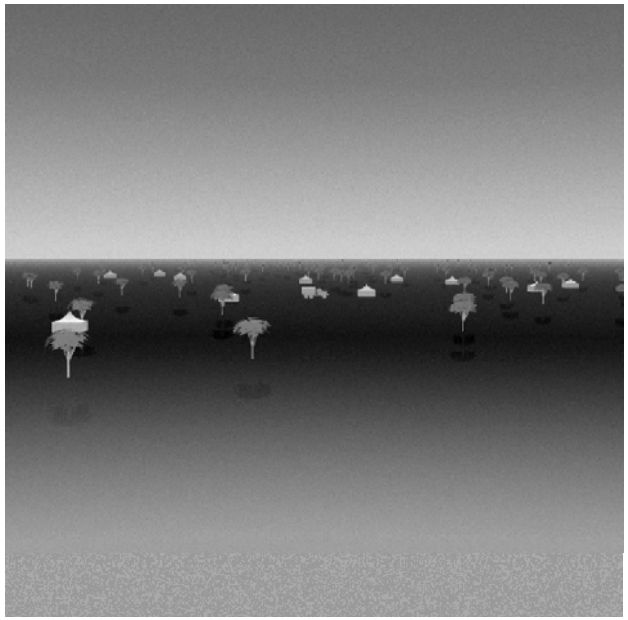


Analysis of clutter features that resulted in optimal Correlation Coefficient for different No.s of features

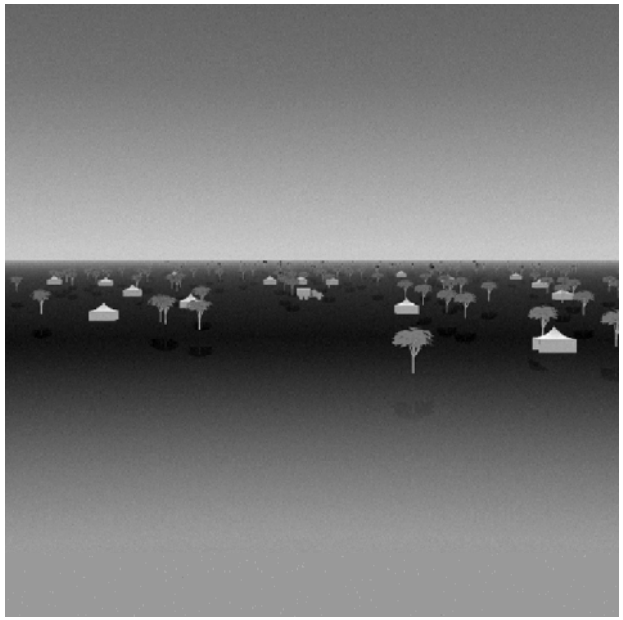
Image Classification by Derived Clutter Complexity CCM



Low Complexity, CCM = 7.67



Medium Complexity, CCM = 14.39



High Complexity, CCM = 17.11

2004 Research Summary

- Clutter complexity of hyperspectral bands
  - Training and testing of the clutter complexity measure
  - Analysis of band selection for target detection
- Synthetic scene generation
- Clutter complexity of hyperspectral scenes
  - Developed 3-D features
  - Demonstrated correlation between the clutter complexity measure and ATR performance

2004 Research Accomplishments

- Improved synthetic scene generation
  - Use a larger emissivity library for each material
  - Randomize the shape of clutter objects
  - Develop scenes representative of the MURI problem scenarios
- Develop more 3D features
- Evaluate CCM over synthetic landmine imagery provided by RIT
- Evaluation of CCM over real imagery (GTRI)
- Rank features in terms of their contribution of a CCM

Future Work

- Clutter complexity measure for alternative targets
  - Obscured ground vehicles
  - Chemical/biological agents
- Automated scene generation
  - Effect of operating conditions on hyperspectral imagery
  - Sampling of the operating conditions
- Feature pruning
  - Factor analysis
  - PCA
  - ICA

Clutter Complexity for Hyperspectral Imagery

2005 Research Efforts

Research Progress

1. **Investigated** and tested additional hyperspectral image clutter metrics
2. **Factor analysis** to identify significant components in clutter metric space and reduce redundancy
3. **Experimented** to determine effectiveness of subset of clutter metrics to predict clutter complexity in hyperspectral images
4. **Classified** images using derived clutter complexity measure
5. On-going work on methods to automate generation of hyperspectral image database with maximal diversity in clutter levels based on identified clutter metrics

(1) Hyperspectral Image Clutter Metrics – Selection Criteria and Categories

- Candidate metrics have to satisfy the following criteria:
  - "Simplicity" of computation compared to ATR algorithm
  - Computation of clutter metric may only use target size information
- All the metrics fall under these categories:

a) Anomaly metrics

b) Information measure in hyperspectral image

c) Features derived from Gray Level Co-occurrence Matrix (GLCM)

d) Single band image statistics

e) Parameters from filtered single bands (Gabor filtering)

(1) Present Hyperspectral Image Clutter Metrics – Description

a) Anomaly detectors:

- Relative entropy: $(X/\|X\|) \bullet (Y/\|Y\|)$

where X and Y are the pixel vector under test and surrounding pixel respectively, both of length = No. of bands.

- Dot product: $\sum (p(x) \times \log \frac{p(x)}{q(x)})$

  where p(x) and q(x) are the distributions of the pixel under test and surrounding pixel respectively.

Both test statistics above are averaged over surrounding pixels considered.

b) Information measure in hyperspectral bands:

- Correlation between all combinations of hyperspectral bands are used as indicator of mutual information

(1) Additional Hyperspectral Image Clutter Metrics – Description

c) GLCM: we experimented with two variants of this, matrices were formed by making

- scatter plots of image pixel intensities at offsets based on the size of the target
- scatter plots of image intensities at random locations

d) Single band image statistics*:

- Image statistics are computed for each band of the hyperspectral image.
- The computed image statistics are: Standard deviation, Energy, Entropy, Edge, Target Interference Ratio, Schmieder and Weathersby clutter metric, FBM Hurst parameter, Generalized Gaussian Decomposition metrics (GGDM 1- 5).
- The Mean, Median, Min., Max., and Range are computed over the distribution of these values in the hyperspectral bands. These are the clutter metrics.

Namuduri, K.R., Bouyoucef, K., Kaplan L.M., "Image metrics for clutter characterization. IEEE Proceedings in Image Processing, Vol.2, pp.467-470, Sept. 2000.

(1) Additional Hyperspectral Image Clutter Metrics – Description (Contd.)

e) Parameters (p & c) from Gabor filtered images**

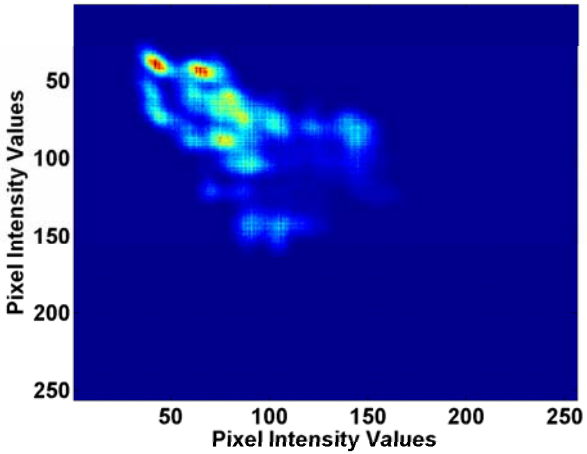
- Gabor filtration of single bands results in images with edges extracted at different orientations
- ***P*** is an indication of distinctness and frequency of edges and, ***c*** is related to the range of pixel values in the filtered image
- The Mean, Median, Min., Max., and Range are computed over the distribution of these parameter values in the hyperspectral bands. These are the clutter metrics.

- Summary:
  - Total of 129 clutter features

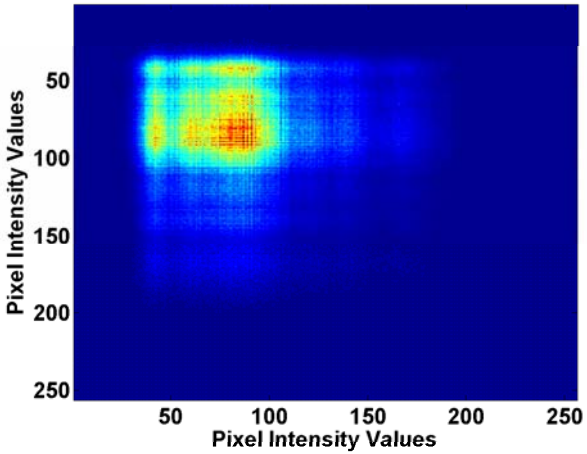
***Anuj Srivastava et al, "Universal Analytical Forms for Modeling Image Probabilities". IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.24, No.9, Sept. 2002*

(1) Hyperspectral Image Clutter Metrics – Examples

Single band from test hyperspectral image



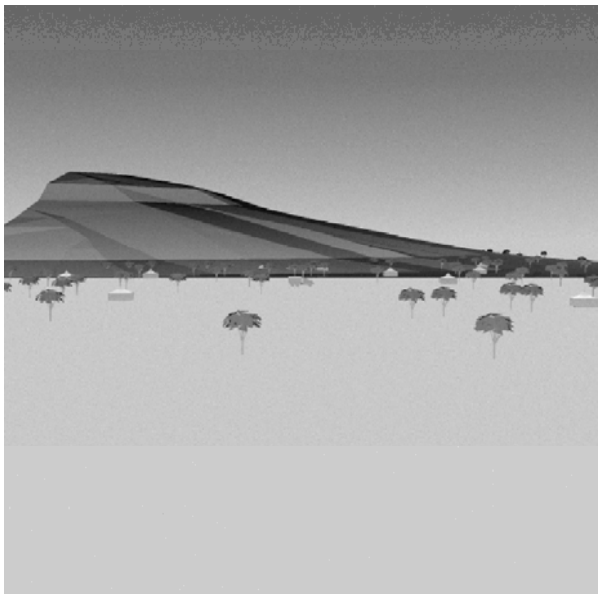
GLCM using offsets based on the target size



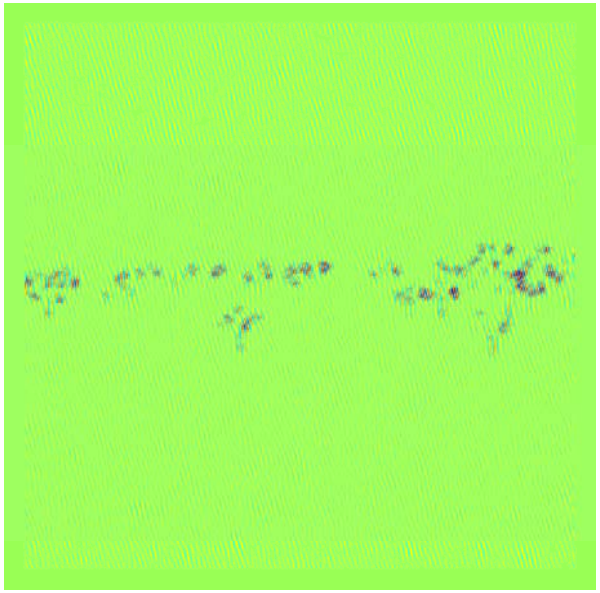
GLCM using random pixel locations

We experimented with both methods of computing the GLC Matrices. Experiments showed a difference in the derived clutter metrics and in both cases, a significant correlation between these metrics and our measure of complexity in images.

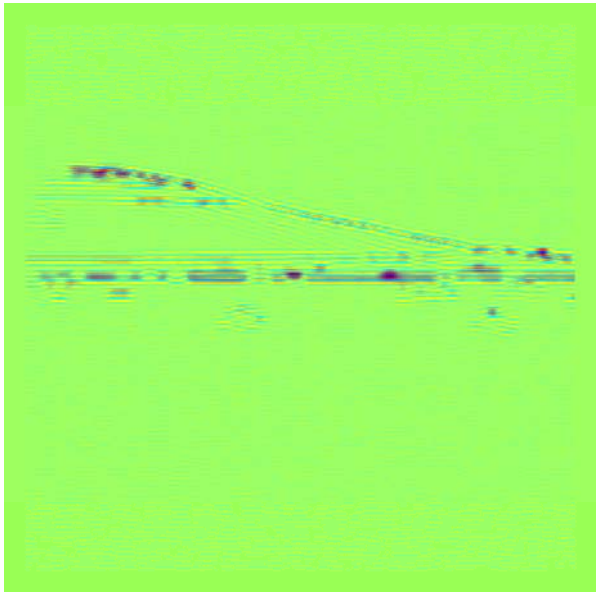
(1) Hyperspectral Image Clutter Metrics – Examples



Single band (λ = 10.28 microns) from hyperspectral image



15 deg. Gabor filter orientation extracts near-vertical edges



90 deg. Gabor filter orientation extracts horizontal edges

(2) Factor Analysis for Dimension Reduction

- Reduction of clutter metric space dimensions is in line with objective to keep Clutter Complexity Measure (CCM) computation simple
- Principal Component Analysis (PCA) and other traditional algorithms for dimension reduction ***will not*** identify the principal factors in terms of the original clutter metrics
- Employed dimension reduction algorithm seeks to:
  - Discard metrics with no significant correlation with image clutter ground truth (false alarm (FA) count)
  - Reduce redundancy

(2) Factor Analysis for Dimension Reduction

- Algorithm:
  - Remove non-correlating metrics: this results in the 'significant' metrics
    - *Compute correlation coefficient (CC) between single clutter metrics and image clutter ground truth (false alarm (FA) counts)*
    - *Discard metrics with 'insignificant' CC*
  - Reduce redundancy: this results in the metrics after 'complete' factor analysis
    - *Compute Correlation matrix of remaining metrics*
    - *Do for all combinations of two metrics in the subset:*
      - *If CC between metric A and B is significant*

        {
      - *if CC between A and FA is > CC between B and FA, discard B otherwise, discard A*

        }

(3) Experiments

- Test data – database of 216 synthesized hyperspectral images
- Image database is randomly divided into 3 equal partitions
- Factor analysis is performed on each partition independently
- Metrics with significant CC are used for train and test experiments
- Resulting metrics after complete factor analysis are used for train and test experiments
- Metrics common to at least 2 of the 3 image partitions after complete factor analysis are used for train and test experiments. These are called 'common' metrics
- Same common metrics above are used for image classification

(3) Result of Factor Analysis

The resulting common metrics from our experiments are:

1. ***Homogeneity*** derived from GLCM with arbitrary pixel location offset
2. ***Entropy*** derived from GLCM with random pixel locations
3. Range of ***p parameter*** values over all bands (90deg. orientation gabor filtered images)
4. Median of ***c parameter*** values over all bands(60deg. orientation gabor filtered images)
5. Mean of ***2-D standard deviation*** of all bands
6. Range of ***FBM Hurst parameter*** over all bands
7. Minimum of ***Edge measure*** over all bands
8. Minimum of ***GGDM measure 5*** over all bands

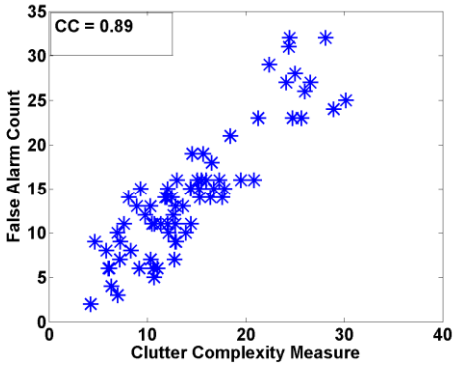
(3) Correlation Plots using Subset of Metrics with Significant CC to FA Count

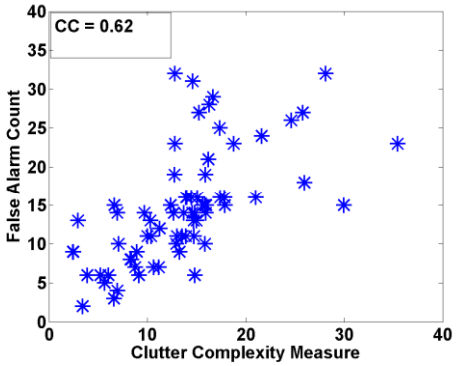
CC Values

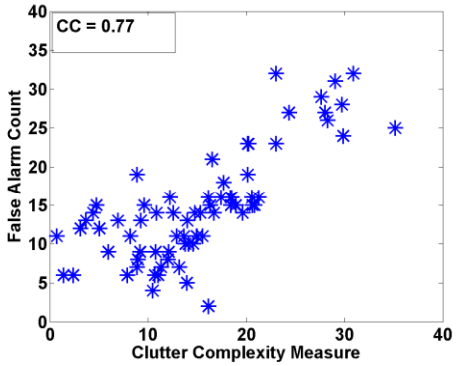
0.89	0.62	0.77
0.75	0.94	0.72
0.67	0.58	0.95

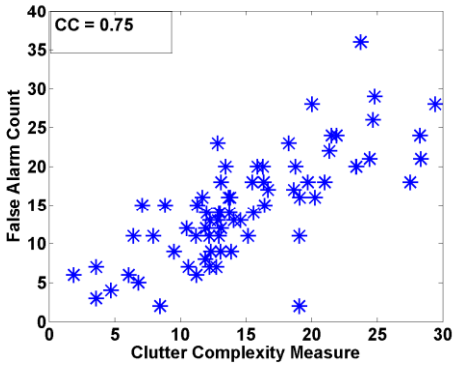
Training Partition I II III

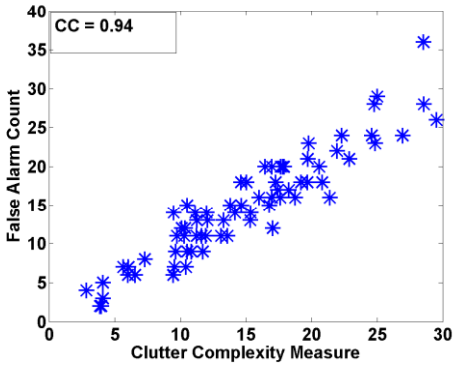
Test Partition I II III

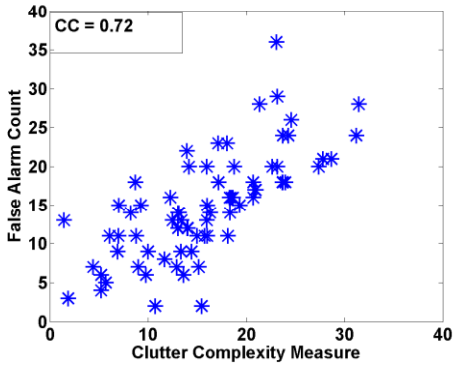


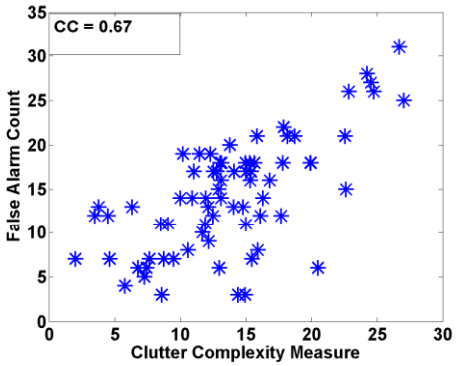


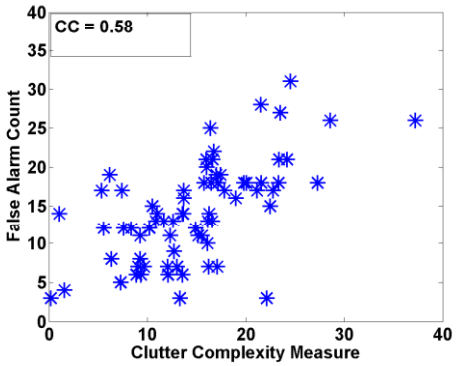


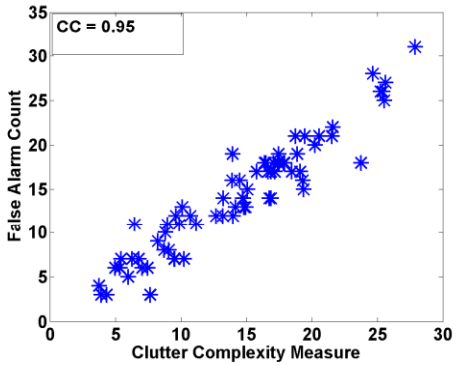












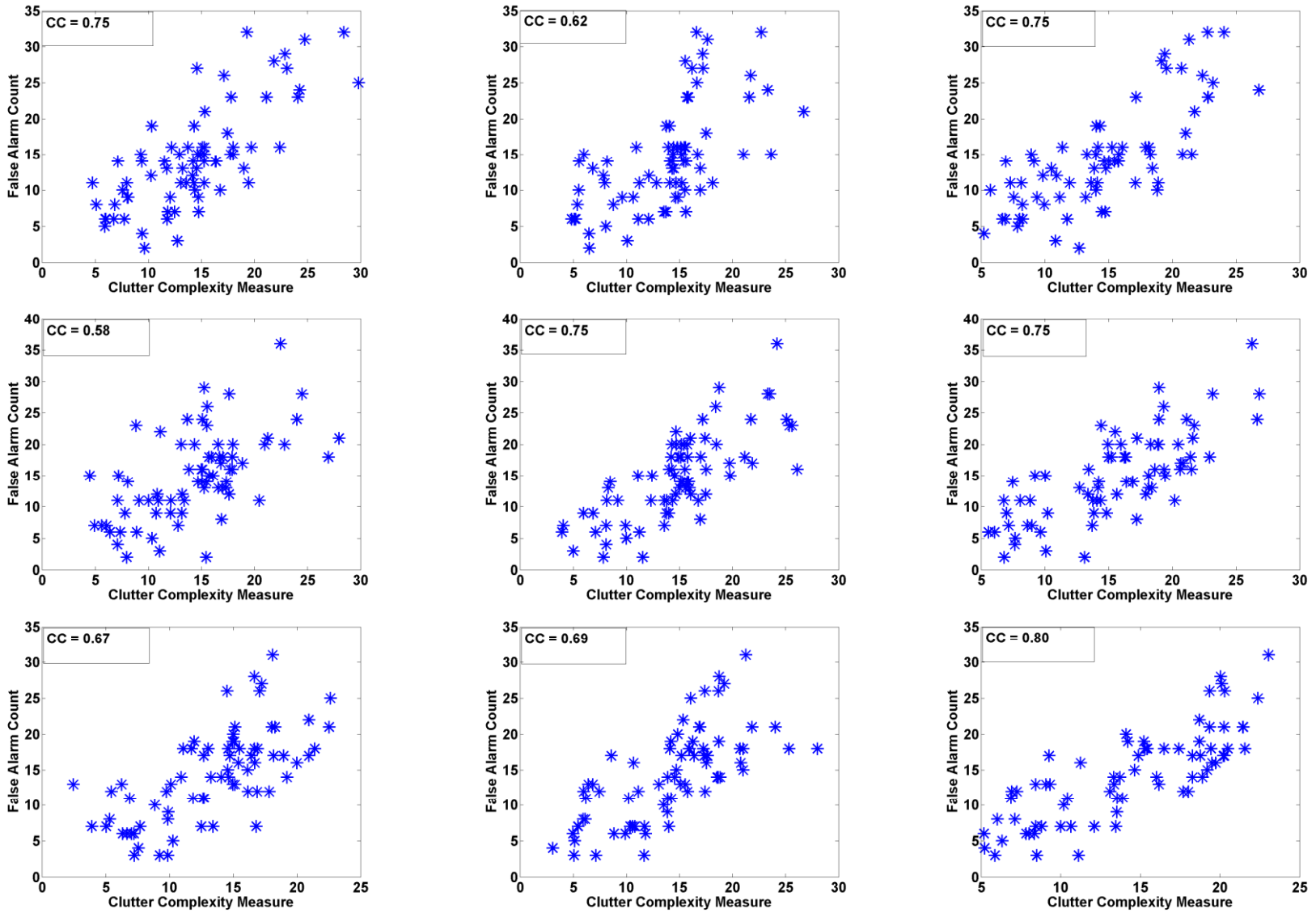
(3) Correlation Plots using Subset of Metrics after Complete Factor Analysis

CC Values

0.75	0.62	0.75
0.58	0.75	0.75
0.67	0.69	0.80

Training Partition

Test Partition



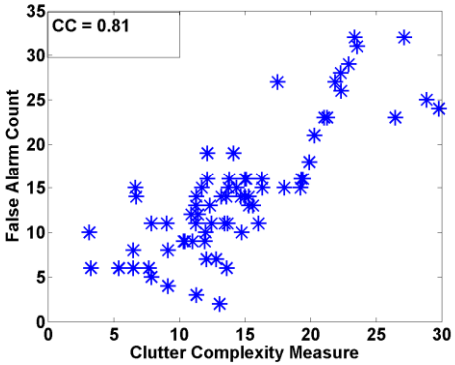
(3) Correlation Plots using Common Metrics

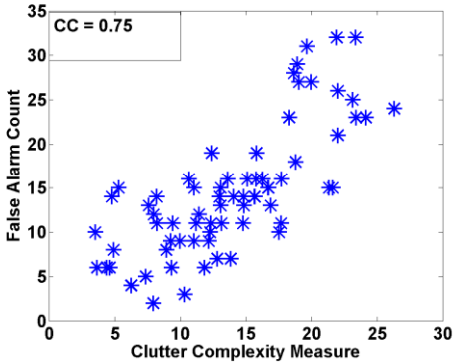
CC Values

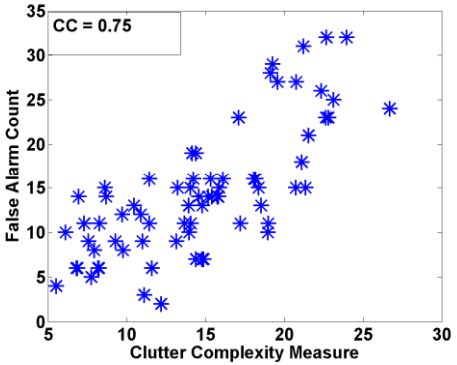
0.81	0.75	0.75
0.70	0.79	0.75
0.73	0.77	0.80

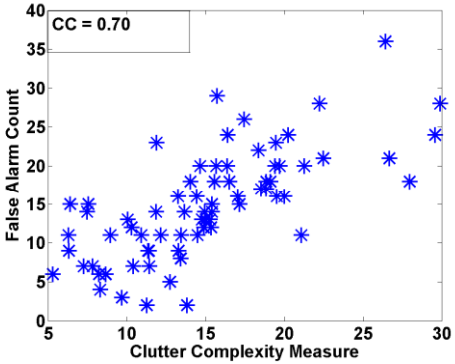
Training Partition I II III

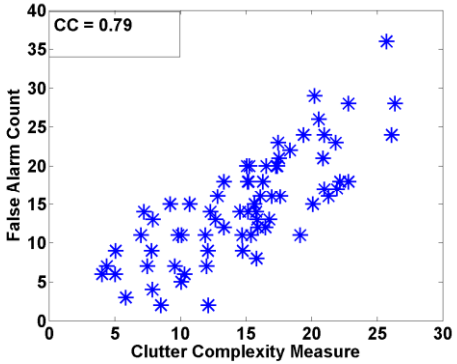
Test Partition I II III

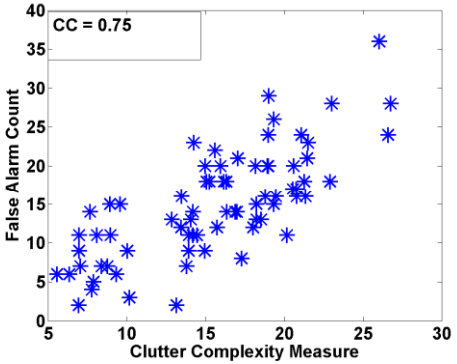


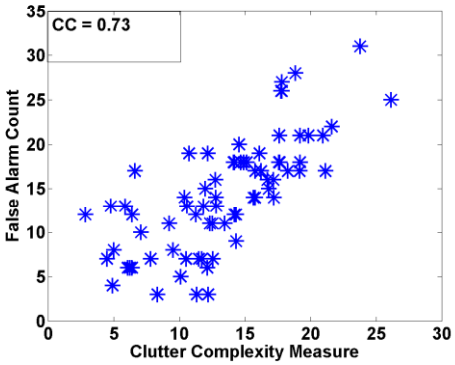


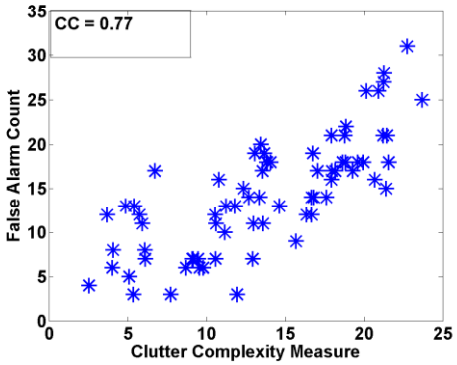


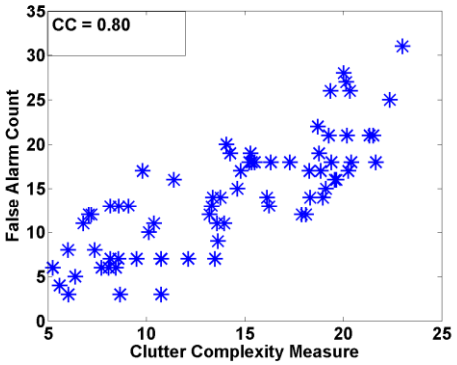




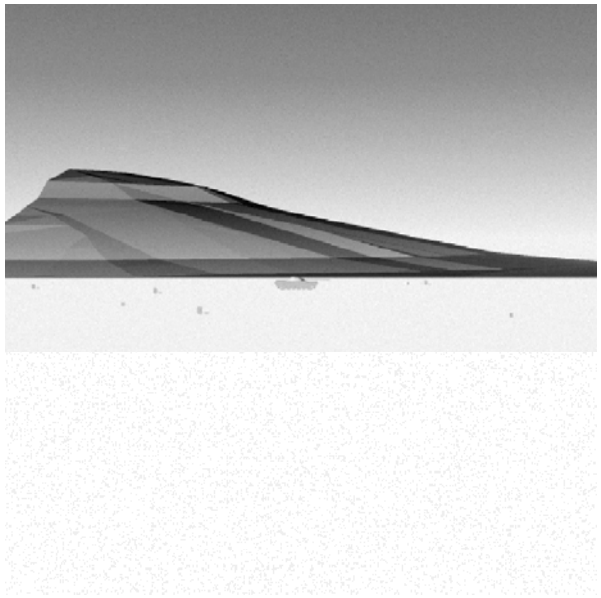




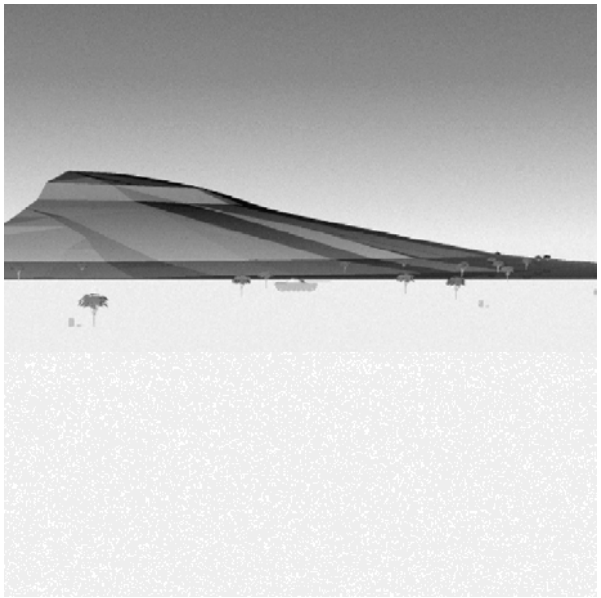




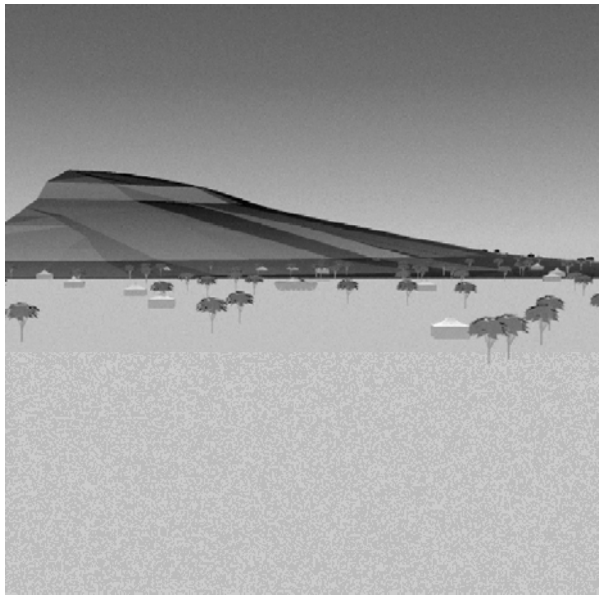
(4) Image Classification using Clutter Complexity Measure (CCM) derived from Common Metrics



(a) Low Clutter Complexity
CCM = 9.52



(b) Medium Clutter Complexity
CCM = 17.39



(c) High Clutter Complexity
CCM = 25.52

Examples showing single bands (λ = 10.28 microns) from hyperspectral images

Implications of a Clutter Complexity Measure

- One, Small (1 or 2 pixels)/Big target:
  - What clutter metrics would be expected to be significant based on target size?

- Multiple, Small/Big Targets:
  - What constitutes clutter?, similar targets, other objects, or combination?

- Combination of Multiple Small and Big Targets:
  - What clutter metrics would be expected to be significant considering multiple target sizes?
  - What constitutes clutter? are small objects clutter for big targets and vice-versa?

Implications of a Clutter Complexity Measure

- One, Small (1 or 2 pixels)/Big target:
  - What clutter metrics would be expected to be significant based on target size?
- Multiple, Small/Big Targets:
  - What constitutes clutter?, similar targets, other objects, or combination?
- Combination of Multiple Small and Big Targets:
  - What clutter metrics would be expected to be significant considering multiple target sizes?
  - What constitutes clutter? are small objects clutter for big targets and vice-versa?

Further Work

- More involved determination of metric weightings
- Repeat complete process performed on synthesized imagery on real hyperspectral image database.
- Continue work in automatic generation of hyperspectral image database with maximal diversity in clutter complexity

Clutter Complexity for Hyperspectral Imagery

2006 Research Efforts

Clutter Complexity Measure (CCM)

- **Goal:** to derive an objective measure of clutter complexity in hyperspectral images
  - as indication of the inherent difficulty of target recognition by ATRs,
  - for upper bound on the performance of ATR algorithms.

- **Approach:** find the aggregation of image metrics and statistics that correlate best with ATR baseline performance.

- **Properties of the CCM:**
  - independent of any particular ATR,
  - obtained with much lower computational complexity compared to a typical ATR.

- **Possible applications:**
  - objective basis for comparing different ATRs,
  - quick pre-assessment of image quality.

Definition of CCM

- The Clutter Complexity ***Measure*** is a function of selected image ***metrics*** that correlate best with baseline ATR performance.

- Metrics are:
  - Descriptive of scene parametric variation and significant for ATR performance
  - Computing them only requires a priori information on the order of spatial extent of the target in the scene at the most
  - Algorithmically uncomplicated, and easy to implement

- They fall into four main categories:
  - Single band metrics
  - Metrics based on band information content
  - Metrics based on anomaly detectors
  - Multi-band metrics

- Total of 129 metrics has been used.

Set of Metrics

Metric Name	Description	No. of Metrics
Single-band clutter metrics[1]		
FBM Hurst Parameter	Texture roughness	5
Standard deviation	Global standard deviation	5
Schmieder Weathersby	Average local standard deviation	5
Homogeneity	Average pixel variation	5
Energy	Average histogram energy	5
Entropy	Average histogram entropy	5
Target Interference Ratio	Average contrast	5
Outlier Ratio	Average percentage of outliers	5
GGABS(5 variations, $I - V$)	Generalized Gaussian Analysis-By-Synthesis	25
Gabor filter(5 orientations)	Parameters p (edge content), c (pixel intensity range)	2×5×5=50
Derived from band information content		
Band correlation	Mean/Median correlation in HSI bands	2
Anomaly detectors		
DotProduct	Average dot product of pixel vectors	1
Kullback-Liebler	Average relative entropy of pixel vectors	1
Derived from GLCM[2]		
GLCM Imax.	Inverse of maximum value from matrix	2×1=2
GLCM Energy	Energy computed from matrix	2×1=2
GLCM Entropy	Entropy computed from matrix	2×1=2
GLCM Contrast	Contrast computed from matrix	2×1=2
GLCM Homogeneity	Homogeneity computed from matrix	2×1=2
	Total	129

[1] 5 metrics - Min., Max., Mean, Median and Range are computed from the distribution obtained from computing these from the HSI image single bands
[2] Same values computed for both implemented variants of GLCM described

Set of Metrics

Metric Name	Description	No. of Metrics
Single-band clutter metrics[1]		
FBM Hurst Parameter	Texture roughness	5
Standard deviation	Global standard deviation	5
Schmieder Weathersby	Average local standard deviation	5
Homogeneity	Average pixel variation	5
Energy	Average histogram energy	5
Entropy	Average histogram entropy	5
Target Interference Ratio	Average contrast	5
Outlier Ratio	Average percentage of outliers	5
GGABS(5 variations, $I - V$)	Generalized Gaussian Analysis-By-Synthesis	25
Gabor filter(5 orientations)	Parameters p (edge content), c (pixel intensity range)	2×5×5=50
Derived from band information content		
Band correlation	Mean/Median correlation in HSI bands	2
Anomaly detectors		
DotProduct	Average dot product of pixel vectors	1
Kullback-Liebler	Average relative entropy of pixel vectors	1
Derived from GLCM[2]		
GLCM Imax.	Inverse of maximum value from matrix	2×1=2
GLCM Energy	Energy computed from matrix	2×1=2
GLCM Entropy	Entropy computed from matrix	2×1=2
GLCM Contrast	Contrast computed from matrix	2×1=2
GLCM Homogeneity	Homogeneity computed from matrix	2×1=2
	Total	129

[1] 5 metrics - Min., Max., Mean, Median and Range are computed from the distribution obtained from computing these from the HSI image single bands
[2] Same values computed for both implemented variants of GLCM described

Experiments with Synthesized Data

- Data description
  - Synthesized images generated using the Digital Imaging and Remote Sensing Image Generation (DIRSIG) software
  - Spectral dimensions: 8-13 microns, 40 nm resolution, 126 bands
  - Spatial dimensions: image size = 512 × 512 pixels, target size = 9 × 9 pixels
  - 216 images overall

- Steps to obtain CCM
  1. compute image metrics for all images
  2. obtain subset of metrics by factor analysis
  3. train on a subset of images to obtain CCM as weighted sum of subset of metrics that correlate best with baseline ATR performance
  4. test generalization of CCM on compliment of trained images

CCM Prediction of Baseline ATR Performance

Accuracy of ATR performance prediction for new images.

Training partitions: Sample size	Test partitions: Same as train sample	Test partitions: Test sample
11 images (5%)	0.84	0.40
22 images (10%)	0.77	0.57
32 images (15%)	0.76	0.62
43 images (20%)	0.74	0.64
54 images (25%)	0.73	0.65
65 images (30%)	0.72	0.66
76 images (35%)	0.73	0.66
86 images (40%)	0.72	0.67

CCM predicts baseline ATR performance with over 60% accuracy after training with 32 images.

Efficient Image Synthesis

- Need to augment available real images with synthesized data.
- Goal: to synthesize a set of hyperspectral images to evaluate the performance of ATRs.
- Synthesized images should be diverse with respect to the degree of difficulty for the ATRs under test.
- Images synthesized using the Digital Imaging and Remote Sensing Image Generation (DIRSIG) model
  - synthesizes Multi- or hyperspectral images in the 0.3 to 20 microns region,
  - an integrated collection of first principle based sub-models:
    - Scene geometry
    - Atmospheric contributions
    - Material properties
    - Ray tracing

Synthesizing images with DIRSIG

- Model each generated image as a function of multiple factors. Each image is a point in the multi-dimensional/factor space.
- Single values, indicative of ATR difficulty from each image, forms multi-dimensional surface
- Need for efficient sampling of this surface
  - Generating images from all combination of factors infeasible,
  - Gradient search based sampling and similar algorithms also not feasible,
  - Evenly spaced, and randomly picked combination of factors for generating sample images may not be efficient .

Adaptive Sampling by Histogram Equalization (ASHE)

- Progressive sampling scheme
  - space is sampled in location of active walker,
  - active walkers move according to state of normalized histogram.
- Benefits:
  - Only requires ability to obtain function value at sampled point
  - Neither prior information on the global, nor relative levels of local variation of the function are required.
  - Minimal computational overhead.
  - Particularly useful when obtaining samples of a function is prohibitively expensive .
- Achieves two purposes that are apparently equivalent:
  - efficient distribution of sample points,
  - improved diversity in samples.

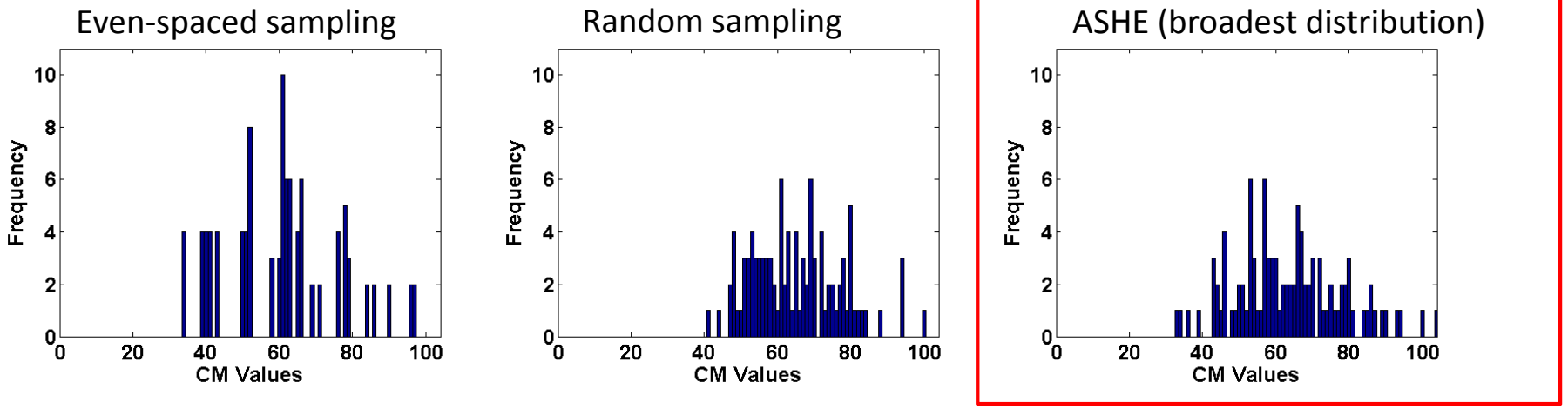
Properties of synthesized hyperspectral images

A single band (λ = 0.56 nm) from the hyperspectral image of the urban scene. The arrow indicates the 3 ×3 pixels region cropped as target.

- Hyperspectral images generated according to urban scene from the DIRSIG tutorial
- 44 equally spaced spectral bands in the visible to near infrared (0.35 - 1.0 nm)
- Spatial size 128×128 pixels

Image Synthesis based on ASHE

- Each synthesized image is modeled as multi-dimensional function, each dimension being a possible contributor to image variability e.g. Hour of day (1-24), Month of year (1-12) Visibility parameter (0-40km)
- Space is sampled using ASHE algorithm in order to maximize diversity in synthesized image with respect to target detection difficulty (Complexity Measure CM values)
- Bin representation: indication of diversity in CM values, which also indicates required diversity in images
- Deviation: indication of evenness in representation



2006 Research Accomplishments

- Computed 129 metrics from hyperspectral images
- Reduced metric space by factor analysis
- Derived clutter complexity measure (CCM) as aggregation of subset of metrics that correlates best with baseline ATR performance
- Derived CCM with this approach for real and synthesized images
- About 15% of our synthesized image databases is sufficient for defining the CCM
- Similar initial results from limited real data
- Comparison of time to run ATR to calculating CCM for synthesized images shows a ratio of about 9 : 1
- Some common image metrics across databases, but generally, CCM is specific to database from which it is derived
- Presented algorithm for efficient image synthesis - Adaptive Sampling based on Histogram Equalization (ASHE)
- ASHE improves diversity in synthesized images with respect to ATR performance

Hyperspectral Polarimetric Data Collection and Analysis

University of Hawaii

Participants

- Dr. Paul Lucey
- Mr. Tim Williams
- Mr. Mike Winter
- Donovan Steutel

Research Objectives

Thrust Area 1: Hyperspectral Polarimetric Data Collection and Analysis

Principal Tasks:

1) Collection of calibrated hyperspectral and hyperspectral-polarimetric field data
2) Search for discriminants
3) Quality assessment of previously obtained data

Principal Task 1: Collection of calibrated hyperspectral and hyperspectral polarimetric field data

Main Asset:

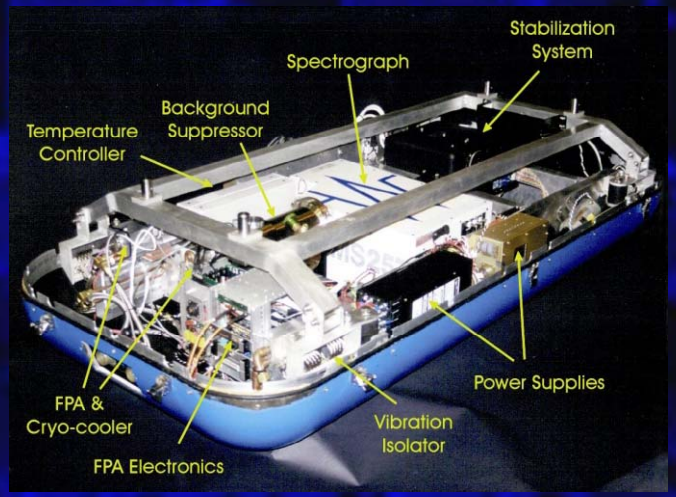
- AHI LWIR Hyperspectral imager
  - Coverage from 7-11.5 micron
  - 256 spatial pixels
  - 256 spectral bands
- Undergoing modification for polarimetric hyperspectral measurements under DARPA Eyeball program

AHI Spectral Technology Group

Airborne Hyperspectral Imager University of Hawaii

Airborne Hyperspectral Imager (AHI)

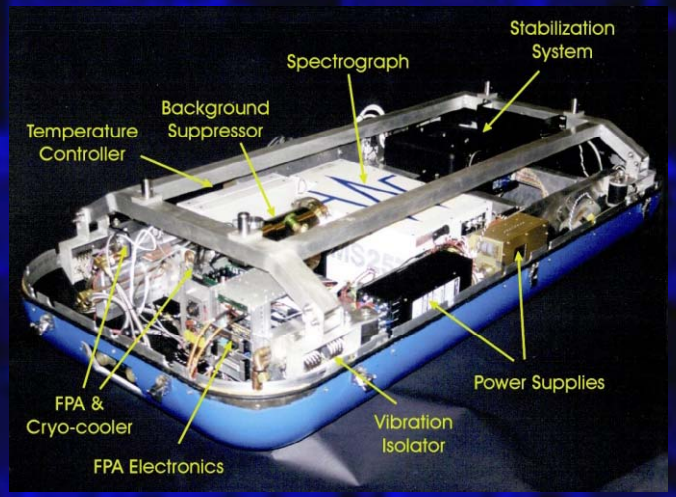
- Developed for DARPA's Hyperspectral Mine Detection (HMD) Program
- LWIR pushbroom hyperspectral sensor (7-11.5 μm)
  - •Rockwell 256x256 element HgCdTe
  - •150Hz frame rate
  - •Real-time two point radiometric calibration



AHI Spectral Technology Group
Airborne Hyperspectral Imager University of Hawaii

AHI (cont)

- Operated in the air and on the ground
- Supported customers from DoD, NASA, EPA and allied military partners
- Airborne data collections have focused on buried mine and concealed target detection



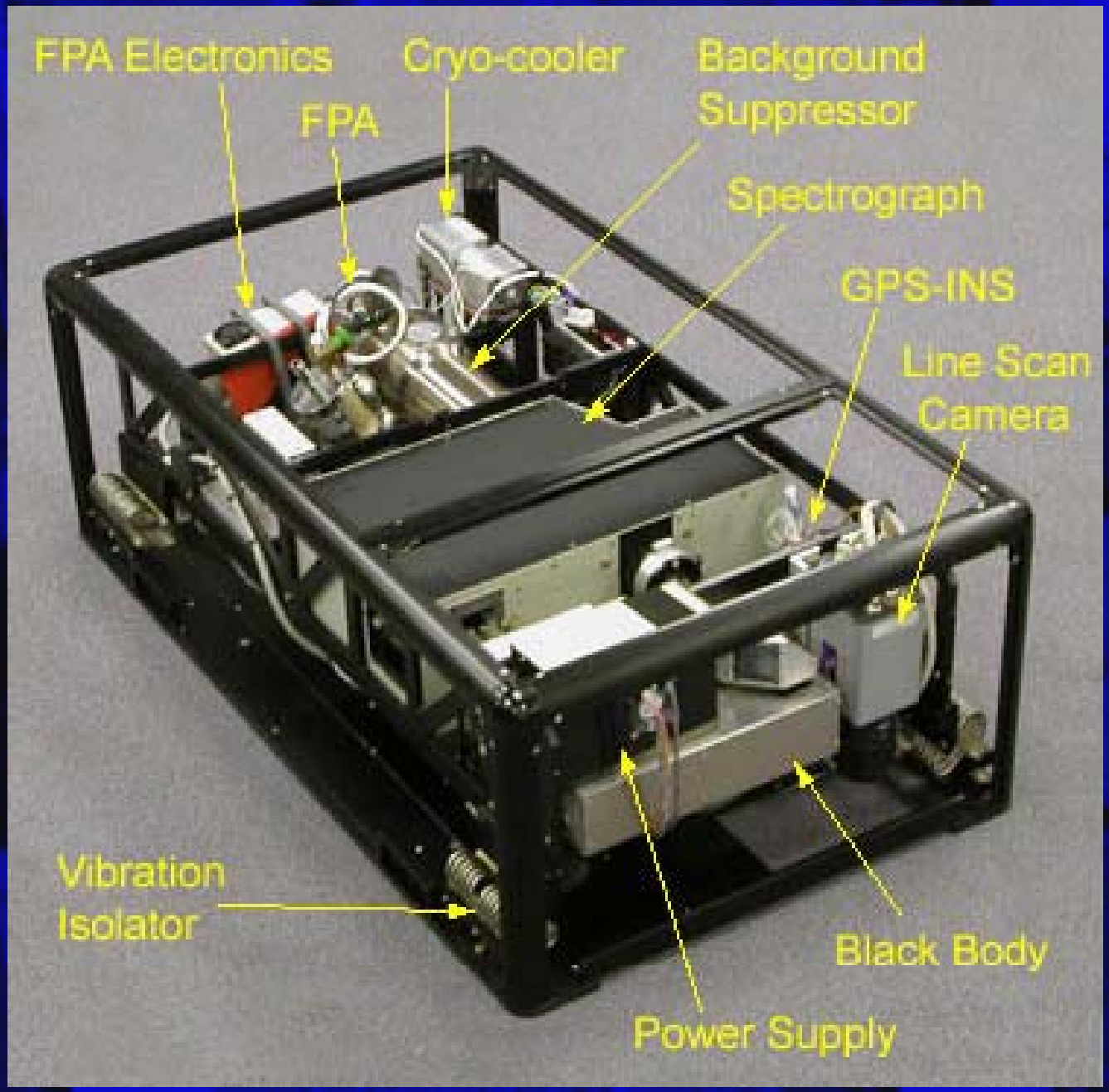
AHI Specifications

- LWIR Push broom Imaging Spectrometer
- 7 - 11.5 µm Spectral Coverage
- 256 Spectral Bands
- 256 Spatial Pixels
- 150 Hz Frame Rate
- 0.9 x 2.0 mrad Angular Resolution
- 0.1° K Sensitivity

Recent AHI Modifications

- GUI/Acquisition Software Upgrade
- Three Color Linescanner with NIR (850 nm) Band
- New Electronics Allow Full 256 Band Resolution
- Multi-Point Blackbody Calibration
- 280 Gbyte On-Board Data Storage Capacity
- Polarimetric Measurement Capability
- INS/GPS Geometric Correction and Location
- Smaller Lighter Sensor Package

AHI Sensor


FPA Electronics
FPA
Cryo-cooler
Background Suppressor
Spectrograph
GPS-INS
Line Scan Camera
Vibration Isolator
Black Body
Power Supply

AHI Platforms

Tracking telescope

Ground Based

Helicopter

Fixed Wing

Applications to Date

- Airborne detection of land mines
- Hyperspectral land mine phenomenology
- Concealed target detection and phenomenology
- Gas detection
- Active laser hyperspectral imaging
- Geologic mapping
- Coastal water temperature mapping
- Missile defense intercept test support
- Basic Hyperspectral Research
- HSI/SAR fusion experiments

AHI Users

Defense Advanced Research Projects Agency (DARPA)

Space and Naval Warfare Systems Command (SPAWAR)

National Aeronautics and Space Administration (NASA)

National Imagery and Mapping Agency (NIMA)

Space Applications International Corp. (SAIC)

Night Vision Laboratory, Ft. Belvoir (NVL)

U.S. Geological Survey (USGS)

Environmental Protection Agency (EPA)

Defense Evaluation and Research Agency (DERA), UK

Defense Science and Technology Laboratory (DSTL),UK

Defense Research Establishment Ottawa (DREO), Canada

Defense Science and Technology Organisation (DSTO), Australia

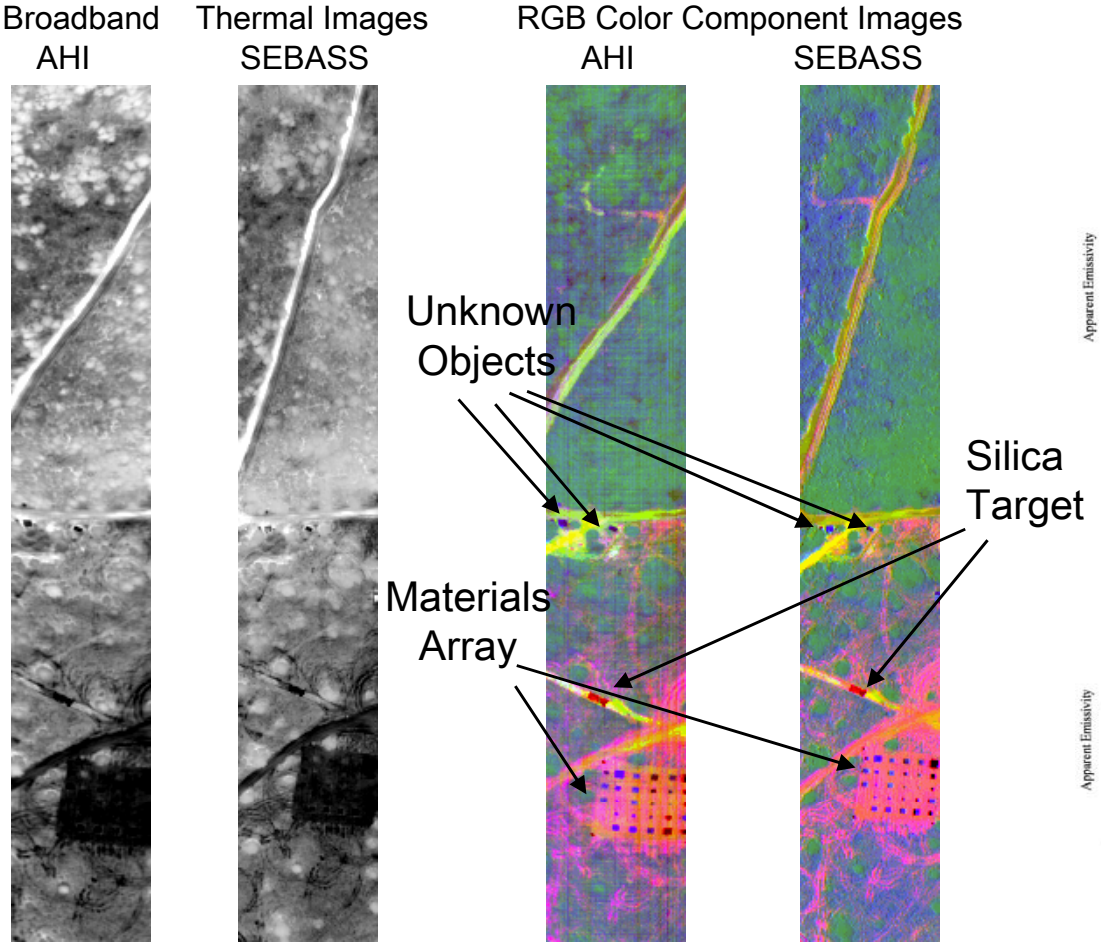
Space Missile Defense Command (SMDC)

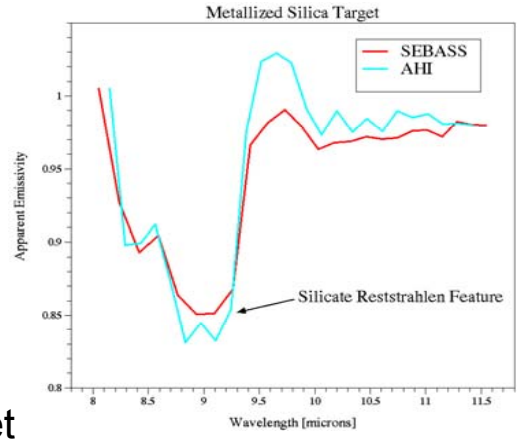
Comparison of AHI and SEBASS

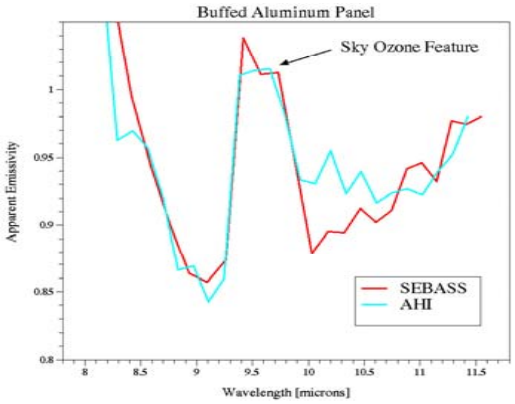
SPACE COMPUTER CORPORATION

- SEBASS spectral resolution reduced 3x to match AHI band width (~0.15 microns)
- AHI spatial resolution reduced 3x to match SEBASS pixel IFOV (~1 meter)

DH2 Site 3 Imagery from HYDRA



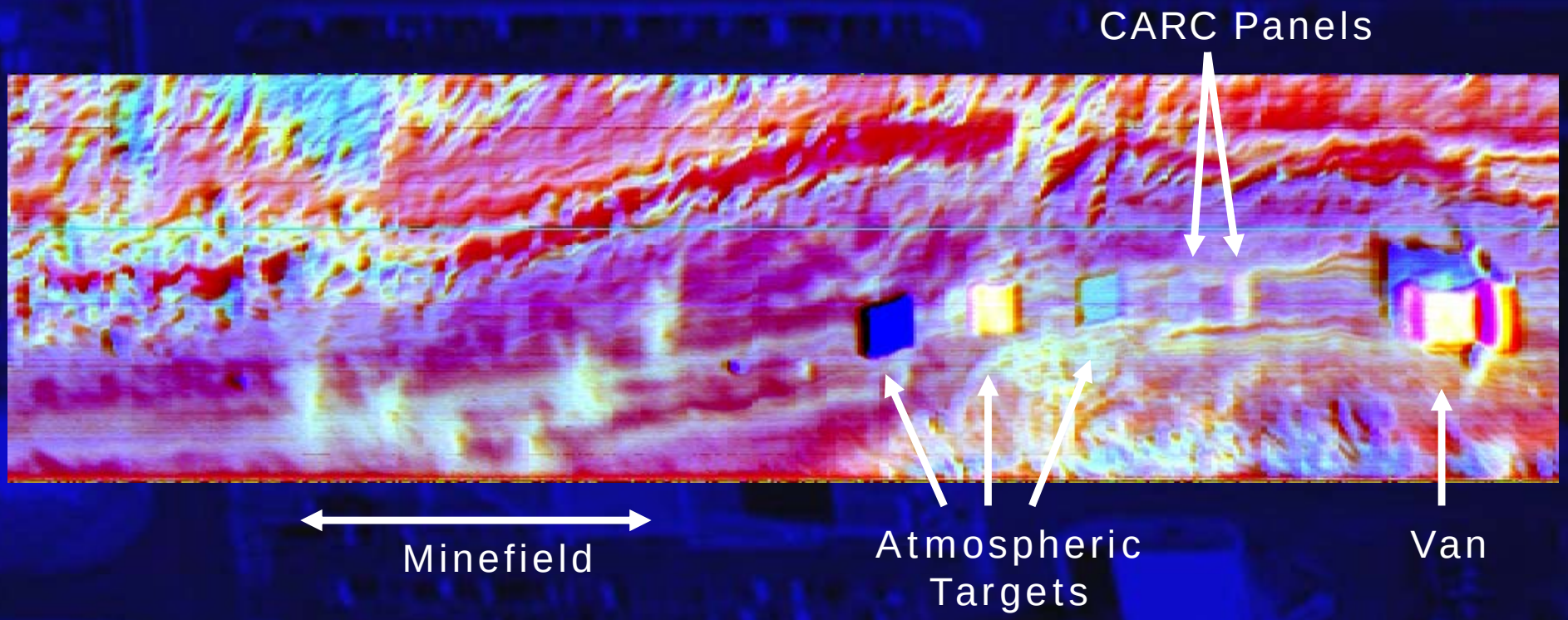




Afternoon Flight Data

3-color composite image of daytime AHI flight on 8/1/97, Ft. Huachuca, AZ.

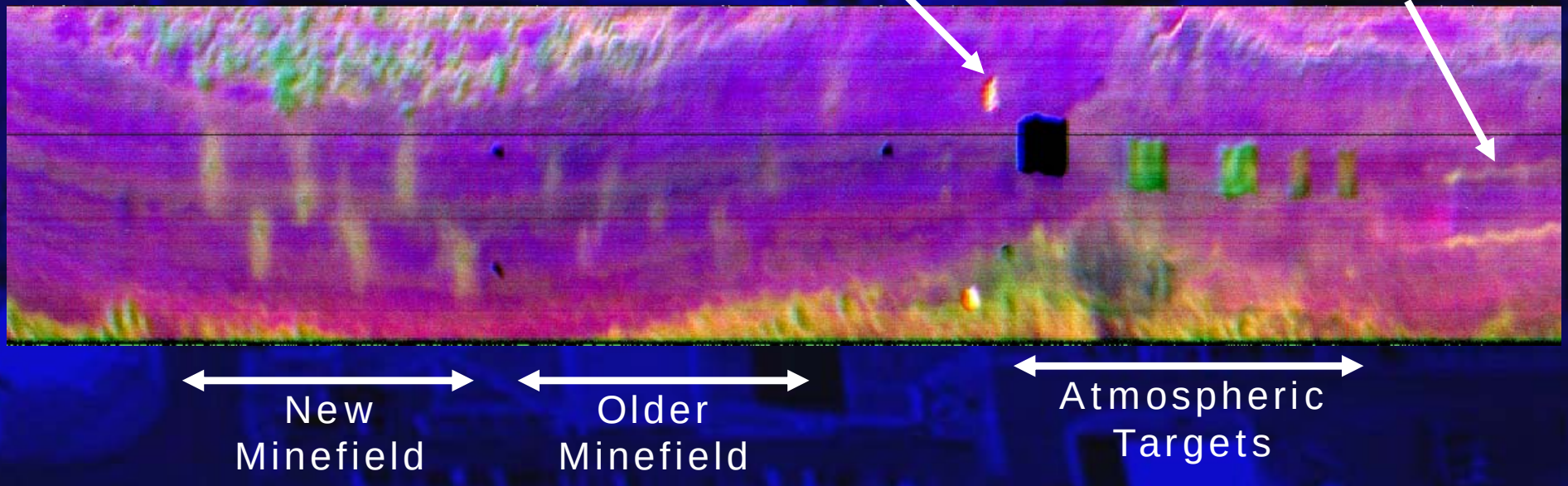
Red: **Average Brightness Temperature**
Green: **Apparent Emissivity at 9.16 μm**
Blue: **Apparent Emissivity at 8.21 μm**



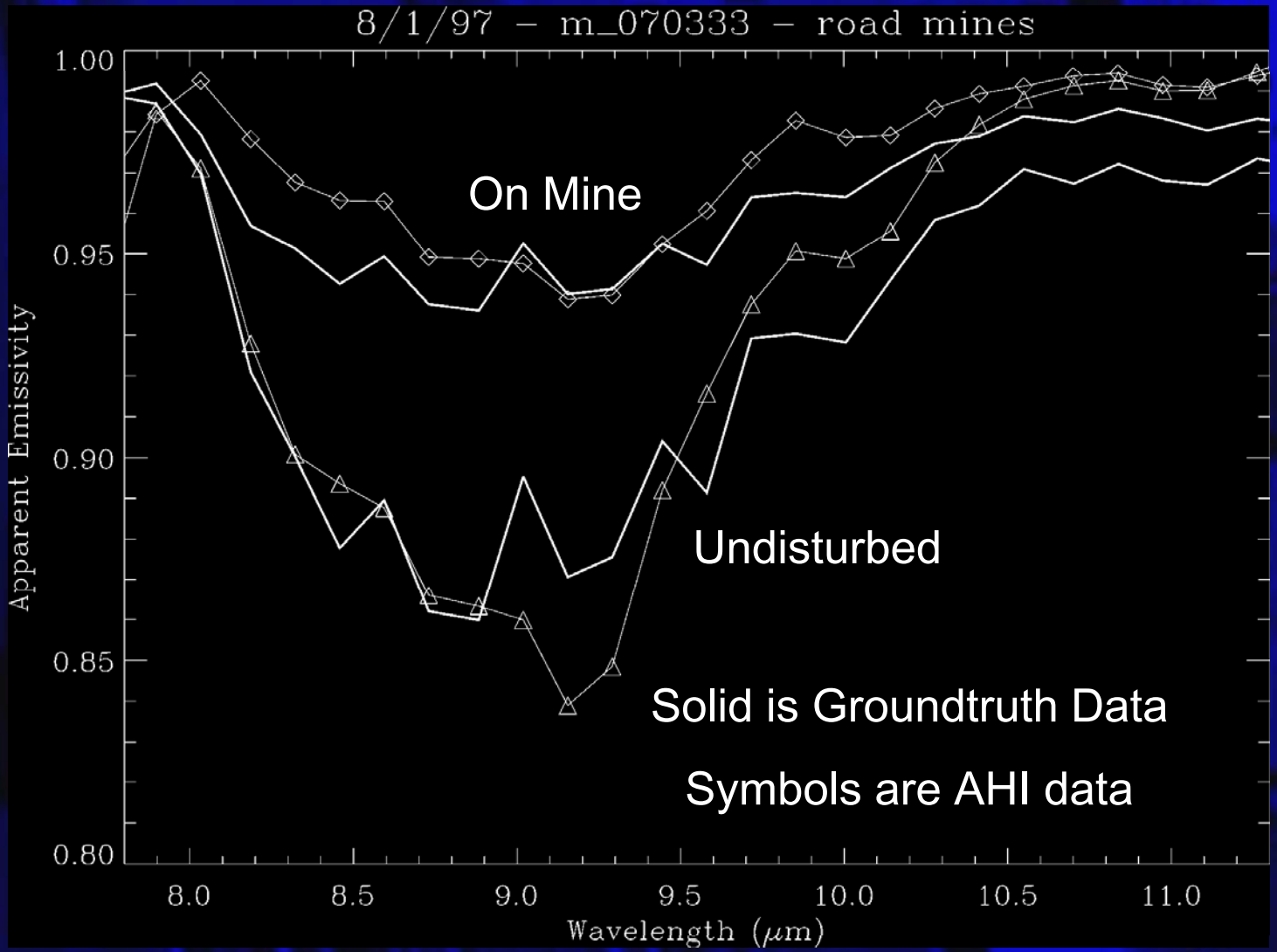
Night Flight Data

3-color composite image of pre-dawn AHI flight on 8/6/97, Ft. Huachuca, AZ.

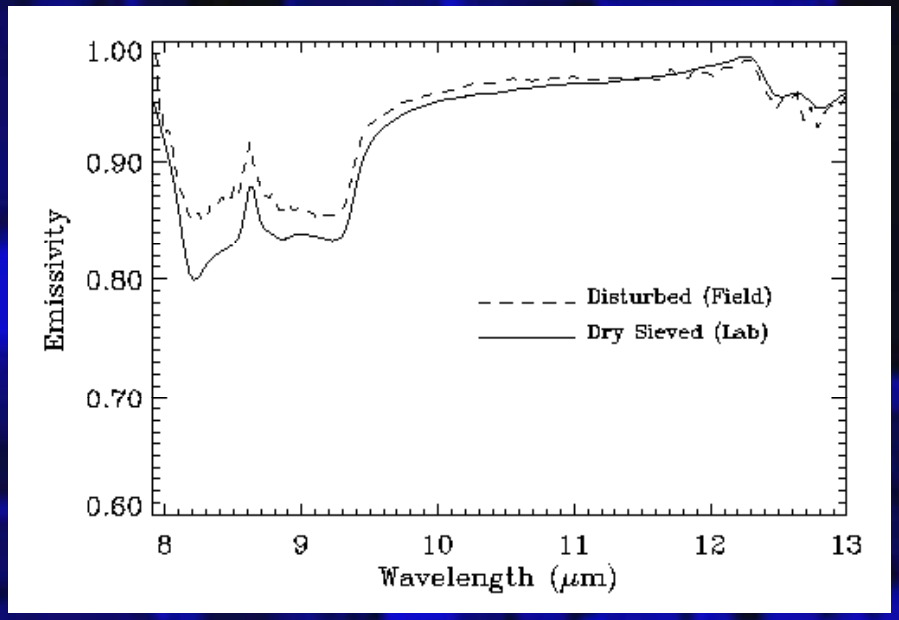
Red: **Average Brightness Temperature**
Green: **Apparent Emissivity at 9.16 μm**
Blue: **Apparent Emissivity at 10.25 μm**

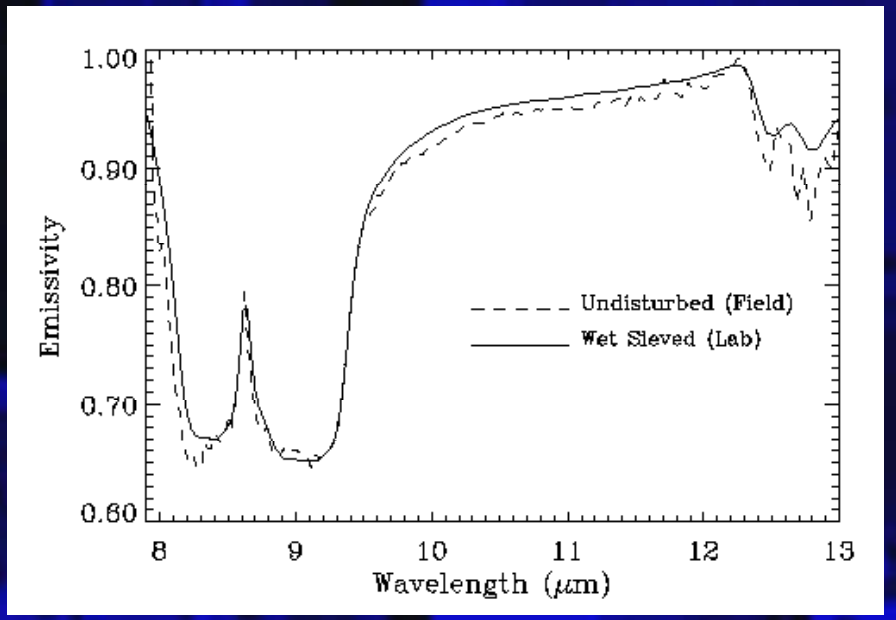


August 1, 1997 - AHI Thunder Road Mines



Geological Mechanism

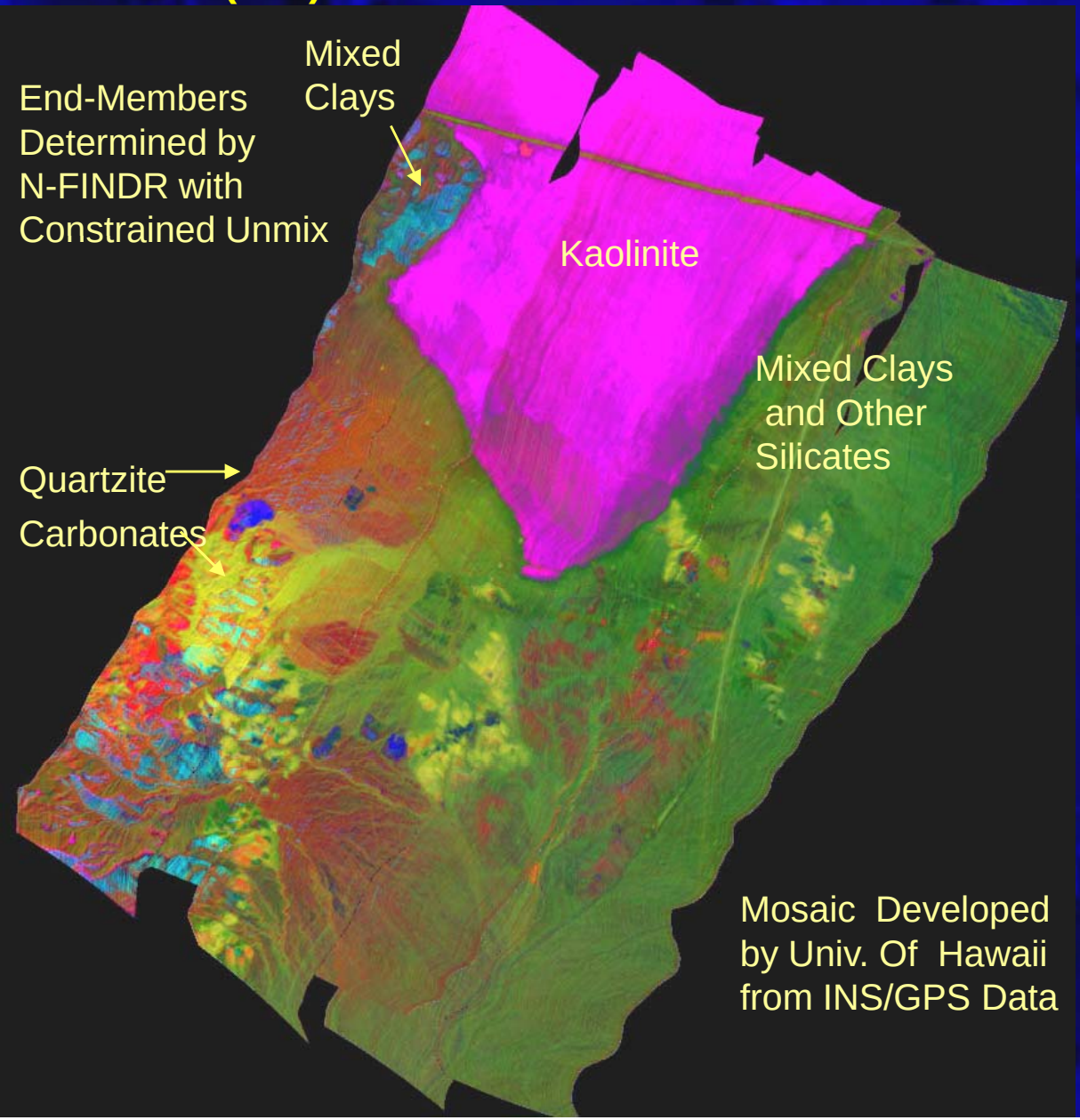




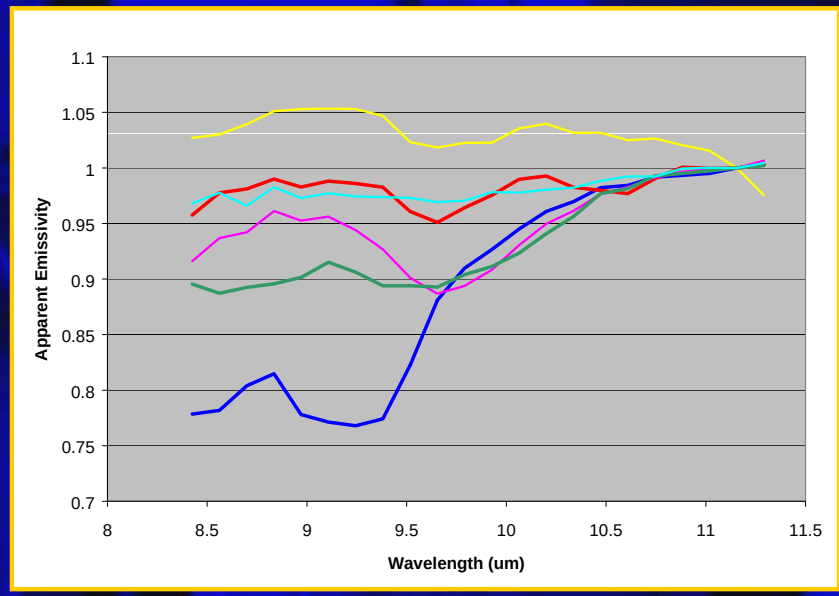
AHI Spectral Technology Group
Airborne Hyperspectral Imager University of Hawaii

Color Fraction Plane Image

Mosaic of (AHI) LWIR Data of Silver Lake CA



Colors in Spectral Plot Corresponds to Colors in Image



Three Primary Colors (Red Blue Green) Along with Mixed Colors Magenta, Cyan and yellow Used to Represent Six Fraction Planes

Principal Task 2: Search for discriminants

Main Asset: in-house statistical analysts

Mike Winter, Post-doctoral Fellow

Donovan Steutel, PhD Student

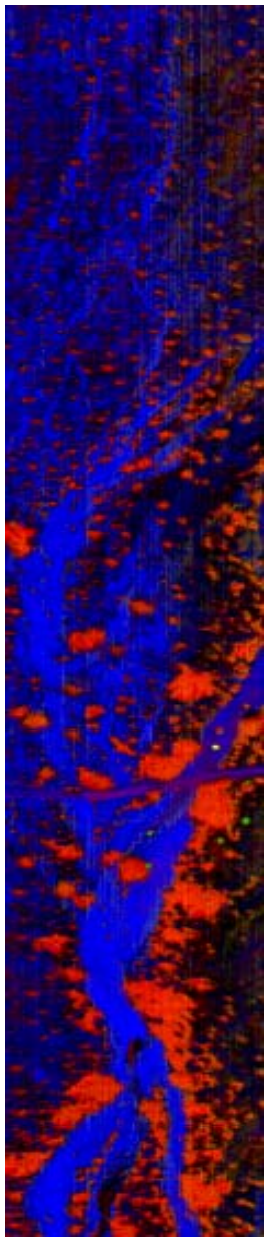
Supporting assets:

RX, Orasis, N-findr spectral anomaly detector

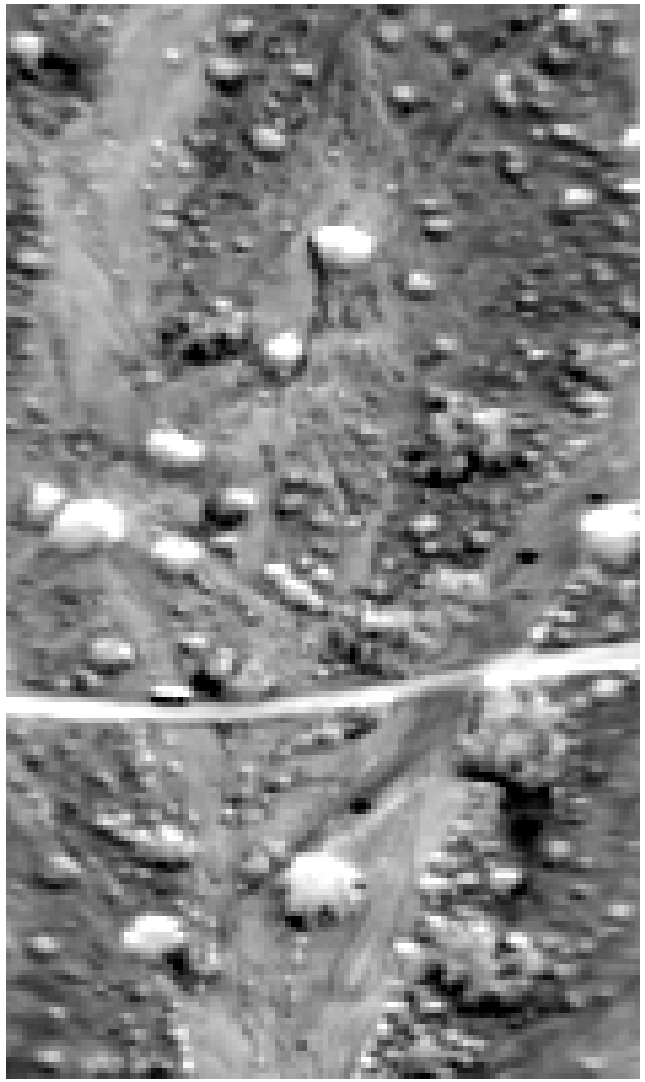
Lechuguilla Panel Experiment

AHI
Airborne Hyperspect

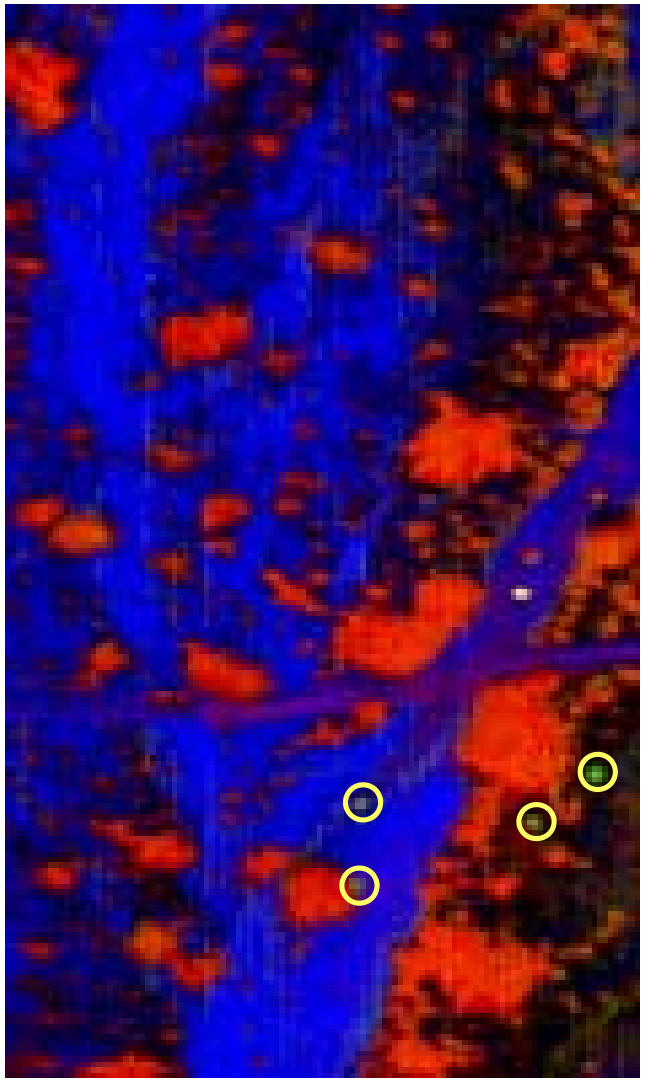
Run 1 - AHI data from Feb 14, 1999



AHI Broad Band IR Image



Fractional Abundance
End-member Image



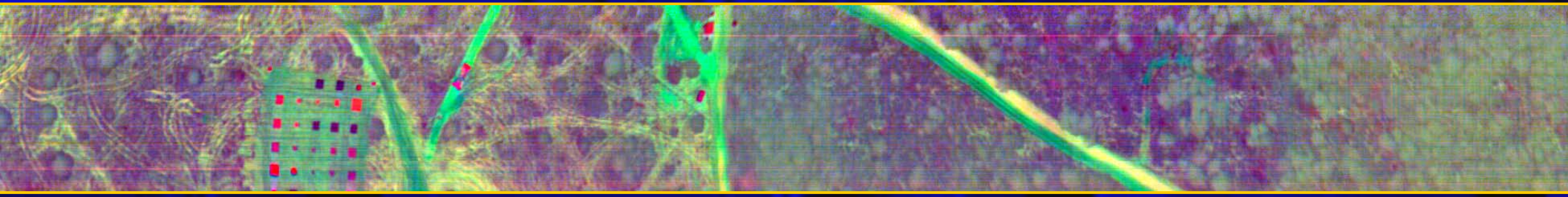
Painted Panels Colored Green

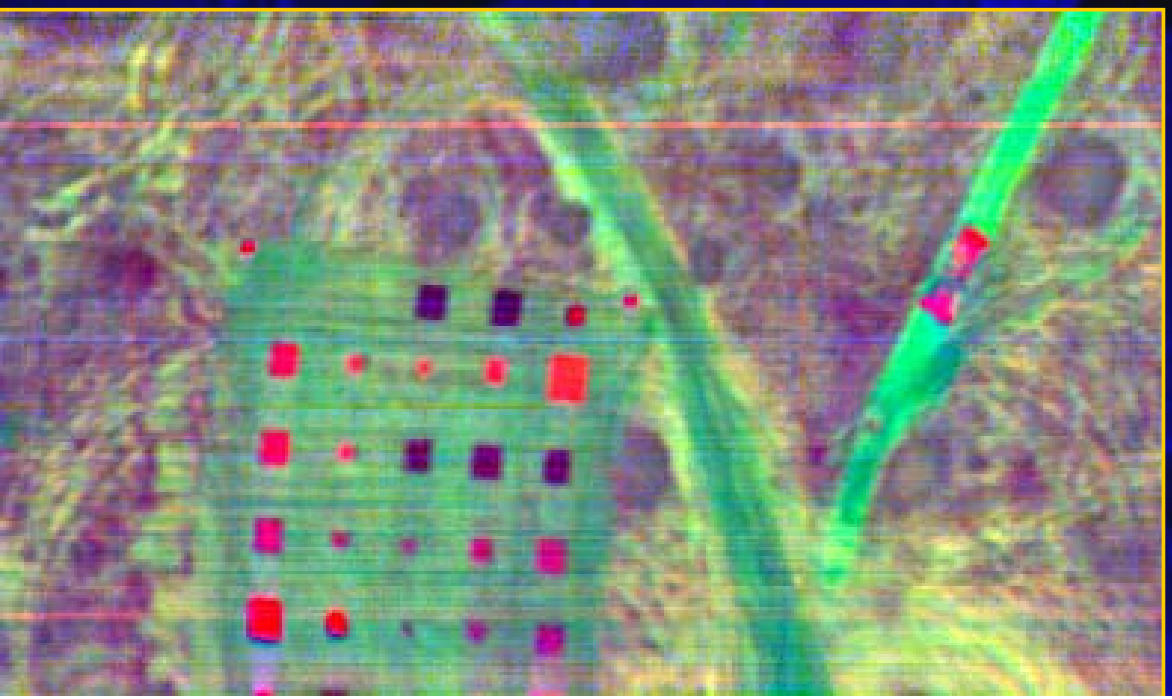
AHI Spectral Technology Group
Airborne Hyperspectral Imager University of Hawaii

AHI Night Mission

Broad Band Temperature Image

Color Principal Component Image (Excluding Temperature Component)



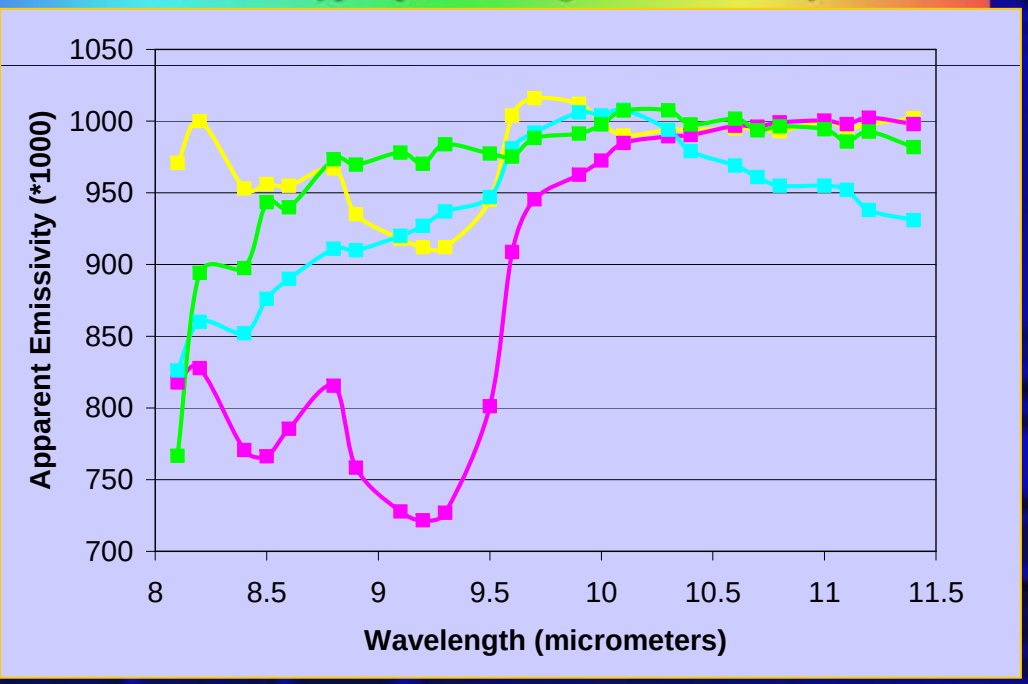


Enlargement of Material Array Area

Color Image Made from PC1,2,3
(first three after temperature removed)

AHI Spectral Technology Group

Airborne Hyperspectral Imager University of Hawaii



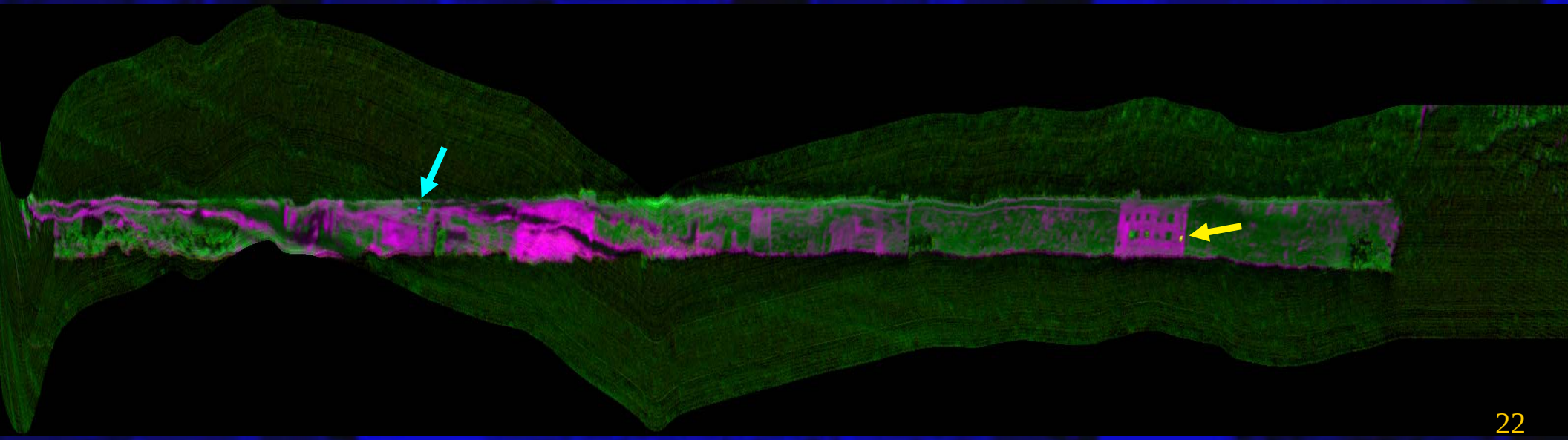
CYAN: metal

Yellow: silica mix (panel)

Magenta: silica mix

Green: ?

Segmented Thermal HSI



Principal Task 3: Existing data and quality assessment

Data set Name or Other Identifier	Data	Sensor	Spectral Region	Spatial Resolution	Number of Bands	Field of View Width (pixels)	Acquisition Platform	Data Classification	Description
CARD SHARP	1996	SEBASS	LWIR MWIR	0.3 – 1m	L 128 M 128	128	300' Tower	Unclassified	Slant View of Targets at Redstone
HYDRA	Nov-98	SEBASS	LWIR MWIR	1m	L 128 M 128	128	Twin Otter 3000 ft	SECRET	Chicken Little
Forest Radiance	1996	HYDICE	VNIR SWIR	0.75 – 3 m	210	320	Convair 3 – 12 kft	Unclassified	Targets at Aberdeen
Desert Radiance	1996	HYDICE	VNIR SWIR	0.75 – 3 m	210	320	Convair 3 – 12 kft	Unclassified	Targets, Backgrounds near Yuma
ASRP NVIS A.P. Hill 1999	1999	NVIS	VNIR SWIR	1 – 2 m	384	256	Twin otter 3 – 6 kft	FOUO	Mil Targets at Fort A. P. Hill, VA

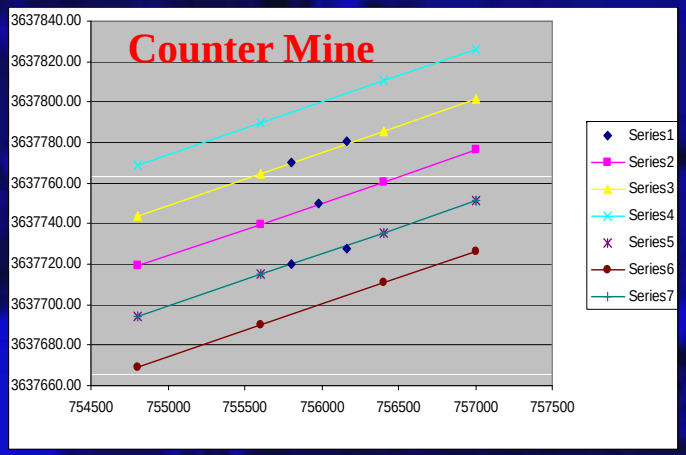
Data set Name or Other Identifier	Data	Sensor	Spectral Region	Spatial Resolution	Number of Bands	Field of View Width (pixels)	Acquisition Platform	Data Classification	Description
ASRP AHI A.P. Hill 1999	1999	AHI	LWIR	0.5 - 1 m	32	256	Twin otter 3 – 6 kft	FOUO	Mil Targets at Fort A. P. Hill, VA
ASRP NVIS A.P. Hill 2000	2000	NVIS	VNIR SWIR	1 – 2 m	384	256	Twin otter 3 – 6 kft	FOUO	Mil Targets at Fort A. P. Hill, VA
Forest Radiance II	2000	SHARP	LWIR	1 m	128	128	RB-57	SECRET	Targets at Aberdeen
Desert Radiance II	2000	SHARP	LWIR	1 m	128	128	RB-57	SECRET	Targets at Yuma
Forest Radiance II NVIS	2000	NVIS	VNIR SWIR	1 – 2 m	384	256	Twin otter 3 – 6 kft	SECRET	Targets at Aberdeen
Greyling SHARP	2001	SHARP	LWIR	1 m	128	128	RB-57	SECRET	Targets in snow at Camp Greyling, MI
Greyling NVIS	2001	NVIS	VNIR SWIR	1 – 2 m	384	256	Twin otter 3 – 6 kft	SECRET	Targets in snow at Camp Greyling, MI

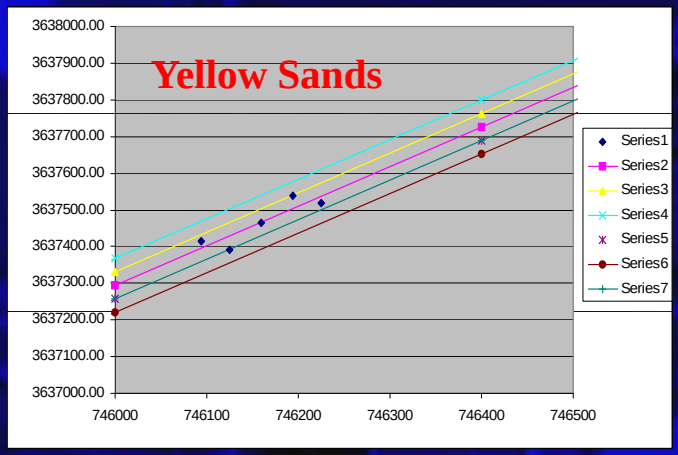
Data Collections Supporting MURI

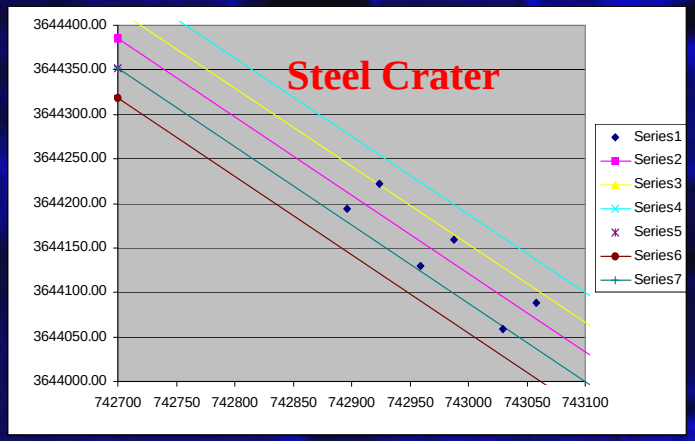
- Wide Area Airborne Minefield Detection (WAAMD), Yuma, April 2003, Ft. Leonard Wood, July 2004
- Sensor Week, Eglin AFB, Multiple Targets, May 2004
- EPA Texas Gas Detection, April 2004

Wide Area Airborne Minefield Detection (WAAMD)

- Yuma Arizona, April 2003
- Ft. Leonard Wood, July 2004
- ~200 Gb data collected







Planned Flight Paths

- Countermine Covered with Five Parallel Flight Paths Each Visit
  - In addition, an Additional Flight Line was Flown Over Older Mines
- Yellow Sands Covered with Three Flight Lines Each Visit
- Steel Crater Covered with Three Flight Lines Most Visits
  - Twice Flown with Five Parallel Flight Lines

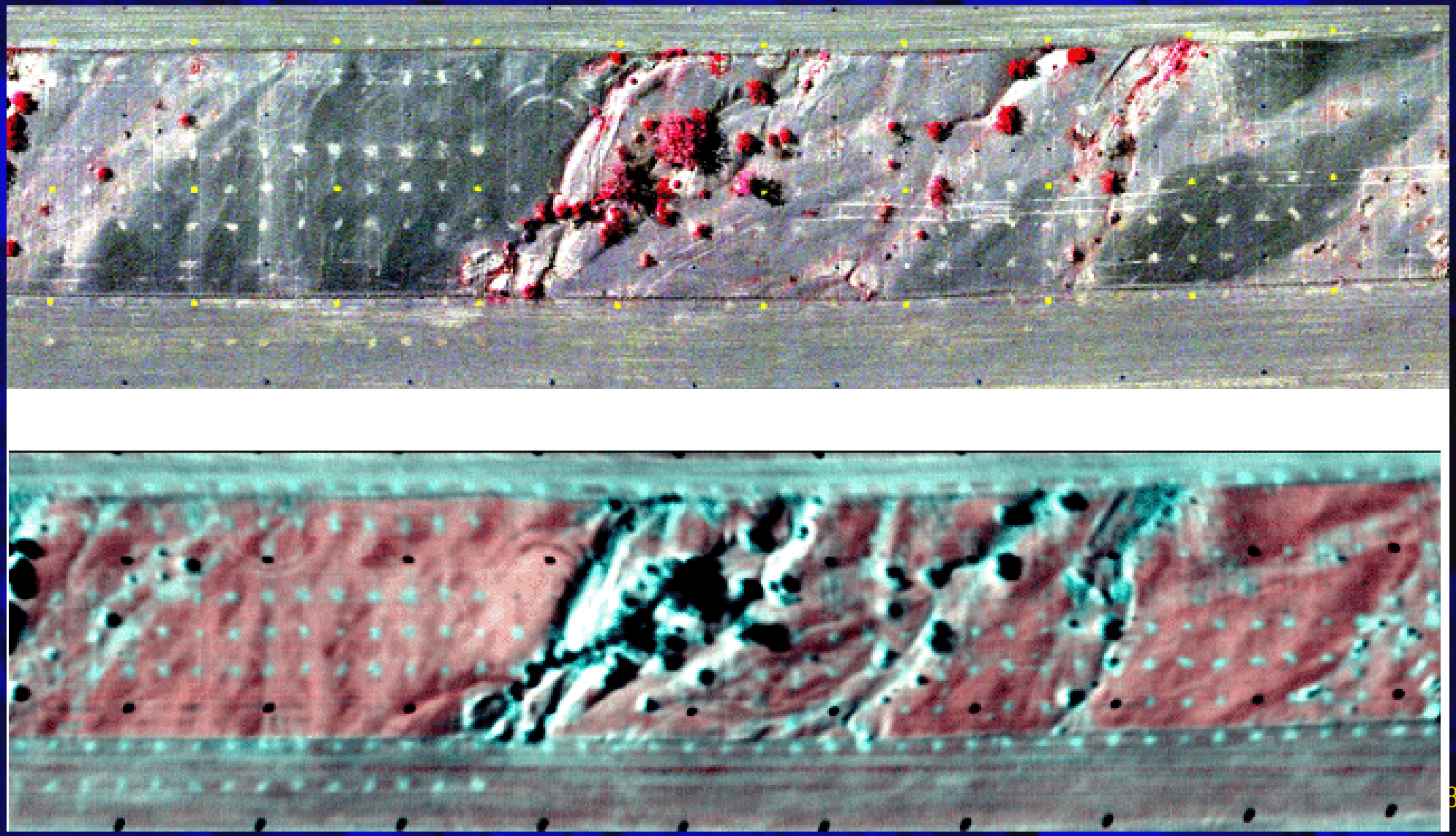
AHI Spectral Technology Group

Airborne Hyperspectral Imager University of Hawaii

Counter Mine

FY7_091520 (1400 ft)

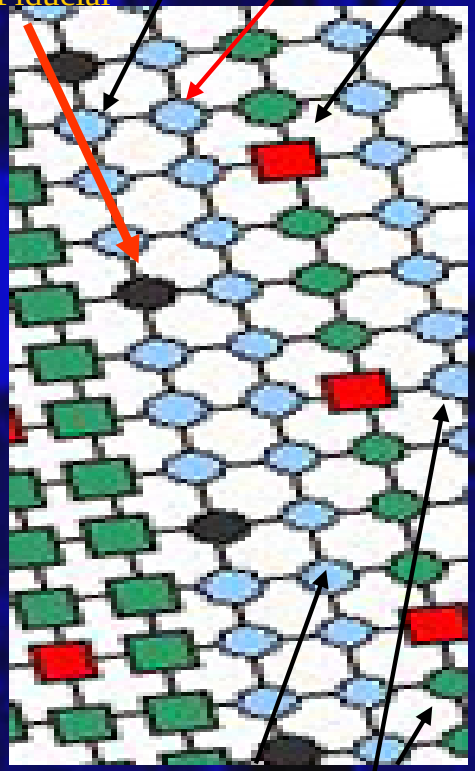
AHI Line Scanner False Color IR



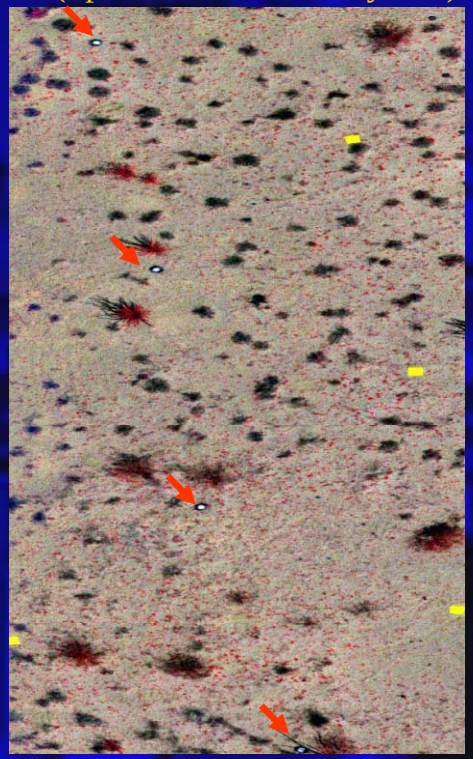
AHI Spectral Technology Group
Airborne Hyperspectral Imager University of Hawaii

Detection of Buried Mines in High Clutter

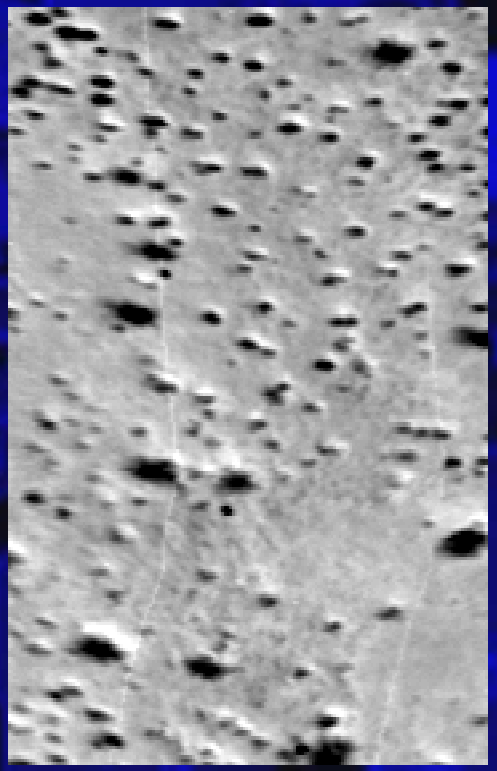
Map of Mine Locations



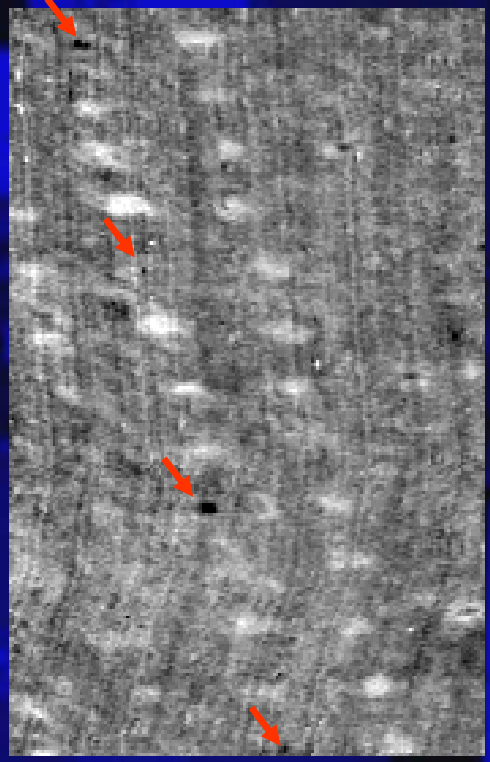
AHI Color Linescanner (Spatial resolution 1 by 3 in)



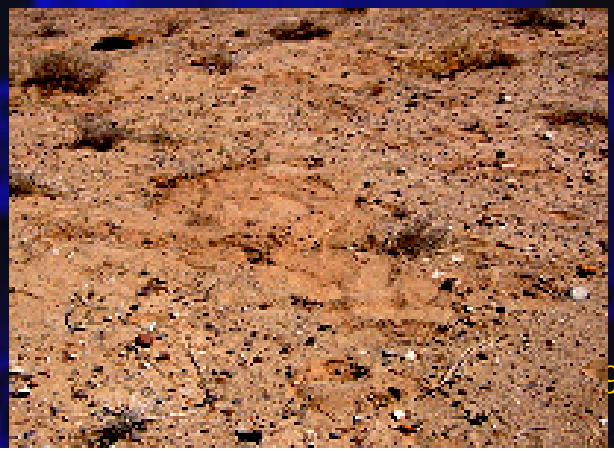
AHI Broad Band IR



AHI Disturbed Soil Fraction Plane



Example from Yuma Yellow Sands



Very Difficult to See Mines in Line Scanner Image
Or Even Close up Photograph

FY8_122646 700 ft

Forest Fusion II Data Collection

- Test location: Fort Leonard Wood, MO
- Dates: 26 Jul – 6 Aug 04
- Sensors
  - LYNX Ku-band SAR
  - Mirage EAGLE GP-SAR
  - COMPASS V-SWIR HSI + Linescanner
  - AHI LWIR HSI + Linescanner
  - HYLITE high altitude LWIR HSI
  - STI VNIR HSI (new cutoff 940 nm) [delayed start]
- Targets
  - 150 large plastic (M19) [Fill?!]
  - 450 large metal (M20) (painted carc- green & brown, filled with playground sand)
  - 90+ small metal (RAAM) (painted non-military green)
  - ~90 medium plastic (VS1.6)
  - 150 disturbed earth “confusors”

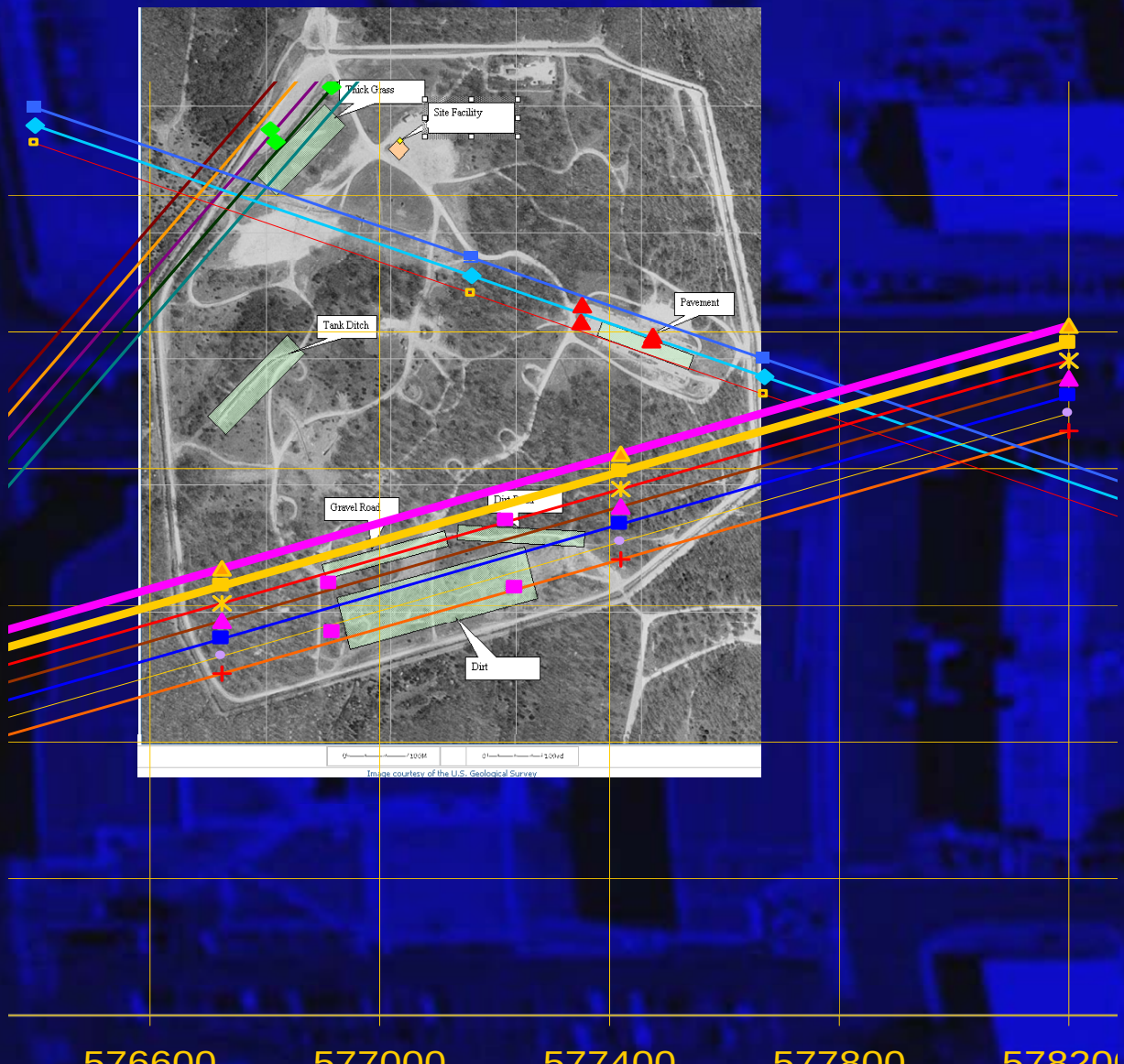
– IEDs (150 mm shells filled with cement, etc, variety of deployments (plastic bags, rock piles, buried wires, wires on surface))

– Concertina wire

– 20 Jersey Barriers

– Tank ditch

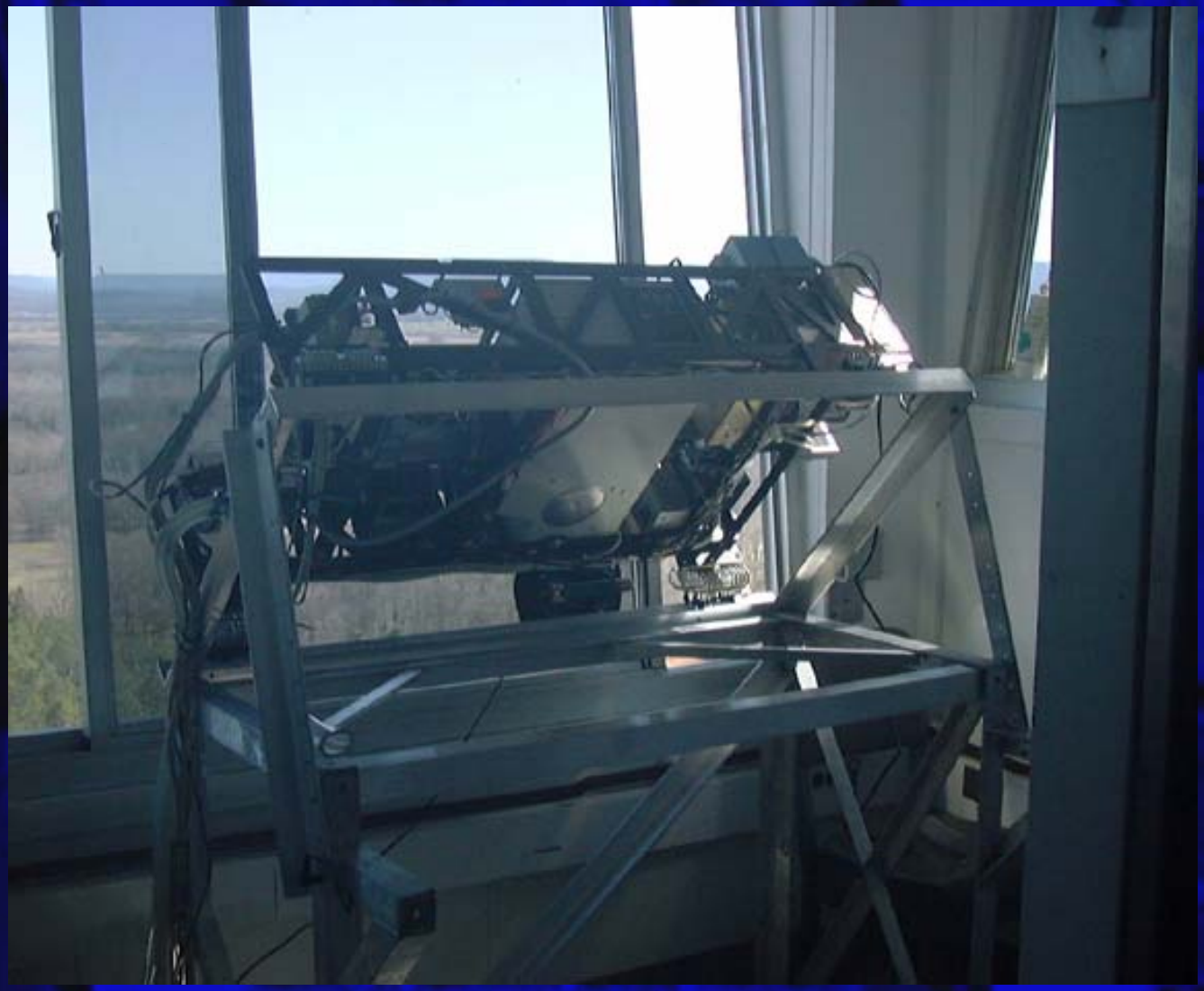
Flight Plans on FLH Map



Tower Test

- AHI Tower Deployment to Support:
  - Land Mine Phenomenology
  - Input to Algorithm Efforts
- Scheduled for Fall 2005

AHI Ground Configuration

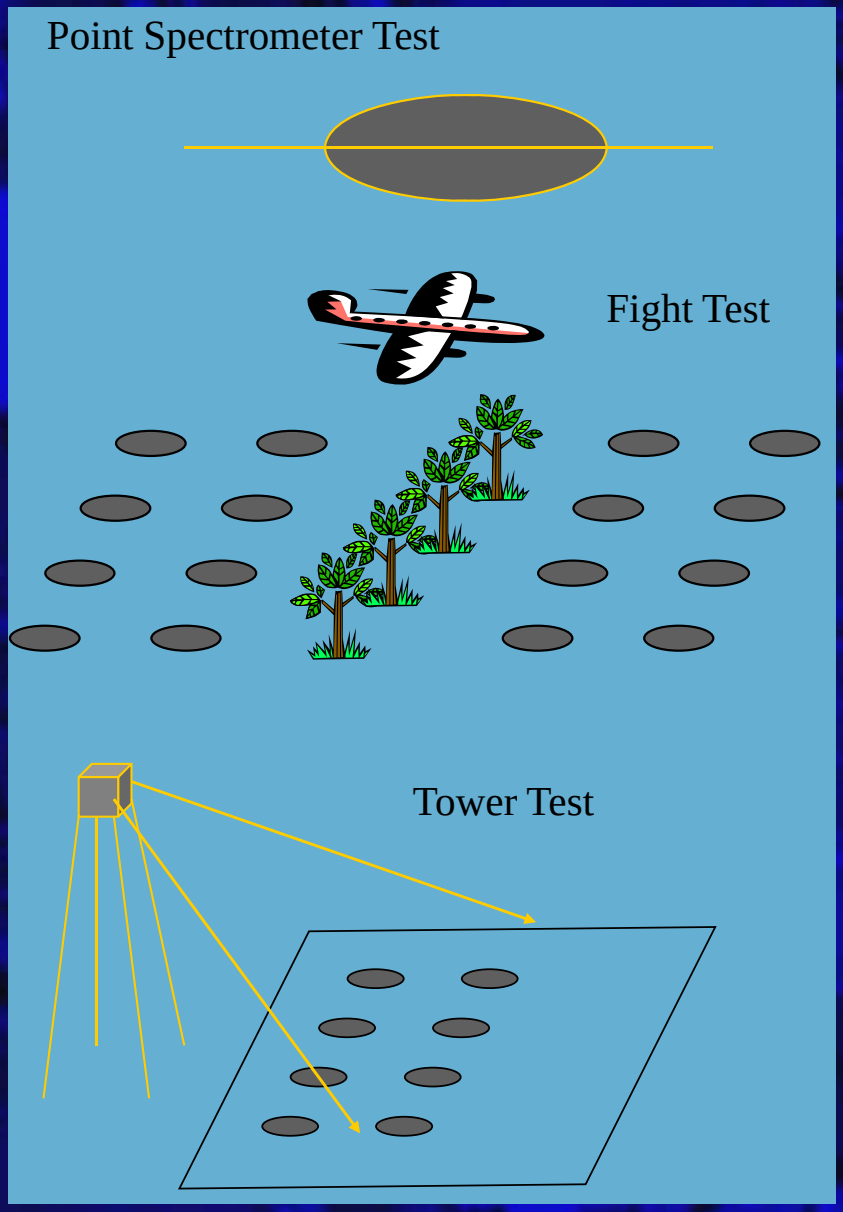


Tower Test Goals

- Purpose:
  - Fill in the Gaps between Buried Mine Point Spectrometer Measurements (ERDC) and Airborne Imaging Spectrometer Measurements (WAAMD). Collect Hyperspectral Polarization Data

- Goals:
  - Study the Diurnal Variation of the Full Spatial Spectral Signature of Buried Mines and Empty Holes
  - Acquire Data for Verification of Buried Mine Modeling
  - Acquire Data on Short-Term Weathering of Buried Mines
  - Investigate Potential Disturbed Soil Signature Counter-measures
  - Acquire Detailed Spatial-Spectral Structure on Buried Mines

Role of Different Tests



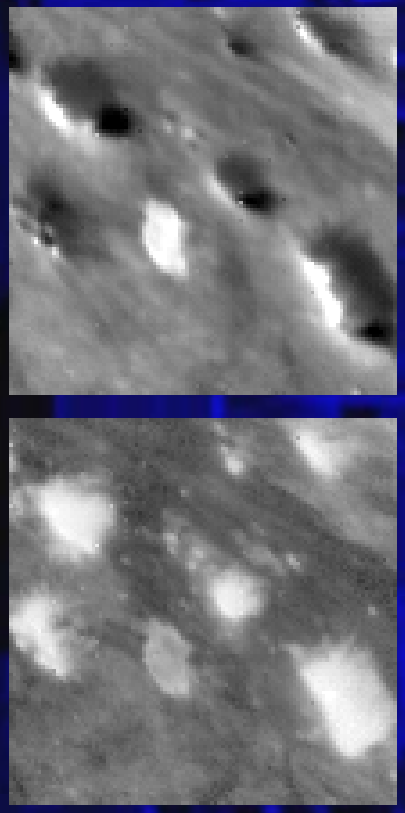
- Point Spectrometer test to study Effect over long Periods of Time
- Very Limited Spatial (single points or a line through a disturbed soil scar)

- Flight Test Acquires Large Quantity of data over Both Minefields and Clutter
- Very Limited Temporal Sampling

- Tower Test Provides the Ability to Study Several Mini-Minefields over the Time of the Deployment
- High Spatial Resolution
- Multiple Times of Day
- Change Detection

Diurnal Measurements

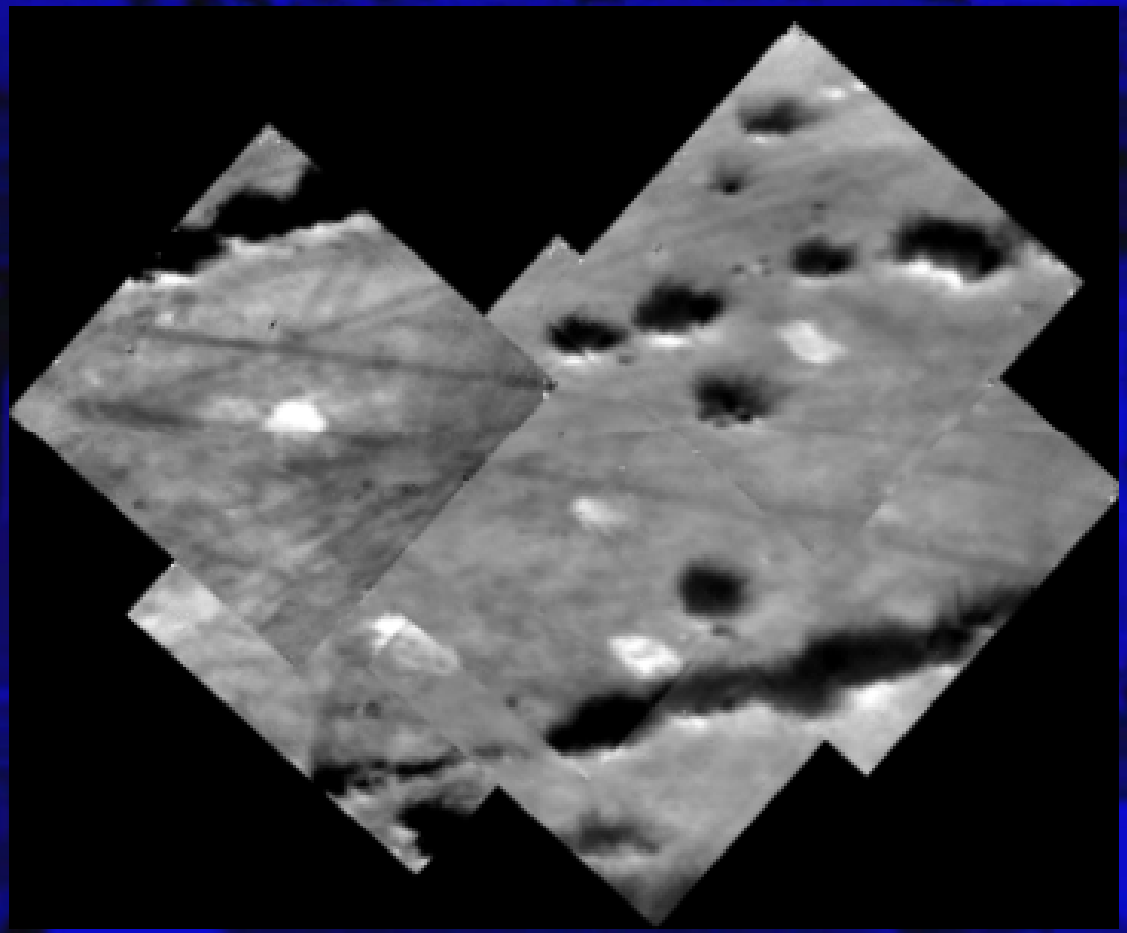
Comparison of Day Image
and Night Image (Mine in Center)



- Tower Provides the Opportunity to measure Mines at many Different Times of Days
- Variation with Cloud Cover
- Make measurements at:
  - Pre-dawn
  - Hour past Dawn
  - Near Mid-Day
  - Hour after Sunset

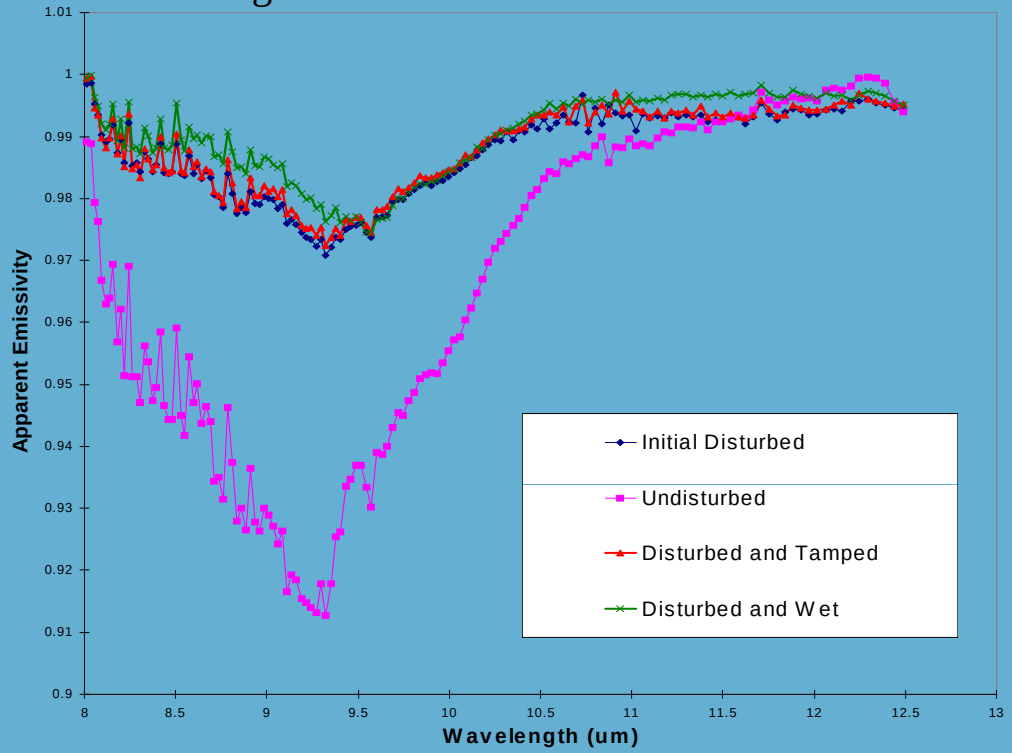
Diurnal Measurements (cont)

Day Mosaic of Fresh Mine Field



Countermeasures

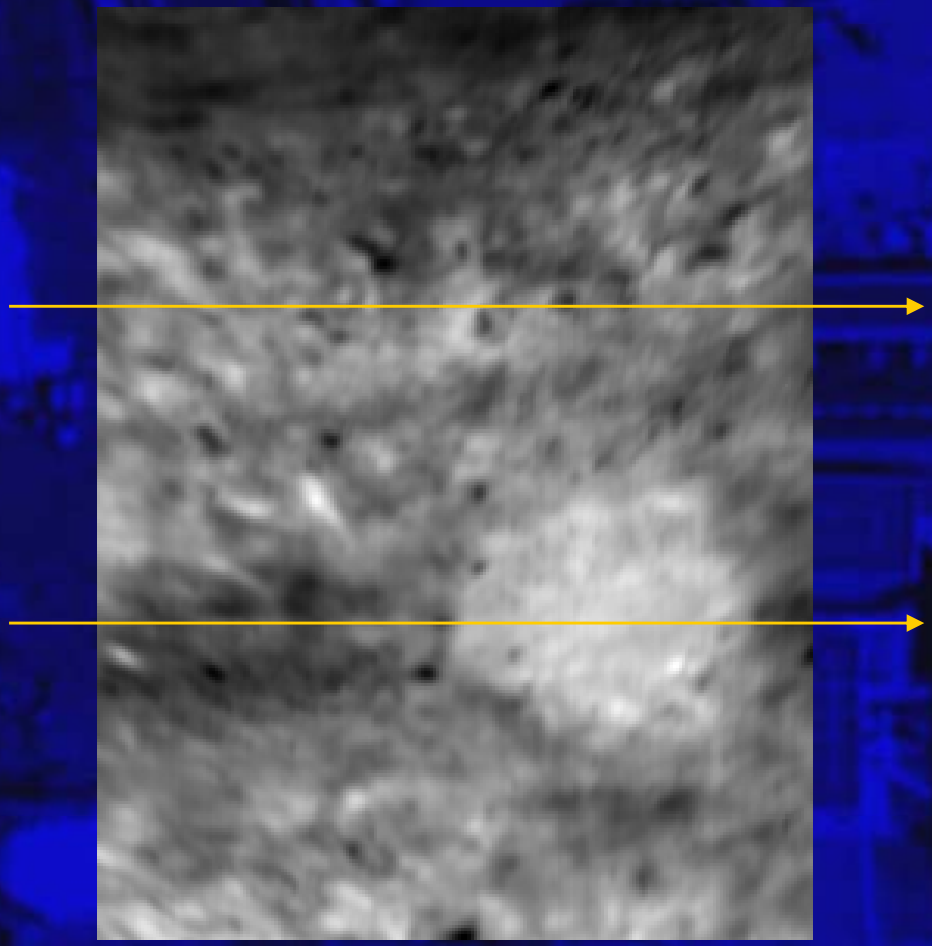
Neither Sprinkling Water or Tamping Removed Signature

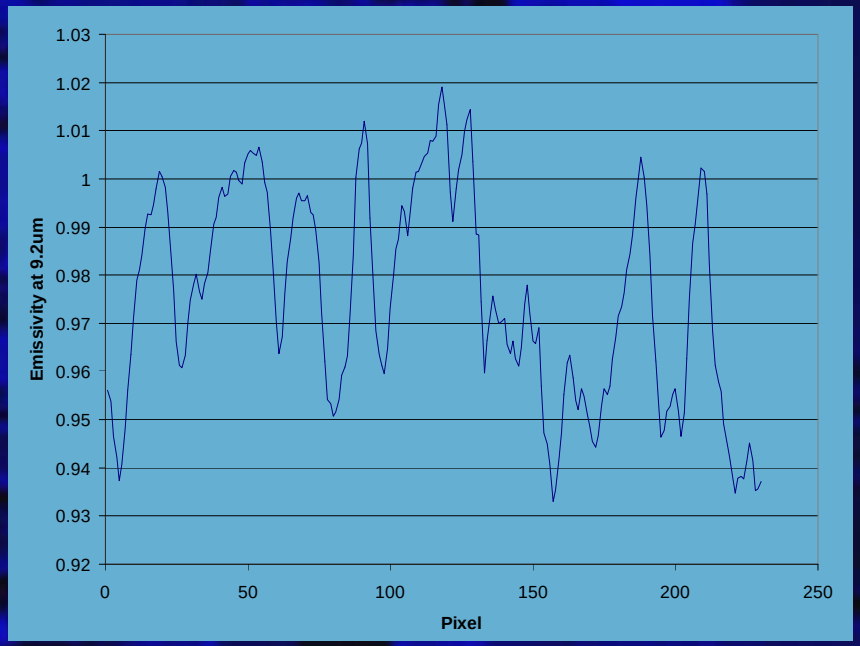


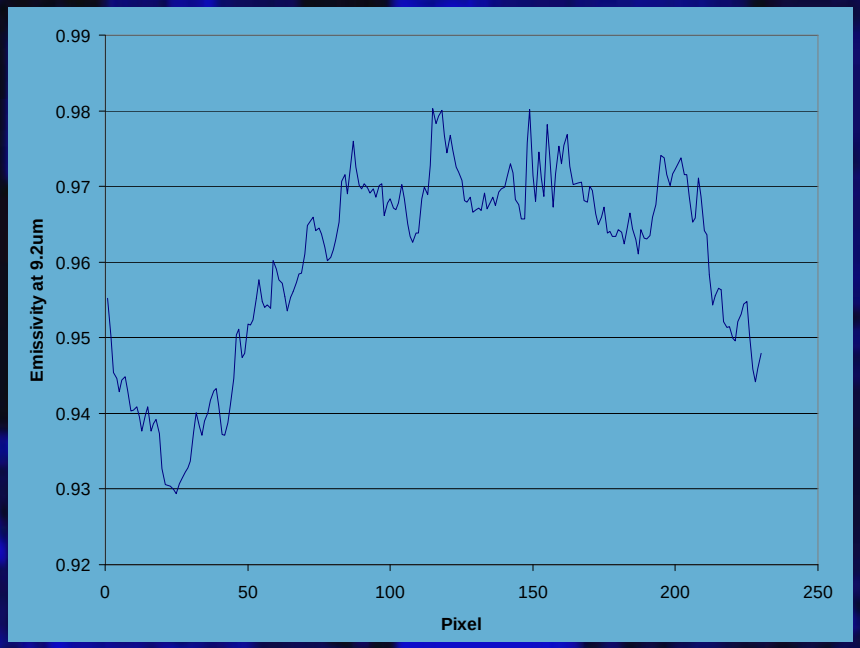
- Measurements in 1995 and 1996 Showed that Suppressing Disturbed Soil Signature is Very Difficult
  - Tamping (boot or Board) had Little Effect
  - Sprinkling Top Soil Only Left More Disturbed Soil
  - Sprinkling Water had Little Effect
  - Full heavy Washing with Water, While removing Emissivity Difference, leaves very obvious Visual Scar
- Opportunity to Conduct Controlled Experiment from Tower
  - Tamping
  - Cover with Surface Dust
  - Cover with leaves
  - Water

Detailed Spatial Spectral Analysis

Transects Through Tower Acquired AHI Image







Tower Test Plan

- Place Mines near Tower Beginning One Month before test
  - One Month
  - Two Weeks
- Deploy Mines in Several Mini-Minefields Around base of Tower
  - Diurnal fresh Mines (mines Put in During test)
  - Aging Mines (One Month, Two Weeks and at Beginning of test)
  - Counter-measures Area
  - Modeling Support Area
    - Possibly including imported Soil Such as sand
    - Inputs from MURI Team
- Two Week Deployment of AHI with D&P Support

Candidate Test Sites

- Ft. Hunter Liggett
- Ft. Huachuca
- Ft. A.P. Hill
- Ft. Belvoir
- Redstone Arsenal

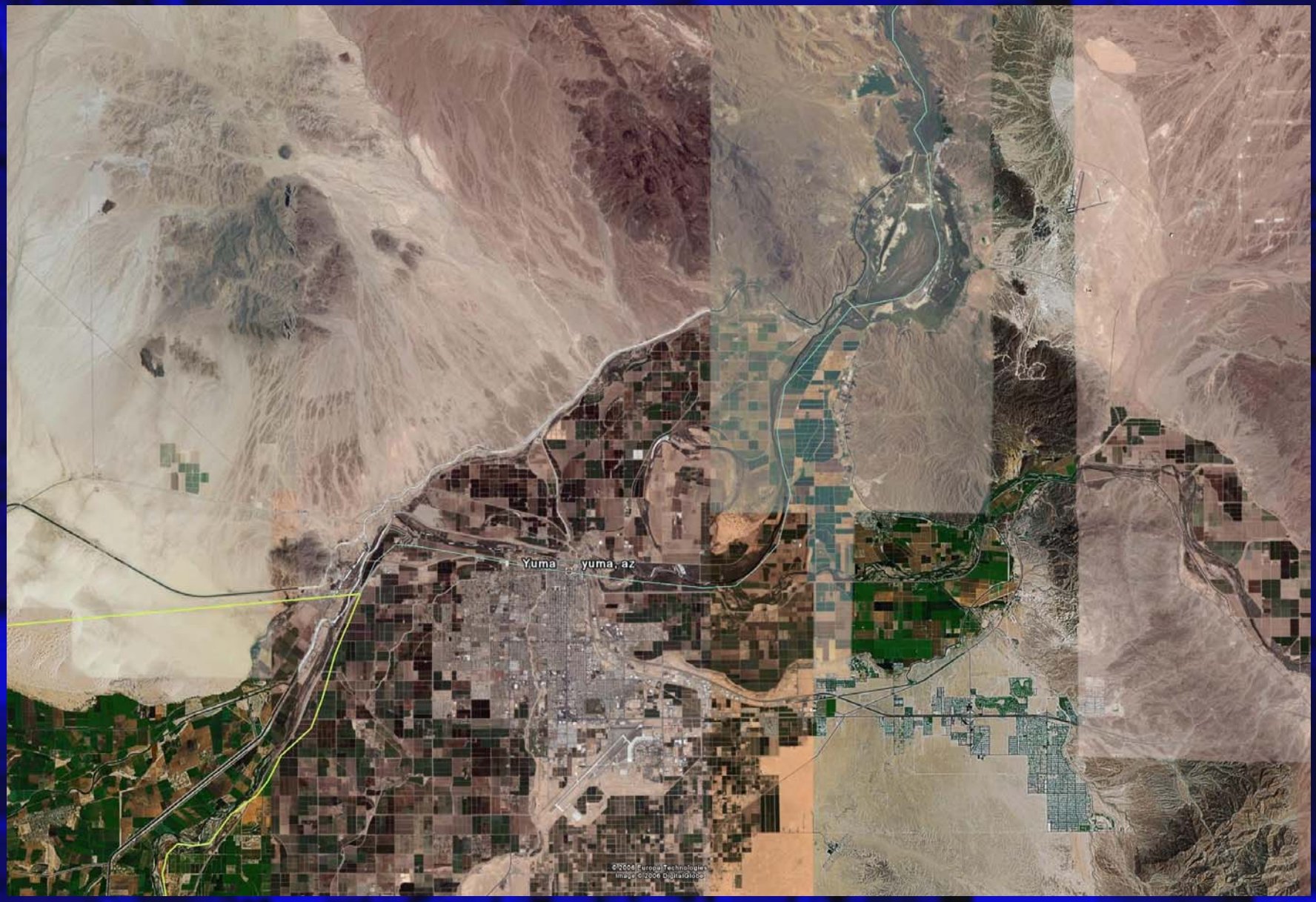
Yuma Data Collection

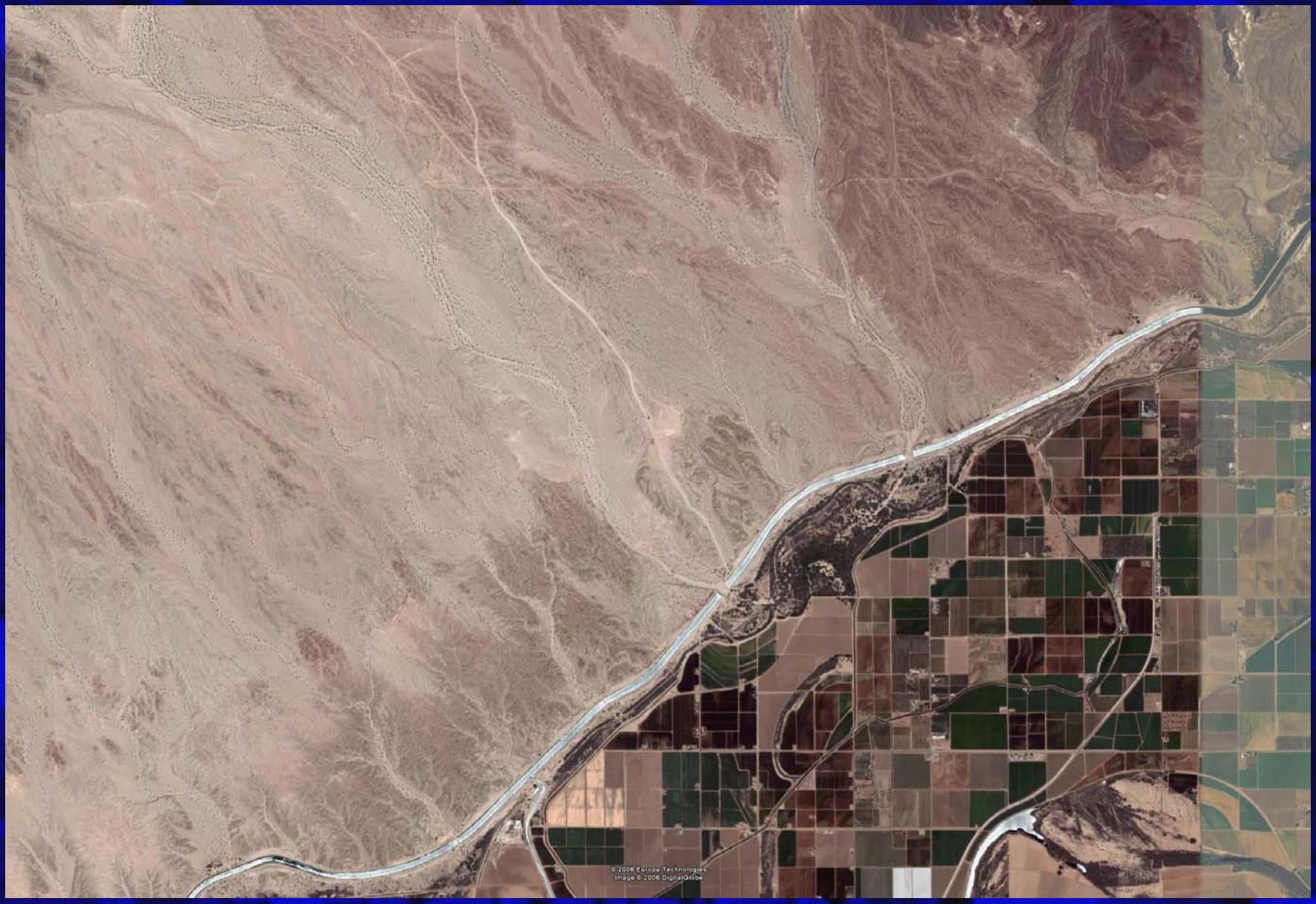
Early February '07

- Change Detection Series
- IED Simulation
- Reststrahlen Confirmation
- Diurnal Comparison
- Algorithm Input
- Yuma Proving Grounds JERC Site

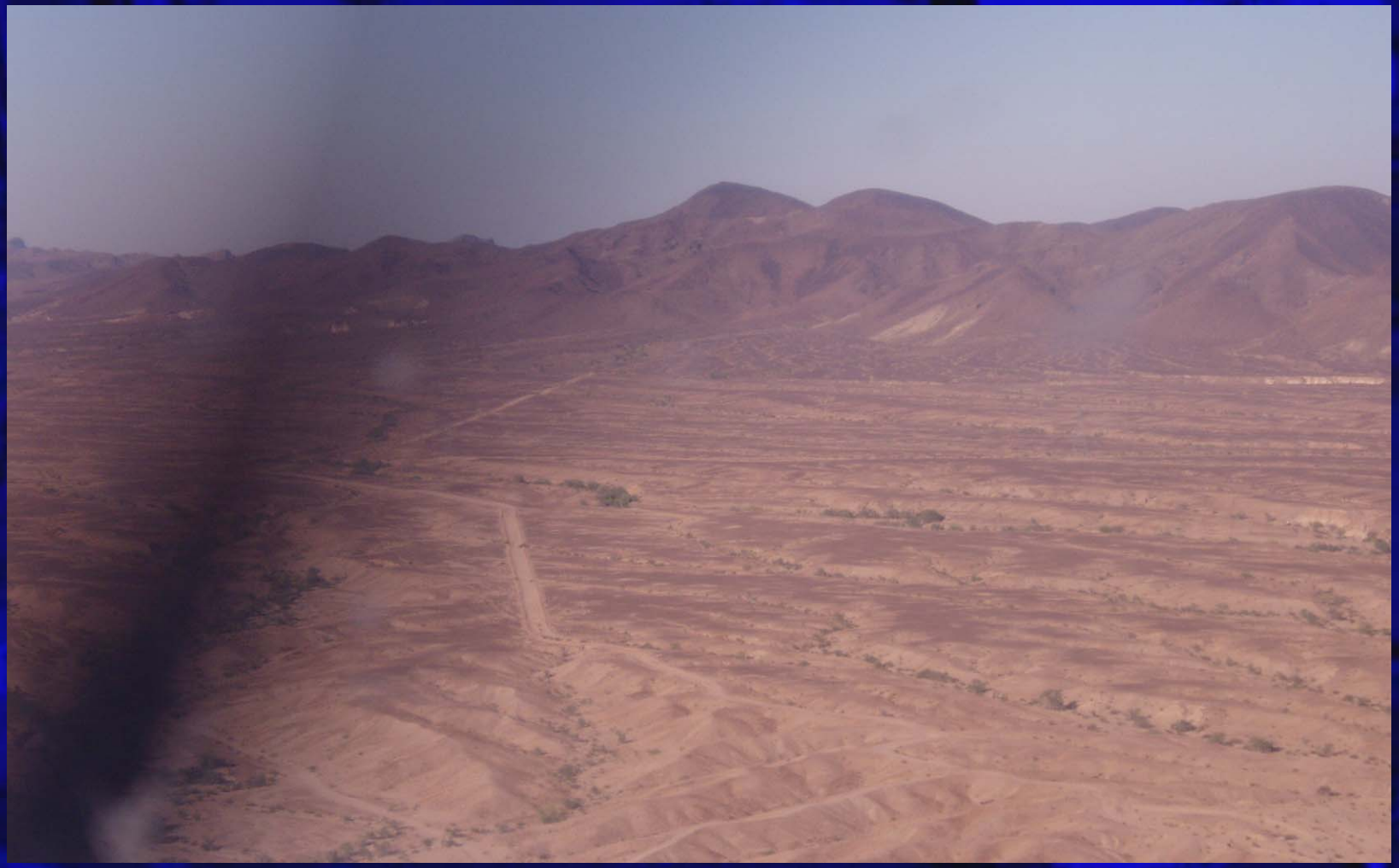
AHI Spectral Technology Group

Airborne Hyperspectral Imager University of Hawaii




© 2006 Europa Technologies
Image © 2006 DigitalGlobe

Proposed Test Area



Aero Commander

Change Detection

- Undisturbed Background Data
  - On and Off Road
- Scene Changes Daily
  - Natural and Man Made materials
  - Simulated Mine/IED Burials

IED Simulation

- 155 Simulants Buried in Road/Shoulder
- Surface Emplacements
- Wires/Wire Trenches
- Surface Obscured Objects

Reststrahlen Tests

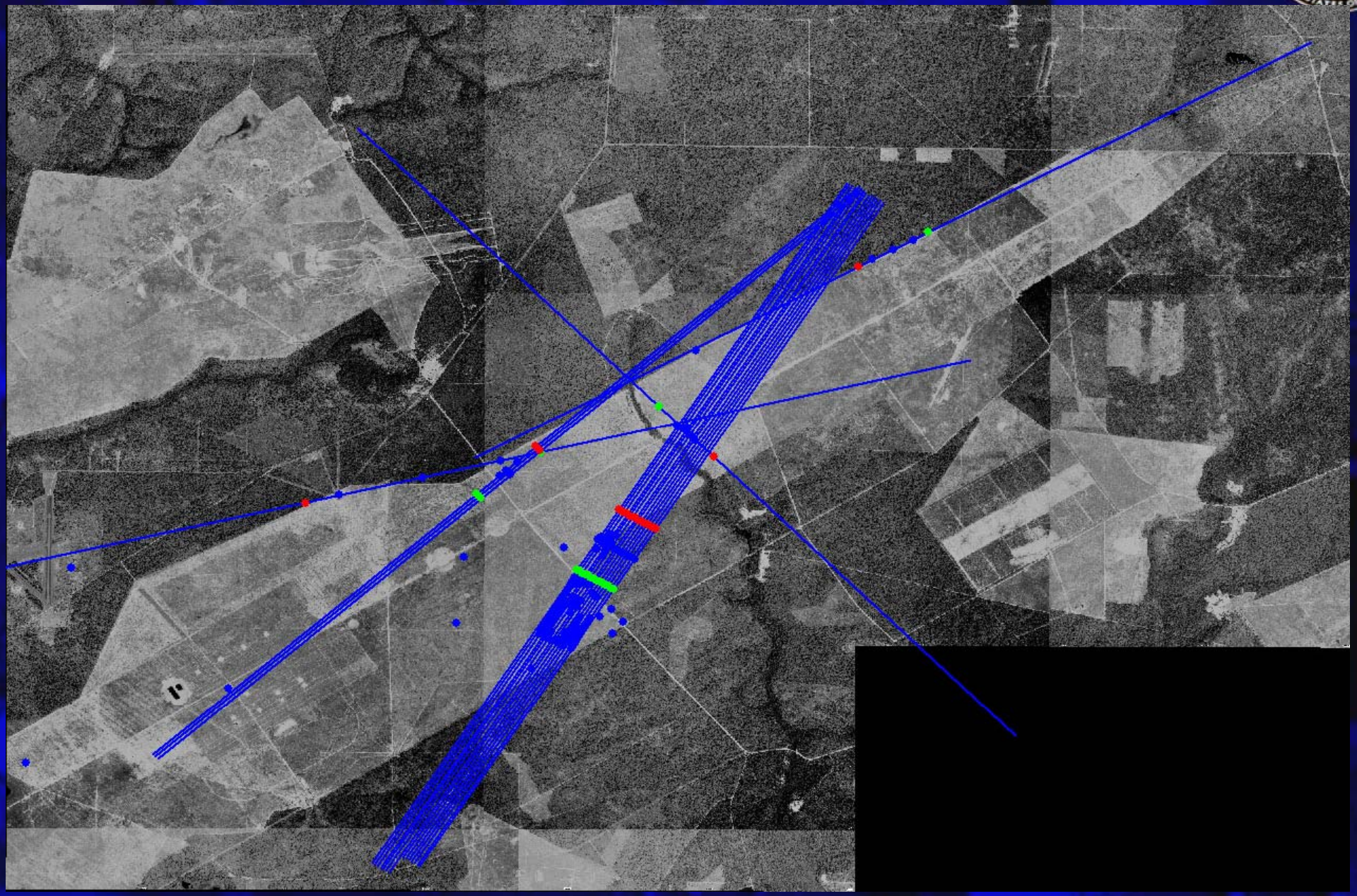
- Goals:
  - Study the Diurnal Variation of the Full Spatial Spectral Signature of Simulated Buried Mines/IEDs
  - Acquire Data for Verification of Disturbed Earth Modeling
  - Acquire Data on Short-Term Weathering of Disturbed Areas
  - Investigate Potential Disturbed Soil Signature Counter-measures
  - Acquire Detailed Spatial-Spectral Structure on Disturbed Areas

Ground Truth Support

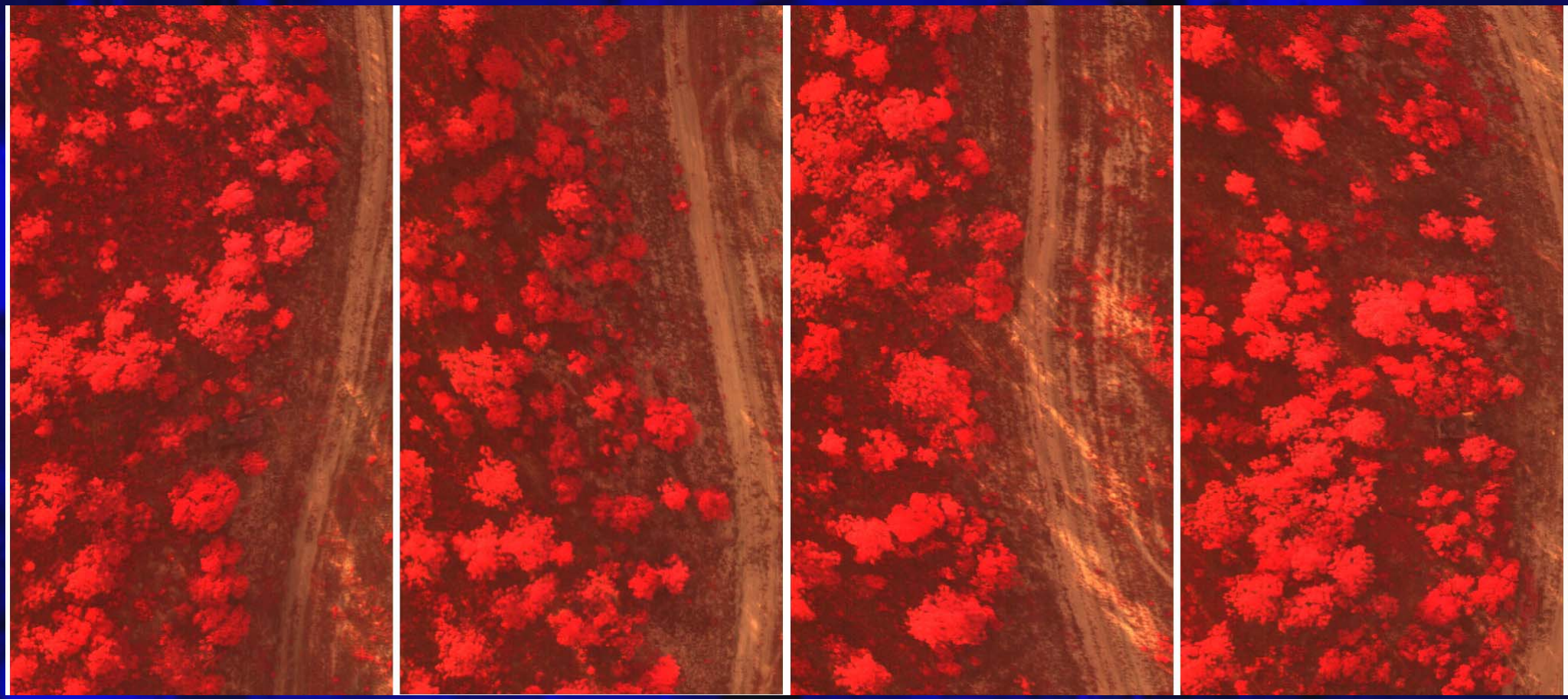
- Designs & Prototypes FTIR
- Photographs All Sites
- GPS Coordinates for all ‘Targets’
- Calibration Panels
- Omega Radiometer Temp Measurements

Sensor Week, Eglin AFB, May 2004

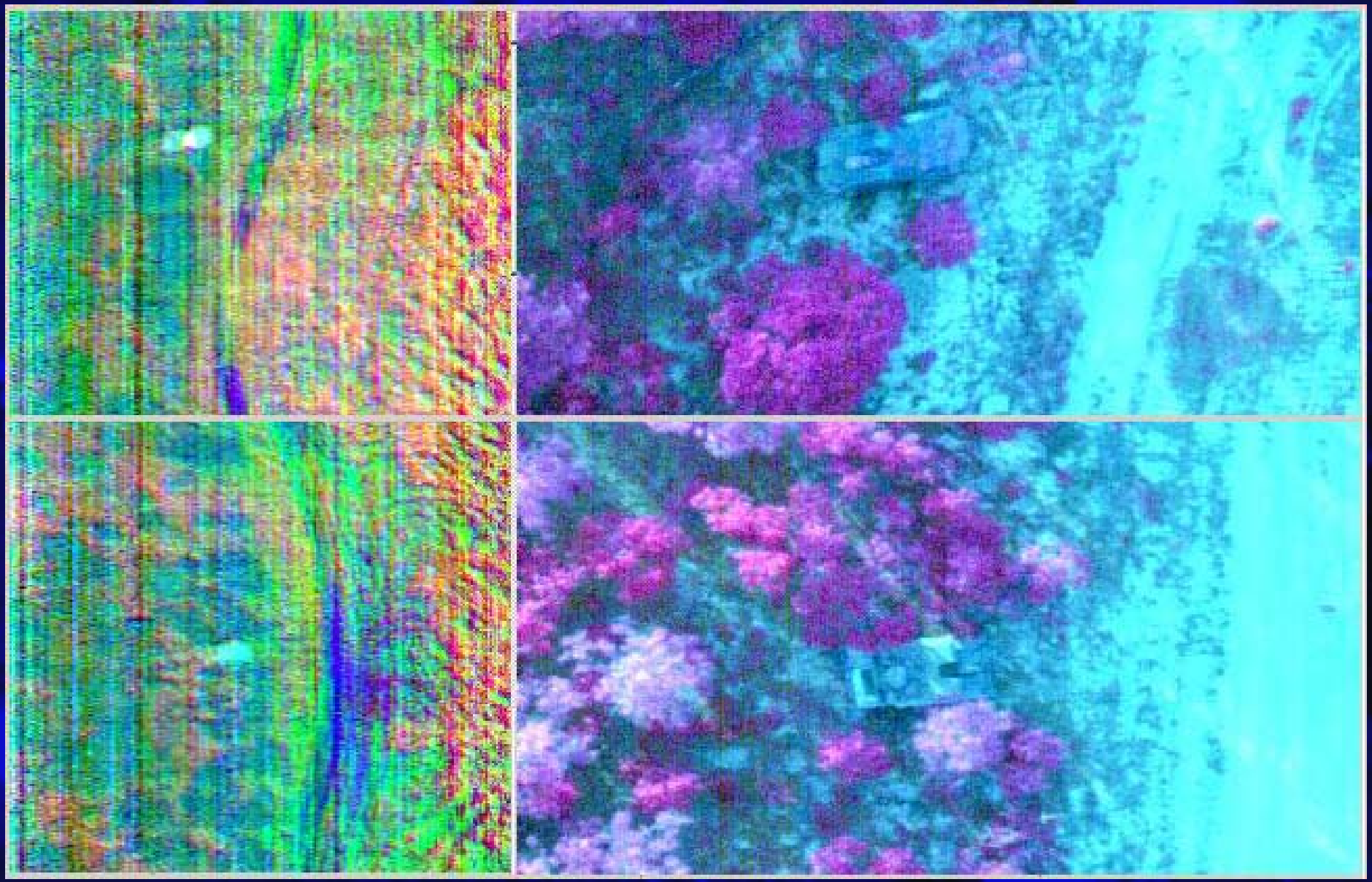
- ~110 Gb data collected
- Day and night operations
- Vehicle and material array targets



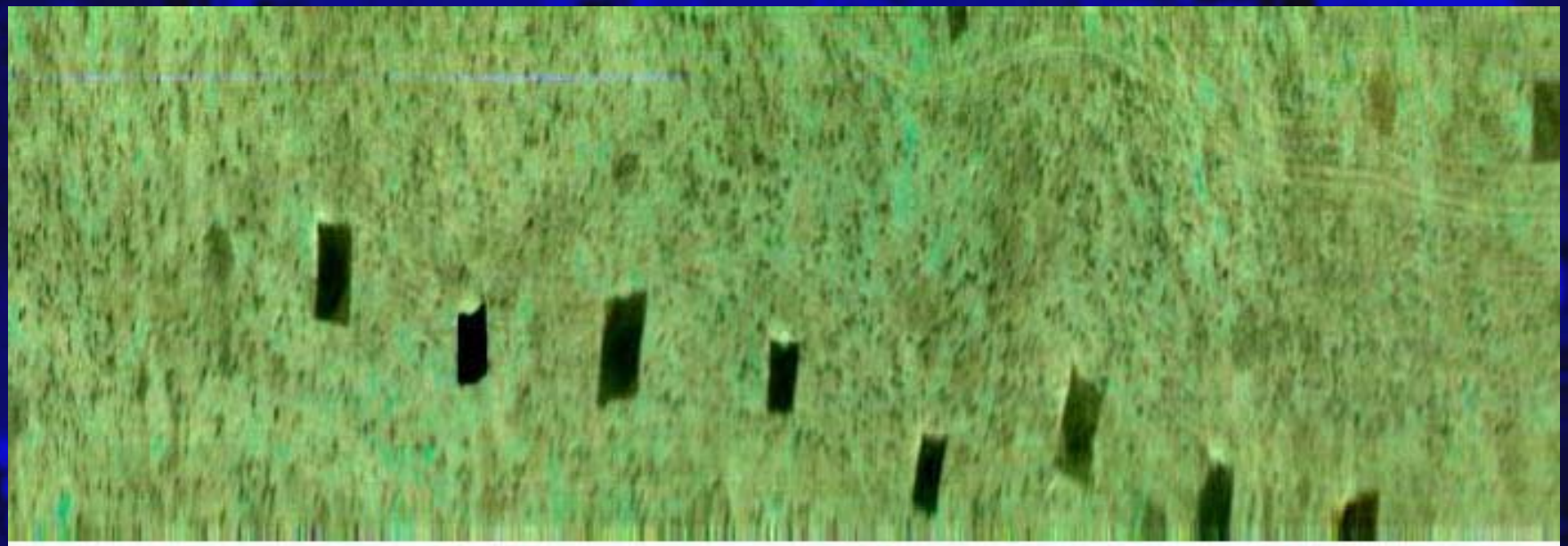
Programmed Flight Lines



Find the Vehicles

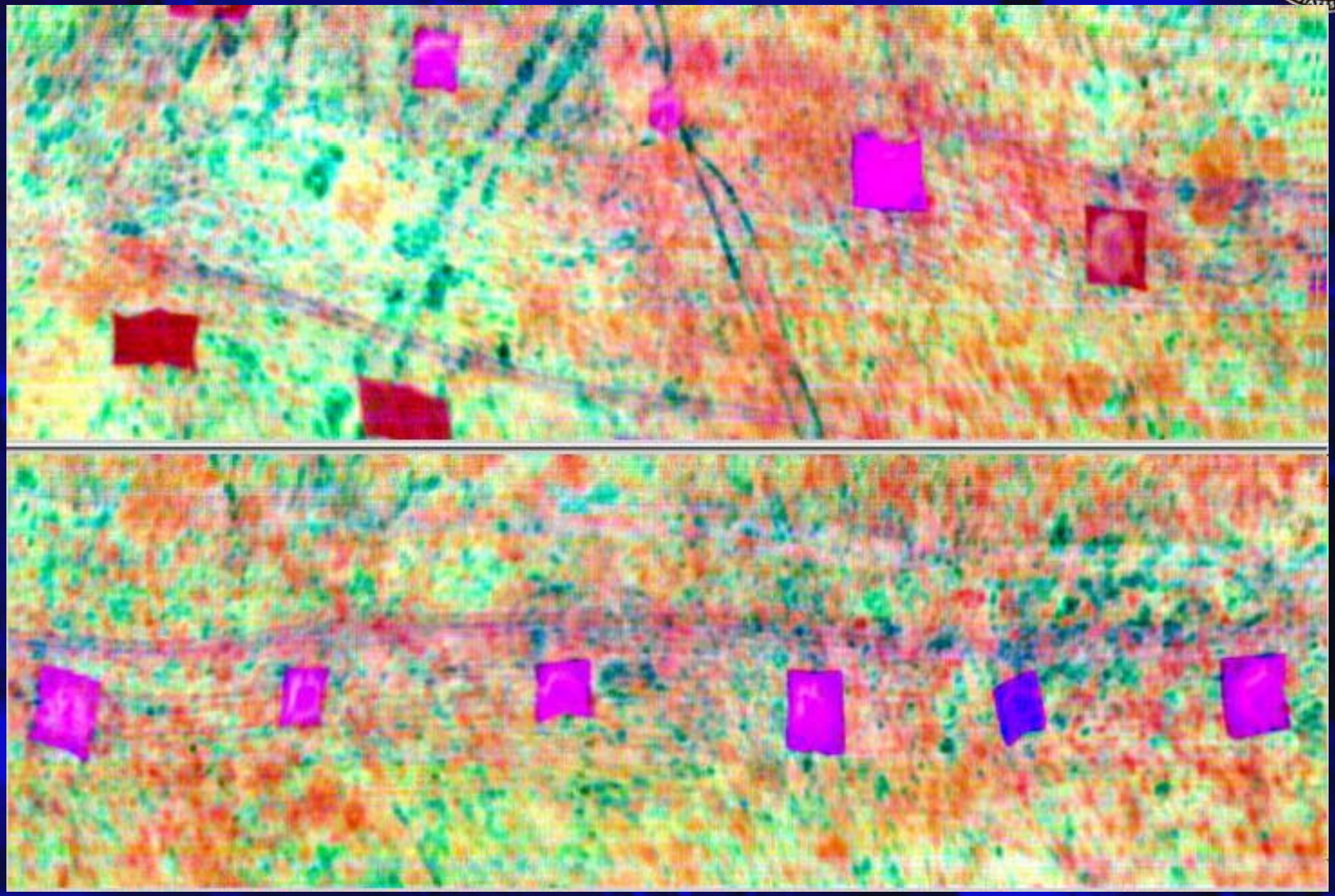


Principal components VIS / NIR

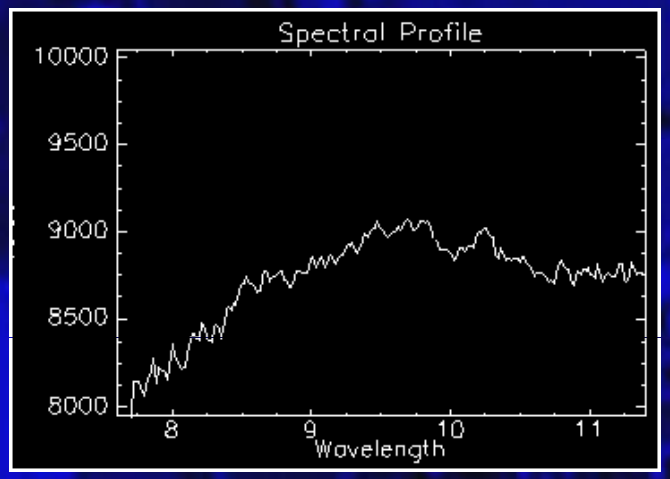


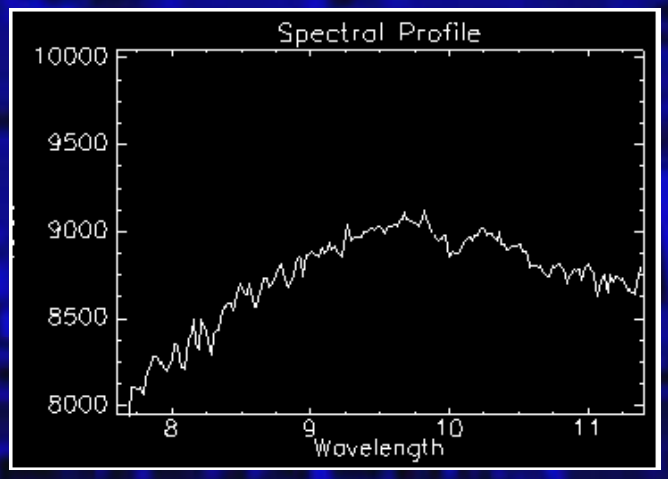
IR

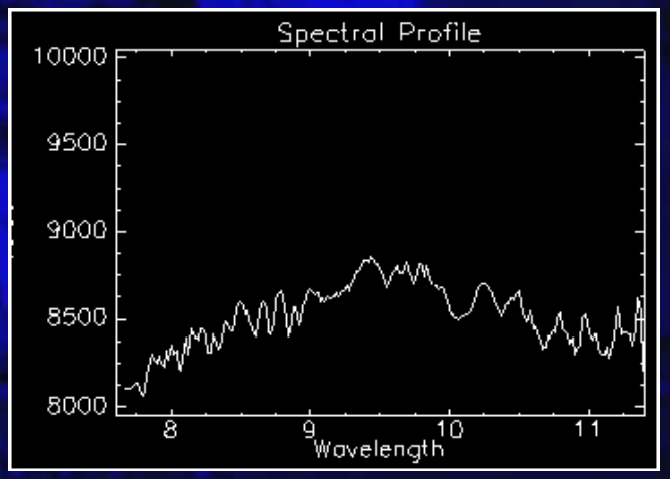
VIS/
NIR

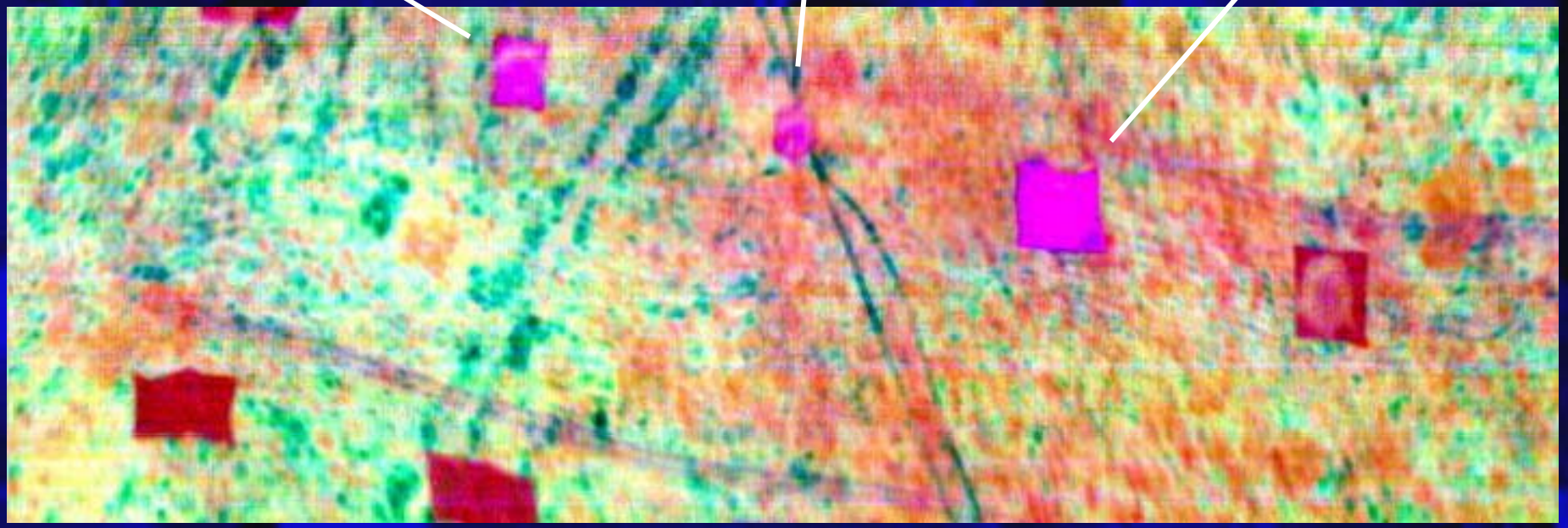


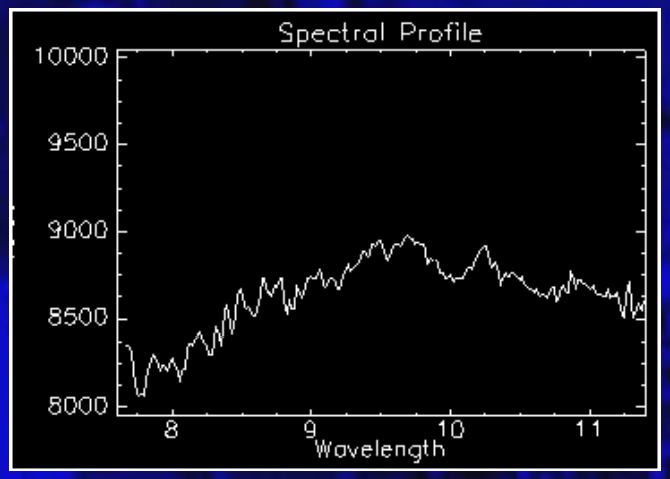
Principal components

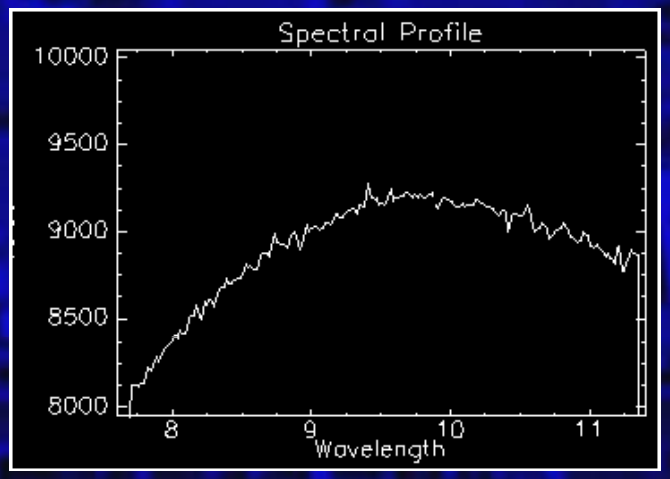

Spectral Profile
10000
9500
9000
8500
8000
8
9
10
11
Wavelength

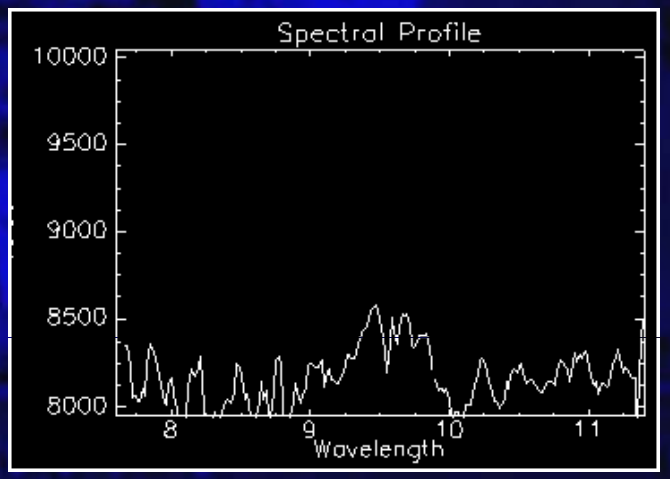

Spectral Profile
10000
9500
9000
8500
8000
8
9
10
11
Wavelength


Spectral Profile
10000
9500
9000
8500
8000
8
9
10
11
Wavelength

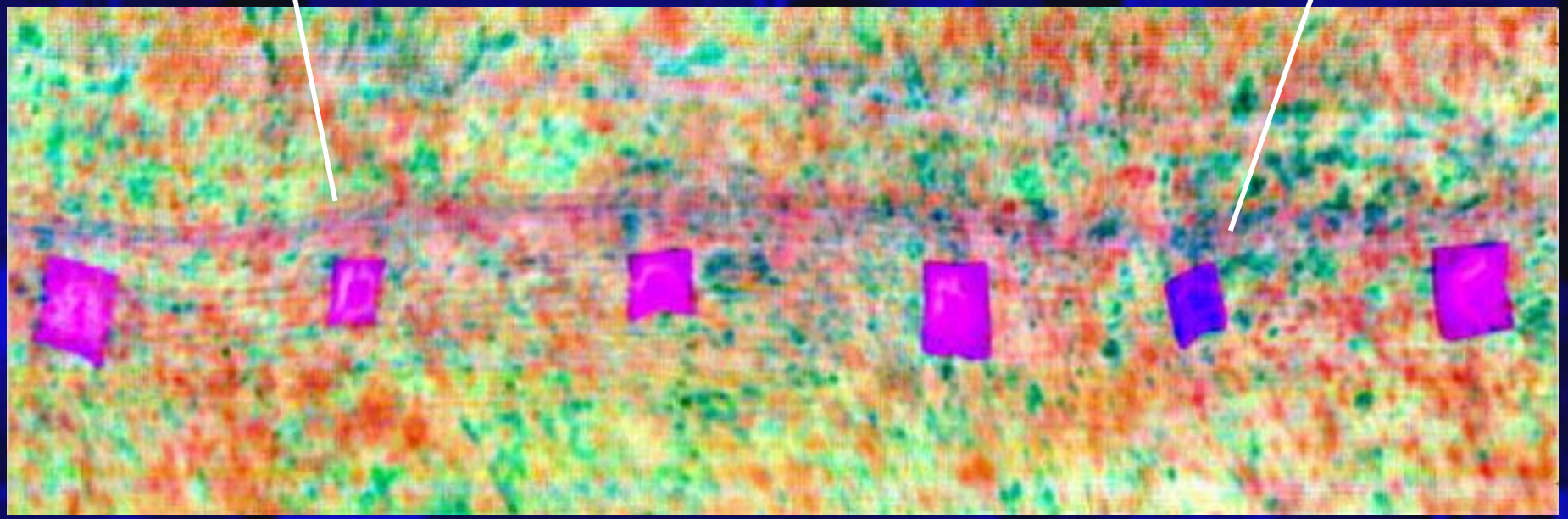



Spectral Profile
10000
9500
9000
8500
8000
8
9
10
11
Wavelength


Spectral Profile
10000
9500
9000
8500
8000
8
9
10
11
Wavelength

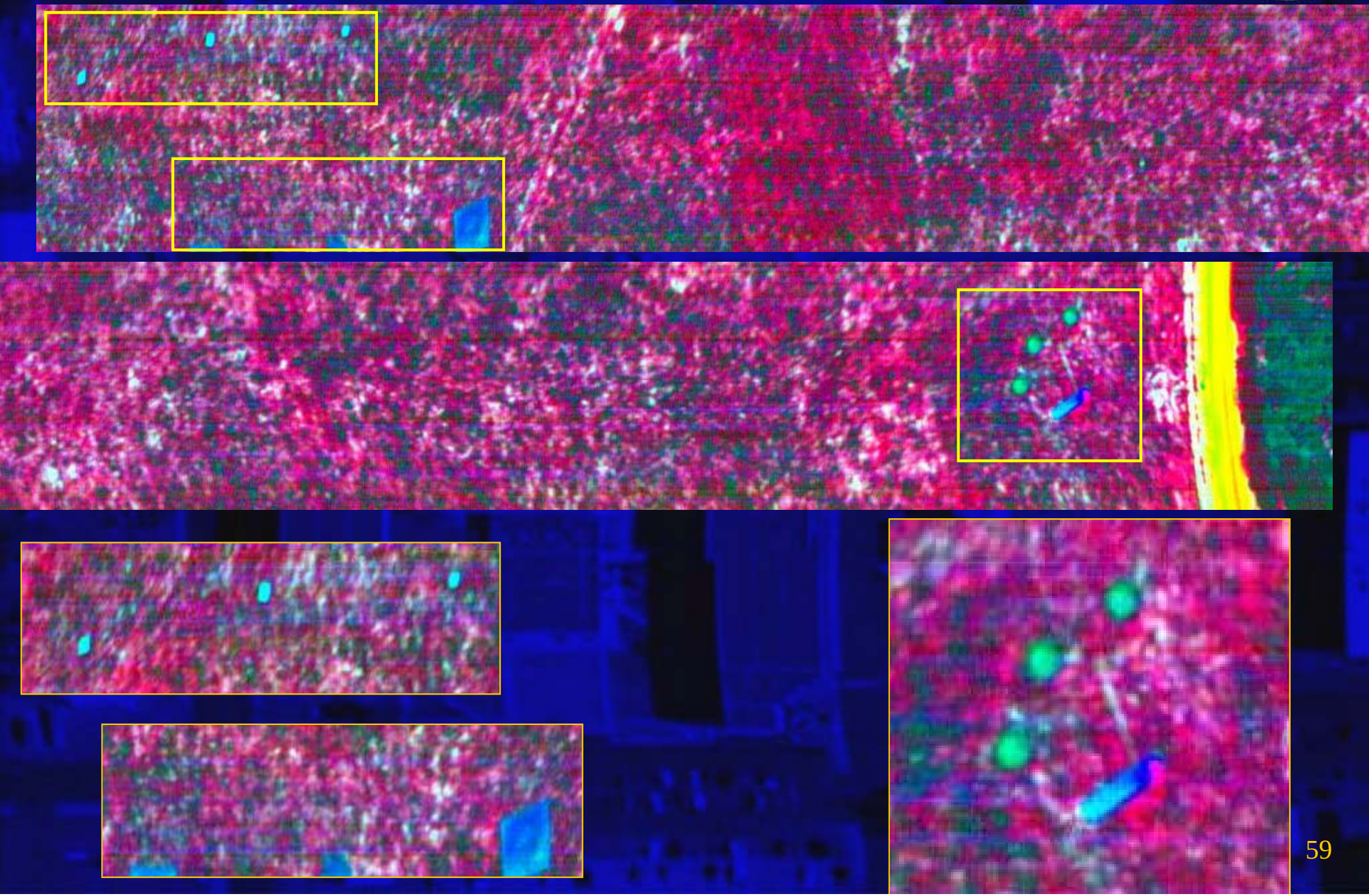

Spectral Profile
10000
9500
9000
8500
8000
8
9
10
11
Wavelength

Background



AHI Spectral Technology Group
Airborne Hyperspectral Imager University of Hawaii

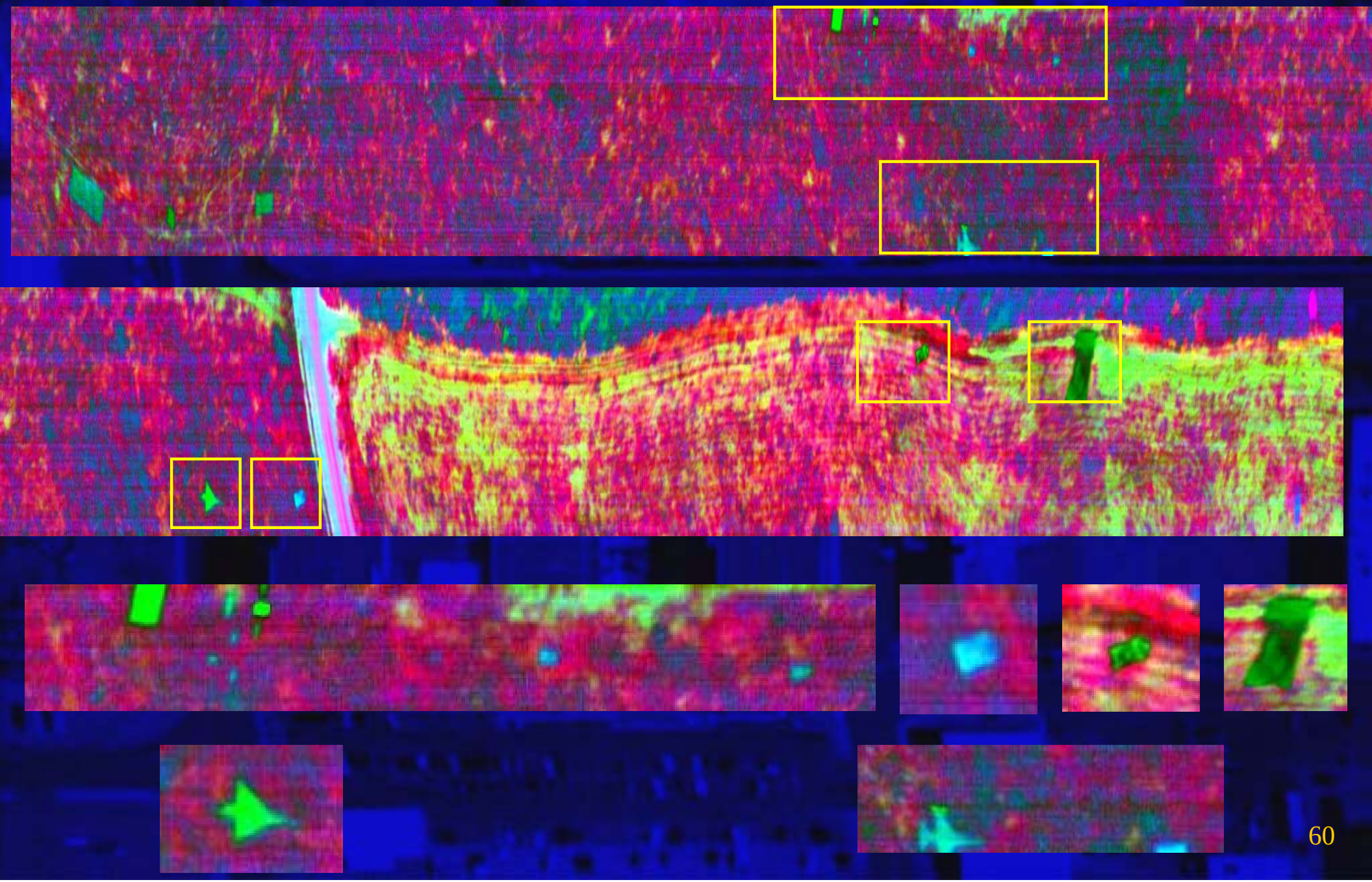
LWIR Anomaly Detection



AHI Spectral Technology Group

Airborne Hyperspectral Imager University of Hawaii

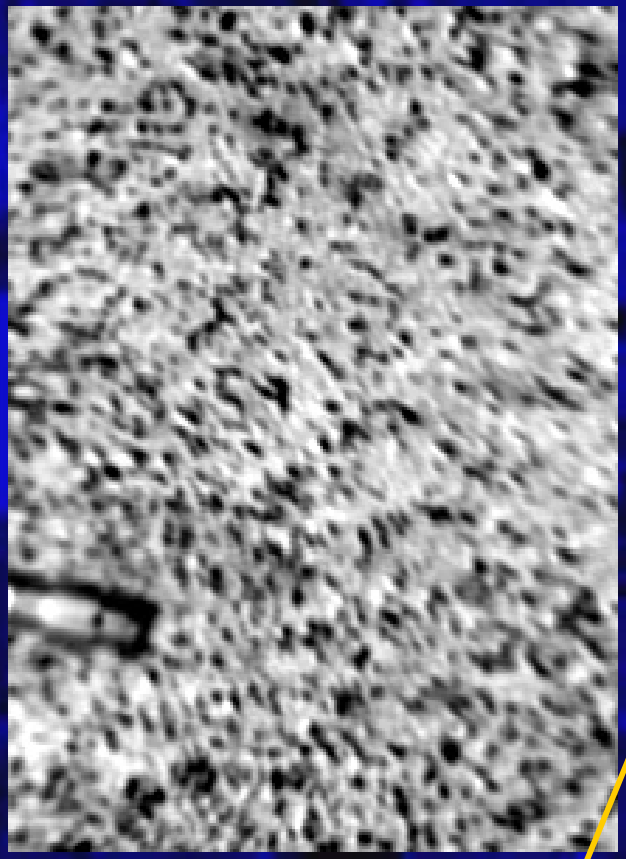
LWIR Anomaly Detection



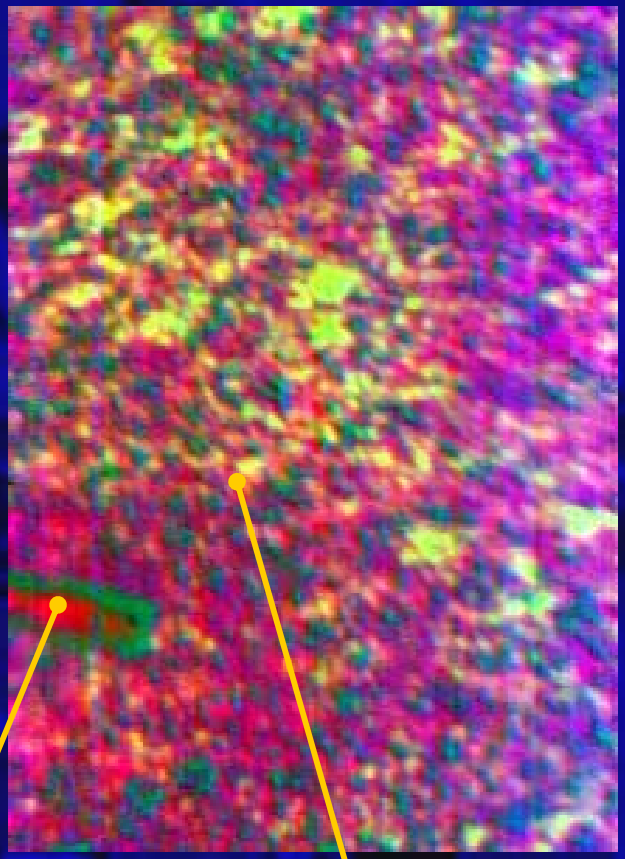
AHI Spectral Technology Group
Airborne Hyperspectral Imager University of Hawaii

eglin_tnk_mat_s300_a2000_r6

Thermal

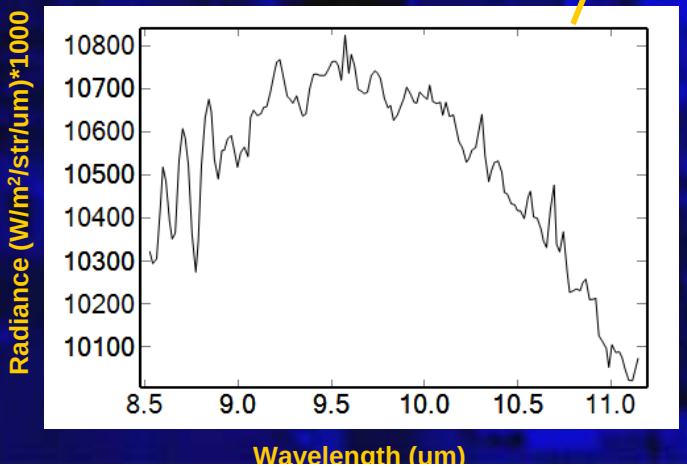


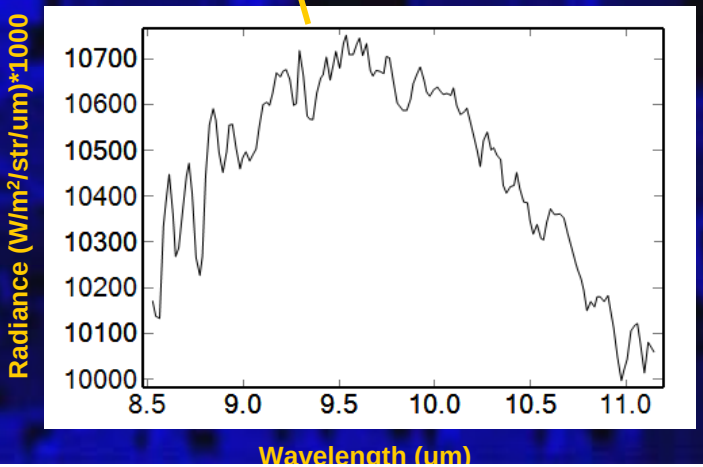
PCA -> RGB

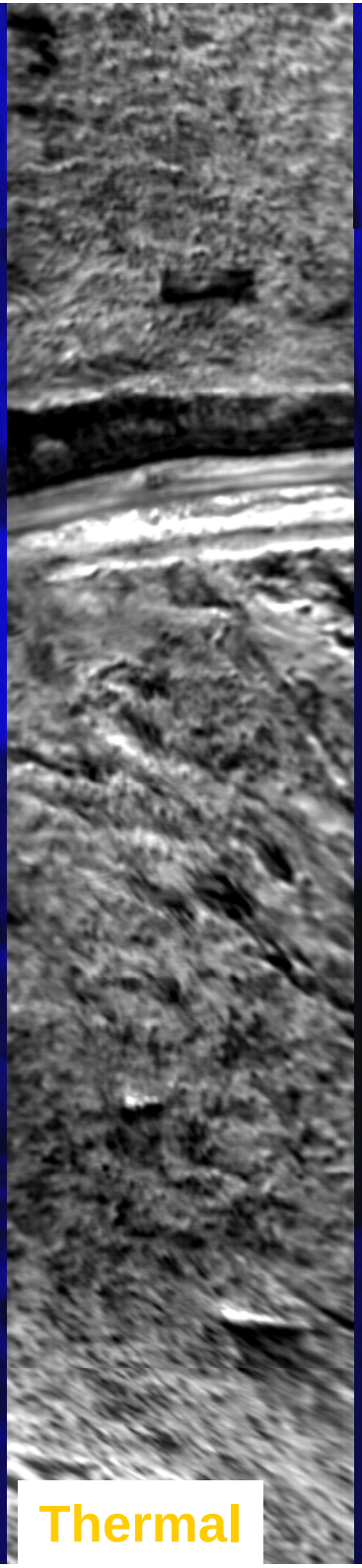


Target appears to be spectrally matched to blackbody thus are difficult to separate from the background

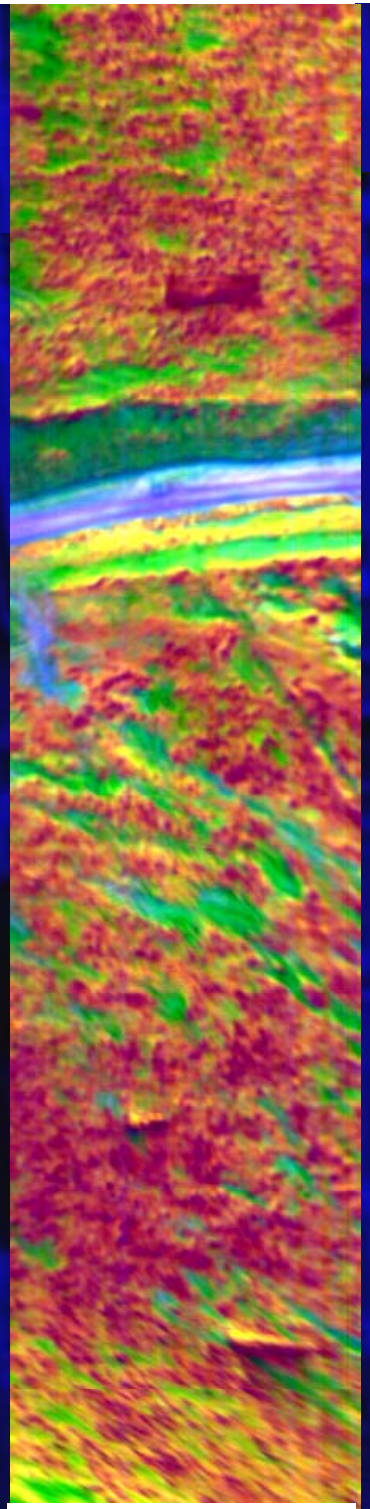
No thermal anomalies, vehicle is within one degree C of background.







Thermal



PCA -> RGB

eglin_other_s300_a2000_r15

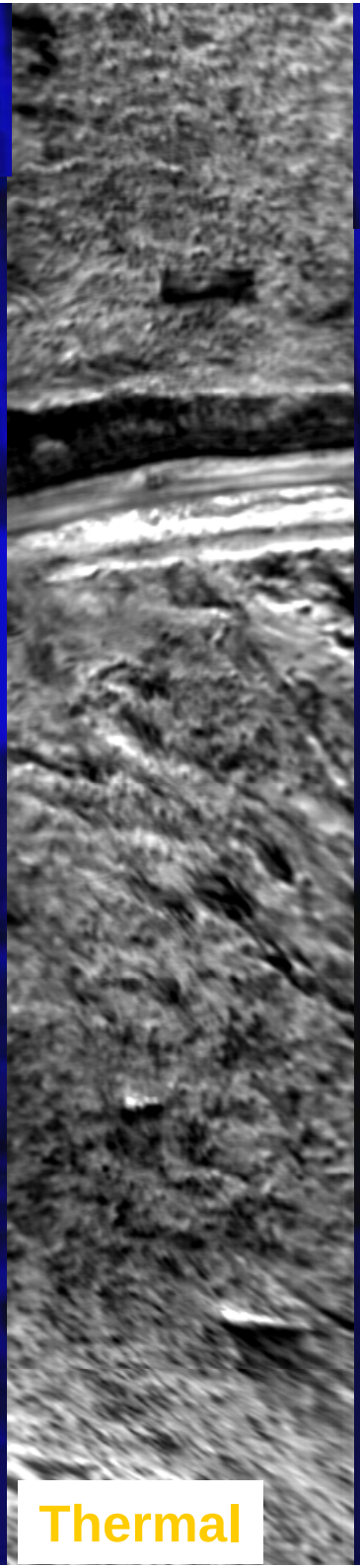
Targets appear to be spectrally matched to blackbody thus are difficult to separate from the background

Some thermal anomalies

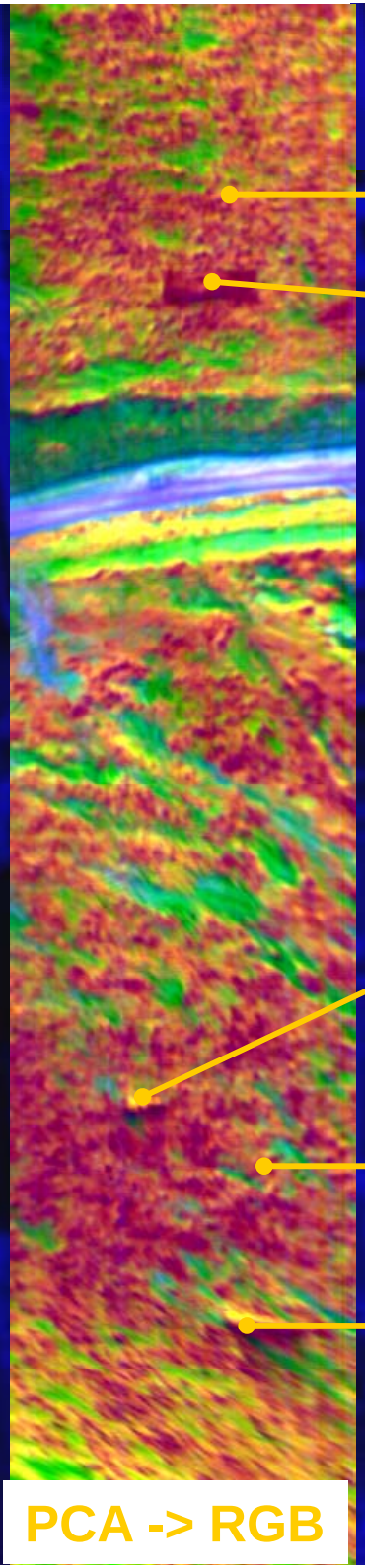
Top target roughly 2 degrees C below background temperature

Bottom target roughly 3 degrees C above background temperature

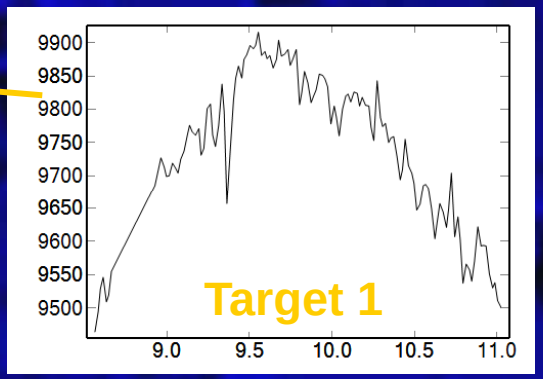
eglin_other_s300_a2000_r15

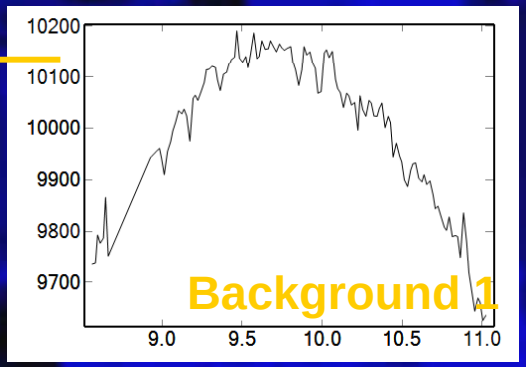


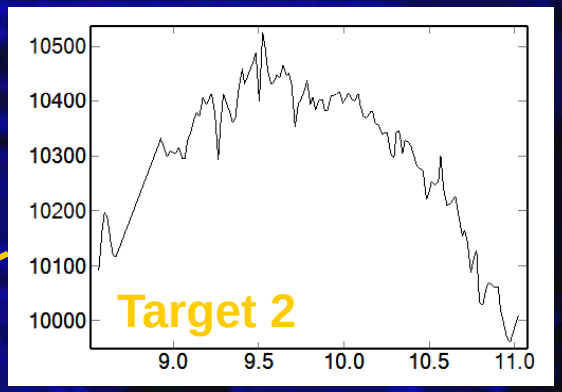
Thermal

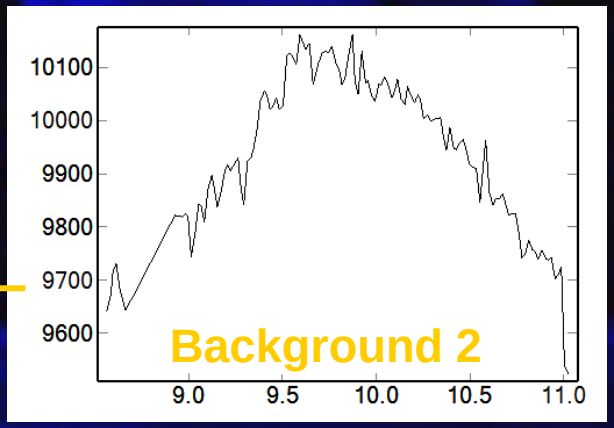


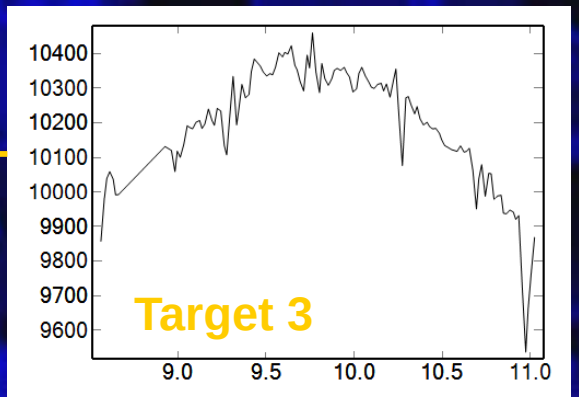
PCA -> RGB



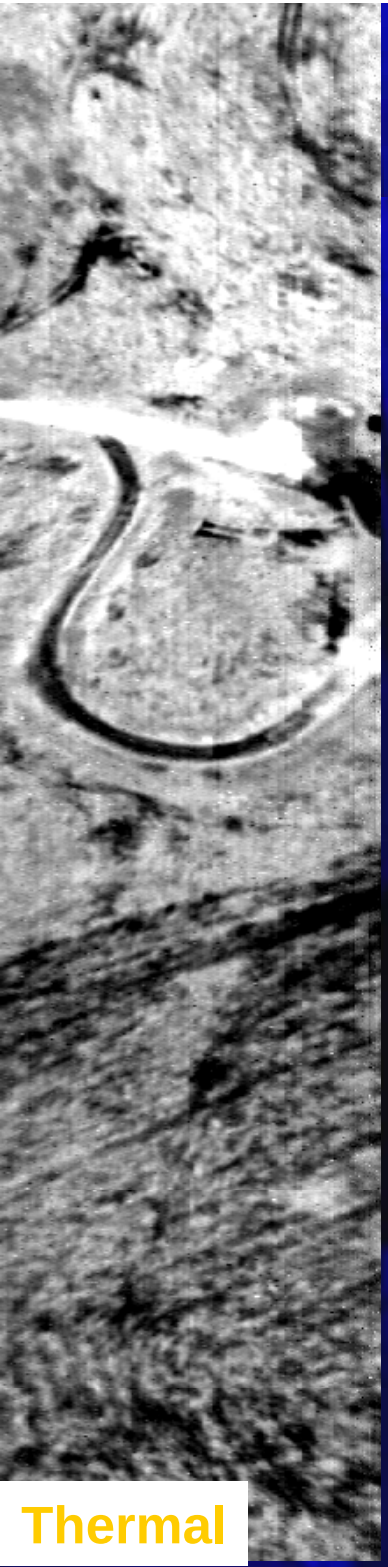






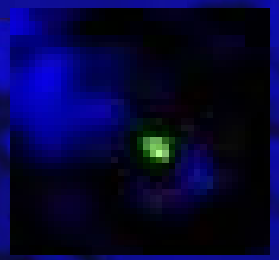


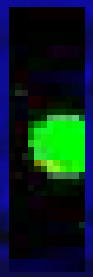
eglin_other_s300_a2000_r16



Thermal

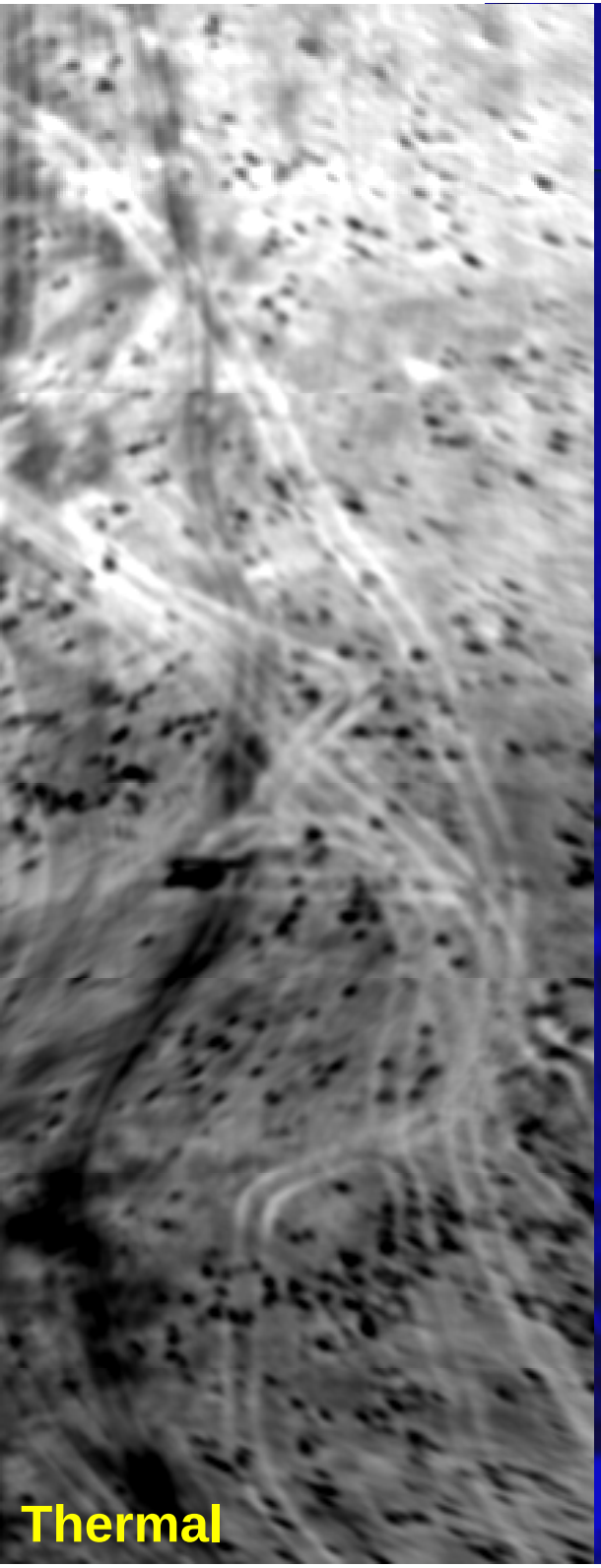
N-FINDR
Matched Filter





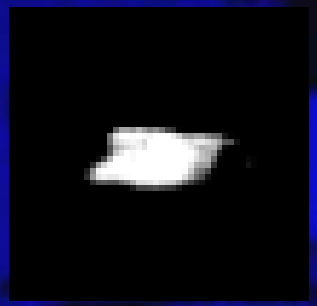
Targets appear fairly easily separable from background.

Targets hard to detect in thermal image due to background clutter



Thermal

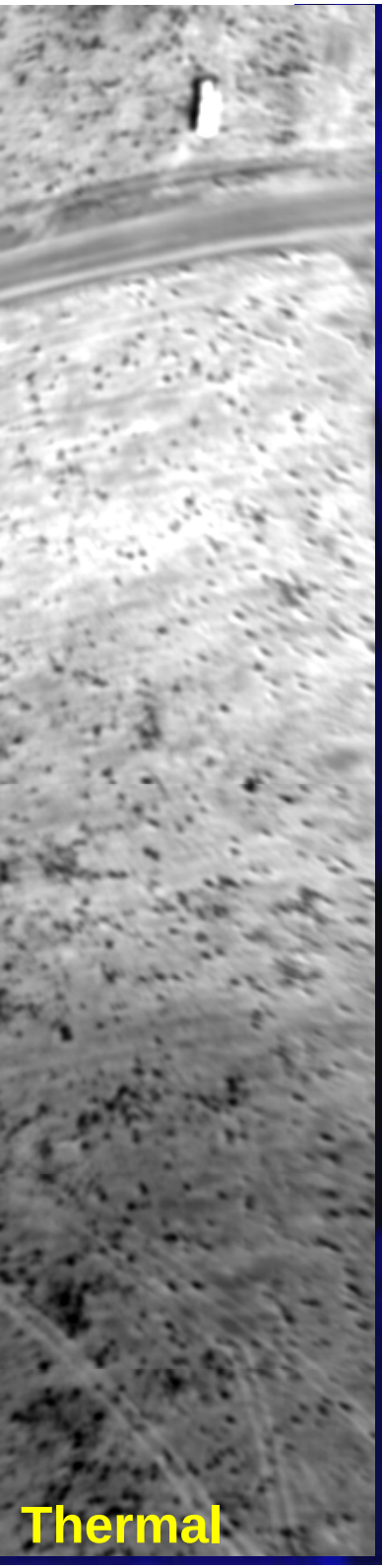
N-FINDR
Matched Filter



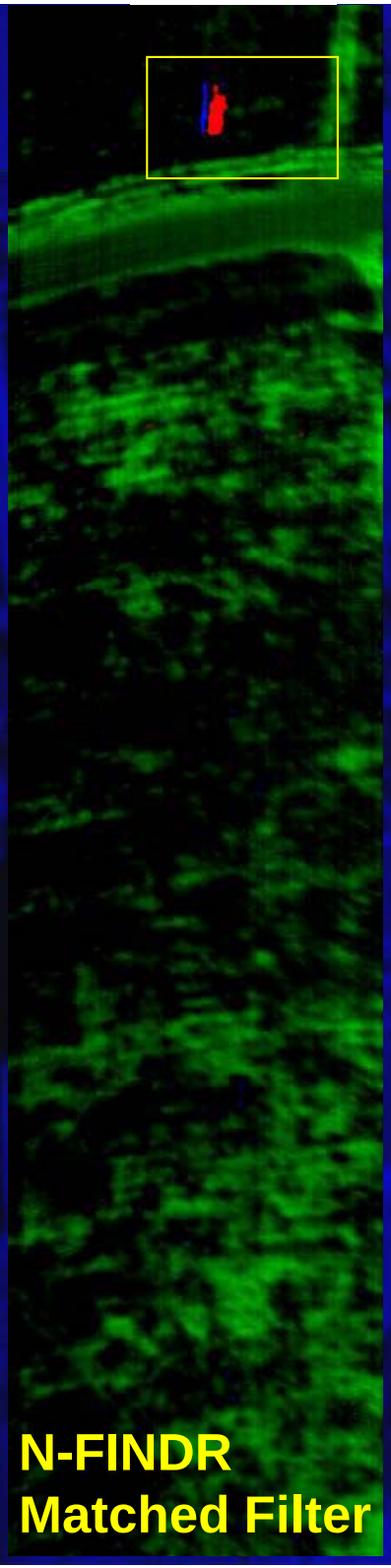
Target is easily separable from background.

Matched filter detects ozone reflection.

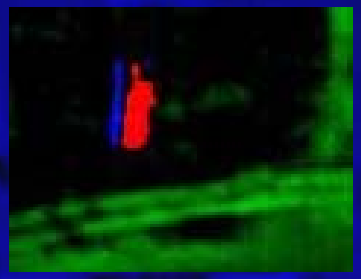
Target is not easily seen in thermal as it is roughly at ambient temperature.



Thermal



N-FINDR
Matched Filter



Target appears fairly easily separable from background.

Target is easily seen in thermal as it is roughly 10 degrees Celsius above the background temperature.

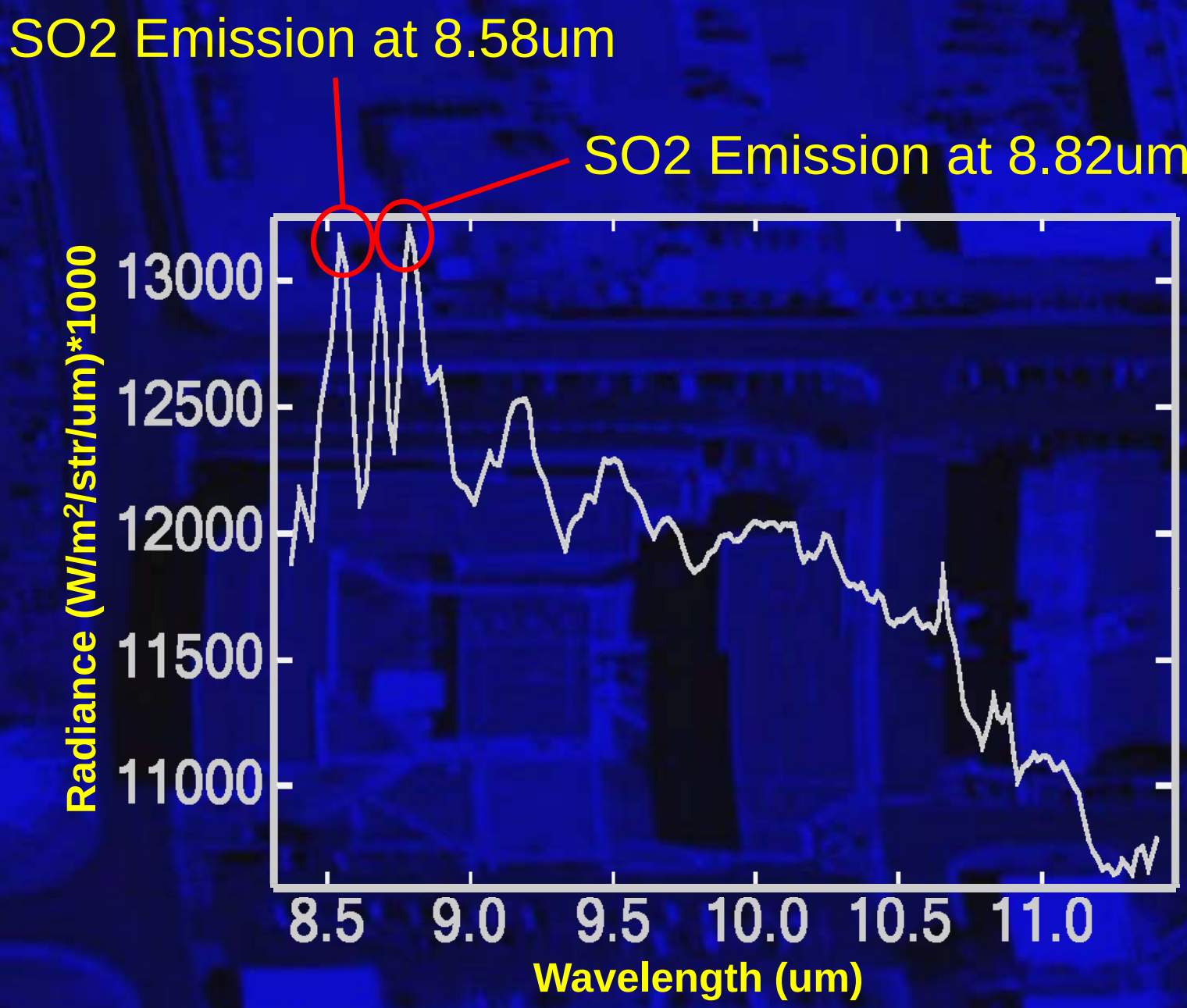
EPA Gas Detection April 2004

- Dow chemical plant
- Agricultural sites
- ~80 Gb data collected
- SO2, NH3, Benzene, others detected

Dow AHI Flight Lines

Dow Emissions Overlay

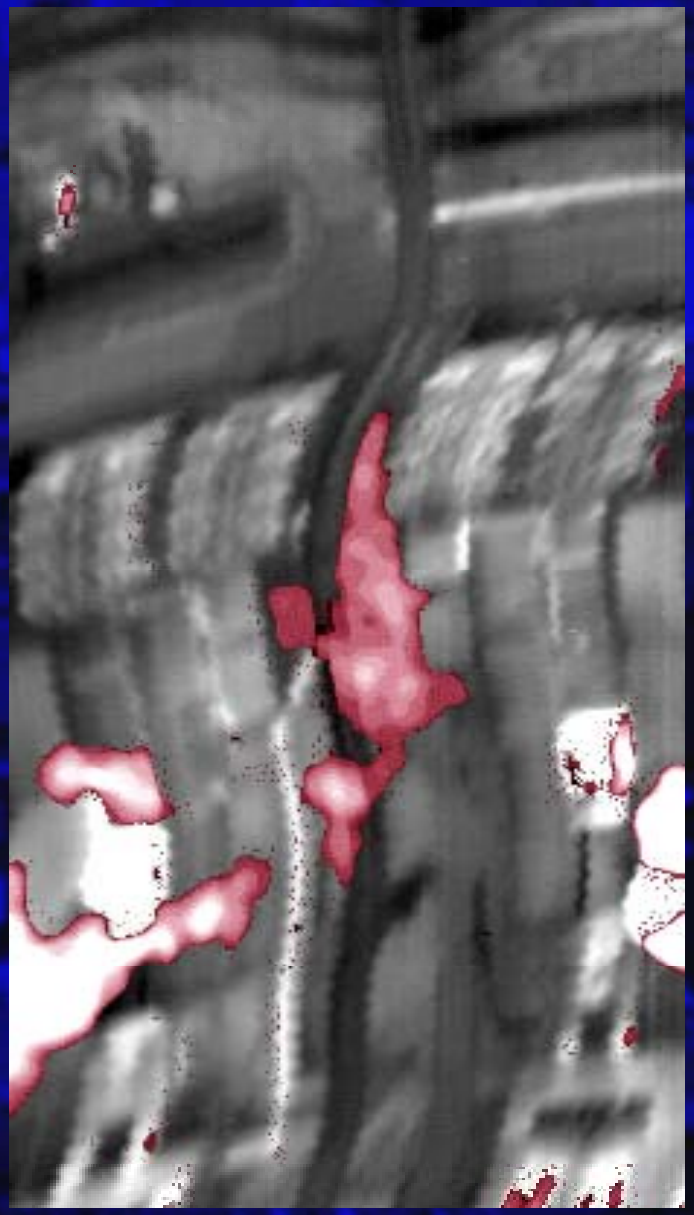
LWIR Gas Detection



AHI Spectral Technology Group
Airborne Hyperspectral Imager University of Hawaii

Gas Data Comparison

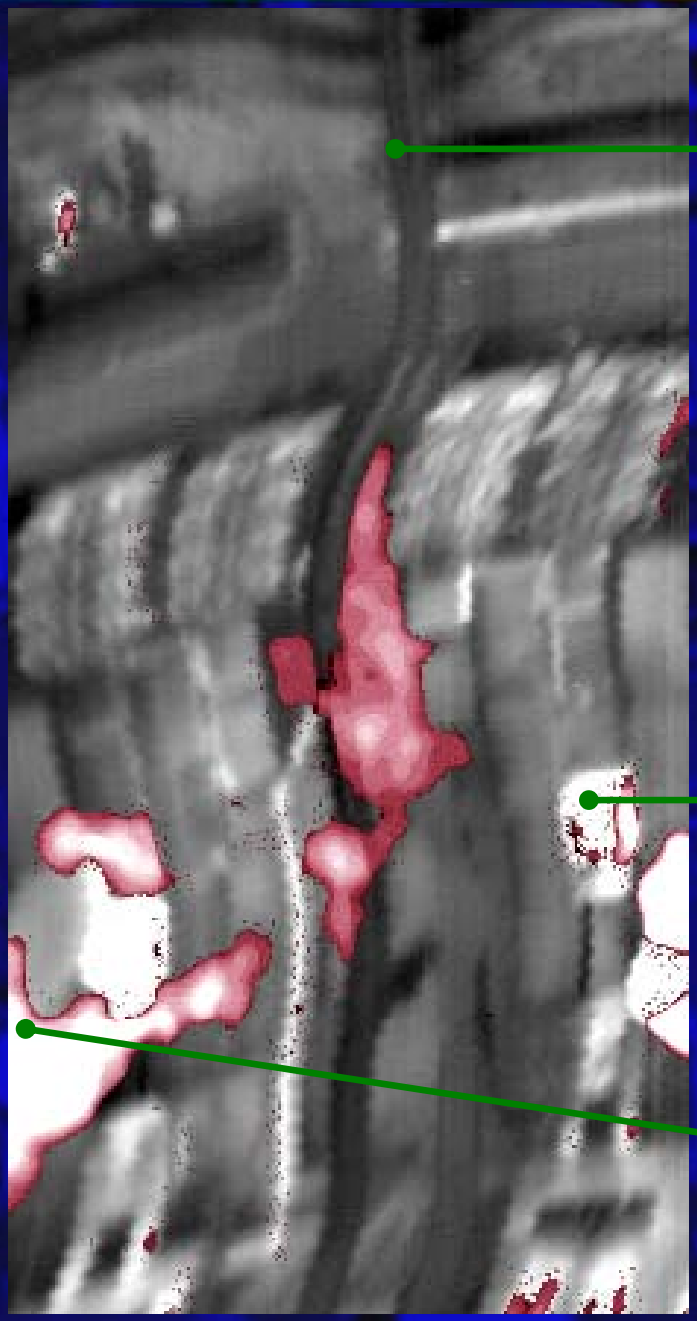
AHI 10um Thermal Image

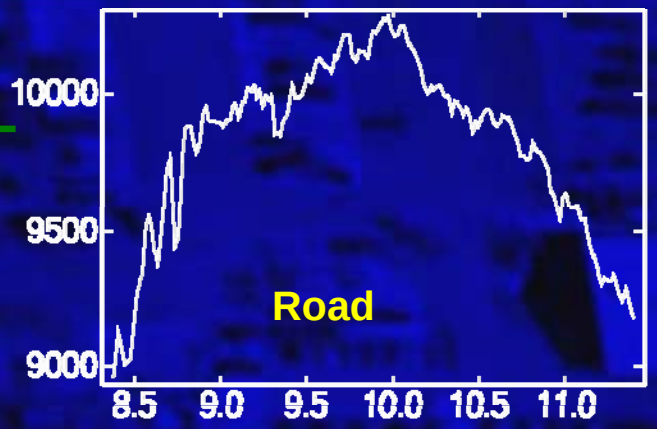


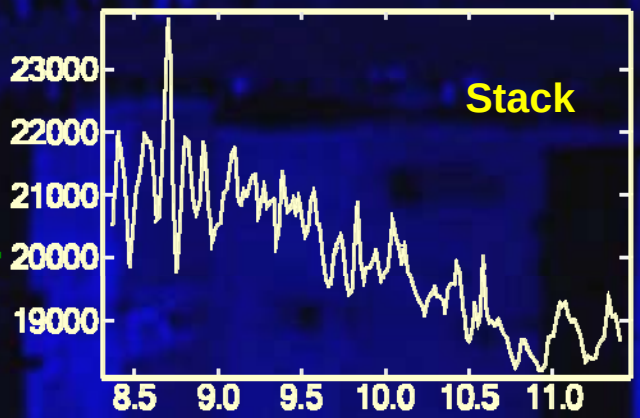
AHI Thermal Overlain with SO2 Detection

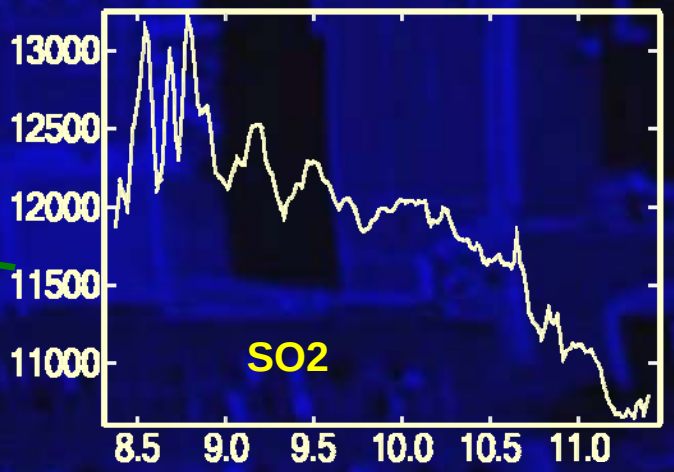
Line Scanner

Spectral detail from gas data









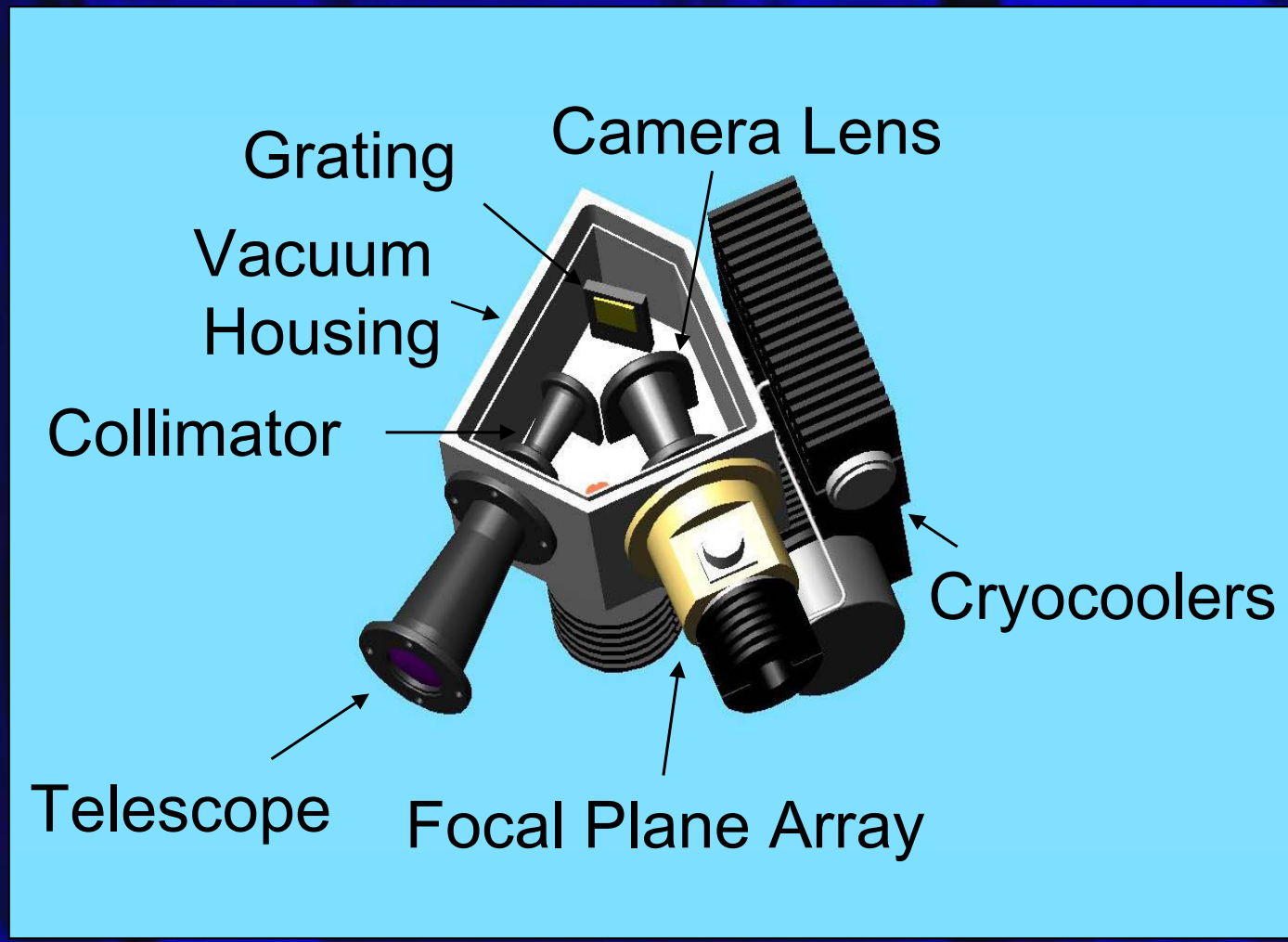
Note: Y-axis units in W/m²/str/um*1000

Next Generation LWIR Hyperspectral Imaging Sensor

AHI II

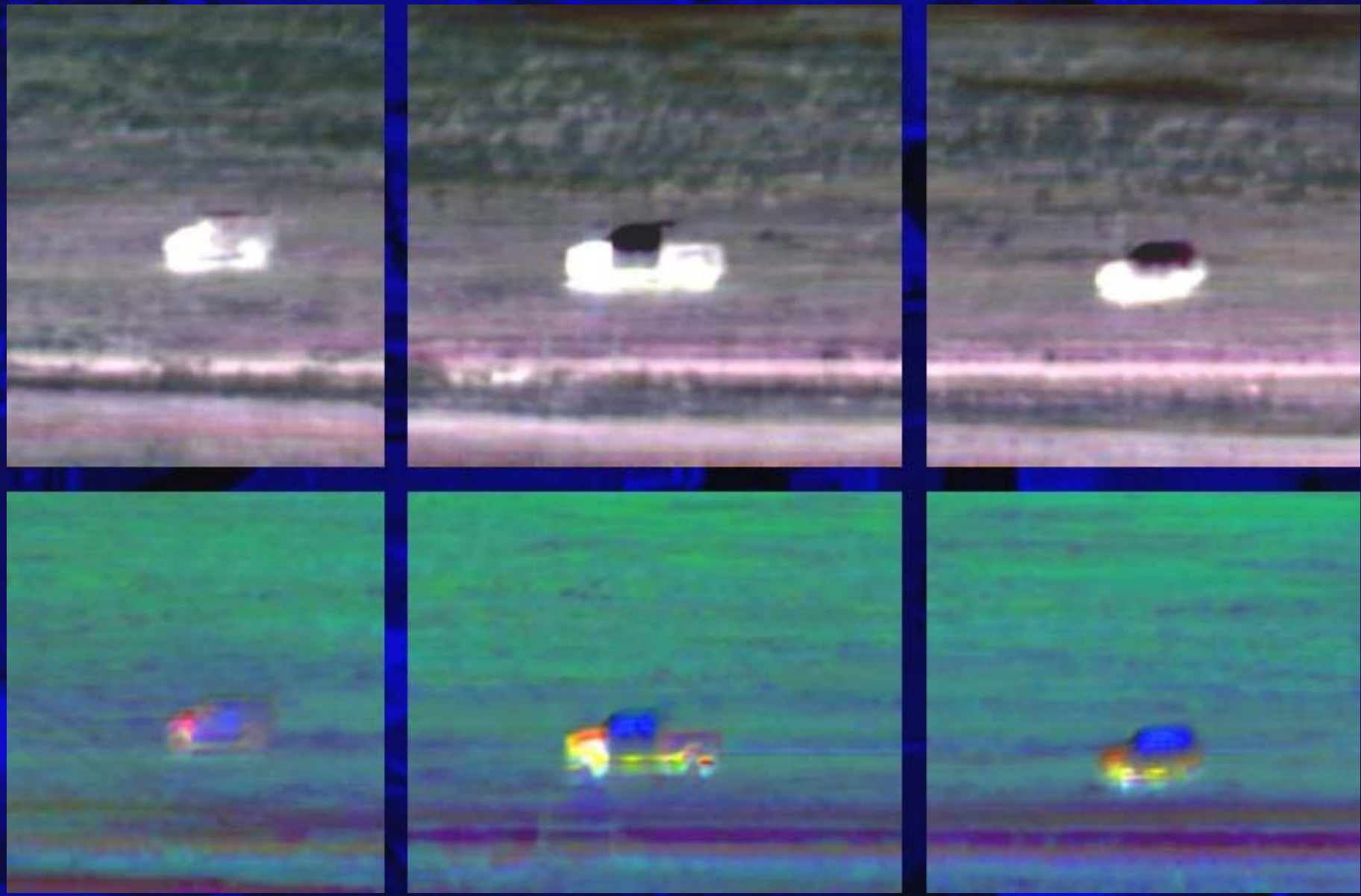
- Spectral Resolution 50nm
- Target Temperature 300C
- System Transmission 53%
- Pixel Size 30 microns
- Final f-no. f/2
- QE 70%
- Frame Rate/Integration Time 150Hz/6ms
- Spectrometer Temperature 110K
- FPA Temperature 56K

AHI II

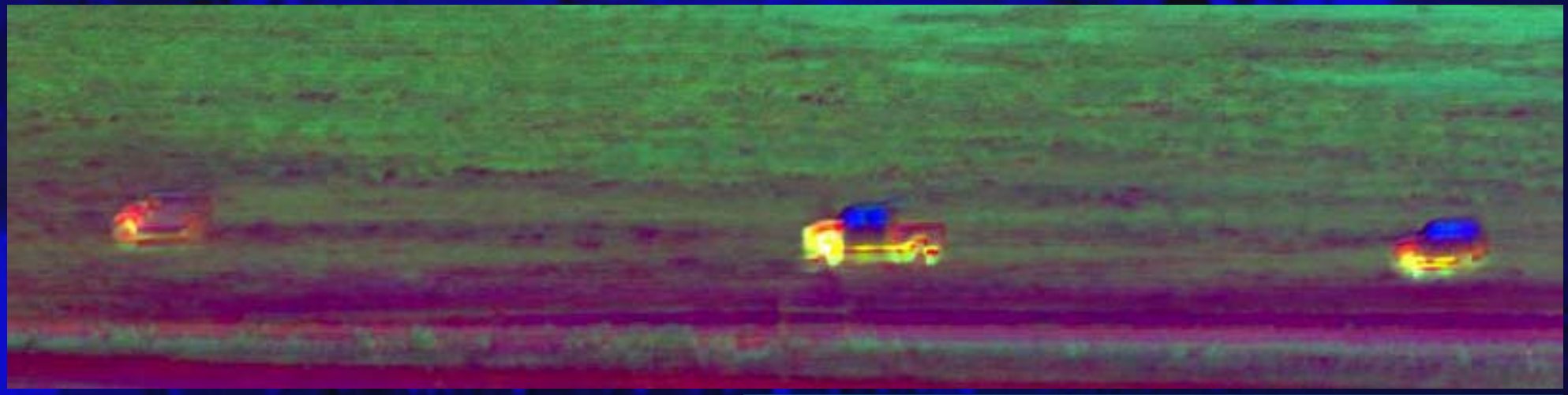

Grating
Camera Lens
Vacuum Housing
Collimator
Cryocoolers
Telescope
Focal Plane Array

Supporting Data Collections & Analysis

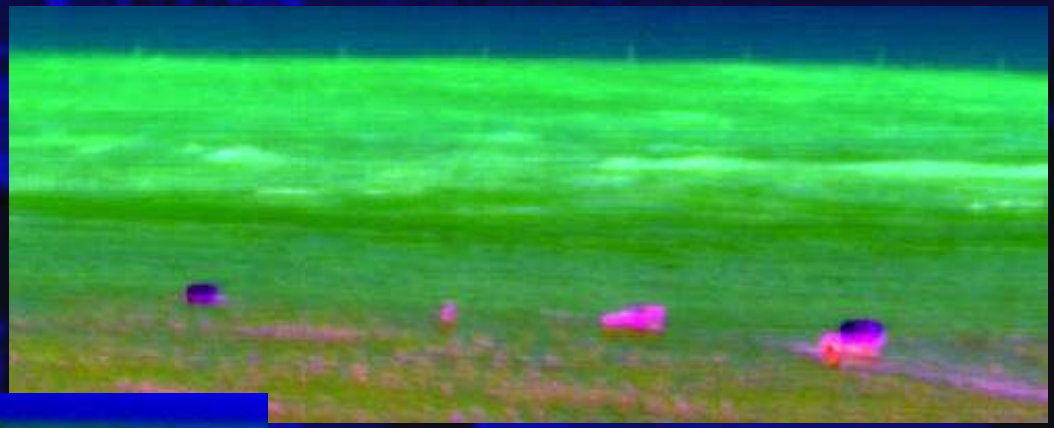
Vehicles

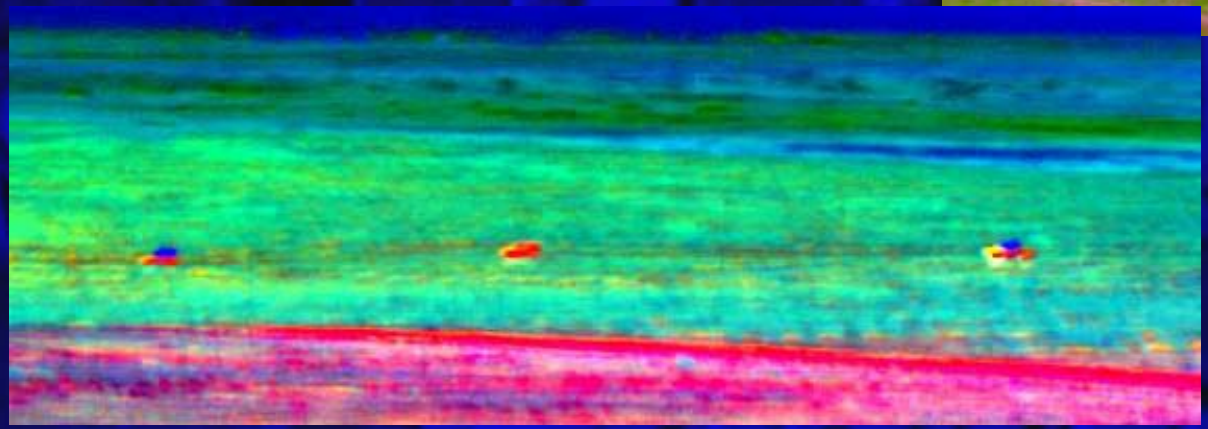


Vehicles

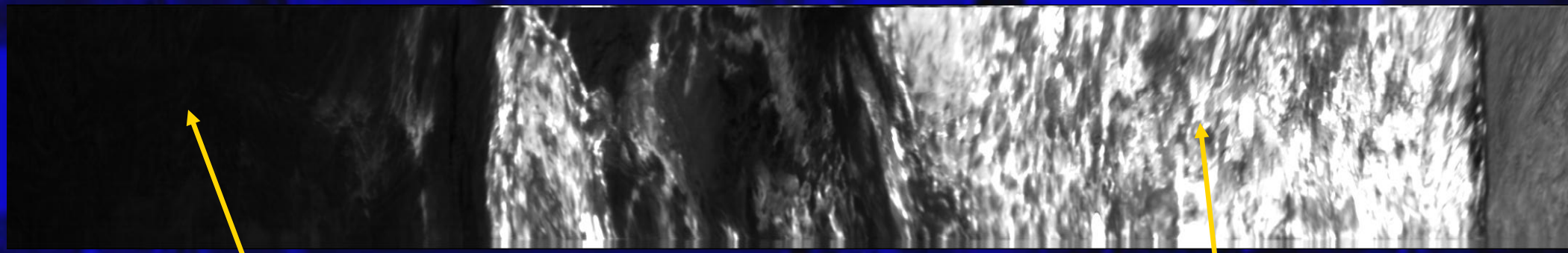


Each frame averaged 16 times and Principal Components calculated

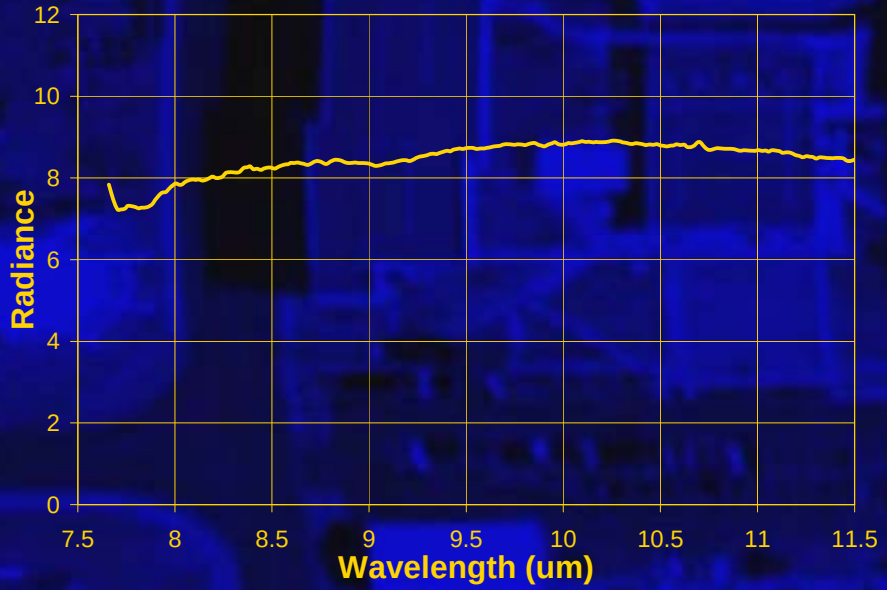




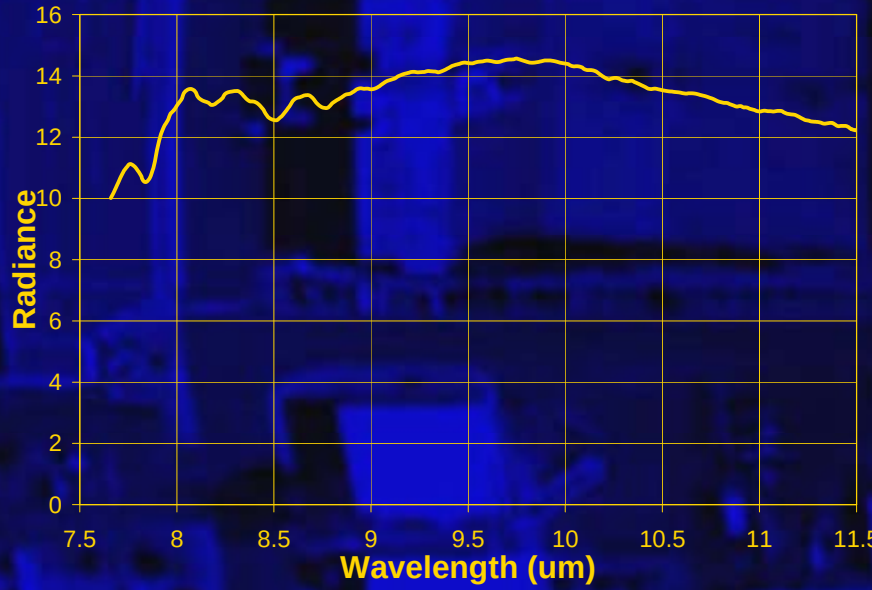
Lava Flow Kilauea '03



Background

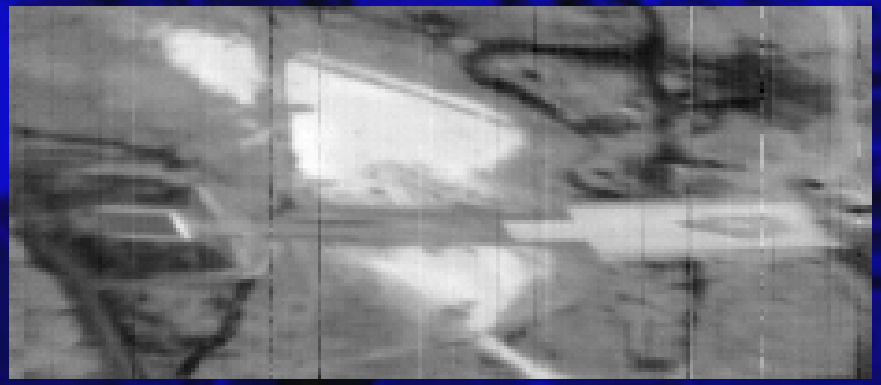


SO_2



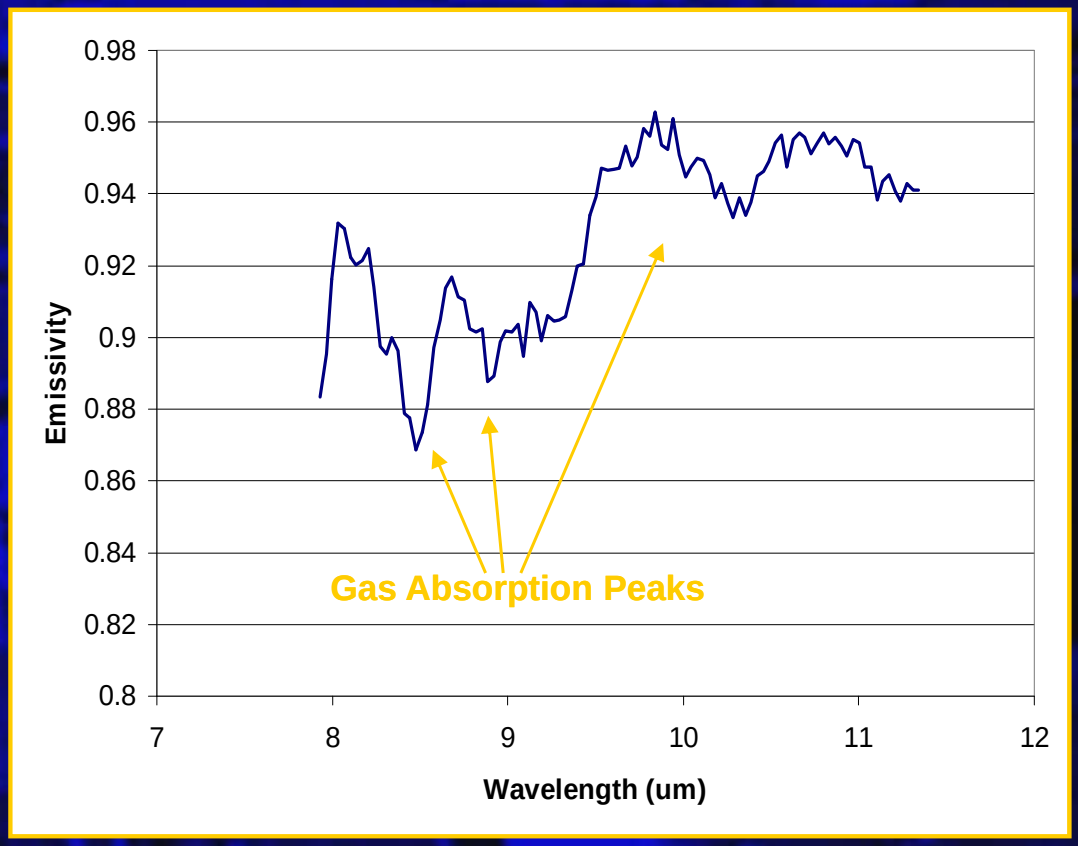
AHI Gas Detection(Freon)

Release 1: 1925 7.4 lbs



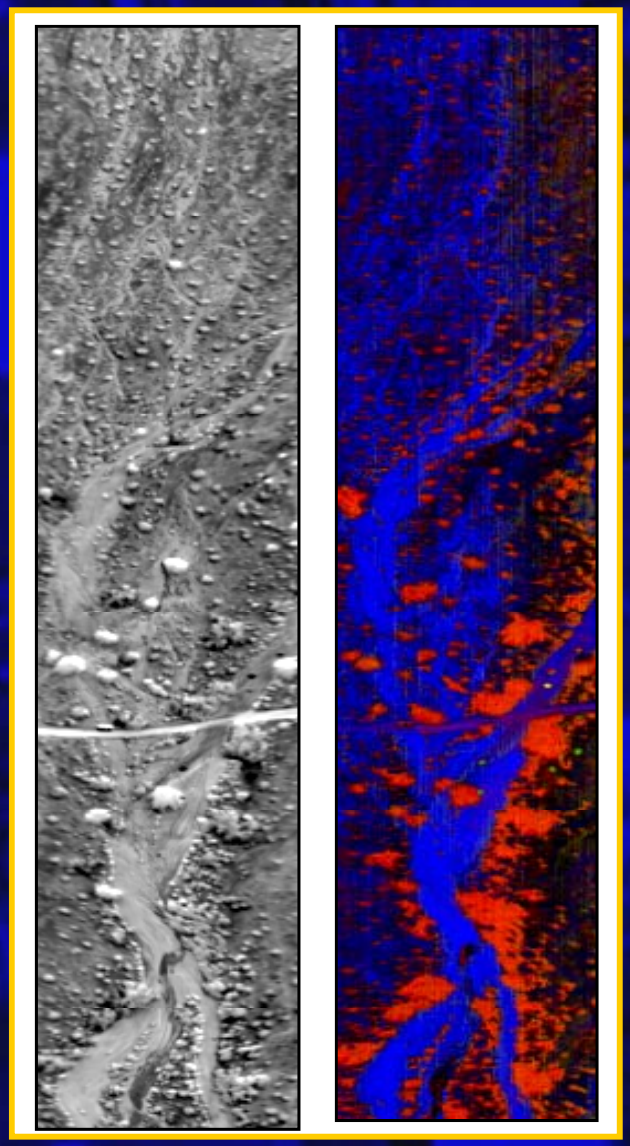
Release 2: 1934 14.8 lbs

Spectra from AHI Data Cube



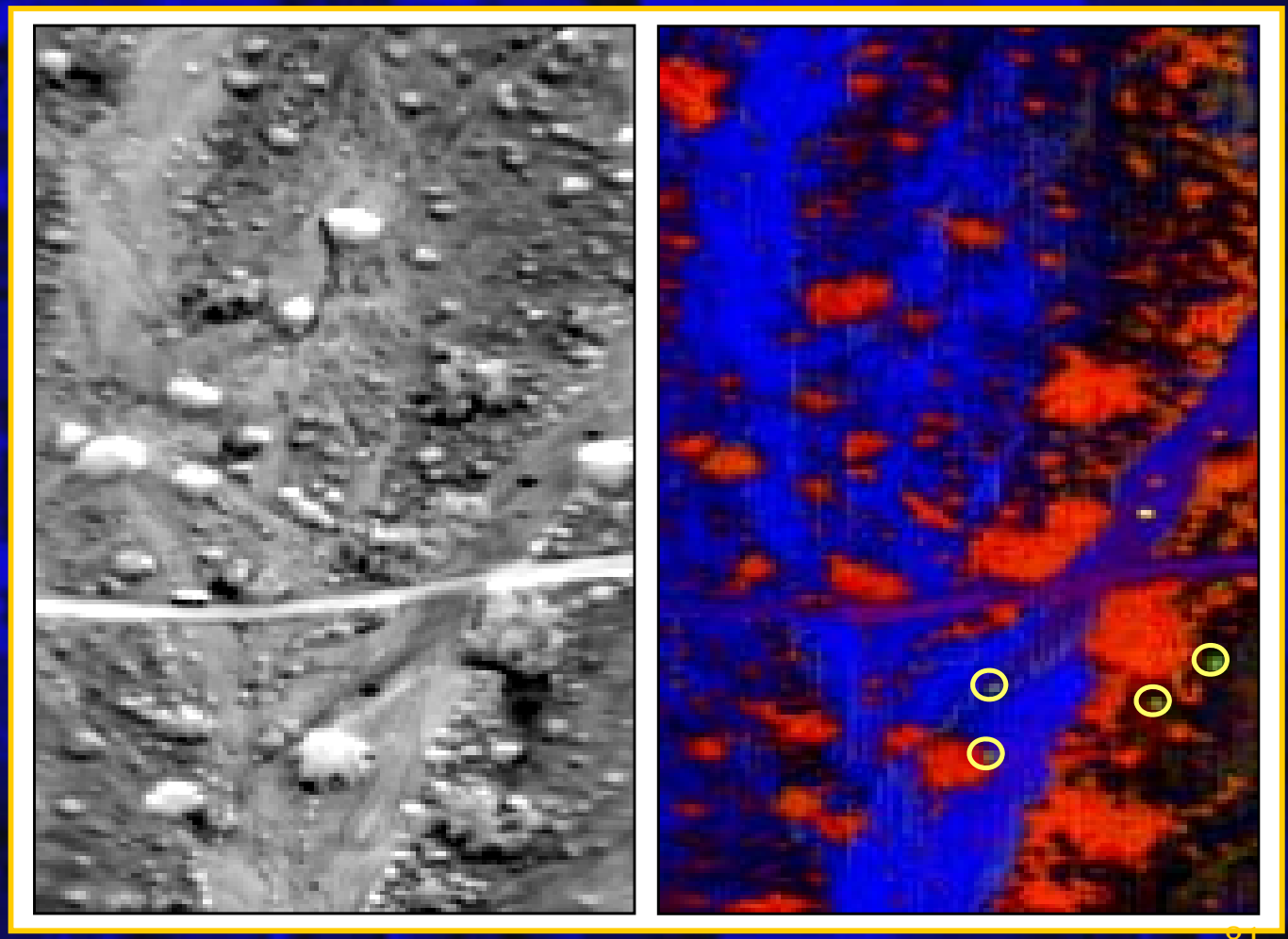
AHI Spectral Technology Group
Airborne Hyperspectral Imager University of Hawaii

Panel Detection Experiment



AHI Broad Band IR Image

Fractional Abundance End-member Image

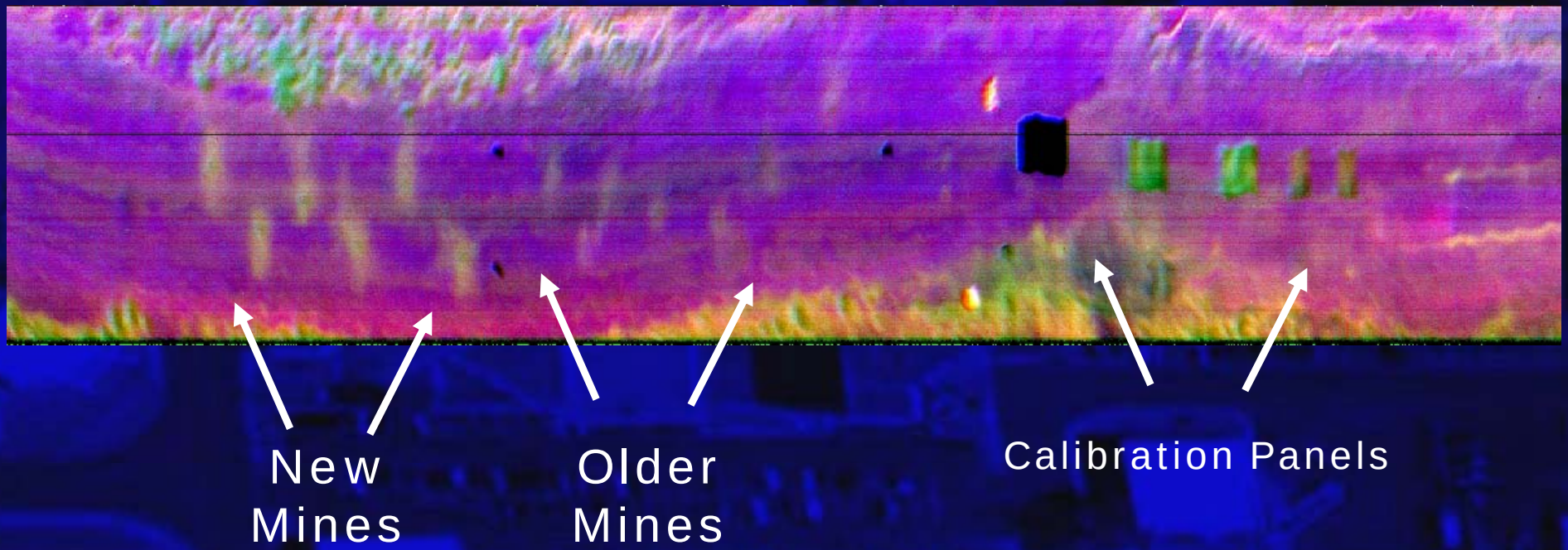


Test Panels Circled in Yellow

Night Detection of Land Mines

3-color composite image of pre-dawn AHI flight

Red: **Average Brightness Temperature**
Green: **Apparent Emissivity at 9.16 μ**
Blue: **Apparent Emissivity at 10.25 μm**

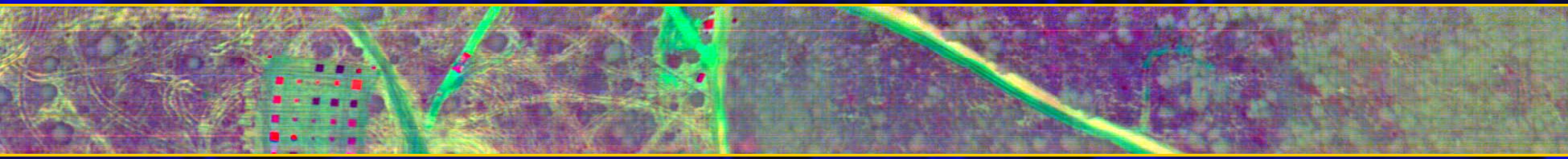


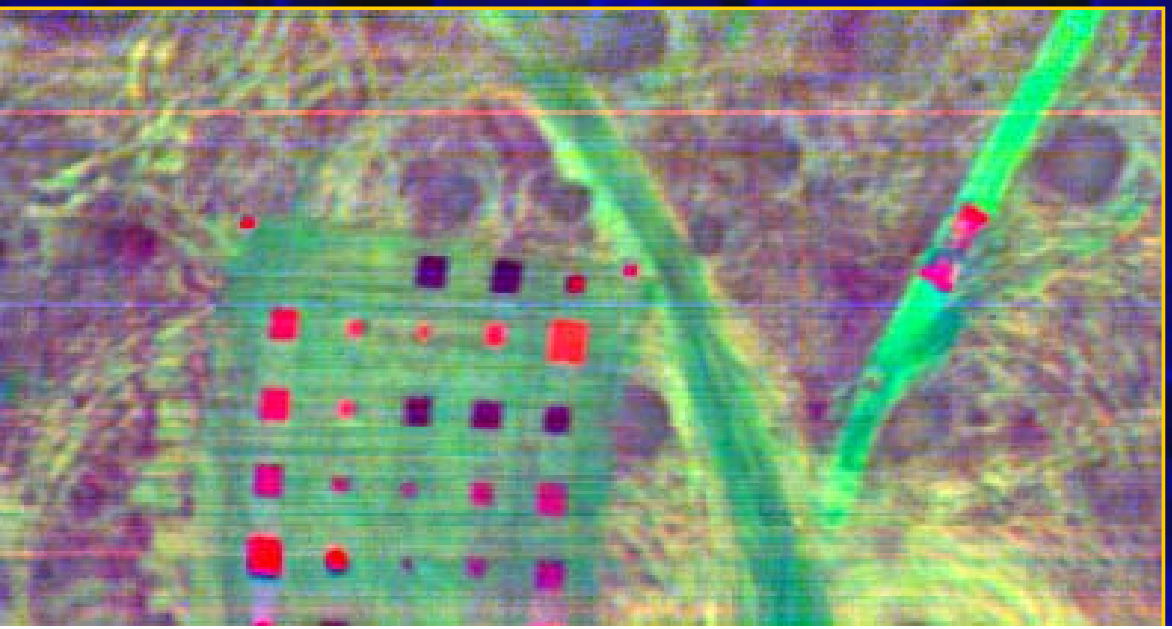
AHI Spectral Technology Group
Airborne Hyperspectral Imager University of Hawaii

Materials Array

Broad Band Temperature Image

Color Principal Component Image (Excluding Temperature Component)





Night Mission

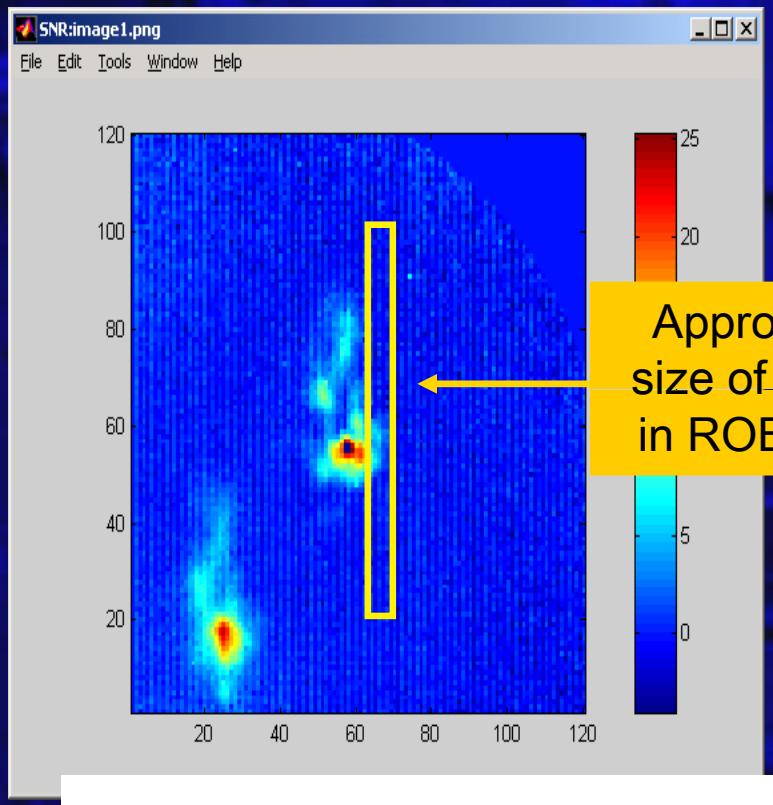
Enlargement of Material Array Area

Color Image Made from PC1,2,3
(first three after temperature removed)

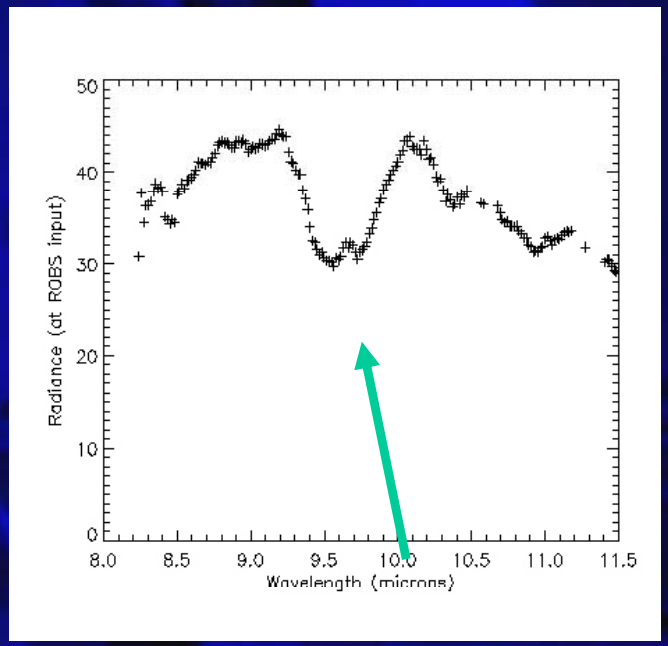
AHI Sensor on ROBS Telescope

AHI Spectral Technology Group

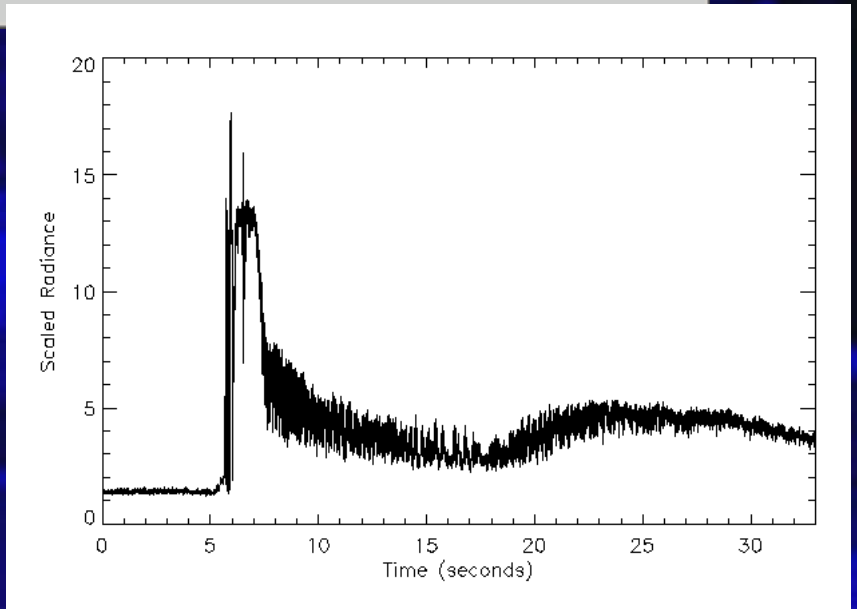
Airborne Hyperspectral Imager University of Hawaii

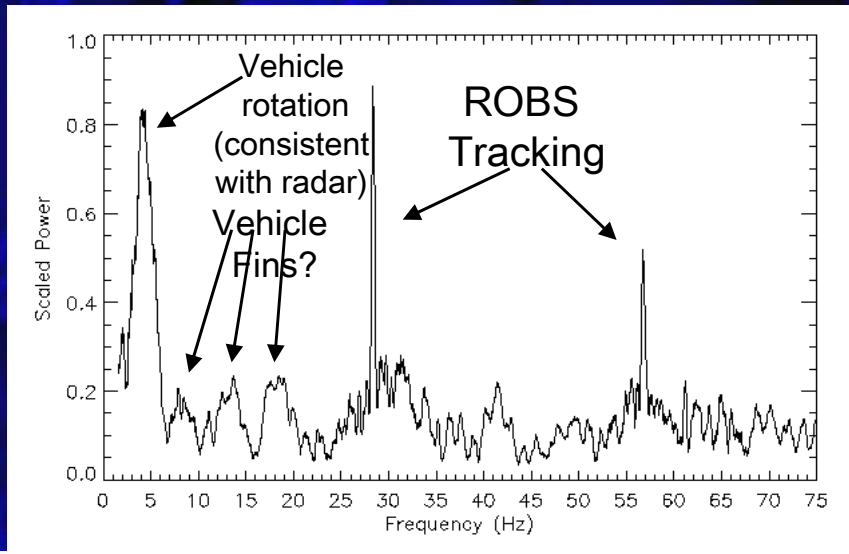


Approximate size of AHI Slit in ROBS FOV



Stratospheric ozone absorption



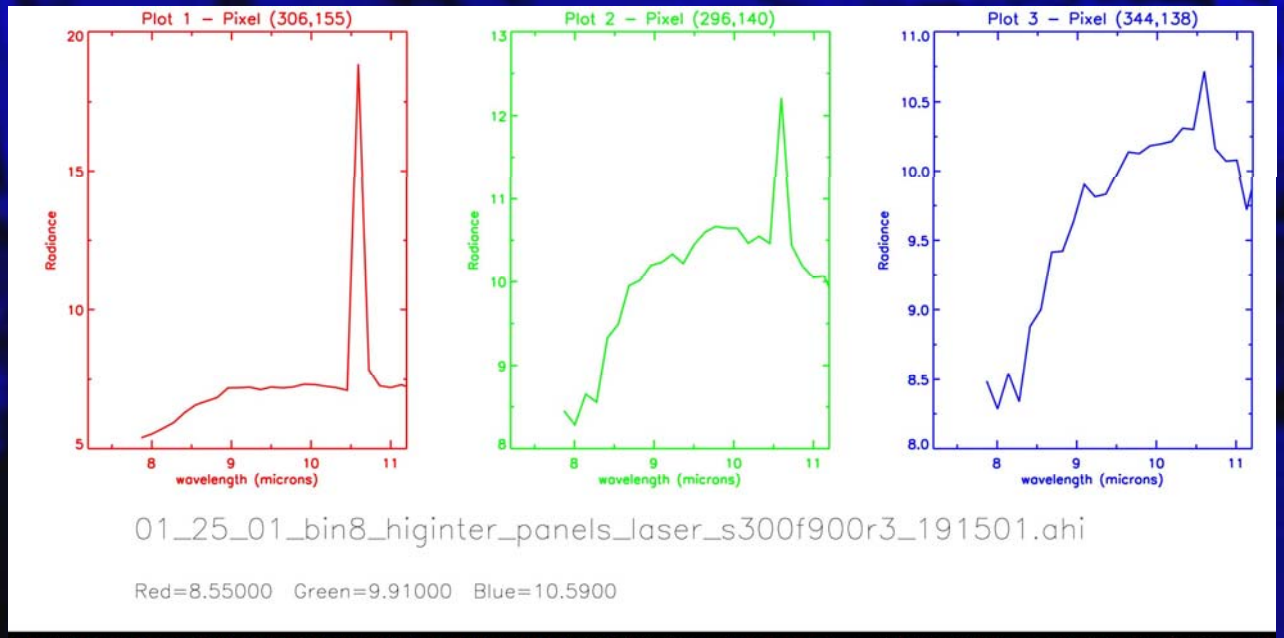


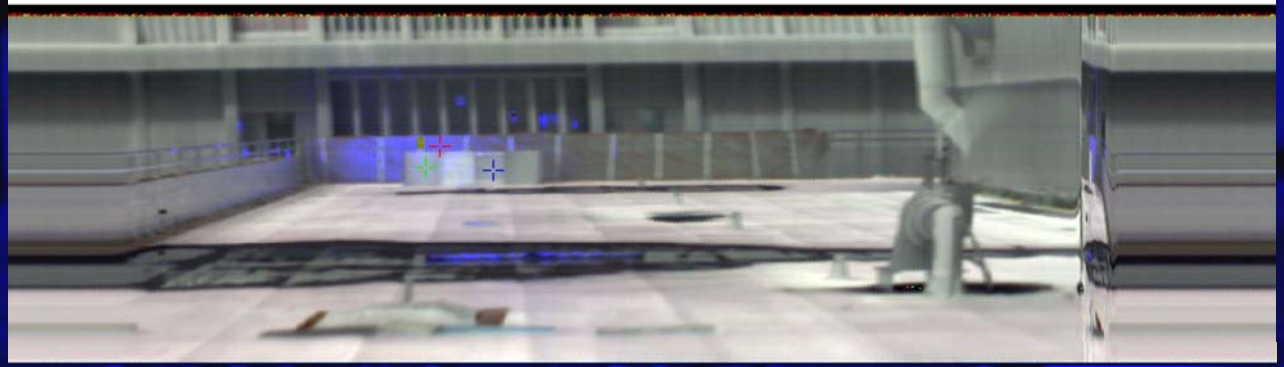
Controlled Active/Passive Hyperspectral Capability

AHI Hyperspectral Imager and 10.6 μm laser illuminator

Target Array

Laser line readily detectable under ambient conditions





Passive hyperspectral

Passive hyperspectral with laser illumination

Isolated laser reflected radiation

Research Accomplishments

- Legacy data review and distribution
- WAAMD data collections
- Sensor Week data collection
- Yuma data collection
- LWIR Spectral Data Analysis Support

Science of Land Target Spectral Signatures: Synthetic Scene Simulation

David Messinger, Ph.D.
Digital Imaging and Remote Sensing Laboratory
Rochester Institute of Technology

Overview of Simulation Activities

- RIT program role: development of synthetic scenes for use by algorithm developers
  - construct scenes
  - validate against existing data
  - incorporate phenomenology from team members
  - provide scene variations to stress algorithms
- Landmine scene
  - completed & validated
  - variants delivered to MURI team
- Concealed Target scene
  - completed

Landmine Target Scene Simulation Construction & Status

Landmine Scene Developments

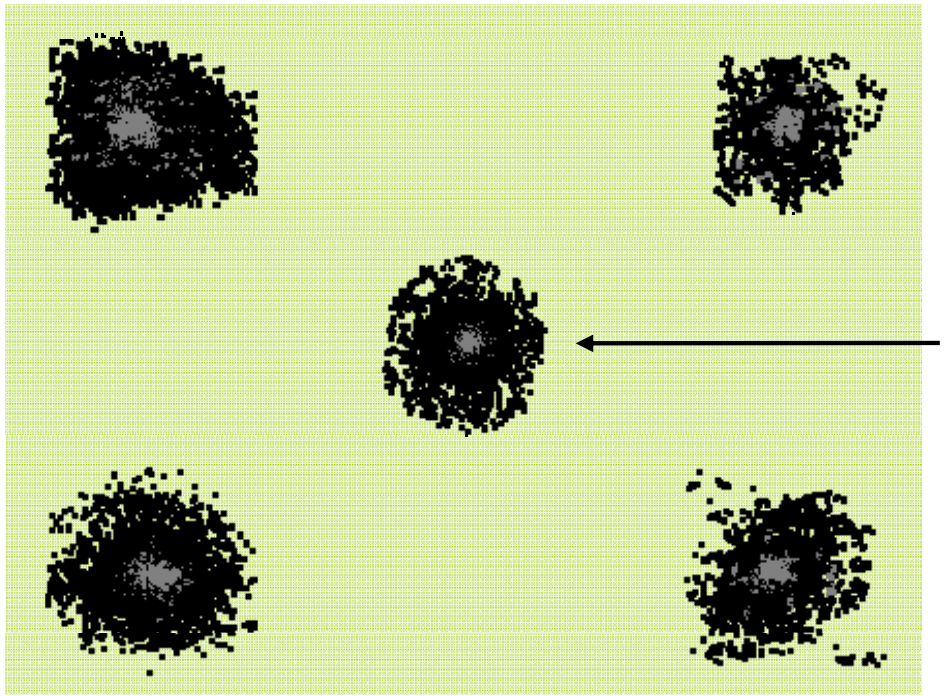
- Scene was created and validated against AHI data
  - delivered to MURI team

- Variations of scene were created:
  - asymmetric targets
  - asymmetric target placement
  - variable atmosphere
  - other altitudes
  - different times of day

Landmine Scene: Disturbed Earth Targets

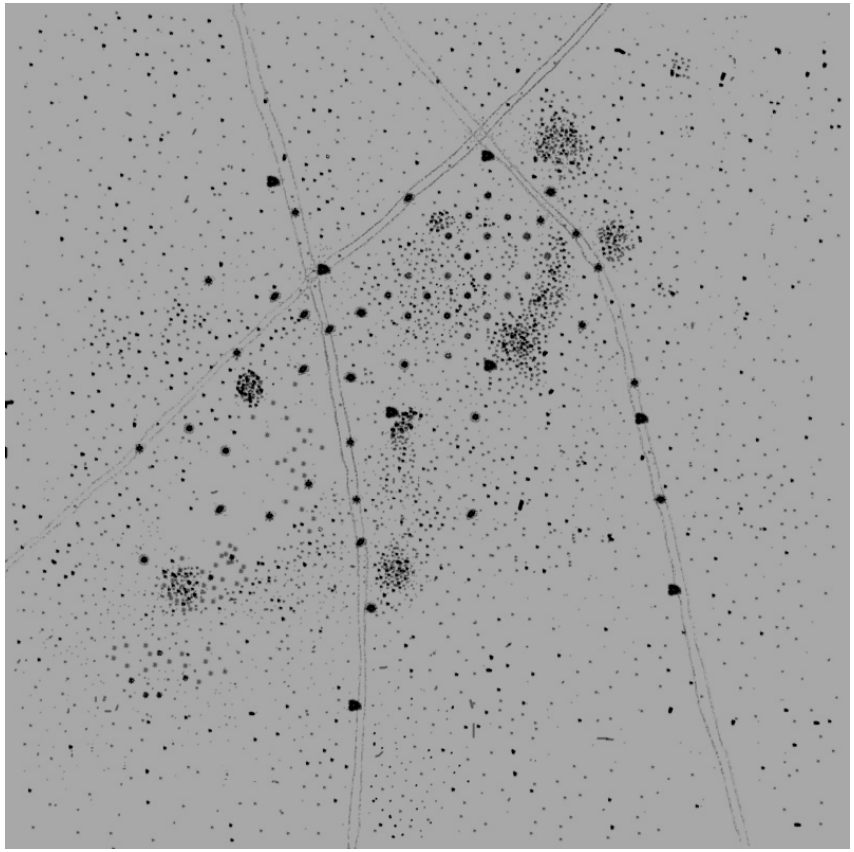
- Original disturbed earth targets were determined to be “too uniform”
- Asymmetric variations of material maps created and placed in scene

gray / black represent different “types” of disturbed soil



Landmine Scene: Target Placement

- Buried targets placed asymmetrically around field
- Several placed along “roads”
- Clutter distribution changed
- Implementation improved for future flexibility
- Combined target & clutter map shown

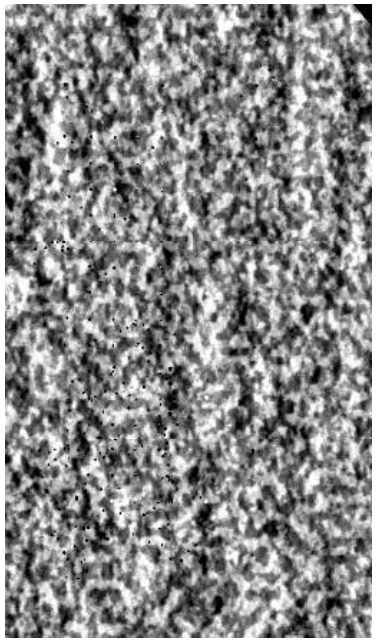


Landmine Scene: Other Variations

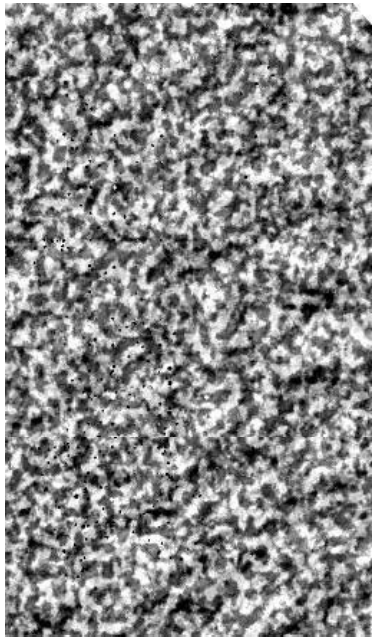
- Atmospheric parameters perturbed from original
  - visibility varied between 50km and 5km
  - column water vapor scaled with visibility
- Scenes rendered from an altitude of 2000ft
- Four other times of day considered:
  - 10AM, 12PM, 3PM, 7PM
- 10 new renderings completed and delivered to MURI team

Landmine Scene: Recent Renderings

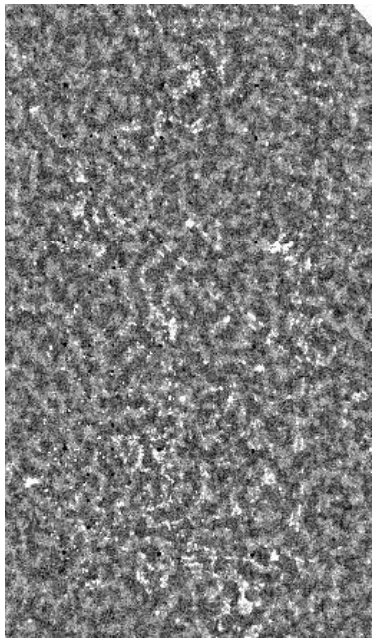
renderings at 9.6 μm



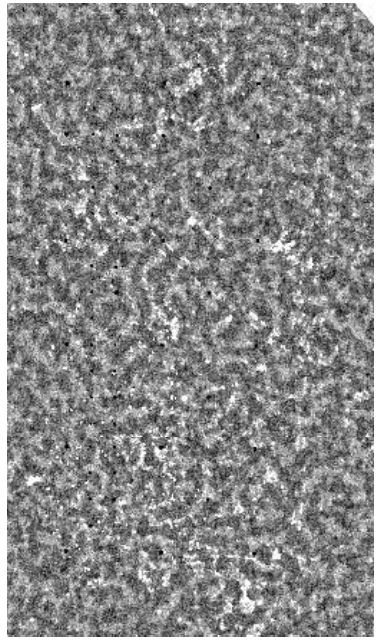
10AM



12AM
5km Vis
1.5X H_2O



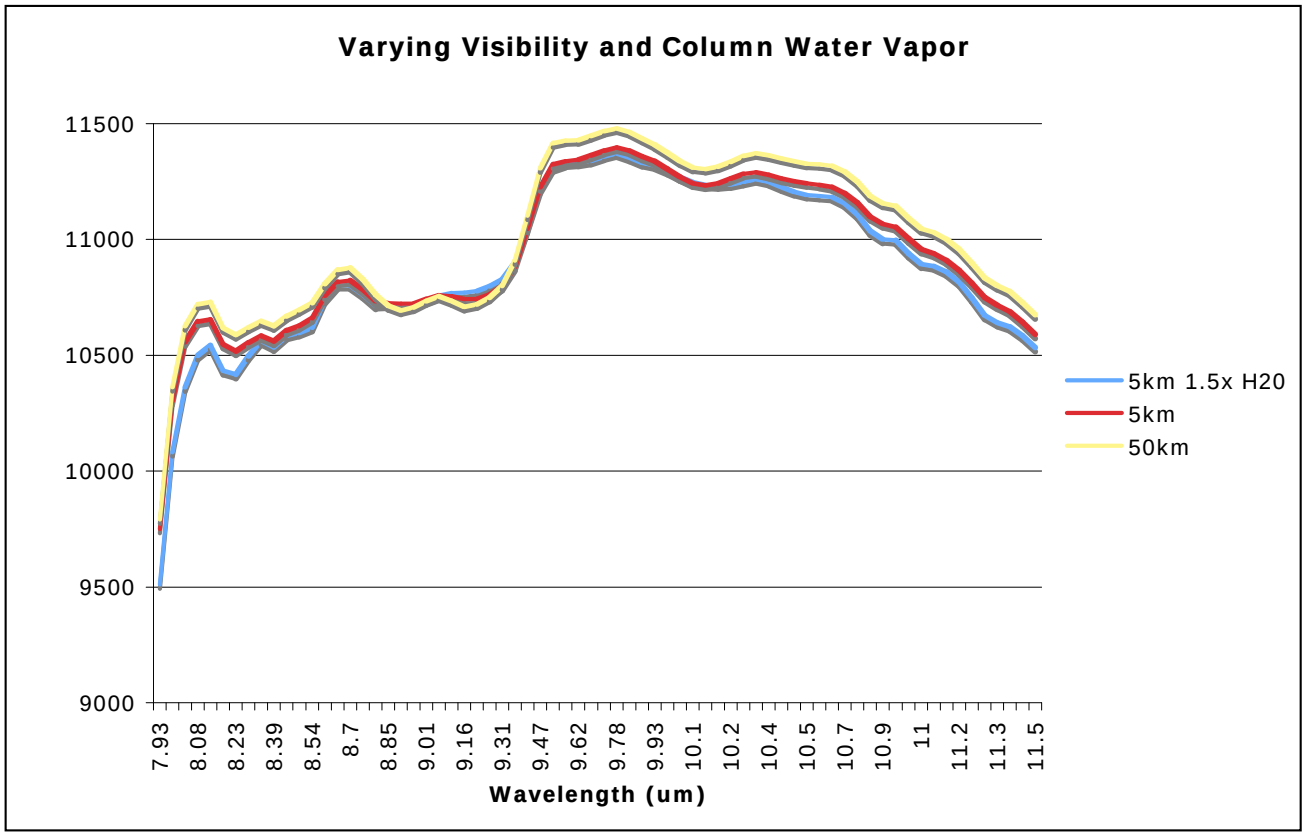
7PM
50km Vis



7PM
Original

Landmine Scene: Spectral Variability

Changes in atmosphere induce spectral variability



spectra taken from a common pixel

Concealed Target Scene Simulation Construction & Status

Concealed Target Scene Overview

- Goal: Develop a synthetic scene for testing of ATR algorithms against concealed targets

- Targets to be considered:
  - vehicles under tree canopy
  - vehicles under camouflage
  - IEDs

- Spectral coverage:
  - originally Vis / NIR / SWIR (0.4 μm - 2.4 μm)
  - extension to LWIR possible if required

Scene Requirements

- Geometry
  - forested area providing concealment for targets of interest
  - roadway
- Imagery Coverage
  - high resolution for spatial context
  - hyperspectral (Vis / NIR / SWIR; thermal if possible)
- Ground Truth Accessibility
  - for spatial layout
  - for spectral measurements
  - for geometric characterization

Required Elements for Simulation

- Accurate geometries
  - DEM
  - in-scene objects (trees, shrubs, targets)
  - material physical properties
  - material spectral properties
    - full-spectrum, simultaneous collection if possible
- Overhead data
  - geo-referenced for accurate placement of objects
  - calibrated for validation
- Weather history
  - for validation
  - obtained from AFCCC

Concealed Target Scene Site Selection

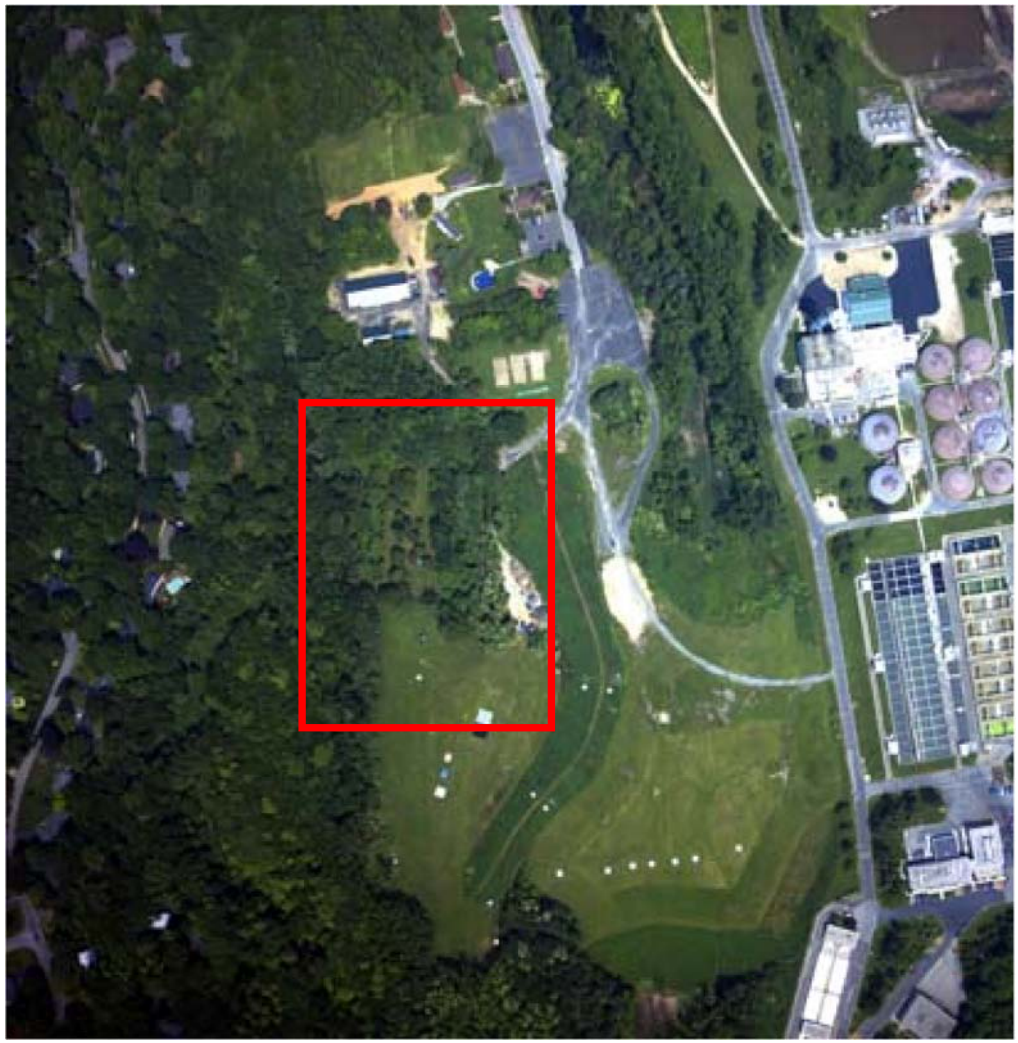
- Several considerations
  - amenable to several types of concealment
  - IED placement (roadsides)
  - site of airborne collection with sufficient ground truth
- Site chosen:
  - Camp Eastman, Rochester, NY
- Subject of extensive collection program in June 2004
  - COMPASS, SEBASS, RIT WASP, RIT MISI
  - extensive ground truth

Scene Specs

- Surface Area: 12,597 m^2 = 3.1 Acres
- Geometric resolution: ~ 3in
- About 10 types of trees identified
- 8 types of surface areas
- Over 300 instances of trees
  - 51 different models created
  - 6 base models of shrubs created
- Several man-made objects in scene, some specifically for experiment in 2004

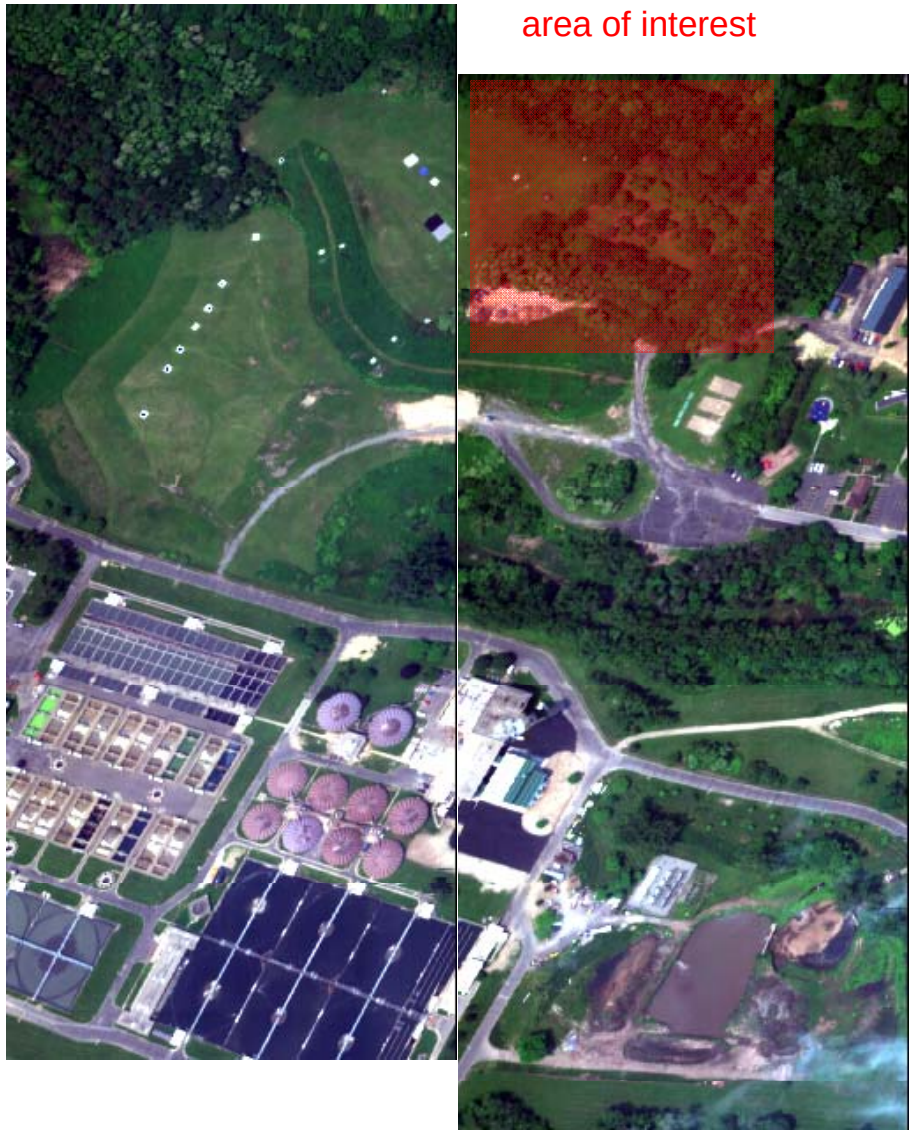
Concealed Target Scene Site

- Near Rochester, NY
- Site of extensive experiment in summer of 2004
- Heavily forested treeline, sparse forest, dirt roadways, etc.
- Coverage from several airborne hyperspectral imagers
- Accessible for measurements campaign



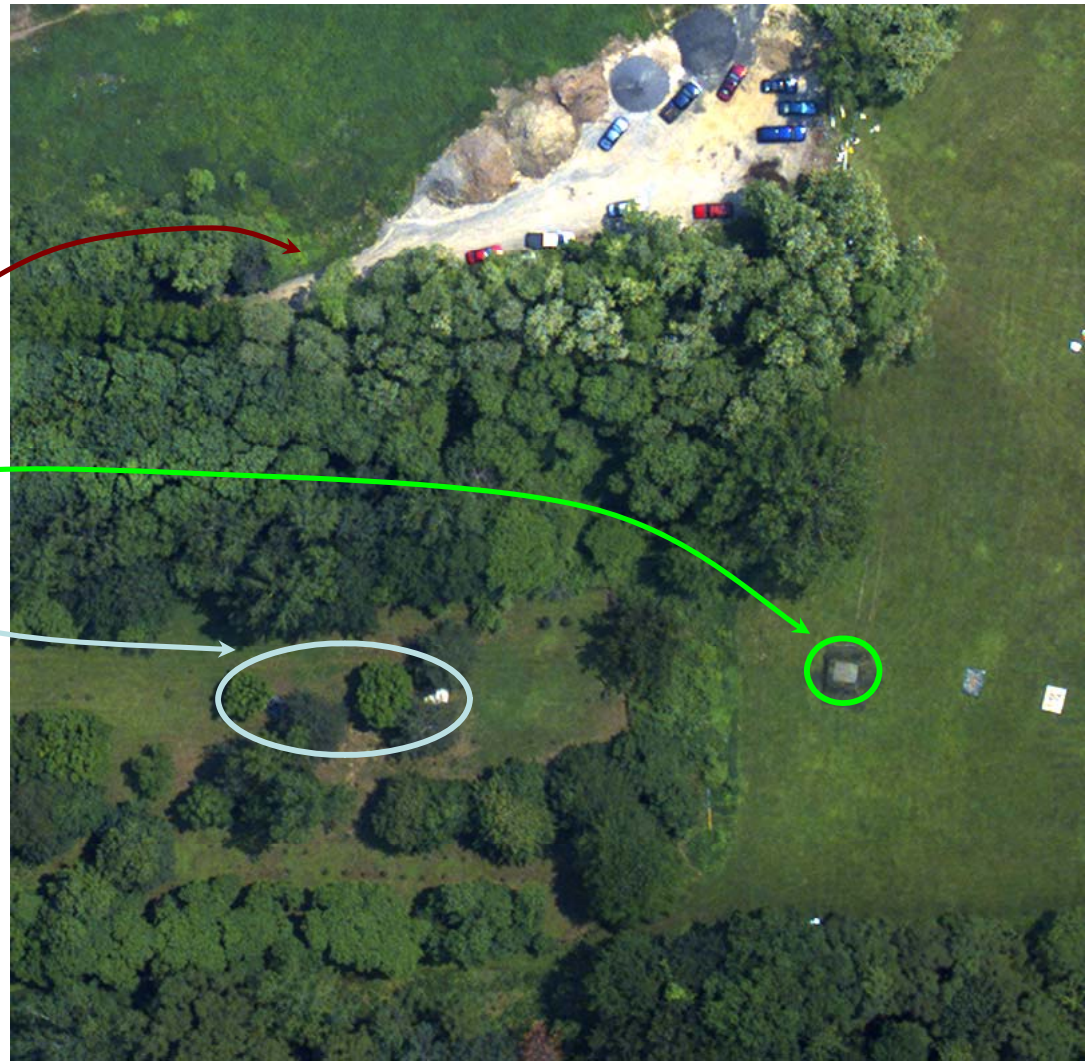
Concealed Target Scene: Site Overview

- Open field near water treatment plant
- Two consecutive COMPASS whisks shown here in RGB
- Several targets in the scene for validation
  - calibration panels
  - tarps under tree canopy
  - tarps under camouflage



Concealed Target Scene: Site Overview

- RIT WASP image of target area to be simulated here
  - GSD ~ 6in
- Dirt parking lot / roadside (IEDs)
- Target under camouflage
- Targets under trees
- Vehicles will be implanted after validation



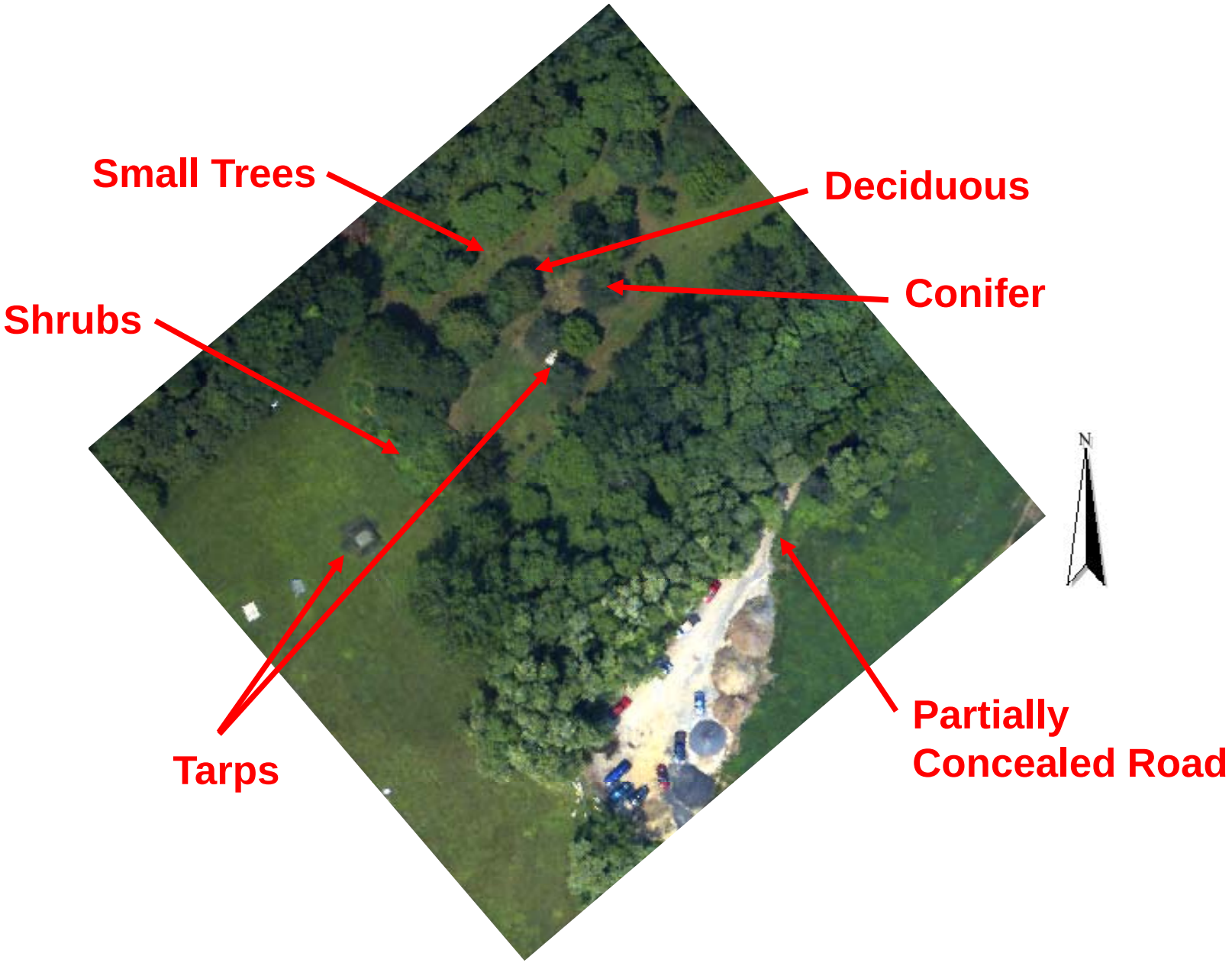
Concealed Target Scene: Site Overview

- COMPASS Image of target area
- Lower resolution
  - GSD ~ 1m
- Concealed targets and roadside visible
  - few pixels on target
- Scene can be rendered at any desired GSD

Synthetic Scene Coverage

high resolution
overhead RGB image

Scene Layout and Features



Building the scene

- 50 types of trees created
- Shrubs created for greater visual/radiometric accuracy – (also LIDAR)
- Models created with regard to ground truth data (pictures, measurements, etc)
- Instancing used for only about 50% of trees
- Man made objects added:
    - Orchard fence, calibration tarps

Scene Ground Shots

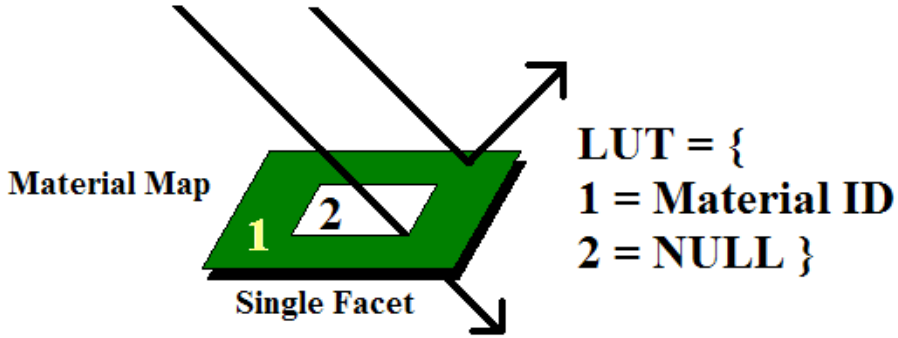
targets of interest can be placed under various levels of concealment by tree canopy

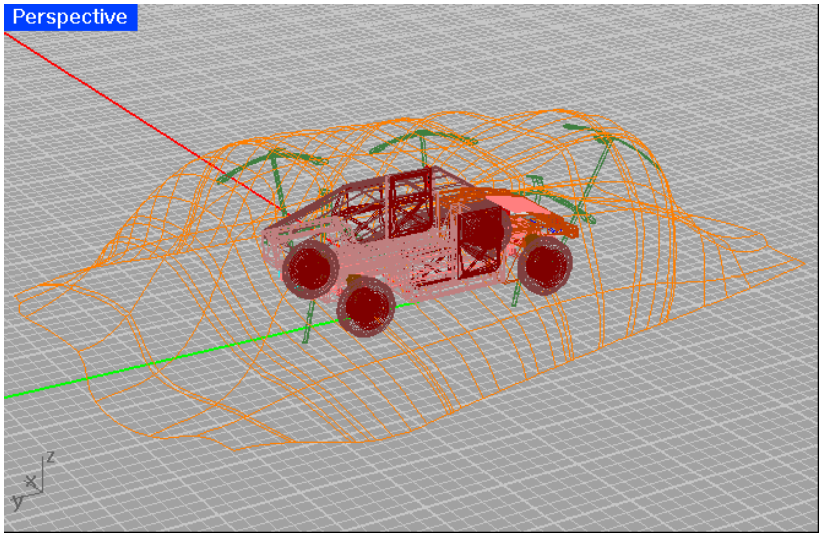
Concealed Targets for Validation

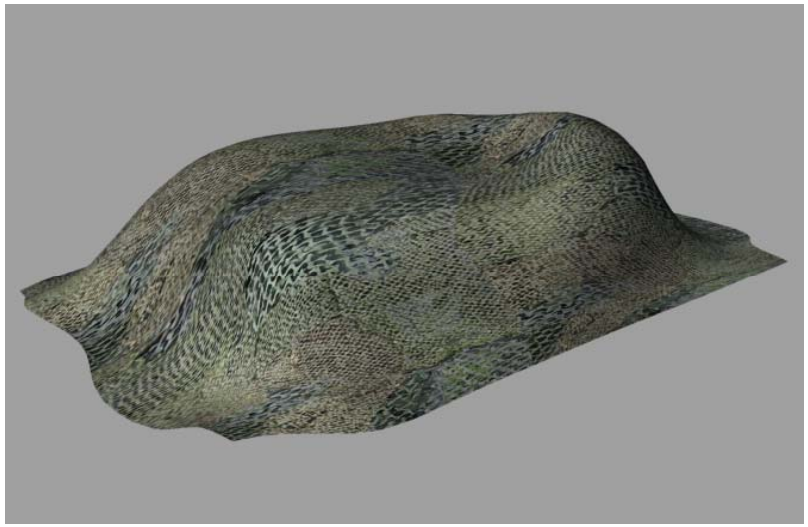
All targets and concealments characterized in the field

Camouflage Design

- Camouflage geometry “draped” over support structures
- Material Map made with NULL material (holes) and several camouflage materials

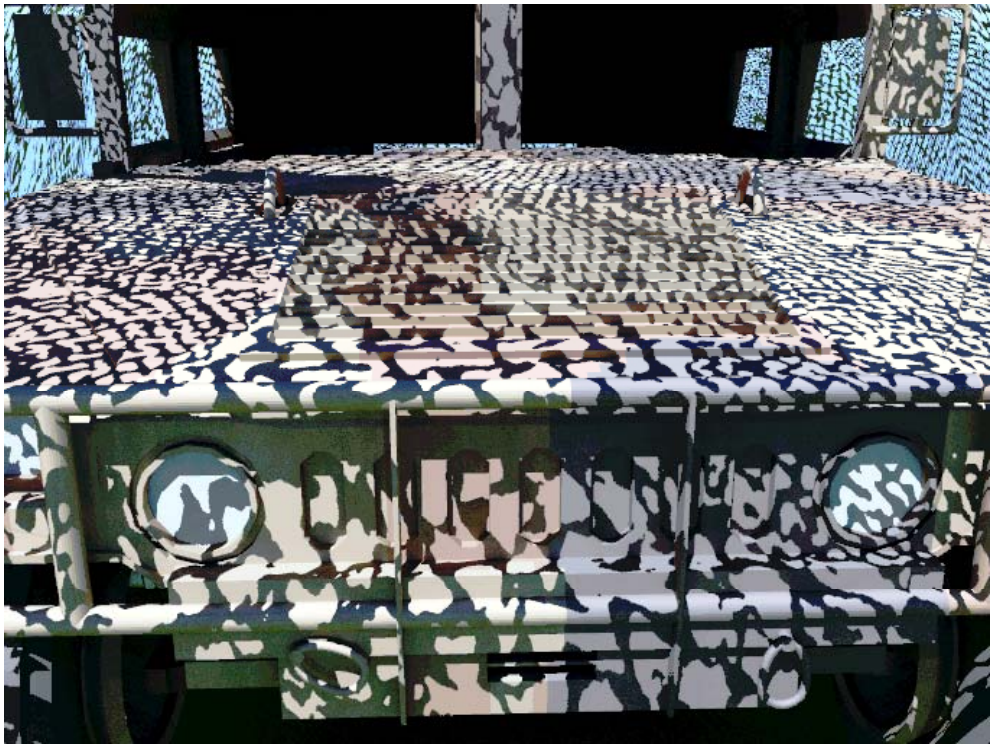






Camouflage Level of Detail

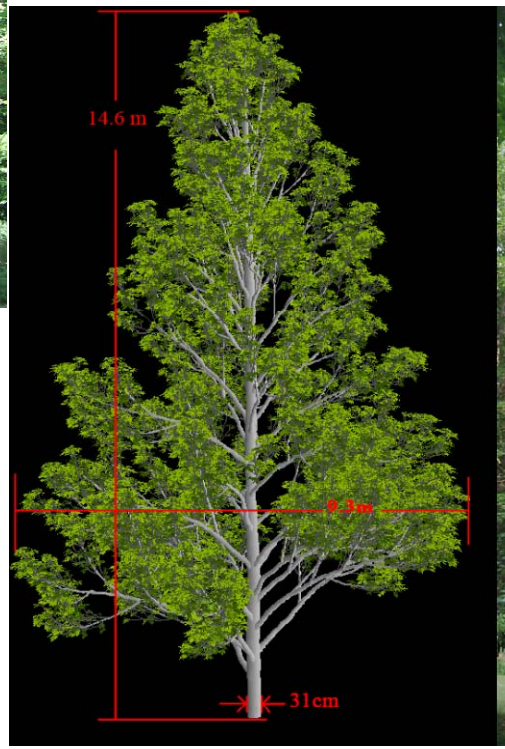
digital photo



simulation

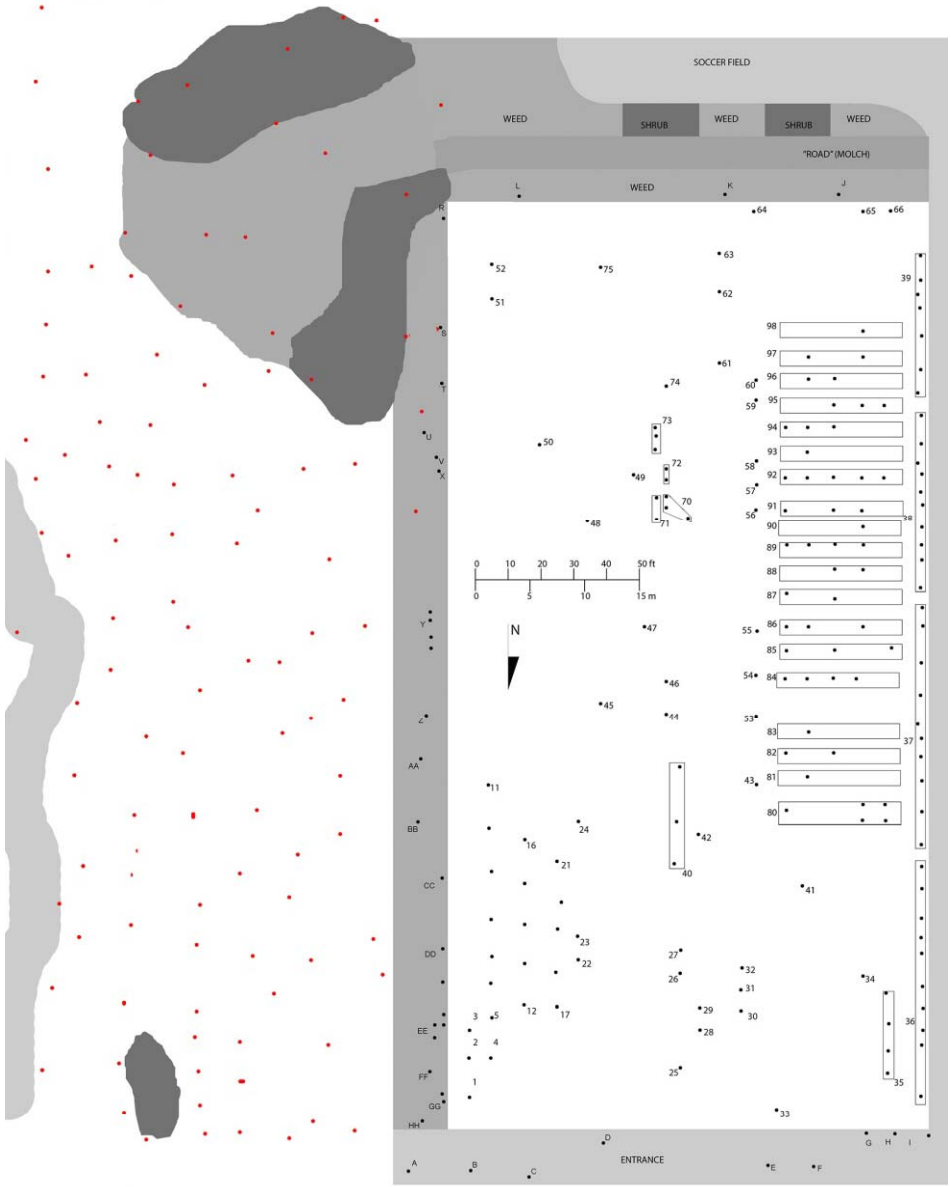
Tree Modeling Detail

Trees are modeled down to the individual leaf level



- Trees modeled in COTS software
- Spectral and geometrically fidelity sought
- Transmission through canopy is critical

Tree Placement Using GPS

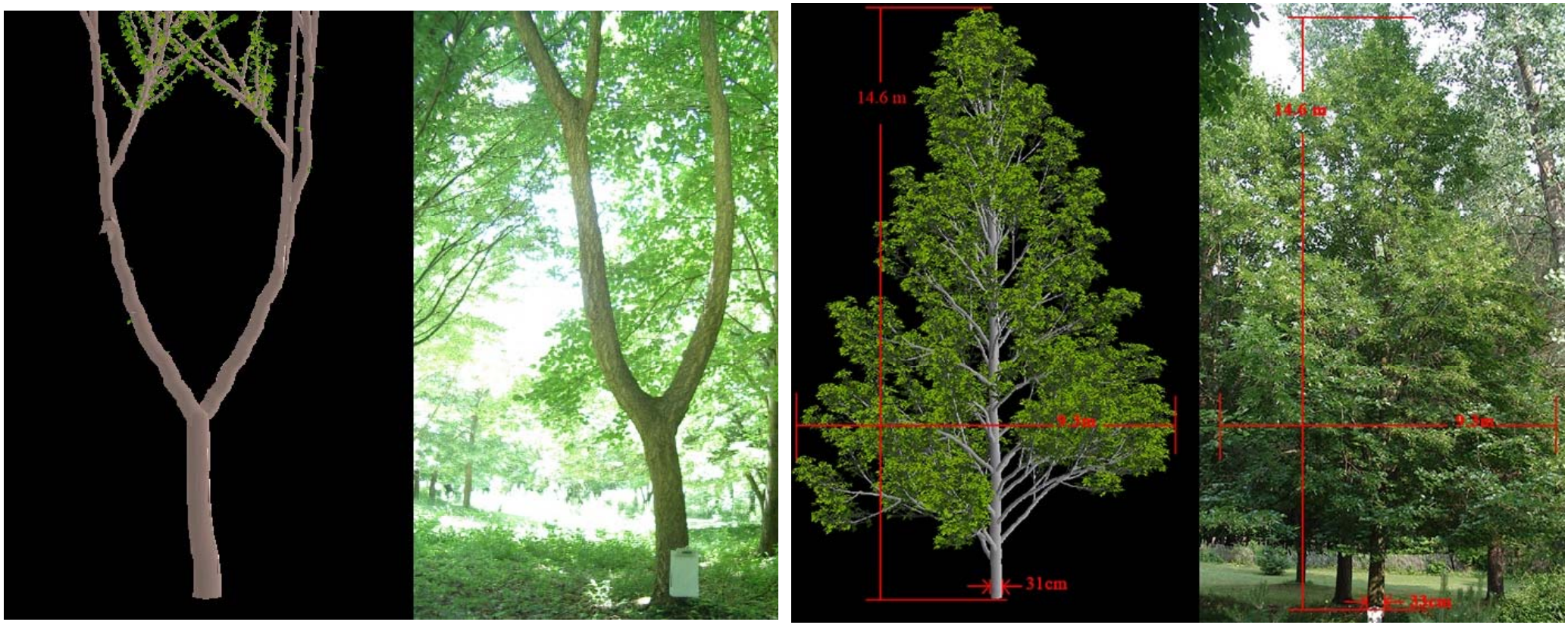


Building the scene

- Accuracy of the scene: examples

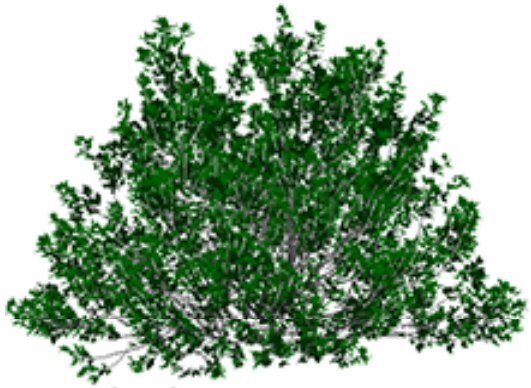
Building the scene

- Accuracy of the scene: examples



Building the scene

- Accuracy of the scene: examples



shrub_near_soccer_
field_1.6m.tif

shrub_near_soccer_
field_2.2m.tif

shrub_near_soccer_
field_(type2)_1.5m.tif

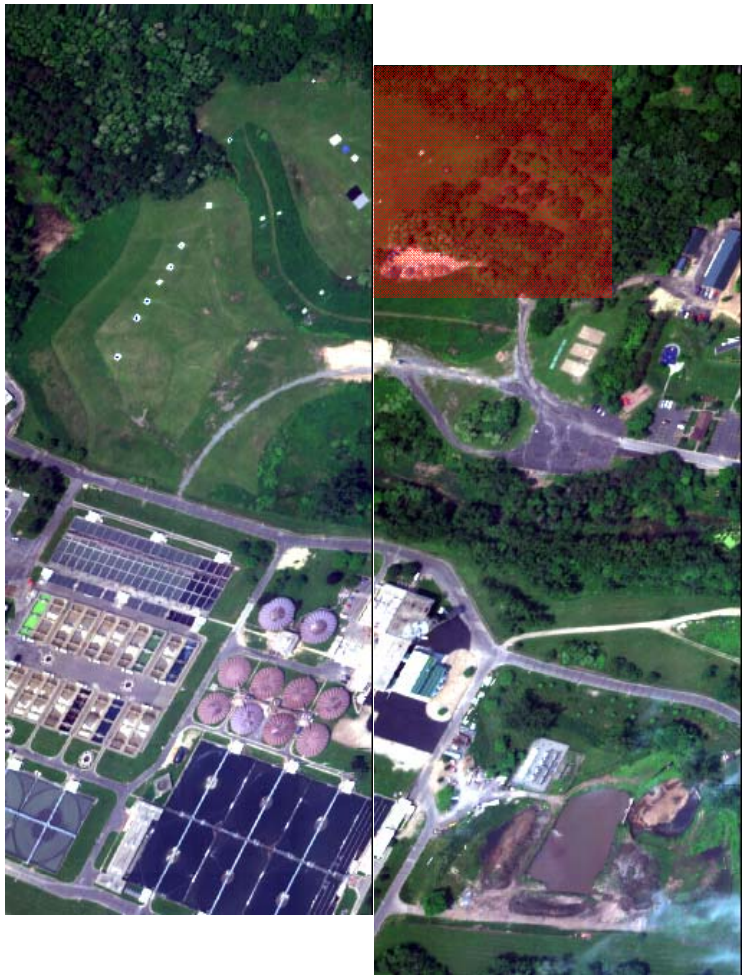
shrub_near_W_
fence_1.1m.tif

shrub_near_W_
fence_1.8m.tif

shrub_near_soccer_
field_(type2)_low.tif

Concealed Target Scene: Current Renderings

area of interest

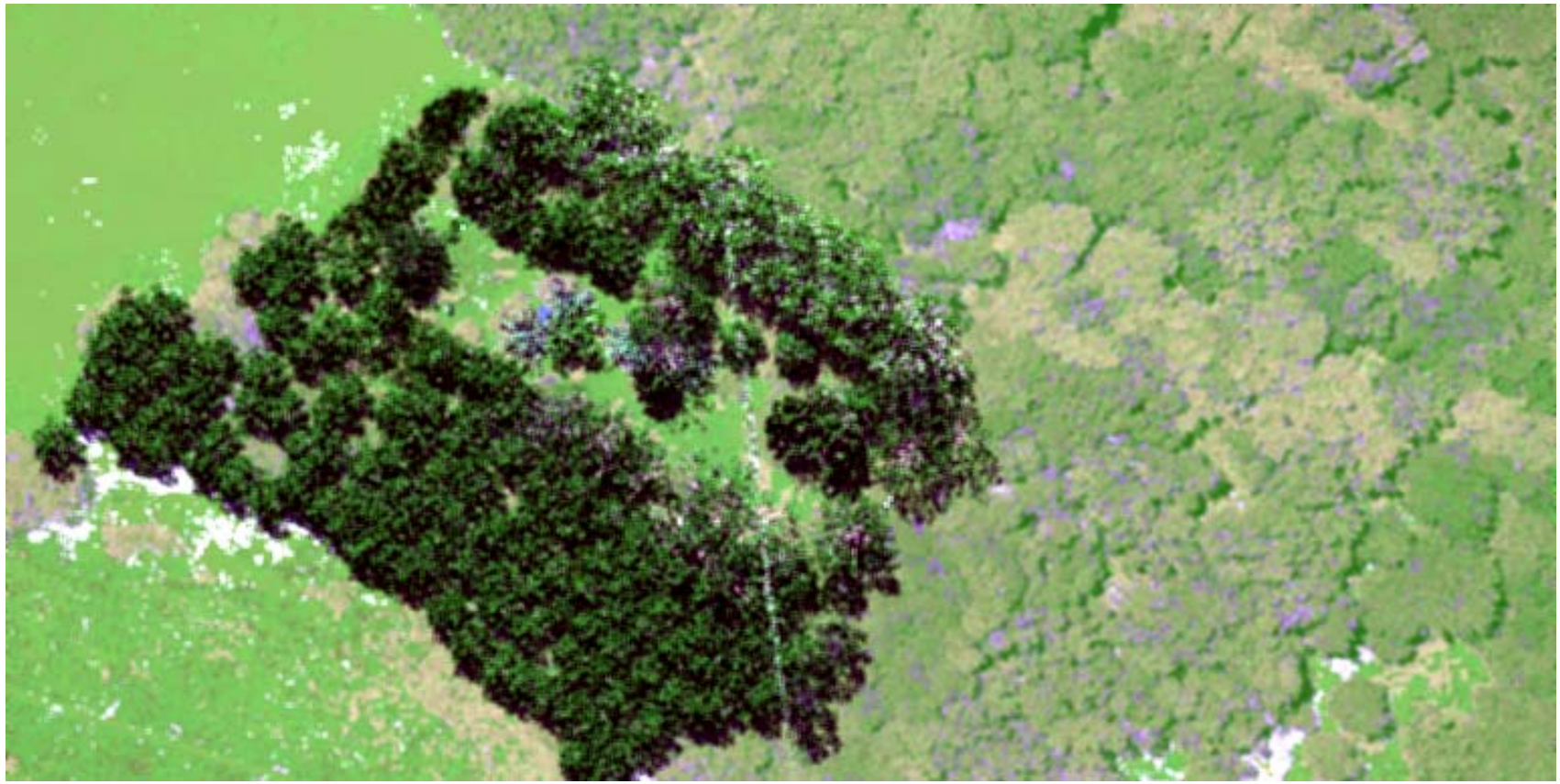


COMPASS

- Synthetic; base geometry
- Low - resolution
- Low - fidelity geometry (~ 1m)

Rendering testing

- Version 1
- Pushbroom sensor



Rendering testing

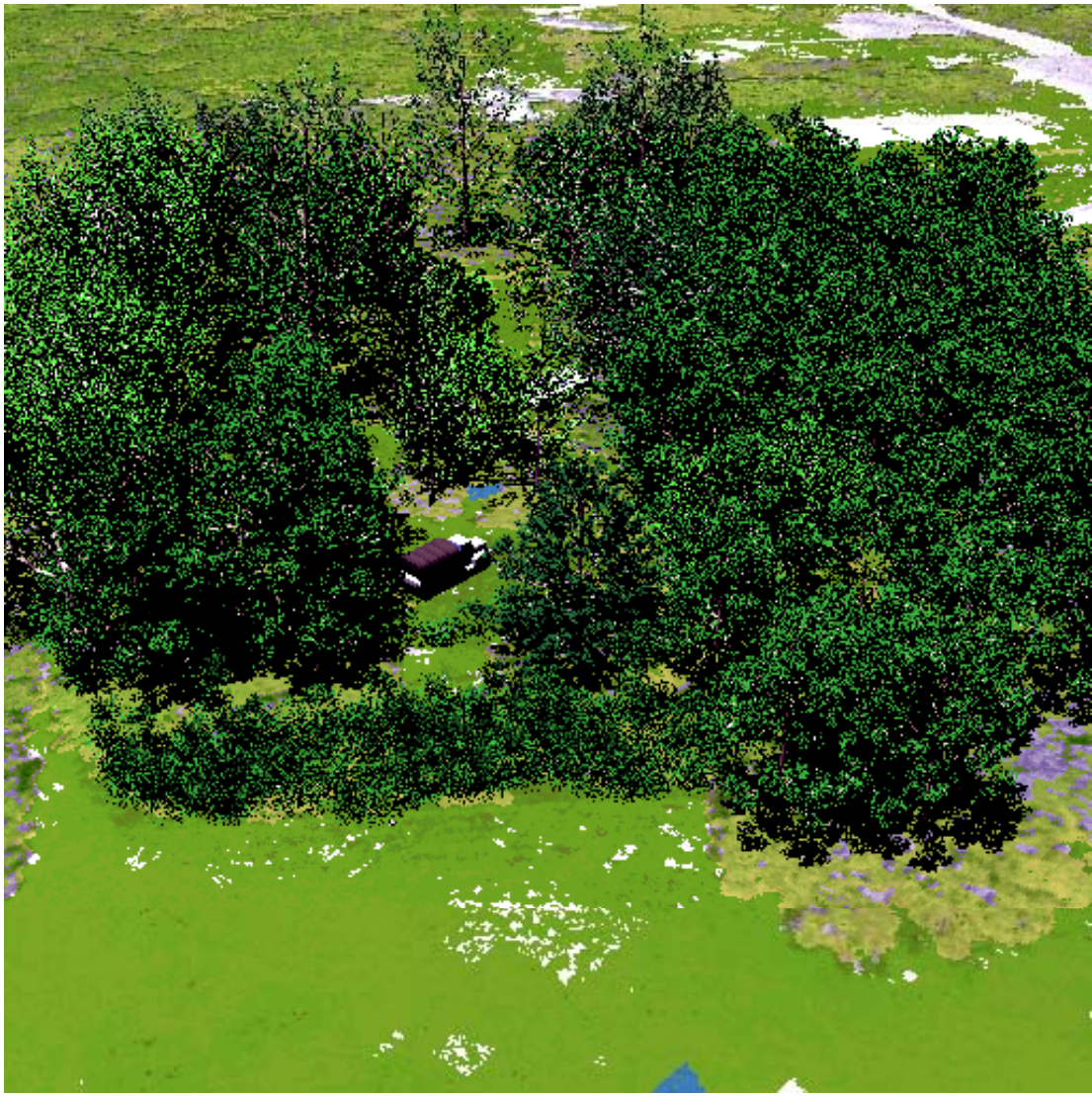
- Version 3
  - improved atmosphere
  - improved sensor modeling
  - improved scene modeling
- Embedded into larger scene

Concealed Target Scene: Current Renderings

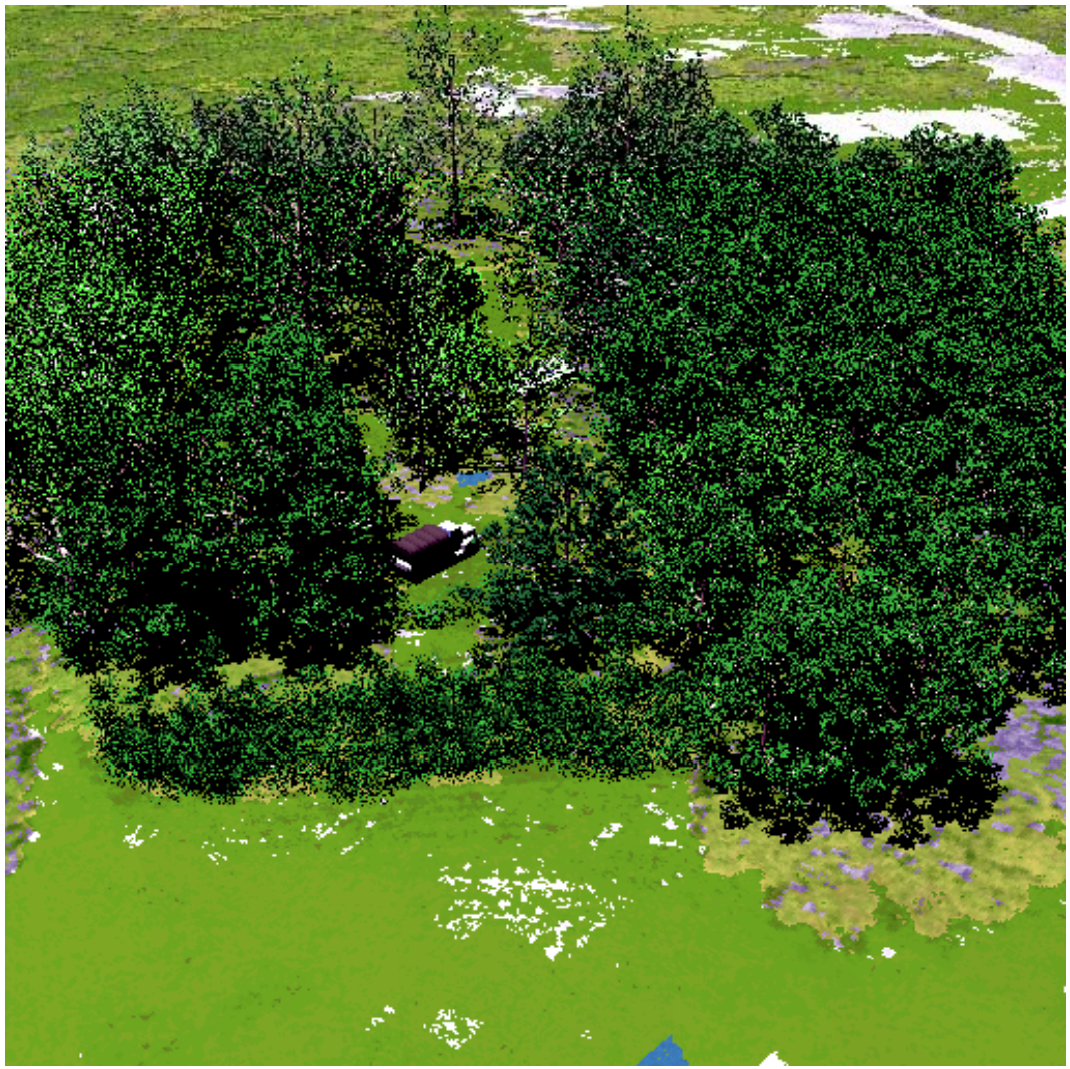
- Off-nadir closeup
- Essentially only some of the trees and surface materials
- Test of the overall geometric layout of scene
- Some trees are too “sparse”

Concealed Target Rendering

- Not placed within larger MegaScene
- Two vehicles placed in scene
  - canvas covered truck
  - tank
- Image is rendered full Vis/NIR/SWIR hyperspectral
  - 156 channels (RGB shown here)
  - 0.4 - 2.4 μm
  - GSD ~ 6in
- Forward - looking from altitude of ~ 500 ft



Concealed Target Rendering



- one target fully visible from this look geometry
- one target is partially concealed

Multiple Views

- Not placed within larger MegaScene
- Eight vehicles placed in scene
  - 4 canvas covered trucks
  - 4 tanks
- Vehicles
  - under trees
  - in field
  - along treeline in road

Multiple Views

Summary

- The Landmine scene was generated and delivered to the team
  - variations meant to stress the algorithms were created
  - obtained feedback about future scene generation
    - targets of variable SCR
- The Concealed Target scene was generated and delivered
  - tree models developed
  - improved texture
  - targets of interest embedded under various levels of concealment
- Validation effort conducted on both scenes

MURI
Science of Land Target Spectral Signatures

MURI – Related Doctoral Theses

Theses

- Manfred Karlowatz, Georgia Institute of Technology
  - *Part I From the Lab to the Field – Recent Developments in Polymer Coated ATR Sensing for the Determination of Volatile Organic Compounds* (2004)
- Joshua Broadwater, University of Maryland
  - *Physics-Based Detection of Subpixel Targets in Hyperspectral Imagery* (2007)
- Alina Zare, University of Florida
  - *Hyperspectral Endmember Detection and Band Selection using Bayesian Methods* (2008)
- Jeremy Bolton, University of Florida
  - *Random Set Framework for Context-Based Classification* (2008)
- Oladipo Fadiran, Clark Atlanta University
  - *Adaptive Sampling by Histogram Equalization: Theory, Algorithms, and Applications* (2007)

PART I

From the Lab to the Field - Recent Developments in Polymer Coated ATR Sensing for the Determination of Volatile Organic Compounds

A Thesis

Presented to

The Academic Faculty

by

Manfred Karlowatz

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy in Chemistry

Georgia Institute of Technology

May 2004

PART I

From the Lab to the Field - Recent Developments in Polymer Coated ATR Sensing for the Determination of Volatile Organic Compounds

PART II

From the Field to the Lab - Investigating IR Signatures for Remote Sensing Applications

Approved by:

Dr. Boris Mizaikoff, Advisor

Dr. Jiri Janata

Dr. Frank Loeffler

Dr. Andrew Lyon

Dr. Thomas Orlando

Dr. Marcus Weck

May 5, 2004

It isn't pollution that's harming the environment. It's the impurities in our air and water that are doing it.

George W. Bush

ACKNOWLEDGEMENT

My dearest thanks and deepest respect go to my supervisor *Professor Boris Mizaikoff* who provided me with the opportunity to work in the ASL group here at Georgia Tech. Amongst many other things, his outstanding motivation abilities, profound knowledge and amazing patience make him a “boss” I dearly enjoyed working with. He and *Dr. Christine Kranz* are an incredible team and their scientific input as well as the social events around the workgroup are highly appreciated. Their guidance and inspiration helped me to grow as a scientist and person (and this it not referring to the pounds I gained throughout my stay...). And of course: Chiarra and Camillo, your beautiful paintings I received all year round and sweets donations at my birthday are highly appreciated. Thank you kids (“morningside rules!”).

Many thanks go to the whole ASL team. In the years I had the pleasure to be a part of it, we have become great friends and undoubtedly all of them contributed to making every day at work an enjoyable one!

A special thanks go to the groups of *Prof. Jiri “Art” Janata* and *Prof. Thom Orlando* for the splendid cooperation, great advice and generous access to their instruments throughout my thesis.

Furthermore, *Dr. Michael Cathcart* is highly appreciated for giving me the opportunity to start working in the exciting field of remote sensing.

I would like to thank my wonderful family, my friends and Angie for the incredible amount of love and support I receive from them. Without you this all wouldn’t have been possible!

TABLE OF CONTENTS

Chapter

ACKNOWLEDGEMENT **IV**

LIST OF TABLES **X**

LIST OF FIGURES **XII**

LIST OF ABBREVIATIONS **XVIII**

SUMMARY **XX**

1. INTRODUCTION 1

1.1. VOLATILE ORGANIC COMPOUNDS IN AQUEOUS ENVIRONMENTS 1

1.2. SCOPE OF PART 1 2

2. BACKGROUND 3

2.1. VOCS DETERMINATION STATE OF THE ART 3

2.2. OPTICAL SENSING OF VOCS IN AQUEOUS ENVIRONMENTS 4

2.2.1. Improving Selectivity and Sensitivity 7

2.2.2. Approaches to Automated Sensor Calibration 9

2.2.3. An Overview of Recent Scientific Contributions to the Field of VOCs Determination in Water by means Optical Sensor Systems 12

2.2.3.1. UV/VIS Sensors 13

2.2.3.2. Near-Infrared (NIR) Sensors 18

2.2.3.3. Mid-Infrared (MIR) Sensors 21

2.2.3.4. RAMAN Sensors 29

2.2.3.5. Laser Fluorescence Sensors 33
2.3. MID-INFRARED SPECTROSCOPY 35
2.4. ATTENUATED TOTAL REFLECTION 36
2.4.1. Principle of Attenuated Total Reflection 36
2.4.1.1. Effective Layer Thickness 39
2.4.1.2. Waveguide Materials 42
2.4.2. Recent Approaches for VOC Determination via Evanescent Wave Sensing 43
2.5. POLYMER SENSOR MEMBRANES 44
2.6. IMPROVED MODEL FOR SIMULATING DIFFUSION-BASED DATA FOR OPTICAL CHEMICAL SENSORS 47
2.6.1. Model-Based Optimized Design of Polymer-Coated Chemical Sensors 51
2.6.1.1. Physical Arrangement and Model Formulation 51
2.6.1.2. Method of Solution 54
3. RESULTS 56
3.1. LABORATORY CONDITIONS – BTX IN WATER 57
3.1.1. Introduction 57
3.1.2. Experimental Setup 58
3.1.2.1. Materials 58
3.1.2.2. Instrumentation 58
3.1.2.3. Mixmaster 59
3.1.2.4. Preparation of the Extractive Polymer Membrane 61
3.1.3. Results 63
3.1.3.1. Water Equilibration 63
3.1.3.2. BTX Enrichment Characteristics 64
3.1.3.3. Test for Field Capability: Continuous Detection of o-Xylene in a Natural Pond Water Matrix 69
3.1.4. Conclusions 72
3.2. SIMULATED FIELD CONDITIONS – VOCS DETERMINATION IN AN AQUIFER 73
3.2.1. Introduction 73

3.2.2. Experimental Setup *74*

3.2.2.1. Materials 74

3.2.2.2. Silver Halide Fibers 74

3.2.2.3. Instrumentation 75

3.2.2.4. Aquifer simulation 78

3.2.3. Sensor Calibration and Validation *80*

3.2.4. Results *84*

3.2.5. Conclusions *88*

3.3. FIELD CONDITIONS – CHLOROBENZENE IN GROUNDWATER 89

3.3.1. Experimental Setup *90*

3.3.1.1. Instrumentation 90

3.3.1.2. Preparation of the Extractive Polymer Membrane 93

3.3.1.3. Sensor System Calibration 93

3.3.2. Results *95*

3.3.2.1. SAFIRA Measurement Surrounding Conditions 95

3.3.2.2. Chlorobenzene Enrichment Behavior 97

3.3.2.3. Sensor Regeneration 102

3.3.2.4. Long Term Stability 104

3.3.2.5. Dynamic Sensor Behavior 106

3.3.3. Conclusions *107*

3.4. MODELING THE DIFFUSION BEHAVIOR OF CHEMICAL SENSORS – HOW ACCURATE ARE EXISTING MODELS? 108

3.4.1. First Case Study: A Simple Diffusion Model applied to Experimental Data *110*

3.4.2. Second Case Study: A Numerical Simulation Model based on Fickian Diffusion applied to Experimental Data *115*

3.4.3. Third Case Study: CFD Simulations (FLUENT) applied to Experimental Data ... *120*

4. CONCLUSIONS AND OUTLOOK 122

5. FROM THE FIELD TO THE LAB – INVESTIGATING IR SIGNATURES FOR REMOTE SENSING APPLICATIONS 127

6. INTRODUCTION 127

6.1. LANDMINES – A GLOBAL PROBLEM 127

6.2. COMMONLY APPLIED LANDMINE DETECTION METHODS 129

6.2.1. Ground Penetration Radar 129

6.2.2. Ultrasound 130

6.2.3. Infrared Sensor 132

6.3. CURRENT DEVELOPMENTS IN REMOTE LANDMINE DETECTION – THE DISTURBED SOIL APPROACH 133

6.3.1. Remote Sensing 135

6.3.2. Hyperspectral Imaging 136

6.4. SCOPE OF THIS THESIS 138

7. BACKGROUND 140

7.1. MID-IR SPECTROSCOPY OF MINERALS 140

7.2. MID-IR SPECTRA OF GLASSES 145

7.3. MODE SPLITTING IN MIR-SPECTRA OF CRYSTALS AND GLASSES 146

7.3.1. Impact on this Thesis 152

8. EXPERIMENTAL 153

8.1. SAMPLES 153

8.2. LABORATORY SETUP 154

9. RESULTS 156

9.1. ATR SPECTRA OF POLYDISPERSE NATURAL QUARTZ 156

9.1.1. Experimental 156

9.1.2. Wetting / Drying 157

9.1.3. Polarized Light 162

9.1.4. Conclusions 165

9.2. ATR SPECTRA OF MONO-DISPERSE SODA LIME GLASS SPHERES 166

9.2.1. Samples 167

9.2.2. Experimental 169

9.2.3. Wetting / Drying 169

9.2.4. Particle Size Dependence of Absorption Features of Soda Lime Glass Spheres in ATR Spectra 173

9.2.5. Polarized Light 179

9.2.6. Conclusions 181

10. CONCLUSION AND OUTLOOK 182

10.1. ARE ATR SPECTROSCOPIC STUDIES SUITABLE AS SUPPORTING METHOD FOR REMOTE SENSING? 182

APPENDIX 185

REFERENCES 189

LIST OF TABLES

Table 2.1 Volatile organic compounds and their maximum concentration levels for drinking water as recommended by the EPA. Reproduced from [17]. 7

Table 2.2: Waveguide materials applied at ASL for MIR evanescent field sensing. 42

Table 2.3 Comparison of $t_{95\%}$ data (time required to extract 95% of analyte amount at equilibrium conditions) for extraction of different hydrocarbons from an aqueous solution with the use of a PDMS coated NIR fiber-optic sensor system. Reproduced from [42]. 50

Table 3.1 Overview on relevant spectroscopic approaches to BTX detection in water ... 68

Table 3.2 Statistical data of the sensor calibration. Intercept *a*, slope *b*, product-moment correlation coefficient *r*, prediction error SS, standard deviation of slope s_b, intercept s_a and limit of detection LOD. n=9. 83

Table 3.3 HS-GC validation measurements of groundwater sample from shaft 5 at the SAFIRA site. 96

Table 3.4 Comparison of the ATR measurements to the HS-GC measurements 101

Table 3.5 Calculated diffusion coefficients (cm^2/s) for CB in E/P-co 114

Table 3.6 Parameters from the Bitterfeld experiments included in the Friedson and Barbari algorithm 117

Table 3.7 Calculated diffusion coefficients for the Bitterfeld field measurements via two different methods 119

Table 3.8 Comparison of the parameters of the basic flow cell used in CFD simulations (Figure 2.13 and [43]) and the flow cell used during the Bitterfeld measurements. 120

Table 6.1 Worldwide landmine distribution and clearance status 128

Table 6.2 Speed of sound in different media [213] 131

Table 8.1 Measurement Parameters for ATR studies 154

Table 9.1 Relevant properties and chemical compositions of the soda lime glass spheres. 168

Table 9.2 Band assignments for soda lime glasses with similar composition [293,294,295] 170

Table 9.3 Revised band assignments for soda lime glass as suggested by Efimov [285].. The revised bands have been shaded for clarity. 172

LIST OF FIGURES

Figure 2.1 Transmission UV spectra of 50 mg/L of the BTX compounds in aqueous solutions in a 1 cm path length quartz cuvettes [82]. 14

Figure 2.2 Comparison of absorbance measured with a 100 µg/L solution of benzene in water (bottom spectrum, scaled 16 time in the dotted spectrum on top) taken with an optical pathlength of 1 cm and absorbance spectrum measured in PDMS block (0.2 cm optical pathlength) after immersion in the solution for 60 min [55]..... 17

Figure 2.3 Illustration of evanescent wave sensing principle (circle at the right bottom) and instrumental setup of coiled fiber-optic sensor and NIR bandpass filter photometer unit (fiber sensor element installed in a flow cell) [41]. 20

Figure 2.4 Response signals vs. time obtained with the NIR evanescent wave fiber-optic photometer system (fiber length 30m) for measurements of aqueous solutions of technical grade xylene (laboratory calibration of sensor system). Absorbance over time data are given for both measuring channels of the photometer located at central wavelengths of 1715 and 1645 nm [41]. 21

Figure 2.5 The most commonly used mid-IR transparent fiber-optic materials and their relevant properties for fiber-optic sensing. It can be seen that attenuation values for the long-wave range need to be further reduced for remote sensing applications. Transmission and attenuation data are the average of reported values [46]..... 23

Figure 2.6 Trace of carbontetrachloride (CT), trichloroethylene and tetrachloroethylene (PCE) in the effluent of a dechlorination reactor measured with evanescent wave ATR-FTIR spectroscopy. The ZnSe ATR element was coated with a polyisobutadiene extraction membrane. The lines show the predicted concentration values after PLS treatment of absorption spectra. Validation with GC shows good agreement [99]..... 26

Figure 2.7 3D illustration of the FT-IR underwater instrument. Main optics and electronics originate from a Bruker Vector 22 FT-IR spectrometer. IR radiation is launched into a flexible, polymer coated AgX fiber, which penetrates the aqueous medium facilitating highly sensitive and selective IR measurements in the marine environment [101]..... 27

Figure 2.8 Sensor dynamics shown for repetitive analyte injections into the flume tank 0.62 mg/L tetrachloroethylene, 0.77 mg/L 1,2-dichlorobenzene and 3.48 mg/L of the xylene isomer mixture were injected at t = 1 min and then every 30 min. The decreases of the readings after the maxima are attributed to dilution of the analyte plume in the solution and to analyte evaporation into the

surrounding atmosphere during measurements in the open flume tank system [101]..... 28

Figure 2.9 Background-subtracted Raman spectrum of a mixture containing 70 mg/L benzene, 100 mg/L toluene and 100 mg/L p–xylene. Spectrum recorded with a liquid core waveguide setup [116]..... 31

Figure 2.10 Schematic of a fiber-optic laser induced fluorescence instrument for in-situ detection of water pollutants. PD: Photodiodes. FI: edge filter. MO: monochromator. MCP: multichannel plate image intensifier. CCD: slow-scan CCD-line camera. ADC: analog-to-digital converters [132]. 35

Figure 2.11 Illustration of the ATR principle..... 37

Figure 2.12 Exemplary enrichment curves for CB (30 ppm) into a 10 µm E/P-co layer for different flow-conditions: 10 mL/min (triangles), 100 mL/min (diamonds) and 750 mL/min (squares) [166]..... 49

Figure 2.13 Schematic of the flow cell and coordinate system for the hydrodynamic model [43]. 52

Figure 3.1 Schematic of the combination of the Mixmaster with the ATR setup for dissolved BTX measurements. 59

Figure 3.2 Front view of the Mixmaster..... 60

Figure 3.3 IR absorptions resulting from water diffusion into an E/P-co membrane with a thickness of 4.2 µm coated onto the surface of a ZnSe ATR-crystal over a period of 32 h. After 24 hours (f) equilibrium conditions are reached and no further increase in absorption is observed. 63

Figure 3.4 IR absorption spectrum of a sample mixture in aqueous solution after enrichment into an E/P-co layer. Enrichment time: 20 min, concentration: 500 ppb (v/v) each. 65

Figure 3.5 Typical enrichment curves for the BTX components in water at a concentration level of 1 ppm (v/v) each into an E/P-co coating. Equilibrium of the diffusion process is reached after approximately 18 min of enrichment time. 66

Figure 3.6 Calibration graphs for benzene, toluene and the xylene isomers in the concentration range of 0 – 1000 ppb (v/v) based on peak area integration. The error bars represent the standard deviation of five subsequent measurements 67

Figure 3.7 Trace of the peak area of the absorption band of o-xylene at 740 cm^{-1} with time during enrichment based IR-ATR sensing. Concentration trace: 50 ppm; 80 ppm; 20 ppm (in pond water; the sensor was exposed to each

concentration for approx. 30 to 35 min) followed by neat pond water for sensor regeneration. 71

Figure 3.8 Close up of the sensor head, showing the active transducer section of the fiber (1), the o-ring seal lead-troughs (2), the two stainless steel tubes containing the sensor head/fiber cable interface (3) and the fiber cables (4). For the described measurements the semicircular part covering parts of the active transducer was removed. 77

Figure 3.9 Sketch of the aquifer simulation "Munich North" (side view). 79

Figure 3.10 Linear regression (solid) and confidence intervals (P=95 %; dash dotted) of the sensor response (circles). 82

Figure 3.11 Comparison of concentrations measured by IR sensor (red), reference HS GC (green with dots) and analyte concentration added to the water stream (blue). Confidence intervals (P=95%, n=9) at $\bar{y}$ are ± 0.6 mg/L (DCB), ± 0.7 mg/L (TeCE) and ± 2.3 mg/L (TriCE), respectively. 85

Figure 3.12 During a measurement campaign of three days the single beam spectra changed in shape and overall light throughput as a result of water intrusion. The first spectrum was recorded ten hours before the first experiment (i.e. series 1), the other three at the beginning of each series, respectively. For all spectra the sensor head was immersed in water. Note the strong decrease in light throughput on the third day leading to an impaired signal-to-noise ratio. 86

Figure 3.13 Left: Scheme of the custom made flow cell. Right: Picture of the flow cell (disassembled) 91

Figure 3.14 Schematic of setup for on-site chlorobenzene measurements. 92

Figure 3.15 The C-H out of plane vibration band of CB for 5 different concentrations after partitioning into the E/P-co layer. 94

Figure 3.16 Calibration curve for 3 repetitive measurements of the calibration set of CB. Error bars are derived from the standard deviation for each data point. 95

Figure 3.17 Exemplary spectra of a groundwater sample from shaft 5 (grey line) and a calibration solution of 50 mg/L CB in water (black line). Spectra were recorded after 24 min of exposure time to the polymer coated transducer. The peak area of the band at 740 cm^{-1} is used for data evaluation. 98

Figure 3.18 Enrichment curves of CB from groundwater at SAFIRA site into the E/P-co layer at a flow rate of 4 mL/min. The 3 measurements were performed at 3 different days. 99

Figure 3.19 Enrichment curves of CB from groundwater at SAFIRA site into the E/P-co layer at a flow rate of 23 mL/min. The 2 measurements were performed at 2 different days. 100

Figure 3.20 Exemplary spectra of a groundwater sample from shaft 5 (black line) and a successively recorded spectrum after regenerating the sensor with distilled water (grey line). Enrichment time and regeneration time were both 24 min with a flow rate of 4 mL/min. The peak area of the band at 740 cm^{-1} is used for data evaluation. 102

Figure 3.21 Enrichment and regeneration cycle for CB at 4 mL/min (squares) and 23 mL/min (diamonds). 103

Figure 3.22 Long term stability test for CB measurements in groundwater. The flow rate was set to 4 mL/min. The lack of data in the time period from 400 to 900 min is due to occupancy of the spectrometer by fiberoptic measurements..... 105

Figure 3.23 Simulation of a chemical spill event, by adding a significant amount of CB to the groundwater sample (at t=14min) after the sensor system was equilibrated with the groundwater sample. 106

Figure 3.24 Enrichment curves of CB from groundwater at SAFIRA site into an E/P-co layer at a flow rate of 4 mL/min. The 3 measurements were performed at 3 different days. 111

Figure 3.25 Enrichment curves of CB from groundwater at SAFIRA site into the E/P-co layer at a flow rate of 23 mL/min. The 2 measurements were performed at 2 different days. 112

Figure 3.26 $Abs_{(t)} / Abs_{max}$ versus $t^{0.5} / d$ plot for the 5 CB enrichment experiments. The series marked with diamonds and circles refer to the high flow rate (23 mL/min) the other 3 series to the lower flow rate (4 mL/min). 114

Figure 3.27 Exemplary results from the iterations for one data series obtained with low flow rate (4mL / min, Run3) and one with high flow rate (23 mL/min, Run4). The calculated diffusion coefficients are printed in the graph. 118

Figure 3.28 Time to steady state (equilibrium) vs. flow channel height for 2 different flow velocities modeled with CFD for the basic flow cell shown in Figure 2.13... 121

Figure 6.1 Principle of Remote Sensing Techniques 135

Figure 7.1 Infrared active internal vibrations of quartz. Left: v_3-asymmetric stretch Right: v_4-asymmetric bend []. 141

Figure 7.2 Approximate frequency range of common internal vibrations of silicates, oxides and other functional groups within minerals [245]. 142

Figure 7.3 General scheme for quantitative IR band assignment for glasses. Reproduced from [285]. 151

Figure 7. 4 Scheme for quantitative IR band assignment for glasses starting from IR reflectance measurements. Reproduced from [285]. 152

Figure 8.1 Experimental setup for cyclic wetting/drying studies of quartz sand via ATR spectroscopy in the MIR regime. 155

Figure 9.1 Overview of experimental procedure. 157

Figure 9.2 Exemplary IR-ATR spectrum of pristine quartz sand (Fluka 83340). The broad absorption feature with a maximum at around 1090 cm^{-1} is attributed to asymmetric stretching vibrations, the double peak at around 800 cm^{-1} relates to a symmetric stretching of the SiO_4 unit cell and the peak at 690 cm^{-1} is related to Si-O-Si bending vibrations. The inset shows an optical microscopy image of the sample. 158

Figure 9.3 Pristine (a), dried (b) and disturbed (c) spectra of quartz sand. The sharp band around 670 cm^{-1} results from atmospheric CO_2 present after opening the sample compartment. 159

Figure 9.4 Microscopy pictures of the finest fraction (left) and the coarsest section (right) of the investigated quartz sample. 160

Figure 9.5 Dried Spectra of a mixed (a), fine (b) and coarse (c) fraction of quartz sand. The spectra of the coarse fraction were scaled, due to very weak absorption features. 161

Figure 9.6 Schematic of agglomeration process of ultrafine quartz particles on the crystal surface during the wetting / drying process. 162

Figure 9.7 ATR spectra of pristine quartz sand samples recorded at different polarization states of infrared radiation: (a) unpolarized light (grey line), (b) p-polarized light (black line), (c) s-polarized light (dotted line). 163

Figure 9.8 ATR spectra of pristine quartz sand samples recorded at different polarization states of infrared radiation: (a) unpolarized light (grey line), (b) p-polarized light (black line), (c) s-polarized light (dotted line). 164

Figure 9.9 Optical microscopy images of soda lime glass spheres. The two smallest size fractions were obtained from the MO-SCI corporation (images on top). The larger size fractions were obtained from Whitehouse Scientific (images at the bottom). 169

Figure 9.10 The ATR spectrum of 112 µm soda lime glass spheres. 170

Figure 9.11 The ATR spectrum of 112 µm soda lime glass spheres (a) compared to a transmission spectrum (KBr pellet) of a 150 to 250 µm size fraction of soda lime glass with a very similar composition (b) (spectrum reproduced from [293]). 171

Figure 9.12 Pristine (a), dried (b) spectra of the 112 µm soda lime glass spheres. 173

Figure 9.13 ATR spectra of silica-gel particles with various diameters [296]..... 174

Figure 9.14 ATR spectra of soda lime glass spheres with different diameters: 400 µm (a), 112 µm (b), 25 µm (c). 175

Figure 9.15 ATR spectra of soda lime glass spheres with different diameters: 400 µm (a), 112 µm (b), 25 µm (c), 4 µm (d) and 1 µm (e)..... 176

Figure 9.16 Normalized (at 1040 cm^{-1}) ATR spectra of soda lime glass spheres with diameters of: 400 µm (a), 112 µm (b), 25 µm (c), 4 µm (d) and 1 µm (e) 177

Figure 9.17 FTIR spectra of $Na_2O–CaO–P_2O_5–K_2O–MgO–B_2O_3–SiO_2$ glasses with different SiO_2 content: (i) 66%, (ii) 59%, (iii) 55%, (iv) 50% and (v) 42% [299]. 178

Figure 9.18 S-polarized ATR spectra of soda lime glass spheres with different diameters: 400 µm (a), 112 µm (b), 25 µm (c), 4 µm (d) and 1 µm (e). Data has been normalized in intensity. 179

Figure 9.19 P-polarized ATR spectra of soda lime glass spheres with different diameters: 400 µm (a), 112 µm (b), 25 µm (c), 4 µm (d) and 1 µm (e). 180

Figure 10.1 Schematic of the environmental chamber developed for temperature and humidity studies on ATR measurements for quartz and other minerals. 187

LIST OF ABBREVIATIONS

AgX	Silver Halide
a-SiO2	Amorphous Silica
ATR	Attenuated Total Reflection
BTX	Benzene, Toluene, Xylenes
CB	Chlorobenzene
CFD	Computational Fluid Dynamics
DCB	Dichlorobenzene
E/P-co	Ethylene/Propylene Co-polymer
EM	Electromagnetic
FEWS	Fiberoptic Evanescent Wave Spectroscopy
FIA	Flow Injection Analysis
FID	Flame Ionization Detector
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
GPR	Ground Penetrating Radar
HPLC	High Pressure Liquid Chromatography
HS-GC	Head Space Gas Chromatography
IR	Infrared
IRE	Internal Reflection Element
LO	Longitudinal
LOD	Limit of Detection
MCL	Maximum Concentration Level
MCT	Mercury-Cadmium-Telluride
P&T	Purge and Trap
PCR	Principle Component Regression
PCs	Principle Components

ppb	Part Per Billion
ppm	Parts Per Million
RS	Raman Scattering
SIA	Sequential Injection Analysis
SPE	Solid Phase Extraction
SPME	Solid Phase Micro-Extraction
TCE	Trichloroethylene
TeCE	Tetrachloroethylene
TO	Transversale
US	Ultrasound
VOCs	Volatile Organic Compounds
ZnSe	Zinc Selenide

SUMMARY – PART I

The increasing interest in the research field of versatile chemical sensing systems is governed to a significant extent by the range of in-situ and on-line applications demanded in all aspects of modern instrumental analysis, such as industrial process analysis, environmental monitoring or pharmaceutical and biological/biochemical analysis. Common to these areas is the acceding effort to efficiently monitor and control various environmental, health and process related parameters with high molecular specificity. The increasing number of environmentally relevant pollutants and the demand for efficient methods to control industrial processes serve as a substantial argument for the development of rapidly responding, selective and reliable sensing devices.

Amongst the various physico-chemical transducer principles, optical sensing schemes have a promising potential as they provide the opportunity for remote sensing at a wide variety of conditions. Sensor systems operating in the mid-infrared (mid-IR) spectral region (approx. 2 – 20 µm) of the optical spectrum allow the availability of reliable and robust sensing systems with high inherent molecular specificity. Sensing applications in this spectral regime are particularly facilitated by direct evaluation of well-structured, molecule specific absorption bands resulting from the excitation of fundamental vibrational and rotational transitions of the analyte molecules.

Most mid-IR sensing approaches rely on a well established spectroscopic technique known as *attenuated total reflection* (ATR) spectroscopy probing analyte concentrations via interactions with the *evanescent field*. Along with continuous progress in the development of mid-IR-transparent optical waveguides, this method has enabled the extension of conventional IR spectroscopy towards field applicable spectroscopic sensing systems. Methods based on direct analyte interaction without chemical

modification of the waveguide surface are generally subject to interferences by IR absorbing sample components or the sample matrix itself, and generally suffer from limited sensitivity prohibiting their application in e.g. environmental trace analysis. Hence, the majority of mid-IR sensing approaches increase selectivity and sensitivity by modifying the waveguide surface with appropriate molecular recognition layers serving as solid phase extraction membrane for the analytes of interest, while simultaneously minimizing interferences of matrix components.

Contamination of drinking water, ground water and seawater with volatile organic compounds (VOCs) poses a significant health risk to humans and awareness of the public towards this matter increased considerably in recent years. Pollutants such as chlorinated hydrocarbons (CHCs), aromatic hydrocarbons (AHCs) and within the latter category especially benzene, toluene and xylenes (BTX) are among the most commonly detected organic contaminants in water. Consequently, benzene, chloroform and trichloroethylene occupy a permanent place among the 20 most relevant priority pollutants in the listings of the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA).

Standard methods for VOCs analysis include purge-and-trap (p&t) and static headspace (HS) gas chromatography (GC) combined with flame ionization detection (FID), among other GC techniques with hyphenated more sophisticated detection systems such as mass spectrometry. Complementary to these methods, solid phase extraction (SPE) techniques have been introduced for pre-concentration of environmental samples and chromatographic analysis after elution of enriched species with suitable organic solvents. Furthermore, the generally necessary sampling step for such laboratory based methods introduces a significant error source into the analysis procedure resulting from the physical properties of VOCs. Volatilization and diffusion losses make specific – and

usually expensive - sampling and storage procedures necessary. Therefore, on-site in-situ sensor systems are of particular demand in this area of environmental analytics. Preliminary works have already demonstrated the potential of evanescent wave MIR sensing for environmental monitoring. Particularly, zinc selenide (ZnSe) crystals and silver halide (AgX) fibers both coated with a thin layer of hydrophobic polymer led to promising results of such sensing schemes in recent years.

In this thesis, considerable efforts have been made to transition these devices from a laboratory environment to real world field applications detecting and quantifying VOCs in water. The presented work is divided into the following components, which ultimately led to the first successfully performed field measurement campaigns of IR evanescent field sensor system:

(i) Improvement of sensor calibration by introducing the “Mixmaster”, an automated mixing system based on sequential injection analysis (SIA) specifically adapted for accurate mixing and handling of dilute solutions of VOCs in water. Introduction of the *Mixmaster* facilitated repetitive evanescent sensor calibrations, along with more reliable and less error-prone preparation of calibration sets. Based on this system, simultaneous, quantitative detection of mixtures of BTX in water during enrichment into ethylene-propylene copolymers (E/P-co) coated onto ZnSe ATR elements has been performed. The obtained results showed accurate detection and quantification to the low ppb concentration region setting a new benchmark for laboratory based spectroscopic measurements for this group of compounds (published in *Analytical Chemistry, 2004, 76(9), 2643-2648).*

(ii) Fiber-optic evanescent field measurement campaigns based on E/P-co coated AgX fibers have been conducted at simulated field conditions at a simulated aquifer system located at the Technical University of Munich. Various VOCs

have been introduced into the water stream of the aquifer system and the concentration gradients of trichloroethylene (TCE), tetrachloroethylene (TeCE) and 1,2 dichlorobenzene (DCB) have been monitored with the sensor system. A fiber-optic sensor head in combination with a 6m long AgX fiber facilitated direct measurements in a borehole in the aquifer system, representing the first demonstration of remote groundwater monitoring by FT-IR based spectroscopic sensors. HS-GC validation measurements were in good agreement with the sensor data, although after 3 days increasing fiber degradation could be observed due to membrane delamination (published in : *Applied Spectroscopy, 2003, 57(6), 607-613* and *Water Science and Technology 2003 47(2), 121-126*).

(iii) For the first test of an ATR based polymer coated sensor system under real world field conditions measurements were performed at the SAFIRA site (German acronym for ‘‘Remediation Research in Regionally Contaminated Aquifers’’), a remediation pilot plant in the region of Bitterfeld / Wolfen (Saxonia-Anhalt, Germany). The applied sensor system consisting of an E/P-co coated ZnSe crystal mounted into a flow-cell designed and developed in the course of this work was used to accurately determine the chlorobenzene concentration in the Bitterfelder groundwater at mg/L levels. Validation was performed with HS-GC measurements. Different aspects of the sensor system including accuracy, repeatability, long term stability and dynamic behavior have been tested. An interesting aspect of these measurements was the experimental proof of the dependence of analyte extraction properties on the flow conditions of the sample matrix surrounding the extractive polymer membrane influencing the response time of the sensor system. These findings are in agreement with extensive computational fluidic dynamics (CFD) simulations, which have recently been presented by our group and collaborators. As a consequence, generally

accepted numerical models solely based on Fickian diffusion, which have been widely adopted to calculate e.g. diffusion coefficients of molecule / polymer combinations are doubted in their correctness for obtaining quantitative results (publications in preparation).

The obtained results demonstrate that MIR evanescent field sensors are suitable for in-situ analysis in environmental monitoring applications at real world field conditions.

SUMMARY – PART II

Landmine detection via remote sensing techniques is a challenging analytical and spectroscopic task. Efforts in detecting small buried objects aim at the combination of various spectroscopic techniques to assess changes in the spectral signatures of soils resulting from landmine insertion. For example, measurements of disturbed soils have shown different spectral contrast in comparison to undisturbed soils [1-5]. To date, these findings are predominantly based on experimental data obtained in real world environments using hyperspectral imaging systems. Hence, it is of great interest to fundamentally investigate the disturbed and undisturbed soil phenomena in a controlled environment. Based on these measurements reliable theoretical models can be established leading to improved interpretation of these features for landmine detection scenarios. In a first step, measurements at controlled laboratory conditions have been performed to investigate individual minerals of the soil matrix and their spectral characteristics at a variety of environmental conditions. Attenuated total reflection (ATR) spectroscopy has been identified as a suitable spectroscopic technique superior to emissivity or reflectance measurements, mainly due to its reproducibility and versatility, while contributing useful data toward fundamental understanding of spectral signatures relevant to remote sensing. Due to the high abundance in natural soils, pure quartz sand (SiO_2) has been selected as the first test matrix. For the investigation of spectral differences between pristine and disturbed quartz sand, a wetting/drying procedure with subsequent sample aerating has been developed, which in a first approximation represent, a sufficient simulation of weathering processes and their impact on related soil disturbances.

This first study could contribute substantial findings which despite of the potential usefulness have not been exploited for remote sensing data evaluation up to now.

Besides the already established differences in spectral contrast of disturbed and undisturbed soil, a strong spectral shift of the maximum of the main absorption feature at 1090 cm^{-1} could be observed. When probed with s- or p-polarized light, the quartz sample showed strong LO-TO mode splitting, which is most likely related to the Berreman effect. These findings advance the variety of spectral characteristics useful to the detection of disturbed soils (i.e. possible landmine sites) with mid-infrared imaging systems. The wetting and drying studies also reveal that the main reason for spectral differences of pristine and disturbed soils eventually relates to changes of the particle size distribution of the sample due to rearrangement of ultrafine particles facilitated by water *(in press 2004: Proceedings SPIE, 5415(Detection and Remediation Technologies for Mines and Minelike Targets IX)*

In a series of experiments mono-disperse soda lime glass spheres have been investigated at the same experimental conditions as the quartz samples in the preliminary study. By the application of these mono-disperse samples the aspect of a possible effect of various particle shapes in case of quartz samples was suppressed. It could be shown that no changes in the spectra during the wetting and drying cycles are apparent if only one size of spheres was applied. This corroborates the assumptions that a changed particle size distribution in the probed volume is the main factor for the spectral differences for disturbed / undisturbed soil systems.

Furthermore, strong spectral shifts and relative band intensity changes are observed when comparing spectra derived from different discrete particle size fractions. The most dominant relative band intensity changes could be assigned to a monotonously increasing non bridging oxygen Si-O stretch vibrational band in accordance with increasing sphere sizes. Measurements performed under linearly polarized light illuminations of the sample could corroborate this finding (publication in preparation).

The presented results advance the variety of spectral characteristics useful to the detection of disturbed soils (i.e. possible landmine sites) with mid-infrared imaging systems.

1. Introduction

1.1. Volatile Organic Compounds in Aqueous Environments

Contamination of drinking water, ground water and seawater with volatile organic compounds (VOCs) poses a significant health risk to humans [6-9] and awareness of the public towards this matter increased considerably in the recent years. Pollutants such as chlorinated hydrocarbons (CHCs), aromatic hydrocarbons (AHCs) and within the latter category especially benzene, toluene and xylenes (BTX) are among the most commonly detected organic contaminants in water [10-15]. As an example for the significance of such compounds as environmental pollutants benzene, chloroform and trichloroethylene occupy a permanent place among the 20 most relevant priority pollutants in the listings of the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) [16].

In a 1999 report of the Environmental Protection Agency (EPA) a quite unsettling finding was that seven of the 21 listed VOCs [17] occur in all 12 States studied in the USA, in either surface or ground water systems. Those were ethylbenzene, *cis*-1,2-dichloroethane, tetrachloroethylene, trichloroethylene, vinyl chloride, 1,1,1-trichlorethane, and xylenes. Many VOCs occur in up to 30 percent of surface or ground water systems in various States [18]. This again conveys the significant need for continuous monitoring of surface and ground waters to ensure the quality of drinking water supplies.

1.2. Scope of Part 1

Important pioneering work in the field of VOCs detection via polymer coated ATR-FTIR spectroscopic sensor systems originates from continuous research in the field of vibrational spectroscopy and chemical sensor technology formerly performed at the "Chemical Sensors Laboratory" at the Institute of Analytical Chemistry, Vienna University of Technology, and now at the "Applied Sensors Laboratory" at the School of Chemistry and Biochemistry, Georgia Institute of Technology. Starting more than a decade ago [19], the first principles of IR chemical sensing systems have been established and evolved into a comprehensive body of research with a substantial diversity of research areas and disciplines involved [36,38,46,59,60,73,76,84,92,93,95,96,100,101,139, ,145,166,174,175,192].

The objective of this PhD thesis was to facilitate the transition from extensive laboratory studies to simulated and real world field measurements with IR evanescent field chemical sensor systems. This challenging task was approached by fulfilling several milestones:

- Improved calibration possibilities by introducing an automated mixing system (*Mixmaster*)
- Quantitative and simultaneous determination of multi-component mixtures of VOCs with a polymer coated ATR sensor systems
- A measurement campaign at an aquifer system under "simulated field conditions" yielding promising results with a polymer coated mid-IR fiber-optic setup
- Accomplishment of accurate determination of the chlorobenzene concentration in a natural groundwater stream by the means of the proposed sensor system

Furthermore, very recent results from CFD simulations, which lead to questioning the generally accepted assumption that the signal generation kinetics for polymer coated evanescent wave sensor systems is solely based on Fickian diffusion of the analyte molecules into the thin extractive polymer layer. Calculations and model comparisons have been conducted in order to verify if nowadays commonly used methods.

2. Background

2.1. VOCs Determination State of the Art

Standard methods for VOCs analysis include purge-and-trap (p&t) and static headspace (HS) gas chromatography (GC) combined with flame ionization detection (FID), among other GC techniques with hyphenated more sophisticated detection systems [20,21]. Complementary to these methods, solid phase extraction (SPE) techniques have been introduced for pre-concentration of environmental samples and chromatographic analysis after elution of enriched species with suitable organic solvents [22,23]. A useful review summarizing analytical techniques for the determination of organic and inorganic chemicals in natural waters, wastewater, and drinking water has been published recently by Dietrich et al [24]. Classical analytical approaches are usually confined to a laboratory environment and require costly, error-prone and time consuming sampling procedures and/or include increasingly restricted usage of organic solvents.

Hence, considerable interest in developing analytical tools for the determination of such contaminants is evident prioritizing continuous operating in-situ devices capable of VOCs detection and/or continuous monitoring as well as quantitative discrimination at trace concentration levels. Continuous water quality monitoring requires qualitative and quantitative measurement of a wide range of adverse compounds in the liquid phase or

in the gas phase. It is estimated that worldwide daily 70.000 synthetic chemicals are used, including approx. 700 different organic constituents, making quality monitoring of for instance drinking water a challenging task [25]. Hence, there is a tremendous demand for continuously operating analytical systems and it is not surprising that chemical sensor technology is among the fastest growing disciplines in modern analytical chemistry. General introductions to chemical sensors can be derived from various books [26-28] and wide variety of applications of chemical sensors has been reviewed by Janata et al. over several years [29-34]. Besides electrochemical transducers, mass-sensitive devices and thermal sensing schemes, robust and versatile optical sensors gain significant importance for environmental monitoring, process control and, the biomedical field.

2.2. Optical Sensing of VOCs in Aqueous Environments

General aspects of optical sensing will be briefly discussed in this chapter. More detailed information on optical sensors and their applications can be found in [35, 36].

Depending on the field of application, optical sensing offers several advantages over other sensing concepts. The signal is optical and, hence, not susceptible to strong magnetic fields, surface potentials, or electrical interferences, e.g. by static electricity. Low-loss optical fibers allow the transmission of optical signals over long distances, enabling remote sensing. Miniaturization allows the development of small, lightweight, and flexible sensing devices. Furthermore, optical sensors are suitable for use in harsh environments like explosion hazard areas as encountered e.g. in mining and petroleum industries.

Over the past 20 years major developments in opto-electronics and fiber-optic communications have revolutionized the telecommunications industry by providing reliable high performance telecommunications links with ever decreasing bandwidth costs. Especially, high-performance silica-based glass fibers are of crucial importance or a rapid exchange of substantial amounts of data. As a consequence of the research in this industrial field component prices have decreased and quality has improved, the ability of fiber-optic sensors to displace traditional sensors for temperature, pressure, rotation, humidity, chemical measurements and other sensor applications has been facilitated.

Two groups of fiber optic sensing systems are generally distinguished:

1. Optical sensor based on direct detection of changes of optical analyte properties or spectral characteristics (*direct sensor*).
2. Chemical optical sensors based on a variety of analyte interaction/recognition/reaction processes at the sensor surface and optical transduction of chemical signals upon interaction of the analyte with the recognition element (*indirect sensors, indicator based-sensors*).

Frequently optical sensors are also classified as follows:

1. *Intrinsic sensors*, where the analyte directly interacts with the radiation transported in the optical fiber.
2. *Extrinsic sensors*, where the analyte affects the light properties while propagating in a medium external to the fiber (in this case the fiber acts only as a waveguide to transmit light to and from the active sensing region).

The combination of these concepts has been realized as so-called *physio-chemical sensors* taking advantage of both principles, e.g. sensors based on enrichment of analytes into a polymer membrane coated onto an optical fiber surface. Such sensor membranes may generate sensor responses, due to bulk changes of optical membrane properties (e.g. refractive index) or may act as a solid phase micro-extraction (SPME) membrane for enrichment of analytes in the vicinity of the waveguide surface utilizing the evanescent field analyte detection [37-41]. Response time and sensitivity depend mainly on the partition coefficients for the respective analyte between aqueous solution and polymer membrane. Hence, thorough investigation of polymer properties is required for fine-tuning and optimization of the sensor behavior. Reversibility of the sensor system is ensured since the enrichment is entirely based on diffusion without any chemical reaction inside the membrane. Thus, concentration fluctuations resulting from a shift of the partition equilibrium conditions can be continuously measured. Recently, interest was focused on improving the response time of such sensors by evaluating diffusion derived data at times prior to reaching equilibrium conditions [42]. Another interesting contribution for improved performance of such coated fiber sensor systems was published by Phillips et al [43] showing that also surrounding flow conditions of the sensor contribute significantly to the diffusion kinetics and should be taken into account during sensor development. Excellent reviews on the various applications of fiber-optic sensors [44-46] and on polymers used for fiber-optical sensors [47] have been published recently.

2.2.1. Improving Selectivity and Sensitivity

The quest for improved selectivity and sensitivity remains the cornerstone of optical sensor technologies or, more generally, for chemical sensors. In this section the requirements for a sensor (system) specifically for the determination of VOCs in water will be discussed. In order to illustrate the challenging task of quantitative and qualitative determination of VOCs in aqueous environments all 21 VOCs listed by EPA as of contaminants regarding drinking water regulations, together with their Maximum Contaminant Levels (MCLs) are shown in Table 2.1. The MCL represents the highest level of a contamination allowed in drinking water and is an enforceable standard.

Table 2.1 Volatile organic compounds and their maximum concentration levels for drinking water as recommended by the EPA. Reproduced from [17].

contaminant	MCL (µg/L)	contaminant	MCL (µg/L)
benzene	5	trans-1,2-dicholoroethylene	100
carbon tetrachloride	5	dichloromethane	5
chlorobenzene	100	1,2-dichloroethane	5
o-dichlorobenzene	600	1,2-dichloropropane	5
p-dichlorobenzene	75	ethylbenzene	700
1,1-dichloroethylene	7	styrene	100
cis-1,2-dichloroethylene	70	tetrachloroethylene	5
1,2,4-trichlorobenzene	70	toluene	1000
1,1,1,-trichloroethane	200	vinyl chloride	2
1,1,2-trichloroethane	5	xylenes	10000
trichloroethylene	5		

A successful sensor (system) should have the ability to discriminate between 230 µg/L of p-dichlorobenzene (3-times the MCL) or the same concentration of o-dichlorobenzene (still below the MCL limit), should be immune to or calibrated against cross-interferences and changing measurement conditions (such as other contaminants, temperature, pH etc.) and deliver robust performance over a long period of application time. It is needless to say that optical sensors technologies have not yet managed to accomplish all these requirements. As recently published works show, performances and diversity of optical sensors in this field are steadily improving (see chapter 2.2.3), however, many relevant analytical tasks still are calling for the need of high-end laboratory solutions. Nevertheless, in the last years optical sensors have reached a development state, which allowed first applications in process control [48], waste water analysis and remediation [41,49] and chemical spill detection [50].

Major improvements in sensitivity and selectivity of sensor systems in the last years can be generally attributed to one of the following approaches (or combinations thereof):

(i) Application of (semi) selective membranes for analyte enrichment directly on the surface of the waveguide [37-41] and detection via evanescent wave sensing, or analyte enrichment and direct measurement in the membrane with spectroscopic techniques other than evanescent wave sensing [51-55]. Thin layers of hydrophobic polymers usually enhance the limits of detection (LODs) for various VOCs by several orders of magnitude. Enrichment into such membranes follows the principles of SPME. Further information on SPME can be derived from [22,23,56]. For a theoretical treatment of the mass transfer of volatile organic compounds into membranes from aqueous solutions refer to [57].

(ii) Methods for evaluating data have improved significantly in the last decade including partial least-squares, principle component regression, cluster analysis, and computational neural networks just to mention some of the powerful chemometric methods nowadays applied for sensor (arrays) data evaluation. Recently published work [58-60] should be highlighted, which represents interesting contributions to the problem of uncalibrated features in data sets, a problem common to many sensing devices. For further information on modern data evaluation techniques please refer to recently published books [61-63] and publications [64-69].

2.2.2. Approaches to Automated Sensor Calibration

In general, well performed calibration procedures are a basic requirement for a reliably and accurately working sensor (system). For the special case of multiple component mixtures of VOCs in water at environmental relevant concentrations this “routine” task can be problematic for following reasons:

- Reproducible preparation of accurate aqueous solutions of volatile compounds at trace level concentrations is difficult due to evaporation losses.
- Storing of standards has to be performed headspace-free and cooled.
- In general, all VOCs have a strong tendency to enrich in any polymer matrix (see sensors based on solid phase extraction principles), which limits the materials used for storing, solution delivery, sealings, gaskets, flow-cell construction etc. to glass and metals.

Unfortunately, literally all existing sensors and sensor systems require extensive calibration (“training”) prior to delivering reliable results, especially in out-of-laboratory conditions. Sensor calibration can be even more tedious in case of multi-component sensing applications usually leading to a need for advanced data evaluation methods, which can cause the calibration set to substantially increase in size. Furthermore, regularly scheduled calibration of sensors and analyzer systems is usually a necessity and tedious. This creates a high demand for reliable methods of creating accurate calibration solutions, in high numbers and if possible automated in a short time, and consequently, computer controlled automated sample preparation systems. Whereas mixing units for gas analysis are readily available, there is a lack of instrumentation for accurate preparation of liquid samples, which is surprising considering the evident need for sensor calibration in VOC analysis. Only few approaches have been presented in the field of automated sample preparation devices suitable for such calibration tasks that can handle liquids at volumes (generally >mL) suitable for optical sensors.

A system using computer-controlled micropumps for automated sample preparation was presented by Lapa et al [70]. For each analyte one pump is required with the outputs confluent at a certain point. Behind, a mixing coil is located providing homogeneous solution of different analyte portions. The setup is based on a rather high repetition rate of complete pump strokes aspirating small volumes. However, the concentrations are not constant across the entire sample volume. On account of stacked analytes, there are obvious steps apparent in the concentration profile, which are considered unsatisfactory. Furthermore, the presented setup is able to handle only up to three analytes at a time. A multi-syringe flow injection analysis (MSFIA) approach which is based on a 4-syringe burette with valve switching between the analyzer side of the system and stock solutions was published by Albertus et al.[71]. There are two main issues limiting the versatility of

this approach:

1. The number of analytes is limited to the number of syringes
2. All syringes are moved by one motor in the same way and only the valves can be switched independently making composition of different analytes difficult.

An example for sensor assessment and calibration with an automated system for handling liquids has been presented by Richards et al [72]. This approach is based on a multi (diaphragm) pump system applied for extensive data generation in conjunction with an electro-chemical sensor for testing and validating calibration models. With this system 1668 experiments were produced in approximately 60 hours in comparison to a duration of more than 2 weeks that would have been required to perform the same amount of experiments manually. However, while such systems may be suitable solutions for certain sensor assessment applications they all lack the possibility for employment for VOCs, due to the extensive use of polymer parts (pumps, tubings, storage vessels etc).
The first system especially designed for sensor calibration tasks for handling VOCs at trace concentration levels was recently presented by our research group [73]. Properties of the mixing system are assessed by mid-infrared (MIR) attenuated total reflection (ATR) spectroscopy of MeOH-acetone mixtures and via multi-component samples containing 1,2,4-trichlorobenzene and tetrachloroethylene, which are enriched into an E/P-co layer (thickness approx. 2 µm). Recorded ATR spectra are evaluated by principal component regression (PCR) algorithms. The presented sample mixing device provides reliable multicomponent mixtures with sufficient accuracy and reproducibility at trace concentration levels. In the development of this mixing system special care has been taken to minimize losses of analytes either via evaporation (head-space free mixing, storing and transport of solutions) or diffusion (full glass syringes, stainless steel

tubings). A broader overview on flow systems and their potential applications is given by Rocha et al [74]].

2.2.3. An Overview of Recent Scientific Contributions to the Field of VOCs Determination in Water by means Optical Sensor Systems

In the following chapter fundamental and recent contributions in the field of optical sensing of VOCs in aqueous environments are discussed. The overview is not intended to be comprehensive or strictly dedicated to optical sensors, but will also include other optical methods which follow the rule of generating the information on the analyte via optical light analyte interaction. The following methods will be described in the review of this analytical field:

- (Fiber)-optical sensors and sensor systems
- (Classical) spectroscopic techniques
- Laser fluorescence techniques
- Other techniques

The criteria to include contributions in this section were merely based on the applicability for (semi-)continuous monitoring, (on-site) sensing and other related application purposes and that the sensor (system) would be able to provide selectivity and sensitivity to a certain extent (no sum-parameter devices). Already established sensor systems are described more detailed, but also designs that potentially will be able to perform under field conditions in near future as well as fundamental works in that field

are briefly mentioned. The listing is structured into traditional subdivisions such as Ultraviolet / Visible (UV/VIS), Near-Infrared (NIR), Mid-Infrared (MIR), Raman, and laser fluorescence methods concluding with a short outlook on future trends. Biosensors have not been considered, as no significant contributions to VOC analysis have not been reported yet, which would fulfill the requirements mentioned above.

2.2.3.1. UV/VIS Sensors

The application of sensor systems operating in the UV/VIS spectral range is usually restricted due to the problem of interference by other components and the fact that not all chemical species have significant absorption features in this wavelength domain. Despite the fact that for the aromatic part of VOCs UV/VIS spectroscopy represents a quite sensitive tool due to strong π-π* transitions of such molecules [75]. The reason for the rather small number of contributions can be found in the very broad appearance of the absorption features in that electromagnetic region, representing a certain lack of selectivity of the spectra.

UV/VIS sensing applications for VOCs in aqueous media are quite rare. However, there are two factors that may contribute to a higher interest in sensor development in the UV/VIS domain:

- Benefiting from the highly advanced telecommunication industry, lasers and specifically waveguide materials transparent in the UV/VIS region are available and rather cheap compared to most other optical sensing techniques and
- As already mentioned above, computational capacities and powerful (multivariate) data evaluation techniques make data utilization of the strongly

overlapping spectral features of organic compounds in the UV/VIS region much easier accessible.

Figure 2.1 shows exemplary UV spectra of the BTX group to illustrate the well pronounced but overlapping spectral features of aromatics.

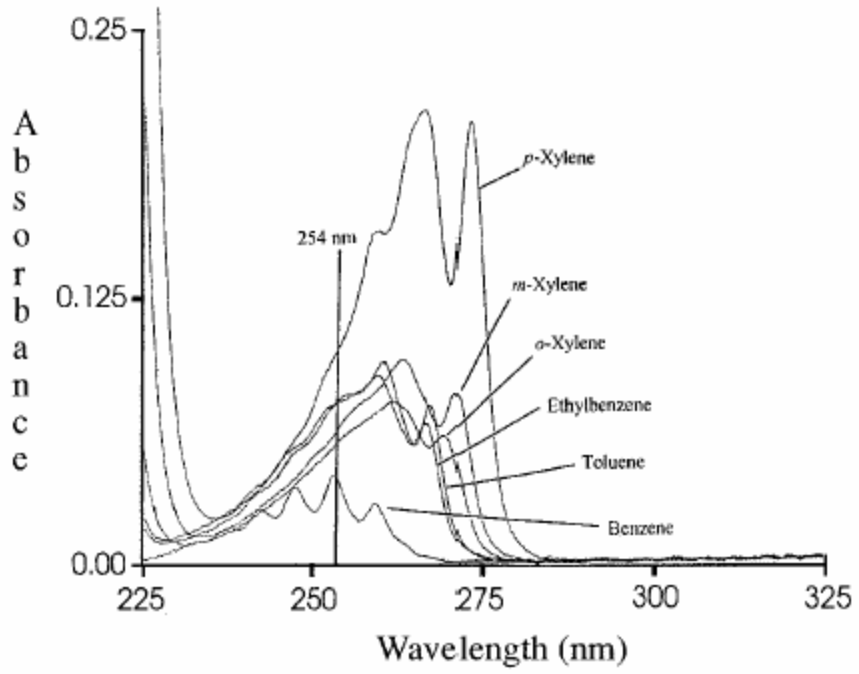


Figure 2.1 Transmission UV spectra of 50 mg/L of the BTX compounds in aqueous solutions in a 1 cm path length quartz cuvettes [82].

One example for successful implementation of innovative data utilization with traditionally obtained UV/VIS spectra of mixtures of BTX in water was published by Vogt et al [76]. An UV spectrometric method based on Ultraviolet Dynamic Derivative Spectroscopy (DDS) [77] is applied gaining selectivity and sensitivity by the use of optically generated first and second derivatives of transmission UV/VIS spectra. The

augmented spectroscopic technique is combined with chemometric algorithms like principal component regression or partial least squares, which are used for calibration of the spectrometer and quantification evaluation of spectra. Measurements were performed on mixtures containing up to 5 compounds including BTX, ethylbenzene, chlorobenzene and gasoline. The authors reported detection limits down to 50 µg/L for each analyte with a 10 cm absorption path-length and a few minutes measurement time. Apart from the inherent problems of transmission measurements including possible turbidity in real-world samples and strong interferences from uncalibrated contaminants, this approach shows the potential for on-line monitoring for all measured contaminants (with the exception of benzene) for drinking water quality [78,79].

Very recently, a miniaturized, submersible UV/VIS spectrometer for in-situ real-time measurements was presented by Langegraber et al [80] utilizing measurements in the spectral range of 200–750 nm for organic matter, suspended solids and nitrates in water. Main specifications of the spectrometric probe (dimensions: 44 mm diameter and approx. 0.6 m length) are as following:

- Measurement times of approx. 15 s, an auto-cleaning system using pressurized air - can be applied in 2″ bore holes (e.g. for groundwater monitoring)
- Utilizes a 2-beam spectrometer with a xenon lamp source - low power consumption (can be battery powered)
- Compromises a data logger on board (facilitates independent operation for one month at measurement time intervals of 30 min)
- Adjustable path length 2–100 mm for in-situ measurements

Applications have yet to be reported and already described problems with turbidity and cross interferences [78,79] may limit the field applicability of the device.

With a rather simple combination of UV/VIS transmission absorption spectroscopy with pre-selective enrichment matrices, as presented by Wittkamp and Hawthorne [52] and later shown again by Lamotte et al [55] problems with turbidity can be effectively reduced. In these contributions SPME related extractions of contaminants such as BTX and ethylbenzene etc. have been performed by applying thin (thicknesses in the mm regime) polydimethylsiloxane (PDMS) membranes as extraction matrices. UV/VIS spectra have been directly obtained with transmission spectra of the enriched contaminants in the PDMS matrix and rather low detection limits of few µg/L for most analytes could be achieved with total analysis times of less than 1 h. Figure 2.2 illustrates the enhancement in sensitivity after the extraction step for benzene as an exemplary analyte and also shows the generally broad spectroscopic features in UV/VIS spectroscopy.

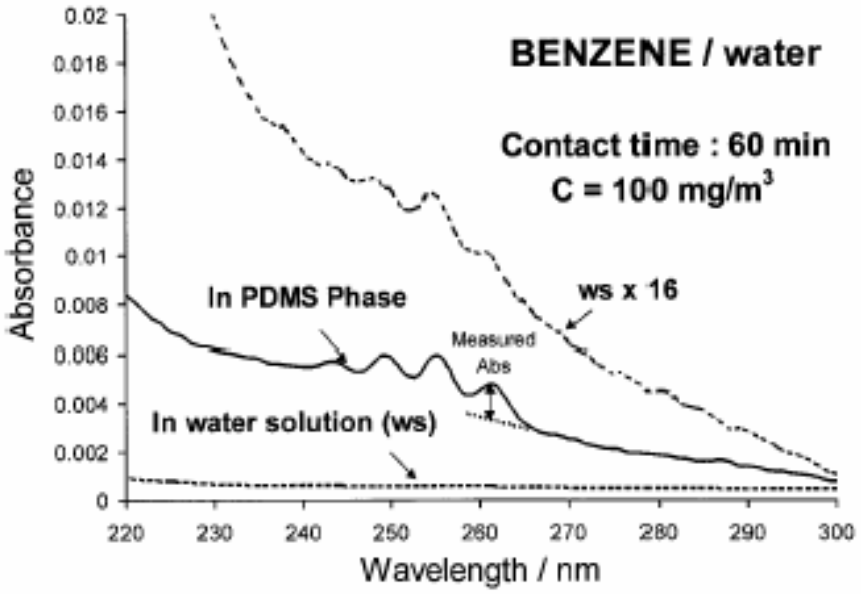


Figure 2.2 Comparison of absorbance measured with a 100 µg/L solution of benzene in water (bottom spectrum, scaled 16 time in the dotted spectrum on top) taken with an optical pathlength of 1 cm and absorbance spectrum measured in PDMS block (0.2 cm optical pathlength) after immersion in the solution for 60 min [55].

Apart from the fact that the extraction step significantly increased the sensitivity of the method, it also represents enhancement in selectivity as only rather hydrophobic and volatile contaminants are enriched in the PDMS matrix. However, a rather eminent disadvantage of this method is that at the presented stage of development only single component analysis can be performed due to the inherent strongly overlapping broad bands of practically all VOCs in the UV/VIS regime.

Few attempts have been described for VOCs determination in water via UV evanescent wave spectroscopy in conjunction with thin extractive membranes directly coated onto a sensing area of fiber-optic sensors based on the fundamental work of DeGrandpre and Burgess [81]. The only reported applications of this methodology for determination of VOCs in water were presented by Schwotzer et al. [40] and Mershman et al. [82]. Both rather similar approaches are based on a silica core plastic cladding optical fiber, where

the cladding was removed from a certain part somewhere in the middle of the fiber. This part of the fiber represents the sensing area, which is then coated with a PDMS layer (thickness in the µm regime). Analytes are enriched into this extractive membrane using a flow-cell setup and can be measured in the evanescent field if radiation is launched into the fiber. In both setups enrichment times were between 30 and 60 min, with reported limits of detection (LODs) of 10 mg/L for toluene [40] and also in the low mg/L regime for all BTX compounds and ethylbenzene as reported by [82]. Multicomponent measurements have not been performed. Due to the availability of low loss and low cost UV transmitting fibers, special applications in remote sensing for e.g. remediation processes or process control can be assumed possible.

2.2.3.2. Near-Infrared (NIR) Sensors

Plastic or silica-based glass fibers have been optimized by telecommunication industry during recent decades. Thus, robust and inexpensive optical fibers are available, which reached their theoretical attenuation limit of approximately 0.3 dB/km already in the late seventies [83]. Hence, fiber optic NIR liquid phase sensing at wavelengths < 2.5 µm utilizing overtone vibrational modes (e.g. C-H, N-H, O-H) for detecting organic compounds is a well established technology. Since overtone vibrations are in general 10-100 times weaker than corresponding ground vibrational modes in the MIR spectral range (3-20 µm), an active fiber/transducer length of 10-30 m is usually required for achieving sensitivities at trace contamination levels (µg/L). Another problem poses the limited discrimination power due to relatively unspecific absorption features in the NIR. These drawbacks are responsible for the lack of miniaturized sensor systems in this spectral region. On the other hand real remote sensing applications such as borehole

measurements can be performed due to the availability of virtually any length of fibers required. A recent general review on water quality monitoring via infrared optical sensors is given by Mizaikoff [84].

The combination of evanescent wave sensing with thin extraction membranes is widely applied in order to pre-concentrate analytes within the probed volume. The extraction membranes do also serve a second important purpose: due to their hydrophobic properties water is effectively excluded from the probed volume, reducing disturbing water background to a minimum as has initially be shown by DeGrandpre et al [81,85].

Major contributions in NIR fiber-optic sensing approaches for VOCs in water have been presented by Buerck et al [86]. PDMS coated multimode silica fibers with a low-OH quartz glass core (diameter 200 µm) have been coiled around a supporting rod for the determination of chlorinated hydrocarbons and aromatics such as trichloroethylene (TCE), toluene and p-xylene in aqueous solution during enrichment in the polymer layer with detection limits in the low mg/L region. The fiber-optic sensor could either be combined with either a NIR Fourier transform infrared (FT-IR) spectrometer or with a low cost filter photometer [87]. Using a similar setup Blair et al [88,89] showed the benefits of principal component analysis and partial least-square analysis as tools for the evaluation of such NIR spectroscopic data. Chemometrics were successfully applied to model the sensor's response to aqueous mixtures of TCE, 1,1,1-tricholorethane and toluene in concentration ranges from 20 to 300 mg/L.

Later, successful measurements of TCE in artificial aquifer systems and at field conditions have been shown with an improved portable NIR fiber-optic system with LODs just below the mg/L level [49]. In-situ measurements with the sensor system were performed in a groundwater circulation well [41], where the contamination with xylene was monitored over a time period of 4 months of continuous measurements.

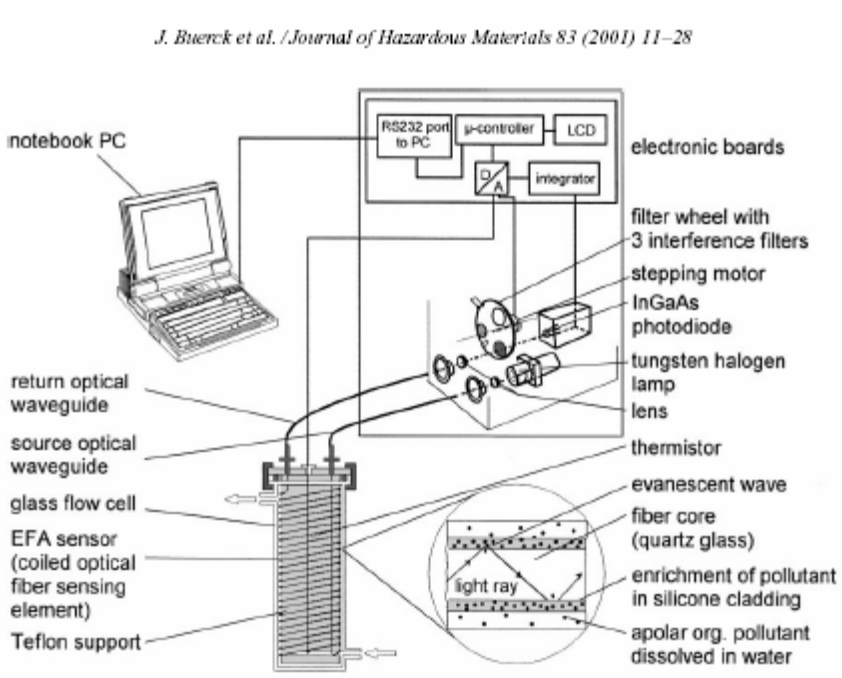


Figure 2.3 Illustration of evanescent wave sensing principle (circle at the right bottom) and instrumental setup of coiled fiber-optic sensor and NIR bandpass filter photometer unit (fiber sensor element installed in a flow cell) [41].

An illustration of the applied NIR fiber-optic sensor can be seen in Figure 2.3 and its dynamic response to technical grade xylene in a laboratory calibration experiment is shown in Figure 2.4. For the given parameters it takes about 20 min to reach equilibrium conditions for enriching xylene in the polymer layer and the analyte can be completely washed out again with water in a time frame of several minutes.

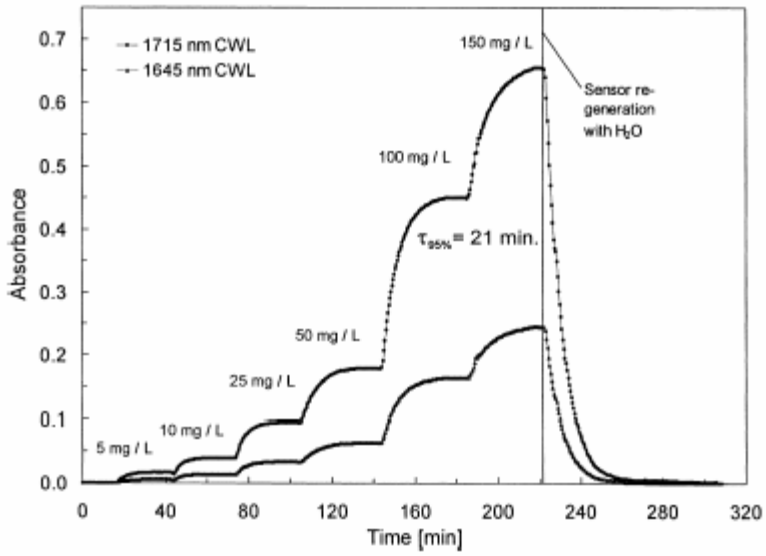


Figure 2.4 Response signals vs. time obtained with the NIR evanescent wave fiber-optic photometer system (fiber length 30m) for measurements of aqueous solutions of technical grade xylene (laboratory calibration of sensor system). Absorbance over time data are given for both measuring channels of the photometer located at central wavelengths of 1715 and 1645 nm [41].

A core-based intrinsic fiber-optic absorption sensor where the distal ends of transmitting and receiving fibers are connected by a small cylindrical section of optically clear PDMS has been developed by Klunder et at [90]. The PDMS acts as both a light pipe and a selective membrane into which VOCs are enriched during measurements. Measurement times of about 30 min and an LOD of 1.1 mg/L for TCE was achieved.

2.2.3.3. Mid-Infrared (MIR) Sensors

MIR spectroscopy operating in the spectral range from 2.3 to 25 µm is recognized as an analytical technique of persistently increasing importance. In contrast to the overtone vibrations in the NIR regime, MIR spectroscopy gives access to comparatively strong, distinct fundamental vibrational / rotational modes of organic molecules. This enables the opportunity to differentiate and quantify components according to their characteristic absorption bands. In respect to approaches for the determination of VOCs in water via

MIR spectroscopic techniques, predominately evanescent wave methodologies have been reported. This is related mainly to the fact that MIR transmission measurements in aqueous media are strongly hindered by broad and intense water absorption bands. Besides the measurement needs, the advancement and increased application of sensors based on optical waveguide technology is strongly coupled to the investigation and optimization of fiber optic materials transparent in the relevant frequency range. The rapid evolution of MIR sensors during the last decade can be mainly attributed to the development of appropriate fiber-optic materials, enabling the utilization of the wavelength range from 2 – 20 μm for sensing applications. Although the current performance of IR fiber-optic materials still requires significant improvement due to the comparatively high attenuation losses, some IR fibers and hollow waveguides are nowadays commercially available. The values presented in Figure 2.5 for the transmission range and minimum attenuation are average numbers retrieved from recently published material and shall provide a brief overview to the current state-of-the-art.

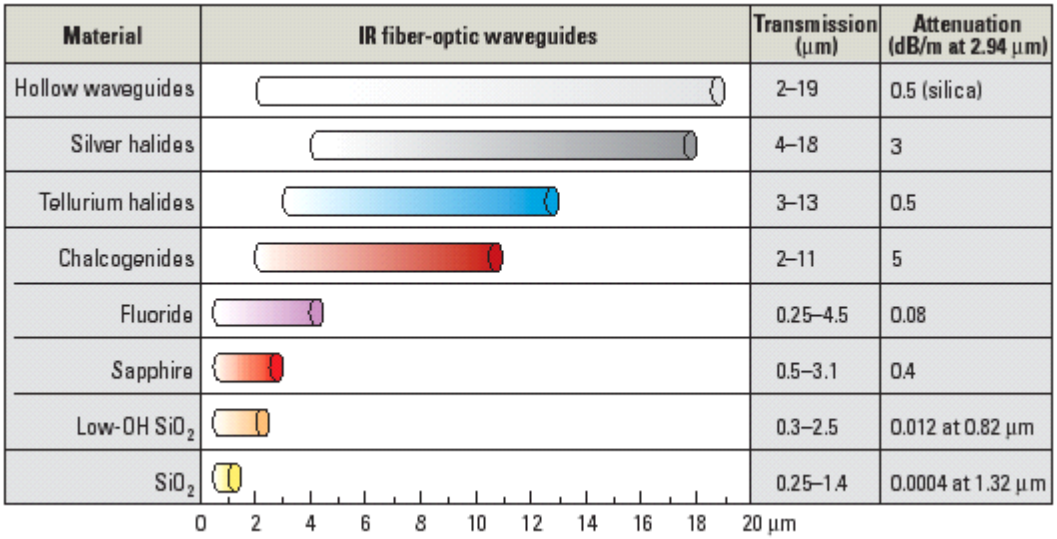


Figure 2.5 The most commonly used mid-IR transparent fiber-optic materials and their relevant properties for fiber-optic sensing. It can be seen that attenuation values for the long-wave range need to be further reduced for remote sensing applications. Transmission and attenuation data are the average of reported values [46].

Variations of these values can be mainly attributed to composition and fabrication variations of the reported materials. More details can be found in a number of reviews focused on the material properties of IR transmitting optical fibers [91,92].

Various fiber-optic based evanescent wave VOCs sensing systems applying polymer coated silver-halide (AgX) fibers (diameters: 700 to 1000 µm) coupled with FTIR spectrometers have been described during the past 15 years [37,38,39,93,94]. All contributions include coupling IR radiation of an FT-IR spectrometer into AgX fibers simultaneously acting as both, waveguide and active transducer. The active sensing area is coated with a thin hydrophobic polymer layer (thicknesses 2 to 10 µm), such as E/P-co, Teflon AF, polybutadiene, etc., enriching volatile organic pollutants. Water is effectively excluded from the measurement, since the selected polymer layer thickness is larger than the penetration depth of the evanescent field guided outside the optical

fiber. Based on this fiber-optic evanescent wave sensor (FEWS) scheme, qualitative and quantitative determination of a wide variety of organic analytes in the mg/l to the low µg/l concentration range has been demonstrated under laboratory conditions.

Very recently the first field application of a prototype MIR sensor system for the determination of VOCs in groundwater was developed and tested by our group [95]. The sensor comprises a portable FTIR spectrometer, coupled to the sensor head via AgX fiber-optical cable. A 10 cm unclad middle section of the 6 m long fiber is coated with E/P-co as extraction matrix, where evanescent wave measurements are conducted. A mixture of tetrachloroethylene, dichlorobenzene, and xylene isomers at concentrations in the low mg/L region was studied qualitatively and quantitatively in an artificial aquifer system filled with Munich gravel. This simulated real-world site at a pilot scale enables in-situ studies of the sensor response and spreading of the pollutants injected into the system with controlled groundwater flow and the analytes were clearly visible in the corresponding IR spectra. The results were validated by head-space gas chromatography using samples collected during the field measurement. The five analytes could be discriminated simultaneously, for two of the analytes the quantitative results are in agreement with the reference analysis. However, factors such as fiber long-term stability and time resolution have yet to be improved. With regard to application in real-world environments the accuracy of this method has been proven to be independent of aqueous sample turbidity, salinity or acidity at expected levels [96].

The most obvious reason for the extensive usage of AgX fibers for sensing applications is, , a transmission range down to the IR fingerprint region, as can be seen in Figure 2.5. However, certain limitations such as e.g. sensitivity to UV light and chemical susceptibility (to e.g. Cl^- ions) have to be taken into account for sensing applications [92] using these waveguide materials. However, evanescent fiber-optic sensing of VOCs in

water with other fiber materials have only been reported by Ertan-Lamontagne et al. [97] using PVC coated chalcogenide fibers and Howley et al. [98] applying PDMS coated sapphire fibers. Both contributions can be considered fundamental contributions in this field, however, suffering from rather high detection limits of tens of mg/L. Furthermore, due to the absence of evaluable data in the fingerprint spectral region, selectivity may be a permanent issue.

To date, field applications of such ATR-FT-IR sensor systems have been rarely reported. Acha et al. [48,99] developed an ATR-FT-IR sensor system for continuous online monitoring of a dechlorination process in a fixed-bed bioreactor without prior sample preparation. The sensor was based on an ATR ZnSe crystal (dimensions: 49 x 9.5 x 3mm) coated with a 5.8 µm thick polyisobutylene (PIB) extraction membrane, facilitating measurements of TCE, TeCE and carbontetrachloride (CT) at low mg/L levels in the aqueous effluent of a fixed-bed dechlorinating bioreactor. Several PLS calibration models were generated to resolve overlapping absorption bands of the chlorinated pollutants. Accuracy of this continuously monitoring ATR-FTIR sensor was validated with GC measurements. A graphical illustration of the results over time (Figure 2.6) shows that satisfying correlation is provided. Furthermore, the dechlorination process could be monitored without perturbation of any kind.

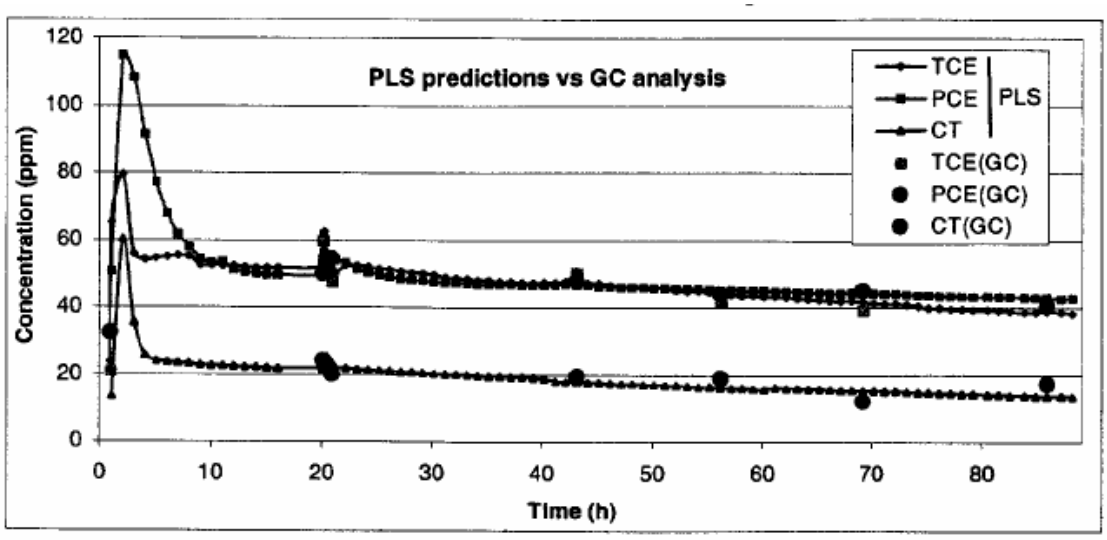


Figure 2.6 Trace of carbontetrachloride (CT), trichloroethylene and tetrachloroethylene (PCE) in the effluent of a dechlorination reactor measured with evanescent wave ATR-FTIR spectroscopy. The ZnSe ATR element was coated with a polyisobutadiene extraction membrane. The lines show the predicted concentration values after PLS treatment of absorption spectra. Validation with GC shows good agreement [99].

In order to overcome the limitations of remote sensing in the MIR regime, which are closely related to the absence of low loss waveguide materials, novel sensor systems have been developed. Kraft et al. and related papers [100,101] describe a sub-sea deployable fiber-optic sensor system for the continuous determination of a range of environmentally relevant VOCs in seawater. A suitable fiber-optic sensor head was developed using an E/P-co coated (thickness approximately 4 µm) 700 µm thick AgX fiber with approx. 38 cm active sensor length. The system was optimized in terms of sensitivity and hydrodynamics, and connected to the underwater FT-IR spectrometer. Figure 2.7 shows a 3D sketch of the underwater sensor system together with some specifications.

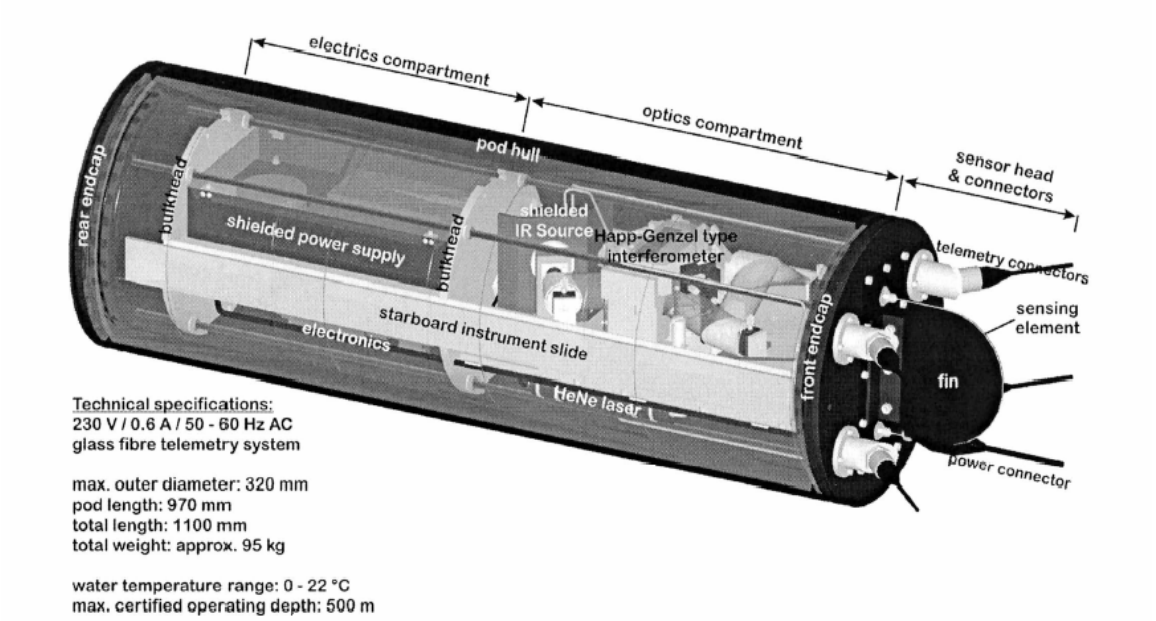


Figure 2.7 3D illustration of the FT-IR underwater instrument. Main optics and electronics originate from a Bruker Vector 22 FT-IR spectrometer. IR radiation is launched into a flexible, polymer coated AgX fiber, which penetrates the aqueous medium facilitating highly sensitive and selective IR measurements in the marine environment [101].

The sensor system was characterized in a series of laboratory and simulated field tests. Sensor characteristics from flume tank measurements with changing concentrations of xylenes, 1,2-dichlorbenzene and TeCE (Figure 2.8) show rapid dynamic response to changing concentrations. The sensor proved to be capable of quantitatively detecting a range of chlorinated hydrocarbons and monocyclic aromatic hydrocarbons in seawater down to the µg/L concentration range, including mixtures of up to 6 components.

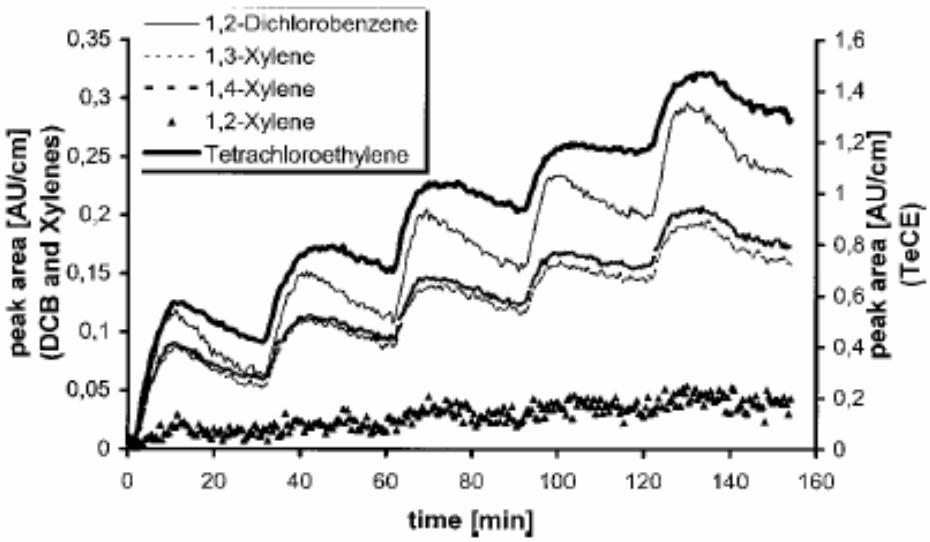


Figure 2.8 Sensor dynamics shown for repetitive analyte injections into the flume tank 0.62 mg/L tetrachloroethylene, 0.77 mg/L 1,2-dichlorobenzene and 3.48 mg/L of the xylene isomer mixture were injected at t = 1 min and then every 30 min. The decreases of the readings after the maxima are attributed to dilution of the analyte plume in the solution and to analyte evaporation into the surrounding atmosphere during measurements in the open flume tank system [101].

It has been demonstrated that varying amounts of salinity, turbidity, or humic acids, as well as interfering seawater pollutants do not significantly influence the sensor characteristics. A certain disadvantage for applications such as groundwater monitoring are the relatively large dimensions of the system. For borehole and other sensing applications, which require small systems, a prototype miniaturized MIR grating spectrometer operating in the wavelength range 8–12.5 µm has been reported [102]. The gain in applicability by miniaturization goes along with decreased performance (sensitivity), when dispersive technologies are applied for wavelength separation. Besides these drawbacks, the system is based on very similar principles as the submersible FTIR sensor system mentioned above. This measurement device may be used for organic contaminants in waste water, leakage fluids, and during remediation of

contaminated soils. However, to date the system was only tested at laboratory conditions where LODs for tetrachloroethylene around 1 mg/L could be obtained.

Besides evanescent field approaches very few alternative concepts have been reported for the MIR region. Heglund et al [103] and Merschman et al [53] applied SPME related extractions of contaminants such as BTX and ethylbenzene etc. by applying either polydimethylsiloxane (PDMS) or parafilm membranes (thicknesses around 130 µm) as extraction materials and followed by direct detection of the enriched organic analytes via FT-IR spectroscopy. LODs have been reported in the region of a few hundreds of µg/L and showed satisfying agreement with GC validation measurements [103]. However, strong absorption features of the SPME matrix renders parts of the IR spectra opaque due to total light absorption and evaporation losses during switching from extraction to detection of the analytes may be inconsistent and difficult to calibrate.

Alternative to the FT-IR approaches, the fiber-optic sensor head can be coupled to a tunable diode laser in order to enhance the sensitivity. Using lead salt lasers emitting in the MIR region a detection limit of around 22 µg/L for tetrachloroethylene has been reported [104, 105].

2.2.3.4. RAMAN Sensors

An alternative analytical method of great promise for VOCs determination in aqueous systems is provided by surface enhanced Raman scattering (SERS) [106, 107] and other variations of Raman spectroscopy with enhanced sensitivity [108]. Compared to NIR and MIR techniques, Raman spectroscopy has the inherent advantage that water only minimally interferes with the measurements. The application of Raman spectroscopy for studying environmental systems is rapidly expanding due to the

molecular specificity of this analytical technique enabling similar chemical identification as IR spectroscopy. Conventional Raman spectroscopy, however, has limited applicability for trace organic detection because of the inherently weak Raman scattering cross section. The SERS effect providing enhancement factors of up to 10^7 on Raman signals obtained from molecules adsorbed at rough metallic surfaces [106,109], has recently generated increasing interest for Raman sensing techniques. However, a limitation are potential interferences of fluorescence by various naturally occurring compounds (e.g. humic acids). Reviewing the contributions in this field would go beyond the scope of this work. More detailed information can be obtained in excellent reviews on this topic [110, 111]. Therefore, only the most relevant recent contributions to Raman based sensing will be briefly discussed.

Raman Spectroscopy

In general, conventional Raman spectroscopic techniques are not sensitive enough for most environmentally relevant concentration ranges of VOCs in water. Hence, efforts have been made developing methods to overcome this limitation. One approach published by Wittkamp et al [51] shows SPME related extractions of contaminants such as BTX and ethylbenzene etc. by applying small (diameters in the mm region) polydimethylsiloxane (PDMS) beads as extraction matrices and subsequent detection via Raman spectroscopy. Raman spectra of the enriched contaminants have been measured directly in the PDMS matrix and LODs from 1 to 4 mg/L for all analytes with a total analysis times of about 40 minutes were obtained. BTX spiked real water samples proofed the applicability of this method for field measurements.

A rather simple but effective method to increase the sensitivity of Raman spectroscopy is to design sample arrangements that increase the interrogated sample volume and thus

increase the intensity of the collected Raman signal. Walrafen et al [112] demonstrated that a hollow optical fiber enabling multiple internal reflections could be used to probe a large volume resulting in sensitivities increased by factors of 100-1000. This technology is also known as capillary waveguide or liquid-core waveguides (LCWs) spectroscopy. After the introduction of reliable, low-loss liquid-core waveguides based on Teflon-AF 2400 [113] applications of LCWs for Raman spectroscopy in general [114] and especially for VOC detection in water have been reported by several groups [115,116] with LODs in the low mg/L [115] or high µg/L region for benzene [116]. An exemplary Raman spectrum of a mixture of benzene, p-xylene and toluene recorded with a LCW setup is shown in Figure 2.9 illustrating the opportunity of multi-component detection. Apparently, the potential of this method is not fully exploited yet, and there is considerable scope that future contributions will present increased sensitivities rendering this method an interesting concept in the field of in-situ VOC analysis.

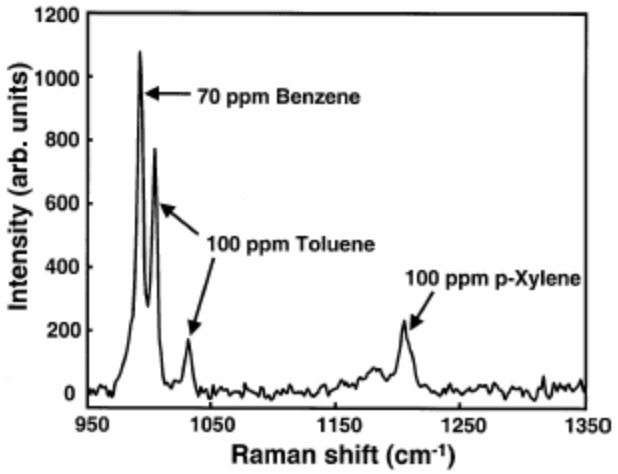


Figure 2.9 Background-subtracted Raman spectrum of a mixture containing 70 mg/L benzene, 100 mg/L toluene and 100 mg/L p–xylene. Spectrum recorded with a liquid core waveguide setup [116].

Surface-Enhanced Raman Scattering (SERS)

A series of contributions present a SERS based sensor system [117,118] applying SERS substrates that have been modified with different thioles in order to promote partitioning of VOCs into close vicinity of the SERS substrate from the aqueous phase. LODs for tetrachloroethylene and benzene of 12.6 mg/L and 7.5 mg/L have been reported. In more recent works, a SERS sensor was coupled with a flow injection analysis (FIA) system for molecular specific water analysis [119]. The flow-through cell incorporates a cascade geometry that is capable of accepting modified SERS substrates and its application to simultaneous detection of BTX applying PLS algorithms is shown. A LOD for benzene of 190 mg/L was achieved with FIA-SERS with improvements expected with future cell designs.

Recently, a laboratory-based system for measurement of organic contaminants in sea-water with sol–gel-derived SERS substrates was presented by Murphy et al [120-122]. By encapsulating silver colloids in a sol–gel-derived xerogel, SERS-active coatings were produced with high mechanical and chemical stability required for underwater field measurements. Photodegradation of the SERS layer was avoided by appropriate choice of optical components and layout. Continuous analysis was performed with two flow-through cells. The first design was a modification of a standard glass cuvette and the second an improved design an in-house-constructed aluminum cell. SERS investigations on samples with turbidities ranging from 0 to 400 NTU were performed with both cells. These tests show the suitability of the developed system for continuous monitoring of real world samples and its potential application in on-line process control. Sensitivity with a LOD of 100 µg/L for chlorobenzene are promising results for these first studies, however, the long-term stability of the substrates (half-life activity of the SERS substrate was 13 days in the best case) has to be improved for on-line monitoring systems.

2.2.3.5. Laser Fluorescence Sensors

Already 1977 Richardson et al. [123] showed that by applying lasers as excitation source for fluorescence measurements very low detection limits such as e.g. 19 µg/L for benzene as a single contaminant dissolved in water can be obtained. However, reviewing all contributions in this field would go beyond the scope of this work. An excellent overview on fluorometric determination of VOCs in water is provided in recent reviews [124-126]. In general only the aromatic compounds (BTX, ethylbenzene, chlorobenzene etc) show detectable fluorescent signatures. Chlorinated hydrocarbons (chloroform, trichloroethylene, tetrachloroethylene etc) have been successfully determined via laser induced breakdown spectroscopy (LIBS) [127] in the gas phase [128]. However, such approaches have been reported to be unsuitable to target those analytes in the liquid phase at relevant concentrations [129]. Furthermore, evident problems when analyzing multi-component mixtures such as quenching effects (e.g. “inner filter” effect [130]), low molar absorptivity and quantum yields of mono-aromatic compounds in comparison to poly-aromatic compounds leads to weak fluorescence emissions from BTX compounds and it is suggested that BTX detection is only useful via fluorometric methods if no other contamination is present [124].

The advantage of time-resolved laser fluorescent spectroscopy [131] for the analysis of environmental relevant aromatic compounds relates to the following beneficial parameters of this methodology:

- (Laser) sources and fiber materials are highly developed and cheap in the excitation wavelength range of (mostly UV).

- Aromatic compounds have a large absorption cross section in the UV and exhibit high fluorescence quantum yields.
- Using pulsed lasers fluorescence techniques, decay curves can be recorded providing additional and in some occasions selective information.

The latter is regarded as very important for environmental field analysis. In combination with effective multivariate data evaluation algorithms access to group- and molecule-specific laser fluorescence spectroscopy is provided.

One recent example for a system based on laser induced fluorescence spectroscopic measurements was selected to briefly describe the concept. A compact and mobile battery-operated laser induced fluorescence system has been presented by Karlitschek et al. [132] and related publications. The system is based on a diode-pumped solid-state laser with UV frequency conversion and a pulse duration of 7 ns. The third (355 nm) and fourth (266 nm) harmonics of the laser can be alternately used. The detection system consists of a polychromator, a gated image intensifier, and a CCD camera, which can acquire time-resolved spectra with nanosecond time resolution. A schematic of this system is shown in Figure 2.10. Fluorescence spectra, decay times, and LODs of 100 µg/L for benzene, 50 µg/L for toluene and 10 µg/L for xylenes have been measured for single contaminants in water.

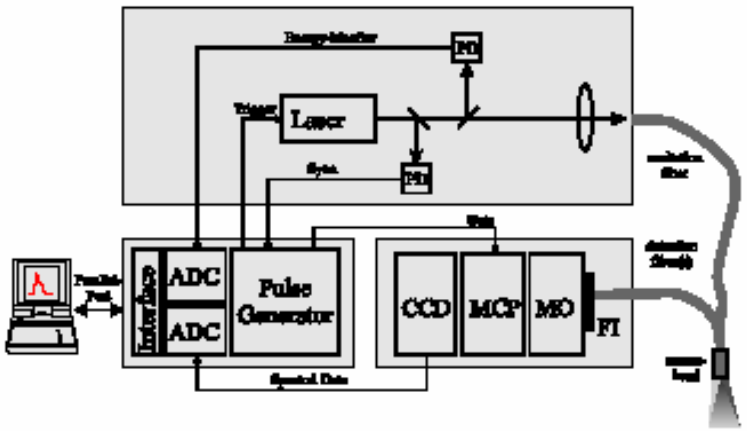


Figure 2.10 Schematic of a fiber-optic laser induced fluorescence instrument for in-situ detection of water pollutants. PD: Photodiodes. FI: edge filter. MO: monochromator. MCP: multichannel plate image intensifier. CCD: slow-scan CCD-line camera. ADC: analog-to-digital converters [132].

Results from contaminated groundwater samples show that molecular specificity can not be obtained with such sensor systems yet. However, the type of contamination can be classified in several groups (BTX, smaller PAHs, larger PAHs etc) due to the wavelength region of the fluorescence decay features and the decay times.

Alternative systems have been tested under field conditions [124,133, 134]. However, none of the approaches seems suitable for on-line sensing of BTX in water for lon-term field analysis. Even though VOCs can not be addressed via fluorescent methods reliably, such sensor systems are suitable devices for specific applications such as e.g. first assessment of contaminated sites or remediation control.

2.3. MID-Infrared Spectroscopy

The mid-IR (MIR) range covers the frequency regime from 4000 cm^{-1} (2.5 µm) to 400 cm^{-1} (25 µm). In this region of the electromagnetic spectrum, radiation stimulates

fundamental transitions between the ground state of vibrational and rotational modes of specific molecular bonds or whole molecules and their excited states. Consequently, information on the chemical functionalities and the type of molecule can be extracted. Depending on the strength of the bond, each mode is excited at a specific energy level manifesting as a characteristic band in the absorption spectrum. As vibrations of whole molecules usually require considerably lower excitation energies,, they produce highly substance specific absorption patterns at longer wavelenghts within the so-called fingerprint region (1200 cm^{-1} – 400 cm^{-1}). MIR spectroscopy is recognized as an analytical technique of persistently increasing importance and is widely used in the analytical community.

However, a major restriction of transmission based MIR measurements of VOCs in aqueous solutions are the broad and pronounced absorption bands of water in this spectral region. Furthermore, field applicability of transmission based methods is restricted due to the potential influence of turbidity in real-world samples. Hence, increasing efforts are focused on optical principles enabling molecule specific determination of organic pollutants in water, such as attenuated total reflection infrared spectroscopy (ATR-IR) [36,] in combination with extractive polymer membranes.

2.4. Attenuated Total Reflection

2.4.1. Principle of Attenuated Total Reflection

The ATR principle derives more generally from internal reflection spectroscopy and has been independently described by Harrick [135] and Fahrenfort [136] in the early 1960ies. Radiation which travels in the high refractive index ATR-element is incident at the

interface between the waveguide and a surrounding medium with lower refractive index with an angle θ. At angles $\theta > \theta_c$, where

$$\theta_c = \arcsin \frac{n_1}{n_2} \qquad (2\text{-}1)$$

is the critical angle, radiation is internally reflected. The principle of ATR spectroscopy is schematically shown in Figure 2.11.

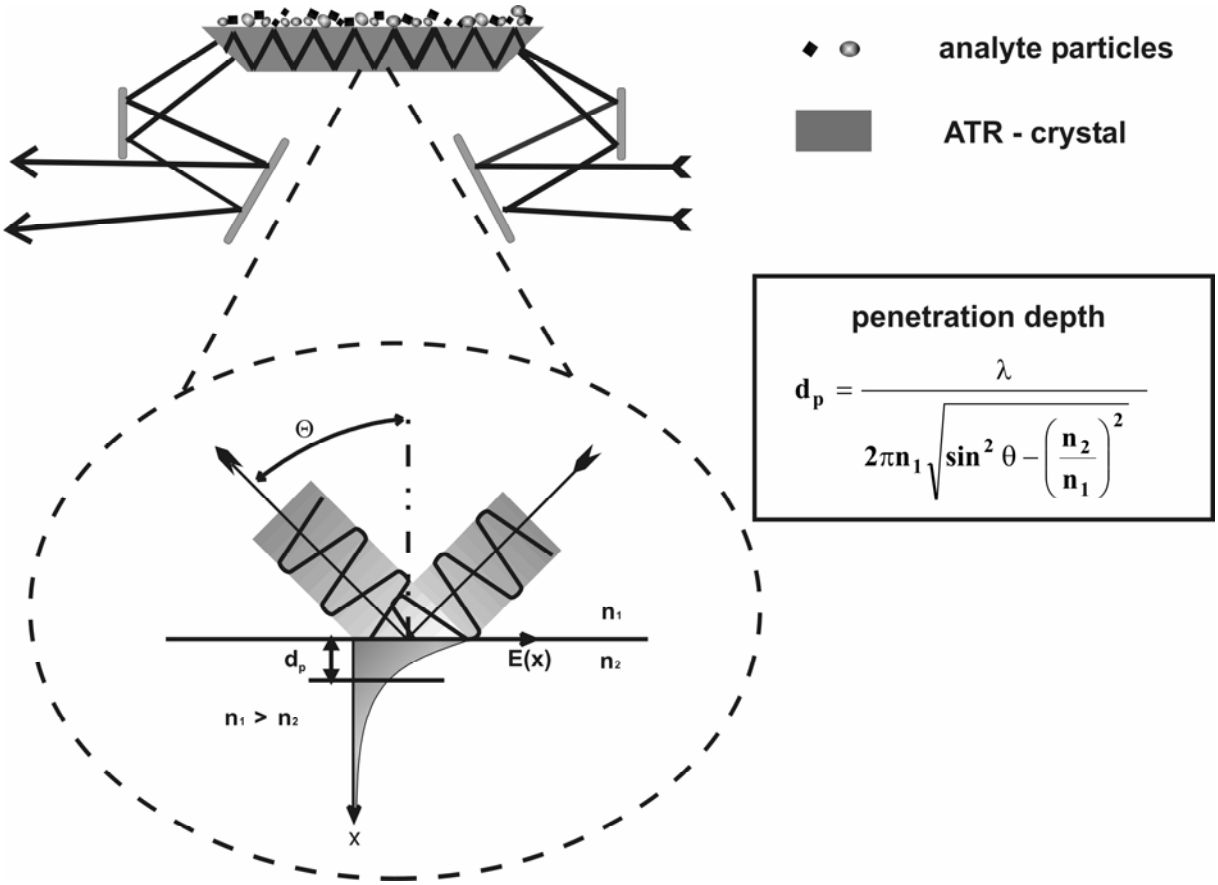


Figure 2.11 Illustration of the ATR principle.

At each internal reflection a certain amount of energy, depending on the wavelength, the refractive indices of the waveguide (n_1) and the surrounding medium (n_2, $n_1 > n_2$) and the angle of incidence, penetrates into the ambient matrix and is guided as leaky mode along the surface of the waveguide. The intensity of this evanescent field decays exponentially with distance from the internal reflection element (IRE) surface:

$$\mathbf{E} = \mathbf{E_0} \cdot \mathbf{e}^{\left(\frac{-z}{d_p}\right)} \qquad (2\text{-}2)$$

E_0 represents the wave amplitude at the interface (z = 0), d_p is the penetration depth and is defined as the distance from the IRE surface where E_0 has decreased to e^{-1} of its value at z = 0. The absorption for a specific wavelength depends on the penetration depth d_p of the associated evanescent field into the absorbing medium. The penetration depth is characterized by the following equation:

$$\mathbf{d_p} = \frac{\lambda}{2\pi \mathbf{n_1} \sqrt{\sin^2\theta - \left(\frac{\mathbf{n_2}}{\mathbf{n_1}}\right)^2}} \qquad (2\text{-}3)$$

where λ is the wavelength, n_1 and n_2 are the refractive indices of the IRE element and the ambient medium, and θ is the angle of incidence.

The penetration depth increases with increasing wavelength (λ), with decreasing angle of incidence, and with decreasing ratios of the refractive indices.

In addition, an effective layer thickness $\mathbf{d_e}$, which corresponds to the same absorption in transmission mode has been described [[137],138].

2.4.1.1. Effective Layer Thickness

In the opinion of Harrick, the interaction of the evanescent wave with the absorbing rarer medium, as calculated by the Fresnel's equations or the Maxwell's equations, does not yield any physical insight into the absorption mechanism or into the interaction of the penetrating field with the absorbing medium [135]. For this reason, he introduced the effective layer thickness as a parameter that expresses the strength of interaction of the evanescent wave with the absorbing rarer medium, and for which simple equations were found.

It is known that in the case of conventional transmission IR spectroscopy, the sample thickness is directly related to the intensity of absorption features of a sample according to Lambert-Beer's law.

From the Lambert-Beer's law, the following expression for the transmittance, $\mathbf{T}$, can be extracted:

$$\mathbf{T} = \frac{\mathbf{I}}{\mathbf{I_0}} = \mathbf{e}^{-\alpha \mathbf{d}} \qquad (2\text{-}4)$$

where $\mathbf{I_0}$ is the incident intensity, $\mathbf{I}$ is the transmitted intensity and α is the absorption coefficient (cm^{-1}), and $\mathbf{d}$ is the sample thickness (cm).

For low absorptions, i.e. $\alpha\mathbf{d} < 0.1$, we obtain:

$$\mathbf{T} = \frac{\mathbf{I}}{\mathbf{I_0}} \approx \mathbf{1} - \alpha\mathbf{d} \qquad (2\text{-}5)$$

Similarly, in the case of internal reflection, the reflectivity, $\mathbf{R}$, can be written as:

$$\mathbf{R} = \mathbf{1} - \alpha \mathbf{d_e} \qquad (2\text{-}6)$$

where $\mathbf{d_e}$ is the effective thickness.

Equation 2-6 is valid for a single reflection. For multiple reflections, the reflected power is given by:

$$\mathbf{R^N} = \mathbf{(1 - \alpha d_e)^N} \qquad (2\text{-}7)$$

where $\mathbf{N}$ is the number of reflections.

For $\alpha \mathbf{d_e}$ << 1, the reflected power becomes:

$$\mathbf{R^N} \cong \mathbf{1} - \mathbf{N}\alpha \mathbf{d_e} \qquad (2\text{-}8)$$

By comparing the low-absorption approximation expressions for transmission (Equation 2-5) and internal reflection (Equation 2-8), it can be observed that the effective thickness represents the actual thickness of a film that would be required to obtain the same absorption in a transmission measurement as that obtained in a reflection measurement. Different expressions for the effective thickness have been derived for two different cases: the semi-infinite bulk case and the thin film case.

Since in this work typical thicknesses of the polymer coatings are in the region of a few μm and therefore exceeding the penetration depth the case of a semi-infinite bulk layer will be described briefly.

Semi-infinite bulk case

For bulk materials, the electric field amplitude (see Equation 3.2) falls to a very low value within the thickness of the rarer medium, $\mathbf{d}$. The sample thickness is larger than the penetration depth ($\mathbf{d} > \mathbf{d_p}$).

In this case, the low absorption approximation for the effective thickness is calculated from the electric field for zero absorption [137,135]:

$$\mathbf{d_e} = \frac{\mathbf{n_1}}{\mathbf{n_2}} \frac{\mathbf{1}}{\mathbf{\cos\theta}} \int_0^\infty \mathbf{E(r)^2 dz} = \frac{\mathbf{n_1}}{\mathbf{n_2}} \frac{\mathbf{E_0^2}}{\mathbf{\cos\theta}} \frac{\mathbf{d_p}}{\mathbf{2}} \qquad (2\text{-}9)$$

Since $\mathbf{d_p}$ is dependent on the wavelength, the effective thickness also increases with wavelength. This is the reason why the internal reflection spectra of bulk materials show absorption bands at the longer wavelengths with relatively stronger intensity. Thus, two bands having the same intensity in transmission spectra will have unequal intensities in internal reflection spectra—the longer wavelength bands appear relatively stronger. This wavelength dependence also results in greater absorption on the longer wavelength side of single absorption band, contributing to band distortion.

2.4.1.2. Waveguide Materials

By using both, IR transparent ATR crystals and optical fibers intrinsic evanescent field sensing systems can be developed. Variations in thickness and length of the applied waveguides determine the number of internal reflections from a geometrical point of view. The refractive indices influence the penetration depth and the optical properties of the waveguide material define the attenuation losses of guided IR radiation.

Table *2.2* gives an overview on IR-ATR waveguides currently investigated at the Applied Sensors Laboratory (ASL). This table is by no means exclusive and represents only a fraction of the available IR transparent materials, as e.g. fluoride or tellurium halides can be processed into mid-IR fibers [36,46,84].

Table 2.2: Waveguide materials applied at ASL for MIR evanescent field sensing.

Material	**Waveguide type**	**Transmission range (cm^{-1})**	**Refractive index**	**General Properties**
Thallium Bromoiodide (KRS-5)	ATR crystal	20,000-250	2.37 (at 10 µm)	slightly soluble in water soluble in bases insoluble in most acids
Zinc Selenide (ZnSe)	ATR crystal	17,000-720	2.41 (at 9.5 µm)	incompatible with acids and strong alkalis insoluble in water and organic solvents
Germanium (Ge)	ATR crystal	5,500-600	4.00 (at 9.72 µm)	insoluble in water insoluble in most bases and acids
Silicon (Si)	ATR crystal	4000-1500 360-70	1.62 (at 5 µm)	insoluble in water insoluble in most acids and bases

Material	Waveguide type	Transmission range (cm^{-1})	Refractive index	General Properties
Sapphire (Al_2O_3)	Fiber	50,000-2500	1.62 (at 5 µm)	insoluble in water incompatible with strong acids and bases
Chalcogenide (AsSeTe glass)	Fiber	10,000-900	2.9 (at 10.6 µm)	insoluble in water incompatible with strong acids and bases
Silver halides - (AgX)	Fiber	2500-500	2.1 (at 10.6 µm)	insoluble in water incompatible with strong acids and bases

2.4.2. Recent Approaches for VOC Determination via Evanescent Wave Sensing

Evanescent field spectroscopy utilizes internal reflection elements based on ATR-crystals or MIR-transparent optical fibers serving as waveguide and optical transducer. Hence, absorption spectroscopy at or near the waveguide surface is enabled via the evanescent field [136,135]. Chemical MIR sensors enrich analytes into a thin polymer membrane coated onto the waveguide surface providing interaction of the evanescent field with enriched analyte molecules. Such sensor systems enable measurements within a period of several minutes instead of comparatively long analysis time for methods based on sampling and discontinuous assessment. Information derived from spectroscopic data allows identification and quantification of a wide range of VOCs at laboratory and field conditions [37,38. With regard to application in real world

environments, the accuracy of this method has been proven independent of aqueous sample turbidity, salinity or acidity at expected levels [96].

Recently, an alternative concept of polymer-coated ATR-IR sensor systems for detecting VOCs in the gas phase has been introduced by Yang et al. [141,142]. In this approach VOCs are detected in the head-space of either a heated or gas-stripped sample solution. Analytes are enriched into a thin PDMS layer, which is coated onto a suitable ATR waveguide and are spectroscopically detected. Although this setup may find its applications for instance in detection of (semi)volatile organic compounds in aggressive environments, it seems preferable to directly measure in the liquid phase for following reasons:

- Compounds with low volatility are still addressable with a direct sensor.
- Direct measurements represent the least complex sensor setup only consisting of a transducer exposed to the sample solution avoiding any prior sample preparation or extraction.
- Less analysis parameters have to be controlled compared to head-space sensing and no stripping system is required.

2.5. Polymer Sensor Membranes

Strong interferences caused by characteristic absorptions of water (O-H stretching band $\nu_{1,3}$ @ 3300 cm^{-1}, O-H-bending band ν_2 @ 1640 cm^{-1}, combination band $\nu_2+\nu_L$ @ 2100 cm^{-1} and the libration band ν_L @ 750 cm^{-1} [143]) render direct ATR spectroscopy of organic pollutants in aqueous sample matrices impossible at low concentrations. To overcome this limitation, hydrophobic polymer layers are coated onto the actively

transducing waveguide surface [37,38] following the general concept of chemical sensors [33,34]. While these membranes serve as solid phase microextraction (SPME) matrix for analyte enrichment, they also exclude water from the analytical volume probed by the evanescent field extending along the IRE surface. Sorbent extraction was developed in the 1980s and is commonly used for the extraction of organic compounds from matrices such as water, air and soil [144,56]. A solid adsorbent layer is exposed either directly to the sample matrix or to the associated headspace. Analyte molecules partition from the sample into the adsorbent layer following a gradient in chemical potential until equilibrium has been reached. In case of very large sample volumes $V_s >> V_p$, the amount of analyte *n* extracted from the sample matrix is dependent on the initial analyte concentration c_0, the partition coefficient of the analyte K_{sp} between the sample matrix and the solid phase extraction membrane and the volume of the membrane V_p following:

$$\mathbf{n = K_{sp} V_p c_0} \qquad (2\text{-}10)$$

Water / polymer partition coefficients of relatively small chlorinated VOC molecules typically range between 100 and 1000 [145]. Thus, analytes are enriched within the adsorbent layer while water is widely excluded. Under the boundary conditions of given penetration depth and fiber diameter the amount of analyte within the evanescent field and subsequently the sensitivity can only be increased by extending the length of the coated fiber, i.e. the length of the active transducer region.

The thickness of the membrane is a critical factor for solid phase extraction applications. Ideally, the polymer membrane thickness is only slightly larger than the information depth d_e, which describes the maximum distance from the waveguide surface from which relevant analytical information can be obtained [135]. Thicker coatings adversely

affect the sensor response time since analytes have to diffuse a longer distance to the waveguide surface until reaching the analytical volume probed by the evanescent field. In contrast, thinner coatings do not sufficiently exclude water from interaction with the evanescent field.

Typical requirements for a polymeric sensor coating for applications in aqueous media include:

- Low permeability for water
- Formation of non-porous layers with sufficient adhesion properties at the waveguide surface.
- Reversible enrichment of the hydrophobic analytes of interest
- Acceptable equilibration times during the enrichment process to minimize the sensor response time
- No or only weak absorption bands in the spectral region of interest (1200 cm^{-1} – 400 cm^{-1})

A number of different coating materials have been tested for their general suitability for sensing applications in aqueous environments [37,38,40,166]. Based on these results E/P-co was selected as suitable membrane material for the detection of VOCs in water. Coating procedures and characterization are described in the respective results sections.

2.6. Improved Model for Simulating Diffusion-based Data for Optical Chemical Sensors

When non-polar organic compounds in aqueous solutions are exposed to a hydrophobic membrane, they preferentially partition from the aqueous phase into the extraction layer. If this membrane is used as the coating for an IRE, the organic compounds will diffuse into the region interrogated by the evanescent field and their presence can be spectroscopically detected if the organic compound provides infrared-active absorption features.

To date, the theoretical description of enrichment processes specifically for polymer coated evanescent field based sensing systems has only been approached in a very simplified way. During the last decade a few models have been developed for different types of chemical sensors, ranging from fiber-optic chemical sensors [88,146, 147] to a dopamine biosensor [148] and thermoelectric gas sensors [149].

During characterization of the analyte transport from the aqueous phase to the probed volume of the sensor system usually following simplifications have generally been adopted:

- Since typical diffusion coefficients for molecules in dilute aqueous solutions are on the order of 10^{-5} cm^2/s [150] while diffusion coefficients for molecules in a bulk polymer are commonly two orders of magnitude lower [151], the assumption is made that transport of analyte into and through the polymer phase is the response limiting process.
- The water / polymer boundary layer has negligible effects on the mass transport rate of the analytes [152].

- Analyte transport in the polymer is exclusively governed by Fickian diffusion in an idealized (defect free) polymer layer.

Following these generally accepted simplifications the mass transport of analytes from the aqueous phase into the volume probed by the evanescent field is entirely independent of the hydrodynamic properties of the system configuration surrounding the active sensor element. Differences in enrichment times of analytes have always been related to different partition coefficients or diffusion coefficients in respect to the particular polymer matrix. Due to the absence of reference data for these values, the inaccurateness of these models has not been considered crucial. Experimental setups of published evanescent sensor systems reveal striking similarities : most devices comprise a flow-cell connected to standard laboratory peristaltic pumps. In such systems the flow-conditions can only be modified within certain limits as standard peristaltic pumps usually do not provide a wide range of flow velocities. Therefore, published flow rates usually range from 1 mL/min to 10 mL/min. Hence, published sensor performances are usually comparable in this respect [153, 154].

However, results reported from pervaporation and ultrafiltration experiments generally suggest that the flow conditions are a major parameter of influence in related applications [155-157]. Some experimental evidence on the influence of flow rates on the signal characteristics of polymer coated evanescent field sensors was given by Roy et al [158], however, without theoretical description of the observed effects. Their work presents a system based on an E/P-co coated ZnSe ATR waveguide used for the detection of various VOCs. Flow-rates are changed from static conditions up to flow rates of 750 mL/min. Exemplary enrichment curves for CB into the extractive membrane for different flow rates are shown in Figure 2.12.

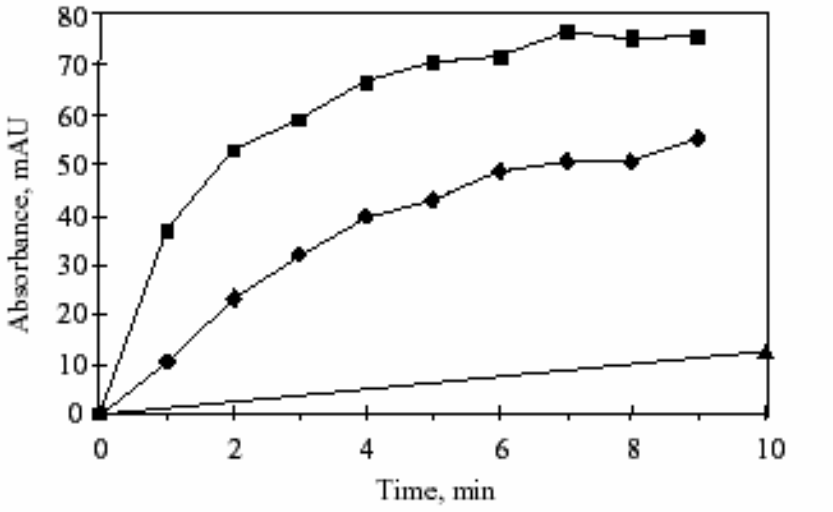


Figure 2.12 Exemplary enrichment curves for CB (30 ppm) into a 10 µm E/P-co layer for different flow-conditions: 10 mL/min (triangles), 100 mL/min (diamonds) and 750 mL/min (squares) [158].

The accelerated increase in signal for higher flow-rates indicates a diffusion limitation at the polymer-aqueous solution interface, which is explained by an immobile water boundary layer at the interface water-polymer. The thickness of this boundary layer can be estimated as a function of flow rate and the extent of agitation, or – more generally - from the flow conditions. Theoretical treatment of the influence of the stagnant surface layer in agitated solutions is given by Louch et al [159] for organic compound extraction via SPE.

Table 2.3 shows different $t_{95\%}$ data (time required to extract 95% of analyte at equilibrium conditions) for the extraction of a variety of hydrocarbons from an aqueous solution with PDMS coated NIR fiber-optic sensors demonstrating the magnitude of the effect of different agitation levels of the solution on the sensor response [42].

Table 2.3 Comparison of $t_{95\%}$ data (time required to extract 95% of analyte amount at equilibrium conditions) for extraction of different hydrocarbons from an aqueous solution with the use of a PDMS coated NIR fiber-optic sensor system. Reproduced from [42].

	$t_{95\%}$ time (min)	
compound	stirred solution	unstirred solution
trichloromethane	1.0	17.0
trichloroethene	4.0	27.2
toluene	9.3	87.0
p-xylene	11.8	236.0
trichlorobenzene	48.0	483.0
gasoline	73.0	794.3

These results lead to the following initial conclusions:

- For virtually all reported data in the field of polymer coated evanescent wave sensor systems the boundary layer (stagnant layer) is the rate limiting factor for the dynamics of the enrichment / extraction process as flow conditions have generally been in the region of strict laminar behavior (Reynolds numbers below 100 for typical flow-cells and flow-rates according to [158]).
- The higher the partition coefficient of a particular analyte the faster the solution depletes of that analyte in close vicinity of the aqueous / polymer interface. Hence, especially for analytes with a high partition coefficient the extraction rate is dependent on agitation of the solution surrounding the extractive layer.
- Only at high agitation levels (flow rates with Reynolds numbers close to turbulent flow) of the solution the boundary layer sufficiently decreases in thickness to a few µm to be neglectable [158]. Hence, only at these conditions the generally accepted results from simplified models based exclusively on Fickian diffusion into the polymer will converge with results

considering flow conditions. However, most results reported in literature to date are far from these flow rates.

2.6.1. Model-Based Optimized Design of Polymer-Coated Chemical Sensors

Based on the findings discussed in the previous section Jakusch [160] has established a fundamental model including the flow-cell geometry into a model describing the enrichment kinetics. In the following, a sound hydrodynamic theoretical model for the simulation of diffusion kinetics for polymer coated evanescent wave sensor systems has been developed in collaboration with the research group Prof. A. Fedorov (School of Mechanical Engineering, Georgia Tech) [43].

2.6.1.1. Physical Arrangement and Model Formulation

A schematic of the modeled flow-cell and coordinate system is shown in Figure 2.13.

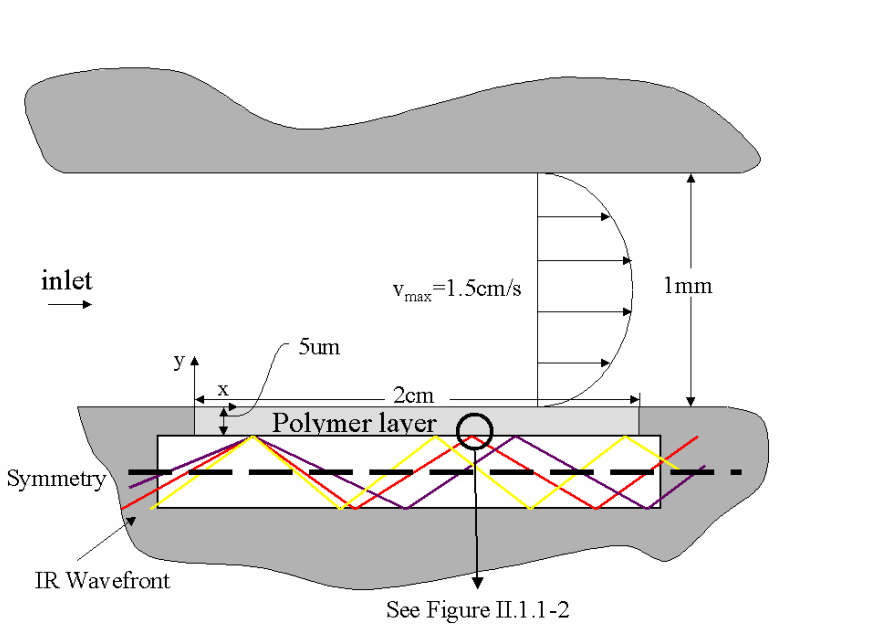


Figure 2.13 Schematic of the flow cell and coordinate system for the hydrodynamic model [43].

Transport of analyte in the flow cell containing the active chemical transducer surface is governed by the following mass and momentum conservation equations and boundary / interface conditions for analyte concentration and flow velocity:

mass conservation

aqueous phase $$\frac{\partial \mathbf{C_a}}{\partial \mathbf{t}} + \mathbf{u}\frac{\partial \mathbf{C_a}}{\partial \mathbf{x}} + \mathbf{v}\frac{\partial \mathbf{C_a}}{\partial \mathbf{y}} = \mathbf{D_a}\nabla^2\mathbf{C_a} \qquad (2\text{-}11)$$

polymer $$\frac{\partial \mathbf{C_a}}{\partial \mathbf{t}} = \mathbf{D_p}\nabla^2\mathbf{C_p} \qquad (2\text{-}12)$$

boundary conditions

$$\mathbf{C_a = C_p} \text{ and } \mathbf{C_p = 0} \text{ at the inlet} \qquad (2\text{-}13)$$

$$\frac{\partial \mathbf{C}}{\partial \mathbf{n}} = \mathbf{0} \quad \text{at all wall boundaries except for the fluid / polymer interface} \qquad (2\text{-}14)$$

$$-\mathbf{D_a}\left.\frac{\partial \mathbf{C_a}}{\partial \mathbf{y}}\right|_{\mathbf{int}} = -\mathbf{D_p}\left.\frac{\partial \mathbf{C_p}}{\partial \mathbf{y}}\right|_{\mathbf{int}} \quad \text{and} \quad \mathbf{KC_a = C_p} \qquad (2\text{-}15)$$

at the fluid / polymer interface.

Here, n is an outer normal at the boundaries, C_a and C_p are the concentrations of the analyte in the aqueous phase and polymer, respectively, D_a and D_p are the diffusivities of analyte in the aqueous phase and polymer, respectively, and K is the partition coefficient for a given analyte and polymer matrix.

momentum conservation for the fluid velocity vector $\vec{\mathbf{v}} = \{\mathbf{u}, \mathbf{v}\}$

$$\text{aqueous phase} \quad \frac{\partial \vec{\mathbf{v}}}{\partial \mathbf{t}} + \vec{\mathbf{v}} \cdot \nabla \vec{\mathbf{v}} = -\frac{\mathbf{1}}{\rho} \nabla \mathbf{P} + \nu \nabla^2 \vec{\mathbf{v}} \qquad (2\text{-}16)$$

$$\text{polymer} \quad \vec{\mathbf{v}} = \mathbf{0} \quad \text{everywhere} \qquad (2\text{-}17)$$

boundary conditions

$$\vec{\mathbf{v}} = \mathbf{0} \quad \text{at all solid walls and fluid/ polymer interface (no slip)} \qquad (2\text{-}18)$$

$$u_{inlet} = u_{max}\left(1-\left(\frac{y-h/2}{h/2}\right)^2\right)$$ fully developed Poiseuille flow profile (2-19)

at the flow cell inlet.

Here, ρ and ν are the density and viscosity of the flowing solution, respectively, P is the hydrodynamic pressure, u and v are velocity vector components in axial and transverse directions, respectively, and u_{max} is the maximum velocity at the centerline (y=h/2) of the flow cell. The assumptions made in the analysis are those of steady incompressible flow, isothermal conditions, and constant fluid viscosity and analyte diffusivity. Also, it is assumed that there are no mass sources present and any heating effects due to the evanescent field present in the polymer layer are negligible.

2.6.1.2. Method of Solution

The governing conservation equations 2-11, 2-12 and 2-16 are of parabolic type and can be effectively solved using an implicit, absolutely stable finite difference numerical integration technique [161]. However, the problem features one jump boundary condition for the analyte concentration at the fluid / polymer interface (equation 2-15), which significantly complicates the simulation procedure. Specifically, it requires a separate solution of the mass-transfer problem in the flow and polymer domains, followed by iterative coupling of these solutions, which is not only very inconvenient but also computationally a very inefficient procedure [161]. In contrast, if one could identify a new “modified” scalar variable equivalent to concentration, which is in itself, as well as the flux associated with this variable, continuous at the interface (i.e. no jump), then

computation can be performed in an efficient non-iterative manner for the combined (fluid and polymer) computational domain. To accomplish this task, the complimentary heat-transfer problem by defining an equivalent “fictitious temperature”, which is continuous at the fluid / polymer interface in the equivalent thermal domain, was solved. For the CFD simulations the commercial available software package FLUENT* was used.

For more detailed description of that transformation please refer to [43]; a short summary of the main CFD simulation results is provided in the following:

- The response time is highly dependent on the analyte diffusivity in the aqueous phase: on average, the time to reach steady state conditions drops one order of magnitude with a two-order-of-magnitude increase in diffusivity.
- For constant volumetric flow rates the optimal flow channel height is the smallest allowable height, which corresponds to the fastest sensor response.
- The least total resistance to mass transfer is achieved when the channel height is equal to or less than the concentration boundary layer thickness at the exit of the channel.
- The sensor response time linearly increases with the thickness of the polymer layer.
- The critical flow channel height for a given flow velocity is independent of the partition coefficient.

* FLUENT CFD Software, Fluent Inc: http://www.fluent.com

- The flow velocity can be used to control the optimal channel height indirectly by altering the concentration boundary layer to approache the height of the channel at the exit.
- Alternative geometries of the sensor flow-cell further improve the response time in comparison to the basic flow-cell design shown in Figure 2.13.

The relevance of these findings to data interpretation of sensing signals recorded with polymer coated evanescent field methods will be discussed in chapter 3.4.

3. Results

In this chapter results from measurements of VOCs in water by polymer coated evanescent field sensor system are presented. In principle, three main measurement series have been performed throughout this thesis at different conditions. Therefore, this chapter is divided into three sections:

i. *Laboratory Conditions:* The simultaneous and quantitative determination of BTX mixtures in water performed at the ASL laboratories at Georgia Tech. E/P-co coated ZnSe crystals were used in conjunction with the automated mixing system (Mixmaster).

ii. *Simulated Field Conditions:* Continuous monitoring of TriCE, TeCE and DCB in a the migrating aqueous phase of an aquifer system at the Technical University of Munich was conducted with an E/P-co coated AgX fiber-optic setup.

iii. *Field Conditions:* A sensor system consisting of an E/P-co coated ZnSe crystal incorporated in an improved flow-cell enabled the determination of CB in groundwater in the Bitterfelder area (Germany) at the SAFIRA remediation site.

3.1. Laboratory Conditions – BTX in Water

3.1.1. Introduction

Developing a chemical sensor usually includes the preparation of numerous solutions for calibration purposes to characterize the sensor performance. The number of solutions that have to be prepared can reach up to hundreds, depending on the analytical problem. Especially when chemometric evaluation has to be applied, a large calibration set is necessary. Due to the fact that the traditional way of preparing calibration sets by diluting stock solutions is error prone and time consuming, it is reasonable to develop automated systems for this task.

Recently, an automated and portable mixing system, based on commonly used components in sequential injection analysis (SIA) [162-164] has been introduced by our research group [73]. In this work the mixing system was applied for the precise preparation of benzene, toluene and the three xylene isomers (BTX) / water mixtures at trace level concentrations (<mg/L regime). An E/P-co coated ZnSe crystal was applied to simultaneously and quantitatively detect individual BTX components in multi-component mixtures by means of MIR-FTIR evanescent wave spectroscopy.

3.1.2. Experimental Setup

3.1.2.1. Materials

Ethylene/propylene co-polymer (60:40) was purchased from Aldrich (Milwaukee, WI). Methanol, benzene, toluene, o-xylene, m-xylene and p-xylene were purchased from Aldrich (Milwaukee, WI) and were of analytical grade. Deionized water was used for preparation of all solutions and for sensor regeneration.

3.1.2.2. Instrumentation

Data was recorded in a spectral range of 600 cm^{-1} to 1400 cm^{-1} using a Bruker Vector 22 Fourier transform infrared (FT-IR) spectrometer (Bruker Optik GmbH, Ettlingen, Germany) equipped with a liquid N_2 cooled mercury-cadmium-telluride (MCT) detector (Infrared Associates, Stuart, FL). A total of 100 scans were averaged for each spectrum with a spectral resolution of 4 cm^{-1}. For ATR measurements a vertical ATR accessory (Specac, Smyrna, GA) in combination with trapezoidal ZnSe ATR elements (50*20*2mm, 45º; Macrooptica Ltd., Moscow, Russia) and a stainless steel flow-cell (custom made, Volume: 280 µl, free contact area to ATR crystal: 5.5 cm^2) were used. A custom made mixing system (*Mixmaster*) [73] designed for handling volatile organic compounds assured accurate concentrations of sample mixtures and continuous flow of the analyte solutions through the ATR cell. A schematic of the experimental setup is shown in Figure 3.1.

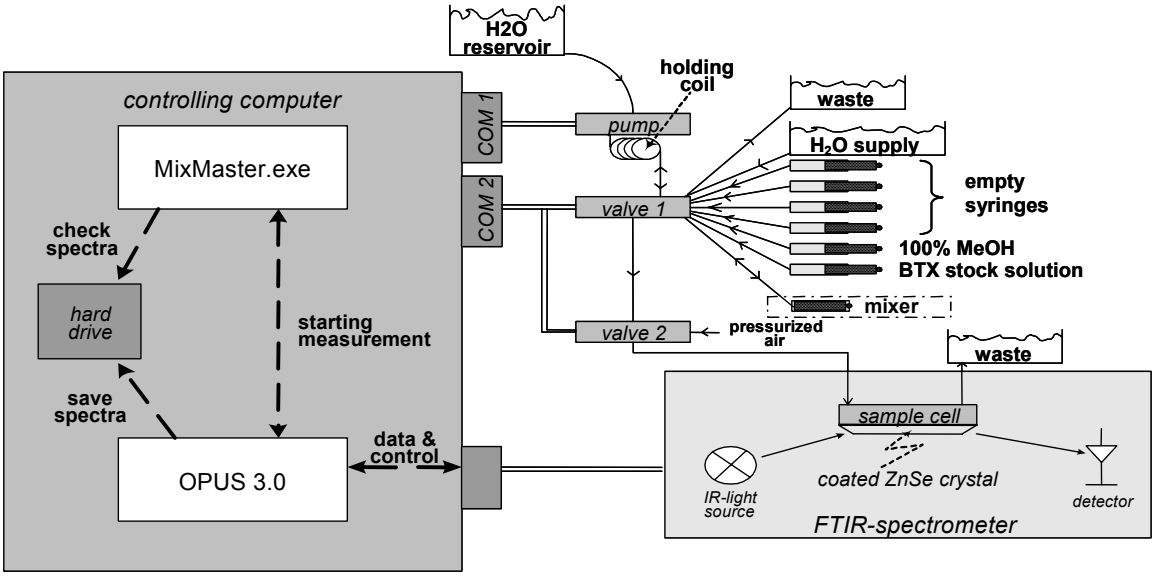


Figure 3.1 Schematic of the combination of the Mixmaster with the ATR setup for dissolved BTX measurements.

3.1.2.3. Mixmaster

In order to ensure a precise set of diluted standard solutions a software-controlled automated mixing system (*Mixmaster*) developed by our research group was applied [73]. The main components comprise a high precision piston pump (syringe volume 25000 µL), which is attached to a selection valve with 10 ports and a 2-way injection valve connected to the ATR flow-cell. A C++ software interface allows controlling all parameters of the system including piston position, pump speed, and positions of the valves. Furthermore, measurements of the FT-IR spectrometer are triggered and synchronized by the Mixmaster control software. Stainless steel tubings connected via bulk-head unions are exclusively used minimizing wall adsorption effects. By avoiding the use of polymer components within the *Mixmaster* system and by providing headspace-free storage and mixing of solutions the *Mixmaster* is especially suitable for

high-throughput investigation of volatile organic compounds, e.g. during extensive sensor optimization and calibration. A detailed description of the mixing system is given elsewhere [73] and a front view can be seen in Figure 3.2

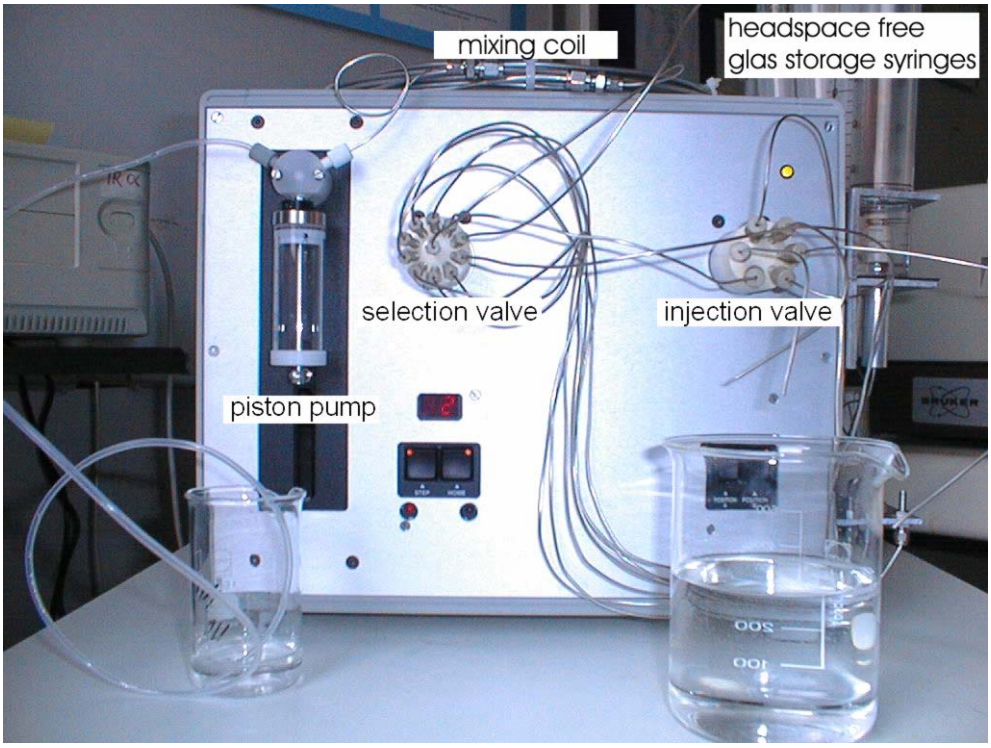


Figure 3.2 Front view of the Mixmaster

A typical measurement cycle for sensor calibration comprises following steps:

- rinsing of the cell with water.
- collecting a background spectrum.
- Preparing the analyte solution.
- Up to 20 min rinsing of the flow-cell with analyte solution while collecting absorption spectra every 2 min; (v) 25 min rinsing of the cell with deionized water to extract the analyte and regenerate the polymer layer.

The sample solution flow rate was held constant at 3 mL/min throughout all experiments. Methanol was used as a solution mediator ensuring that the BTX mixture remains dissolved in aqueous solution. Previous works have shown that this procedure has no effect on the final sensor readings [96]. Dilutions were prepared from a primary stock solution of 200 ppm (v/v) of all compounds of the BTX group in pure methanol. Thorough mixing resulted in a total methanol concentration of 0.5 % (v/v) in the investigated sample solutions. Cross interferences and thus influences on the enrichment properties due to measurement of an analyte mixture, are not to be expected in the examined concentration range presented in this study, as has been shown previously [100]

3.1.2.4. Preparation of the Extractive Polymer Membrane

A 1 % (w/v) coating solution E/P-co was prepared by dissolving 0.5 g of granular polymer under reflux in 50 ml n-hexane. Prior to coating the ATR crystal was thoroughly rinsed with methanol. Approx. 300 µL of clear, hot solution were applied to the surface of the ATR crystal using an Eppendorf pipette. The crystal was kept at room temperature for at least 2 h ensuring evaporation of most of the solvent. Subsequently, the polymer coating was exposed to hot air treatment with a hot air gun at 150 °C for 5 min to remove remaining traces of solvent. The thickness of the layer was determined by differential weighing to be 4.2 µm.

For sensing applications of trace components it is essential to coat the transducer surface with (chemo)selective membranes excluding interfering matrix components that would overlap or mask absorption features of the investigated analyte. This is of particular importance when measuring in strongly IR absorbing matrices such as water. In the present study, a thin layer of hydrophobic E/P-co is coated onto the waveguide

surface. Hydrophobic analytes partition into the hydrophobic membrane while water and other polar components are widely excluded from the analytical volume probed by the evanescent field. Additionally, the polymer coating enhances the sensitivity of the sensor by enriching hydrophobic analytes in the polymer membrane following the principles of solid phase extraction. One approach to roughly estimate the enrichment factor for a particular analyte is to relate obtained absorption peak heights from measurements with uncoated waveguides to results achieved with polymer coated transducers. However, limited solubility of BTX in water and strong absorptions of the water matrix in the fingerprint region of the MIR spectral range prohibit direct ATR measurements with uncoated crystals. Hence, analyte solutions with a concentration of 1 % in methanol have been prepared for an estimation of the achievable enrichment factors. ATR spectra of methanolic solutions have been recorded with uncoated ZnSe crystals. Peak heights for each analyte have been normalized and correlated to peak heights obtained for an enrichment measurement of a 500 ppb (v/v) aqueous analyte solution with an E/P-co coated ZnSe crystal after the partition equilibrium has been established. Following this approach, enrichment factors > 15.000 are estimated for benzene, toluene and for the xylene isomers, respectively.

Based on previous experience, E/P-co proved to be a suitable material for enrichment of a wide range of hydrophobic compounds from aqueous solutions [165,166].

3.1.3. Results

3.1.3.1. Water Equilibration

Despite the hydrophobic properties of the membrane a considerable amount of water diffuses into the polymer coating over a period of several hours causing IR absorptions as shown in Figure 3.3.

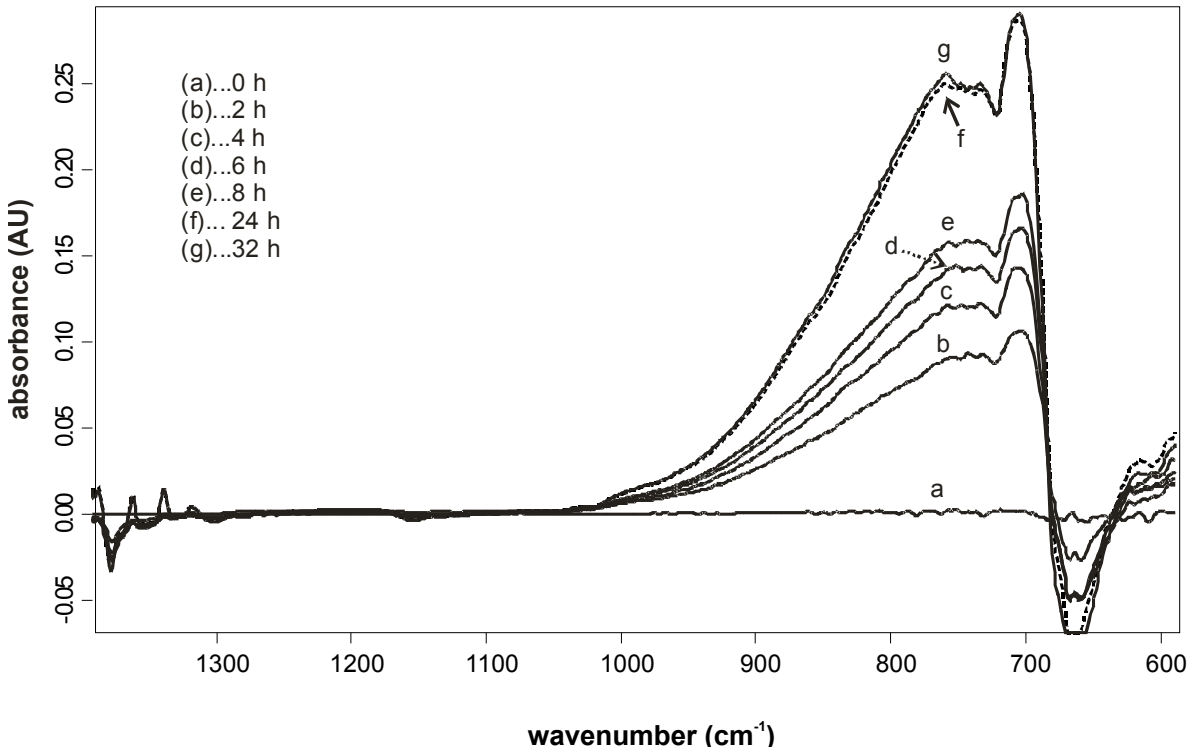


Figure 3.3 IR absorptions resulting from water diffusion into an E/P-co membrane with a thickness of 4.2 µm coated onto the surface of a ZnSe ATR-crystal over a period of 32 h. After 24 hours (f) equilibrium conditions are reached and no further increase in absorption is observed.

A broad absorption band, caused by swelling of the polymer during the water diffusion process, occurring between 1000 cm^{-1} and the cut-off frequency of the detector around 600 cm^{-1} significantly influences spectroscopic measurement due to the resulting

baseline drift. In a simple sample matrix this effect can be compensated by selecting suitable peak integration methods and integration limits. However, a stable baseline increases the reliability of the measurements and enables automated data evaluation, as shown in a recent study by our research group [59,60]. Figure 3.3 illustrates that water diffusion reaches equilibrium conditions before 24 hours of exposure to the aqueous phase. Hence, prior to analysis the coated waveguide has been equilibrated with deionized water for a period of at least 24 h.

3.1.3.2. BTX Enrichment Characteristics

Figure 3.4 shows an exemplary spectrum of a mixture of benzene, toluene and the three xylene isomers with a concentration of 500 ppb (v/v) each after an enrichment time of 20 min into an E/P-co layer. Corresponding absorption peaks have been labeled for clarity.

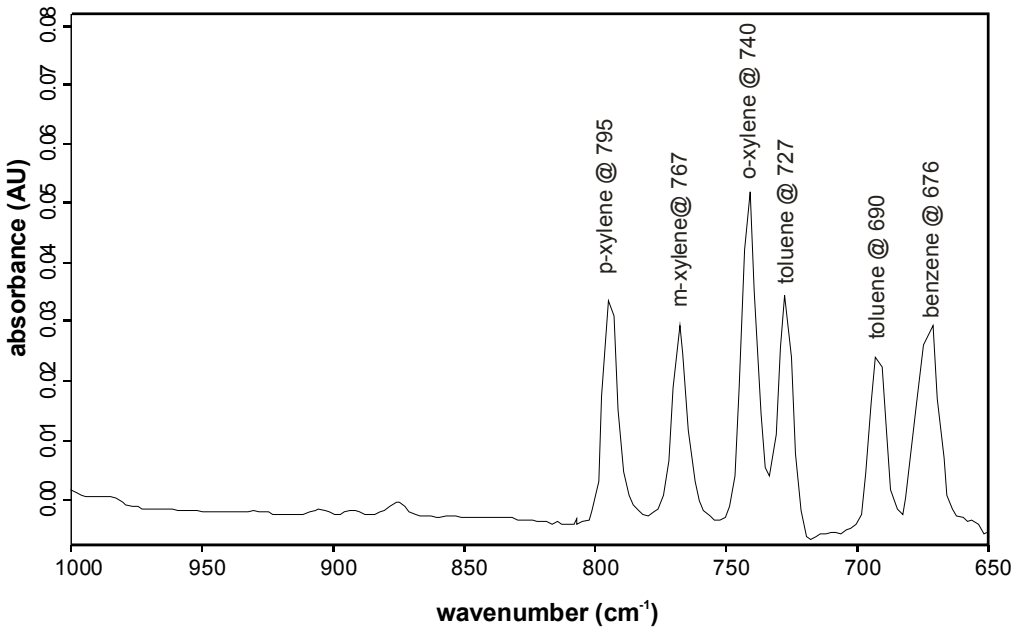


Figure 3.4 IR absorption spectrum of a sample mixture in aqueous solution after enrichment into an E/P-co layer. Enrichment time: 20 min, concentration: 500 ppb (v/v) each.

Typical absorption bands of benzene, toluene and the three xylene isomers can be identified resulting from molecule specific aromatic C–H out of plane vibrations in the fingerprint region of the mid-infrared spectral range. Band assignment has been performed via single component enrichment experiments leading to the following allocation of the absorption features: benzene at 676 cm^{-1}, toluene at 690 cm^{-1} and 727 cm^{-1}, o-xylene at 740 cm^{-1}, m-xylene at 767 cm^{-1} and p-xylene at 795 cm^{-1}. Each analyte shows distinctive absorption features, which are not or only slightly overlapping. Hence, during this first study conventional peak integration of the IR absorption bands was applied. More complex samples will be evaluated using chemometric data evaluation techniques particularly suitable for optical sensors [60].

Recently, it has been shown by Phillips et al. [43] that achieving steady state conditions for polymer coated sensor systems is not only dependent on the partitioning behavior of analytes into the polymer layer. Factors such as analyte diffusion properties within the

aqueous phase, flow channel height and flow velocity substantially affect chemical sensor response. The only tunable parameter of the ATR flow-cell used throughout the experiments was the flow velocity. Evaluating preliminary experiments the flow velocity of the analyte solution was set to 3 mL/min, which enabled measurements in the time regime of several minutes without using excessive amounts of analyte solution.

Figure 3.5 shows typical diffusion curves of the investigated analytes plotting the integrated peak area vs. the enrichment time. After 18 min of enrichment the diffusion process reaches equilibrium conditions for the given analyte mixture and data evaluated at this or at a later time delivers most reliable and sensitive results.

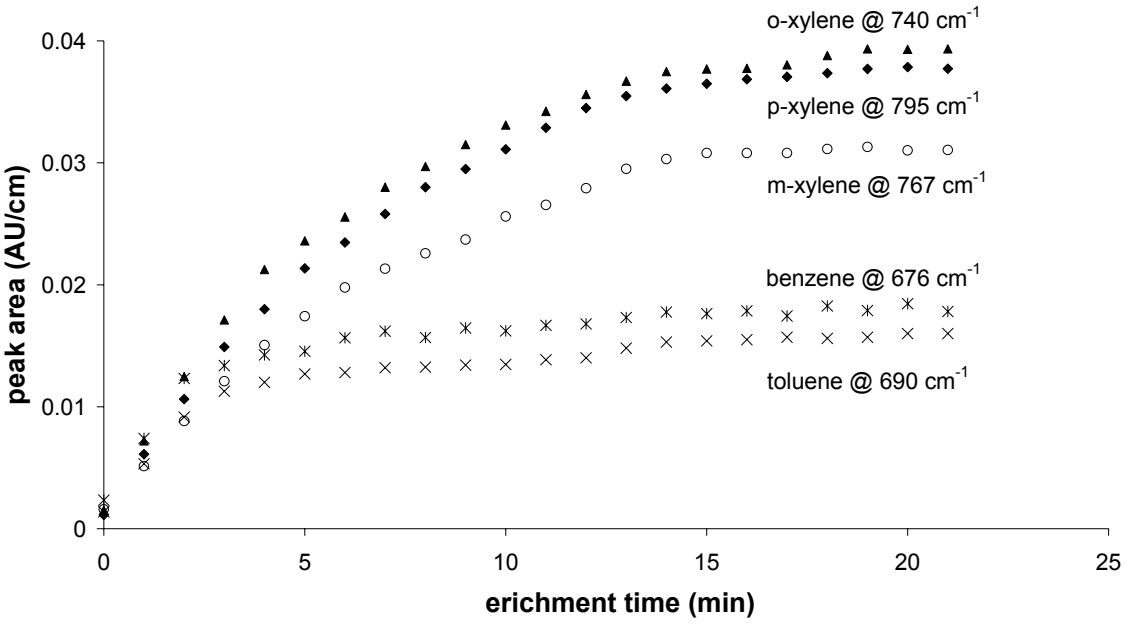


Figure 3.5 Typical enrichment curves for the BTX components in water at a concentration level of 1 ppm (v/v) each into an E/P-co coating. Equilibrium of the diffusion process is reached after approximately 18 min of enrichment time.

Figure 3.6 shows the obtained calibration curves for the investigated analytes in aqueous solution. The calibration of the sensor has been performed by five repetitive measurements of a concentration series ranging from 50 ppb (v/v) to 1 ppm (v/v) for

each analyte in the mixture. Error bars for each data point represent the calculated standard deviation derived from five repetitive measurements each. Prior to each measurement the polymer coating was regenerated by rinsing the flow-cell with water at a flow velocity of 3 mL/min for 25 min, which efficiently removed all analytes from the sensing membrane.

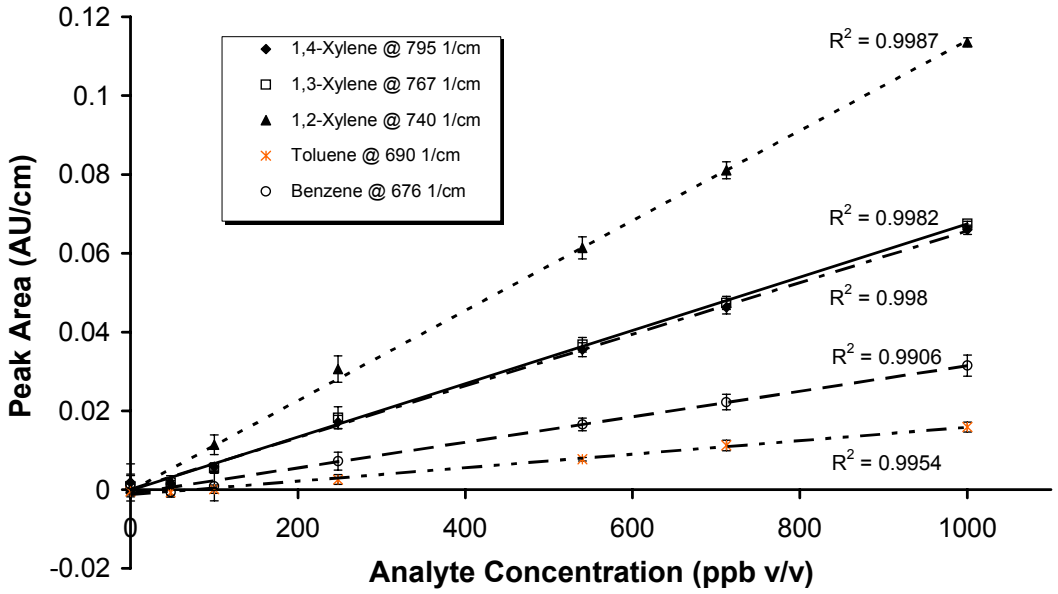


Figure 3.6 Calibration graphs for benzene, toluene and the xylene isomers in the concentration range of 0 – 1000 ppb (v/v) based on peak area integration. The error bars represent the standard deviation of five subsequent measurements

Data evaluation has been performed by peak area integration due to clear separation of the investigated absorption peaks. The integrated areas are plotted vs. the concentration resulting in linear fit functions with R^2-values generally > 0.99 included in Figure 3.6. Detection limits for each analyte in the mixture have been calculated according to IUPAC by the 3 sigma criteria (3 times standard deviation of the peak-to-peak noise related to the slope of the linear regression function) and resulted in LODs in the low ppb range in

mixture of all examined analytes. These values represent a significant improvement compared to previously reported results using head-space IR-ATR measurements for similar analytes yielding significantly higher detection limits [141,142]. The LODs from this work and other relevant spectroscopic approaches to BTX determination in water are given in Table 3.1.

Table 3.1 Overview on relevant spectroscopic approaches to BTX detection in water

reference	this thesis	142	103	55	76	167	42
method	FTIR-ATR	FTIR-ATR	FTIR transmission	UV transmission	UV derivative spectroscopy	Photo-acoustic	SPME-NIR
spme-matirx	EP/Co	PIB	parafilm	PDMS			PDMS
simultaneous detection	yes	no*	no*	no*	yes	no*	no*
time per measurement**	20 min	20 min	>30 min	90 min	1 min	40 min	20 min
benzene LOD ppb (v/v)	45		160	18	+/-50	308	
toluene LOD ppb (v/v)	80	292	652	5	+/-50	954	173
o-xylene LOD ppb (v/v)	10		72	4	+/-50		
m-xylene LOD ppb (v/v)	20		886		+/-50		
p-xylene LOD ppb (v/v)	20		57	3	+/-50		129

*… The author wants to emphasize that this statement does not imply that simultaneous detection of analytes is generally impossible with these methodologies, however, data utilized for LOD determination is derived from single analyte experiments only in the cited references.

**... This timeframe only refers to the actual measurement and doesn't account for other steps such as sensor equilibration time, calibration time, data evaluation time and so on

The sensor system presented in this work shows competitive or preferable performance in respect to LODs and measurement time to other relevant spectroscopic approaches reported for BTX analysis. Furthermore, 20 min per measurement cycle for quantitative simultaneous determination of five components is a reasonable time frame for a multitude of analytical applications including waste-water monitoring, remediation

process surveillance or drinking water monitoring contingent upon improved limits of detection. With the introduction of appropriate chemometric data evaluation techniques remote analysis will be further facilitated [59].

3.1.3.3. Test for Field Capability: Continuous Detection of o-Xylene in a Natural Pond Water Matrix

In the following experiment a preliminary test for the field applicability of sensor systems similar as described in the previous chapter is presented.

A continuous measurement series of various concentrations of o-xylene added to urban pond water is presented. The obtained data demonstrates the potential of polymer coated evanescent field sensors for real-world applications in water quality monitoring.

Materials

Ethylene/propylene co-polymer (60:40) was purchased from Aldrich (Milwaukee, WI). Methanol, o-xylene, were purchased from Aldrich (Milwaukee, WI) and were of analytical grade. Pond water was used for preparation of all solutions and for sensor regeneration. The pond water was sampled from a domestic goldfish-pond in Atlanta, GA, USA.

Preparation of Ethylene/Propylene Co-polymer Thin Film

Coating Procedure

The coating procedure adheres to the description in chapter 3.1.2.4 with following modification: 210 µL of the hot coating solution were applied resulting in a film thickness of approx. 3.3 µm determined via differential weighing. Preparation of the o-Xylene Samples A 1 % (v/v) solution of o-xylene in methanol was prepared and diluted with

pond water to 20, 50, and 80 ppm (v/v) of analyte concentration. Additional methanol was added to keep the amount of methanol constant at 1 % (v/v). The sample solutions have been freshly prepared prior to each measurement ensuring minimal losses due to evaporation.

Instrumentation

Data was recorded in the spectral range of 400 cm-1 to 1600 cm-1 using a Bruker Equinox 55 Fourier transform infrared (FT-IR) spectrometer (Bruker Optics, Billerica, MA) equipped with a liquid N2 cooled mercury-cadmium-telluride (MCT) detector (Infrared Associates, Stuart, FL). A total of 100 scans were averaged for each spectrum with a spectral resolution of 4 cm-1. For this continuous study spectra were recorded every minute for a period of approx. 8 hours. For ATR measurements a horizontal ATR accessory (Specac, Smyrna, GA) utilizing trapezoidal ZnSe ATR elements (72*10*6mm, 45º; Macrooptica Ltd., Moscow, Russia) and a stainless steel flow-cell (custom made, Volume: 2 ml, free contact area to ATR crystal: 7.2 cm²) were used. Solutions were pulled through the flow-cell via an Alitea C8-Midi peristaltic pump (Watson-Marlow Alitea, Wilmington, MA) at a constant flow rate of 4.5 mL/min.

Results

After equilibration with water as described above the sensor was exposed to neat pond water samples for several hours. No significant further changes of the absorption spectra could be observed. Following, the sensor was exposed to pond water samples spiked with o-xylene and an increasing absorption feature at 740 cm^{-1} (aromatic C–H out of plane vibration of o-xylene) could be observed after a measurement time of one minute already.

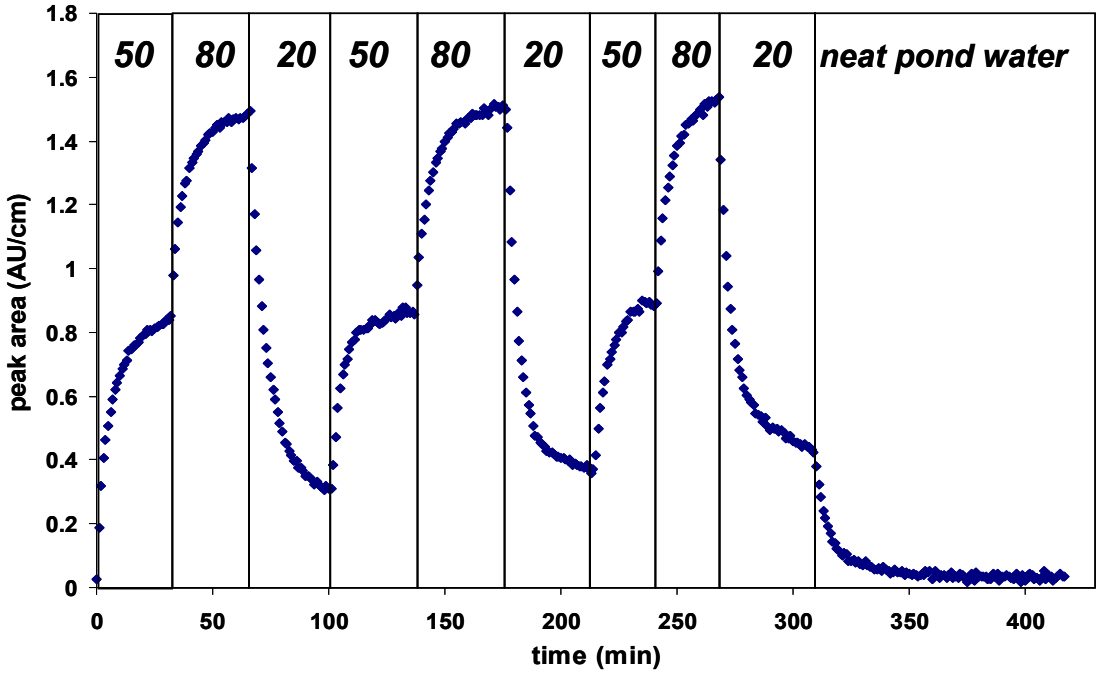


Figure 3.7 Trace of the peak area of the absorption band of o-xylene at 740 cm^{-1} with time during enrichment based IR-ATR sensing. Concentration trace: 50 ppm; 80 ppm; 20 ppm (in pond water; the sensor was exposed to each concentration for approx. 30 to 35 min) followed by neat pond water for sensor regeneration.

Figure 3.7 shows the continuous measurement of o-xylene in pond water over a period of 8 hours for a repetitive concentration trace of 3 different levels (50, 80, and 20 ppm v/v). Concentrations have been changed every 30 to 35 minutes. The trace at 740 cm^{-1} clearly shows that o-xylene partitioning never reaches equilibrium conditions for the selected observation window. However, it is evident that the response time of the sensor to changing concentrations of the sample solution is < 1 min, which is an essential aspect for rapid on-line data evaluation and e.g. threshold monitoring. Appropriate multivariate data evaluation techniques, which should enable prediction of the equilibration concentration of analytes for a calibrated system after very short enrichment times are currently developed in our research group. Slightly increasing peak area values from one repetition to the next along with a minute positive off-set after

regenerating the sensor with neat pond water (see minute 300 to 420) indicate that the broad water absorption band in the spectral region between 1000 cm-1 and the cut-off frequency of the detector (around 600 cm-1) was still slightly increasing throughout the measurement. Recently, our research group has developed a multivariate method for automated recognition and correction of baseline drifts [59,60]. As most chemical sensing systems are affected by baseline drifts due to ageing, degradation, and swelling of the molecular recognition interface, this generic solution enables the application of membrane based sensing devices in real-world environments.

3.1.4. Conclusions

A new approach for simultaneous and direct detection of benzene, toluene and the three xylene isomers (BTX) in aqueous solution based on polymer coated mid-infrared evanescent wave sensors has been presented. Investigated sensor characteristics include the enrichment time, sensor sensitivity and reproducibility. Linear relationships between characteristic absorption peak areas vs. input concentrations with R^2-values > 0.99 could be obtained for each analyte along with high reproducibility for 5 consecutive measurements. With the presented measurement setup equilibrium conditions for this diffusion based sensor were achieved within approx. 18 min, which is comparable to other membrane based chemical sensor systems. Sensitivity in the low ppb (v/v) region for all BTX compounds and during simultaneous detection experiments represent a significant improvement compared to any ATR-IR sensor reported to date for this class of analytes. At the present stage of development the sensor system is already suitable as analytical device for online, in-situ process monitoring of multiple organic components at low ppb concentrations. Further optimization of the presented method includes

aspects such as flow cell design for minimized response time and chemometric data evaluation enabling remote operation. Hence, multi component measurements with FTIR-ATR techniques in the low ppb to sub ppb region are foreseeable in the near future.

3.2. Simulated Field Conditions – VOCs Determination in an Aquifer

3.2.1. Introduction

These measurements were conducted as part of the IMSIS (In-situ Monitoring of Landfill Related Contaminants in Soil and Water by Infrared Sensing, EVK1-CT-1999-00042) project in order to show the applicability of polymer coated ATR sensor systems at in the field sensing tasks. Measurement conditions at such aquifer systems are highly similar to real world conditions with the advantage that analytes can be introduced at known concentrations. In order to directly detect pollutants in the boreholes of the aquifer system a fiber-optic sensing approach was applied. A specially designed sensor head assembly with 6 m long AgX fibers was developed in collaboration with the research group of Prof. Abraham Katzir (Tel-Aviv University). Analytes of environmental significance (TriCE, TeCE and DCB) have been detected and quantified in high agreement with HS-GC validation measurements throughout this test.

3.2.2. Experimental Setup

3.2.2.1. Materials

E/P-co, with 60% ethylene content was obtained from Sigma-Aldrich (Sigma-Aldrich Handels GmbH, Austria); all other chemicals were of analytical grade. Aqueous stock solutions for sensor calibration were prepared with deionized water. The aquifer system was operated with conventional tap water.

3.2.2.2. Silver Halide Fibers

The silver halide (AgX) fibers used during this study have a composition of $AgCl_{0.4}Br_{0.6}$ [168,169]. Core only fibers have been used with a diameter of 900 µm, a refractive index of 2.13 @ 10 µm and an average damping factor of 0.2 dB/m @ 10.6 µm. AgX fibers are ideally suitable for the proposed sensing application due to high mechanical flexibility, an optical window between 3000 cm^{-1} and 500 cm^{-1}, robustness in the required temperature range from -10° C to +40° C and, finally, long shelf- and application lifetime. However, silver halides are chemically instable when exposed to UV radiation (photolysis), base metals (e.g. aluminum; cementation), hydrogen sulfide (formation of insoluble Ag_2S) and halide ions (complex formation). Thus, sensor head, coating and fiber cables have to ensure appropriate protection of the fiber from environmental impact.

3.2.2.3. Instrumentation

FT-IR

All measurements were performed with a Bruker Vector 22 FT IR spectrometer (Bruker Optik GmbH, Ettlingen, Germany) equipped with a LN_2 cooled mercury-cadmium-telluride (MCT) detector (detectivity D^*=3×10^{10} cm $Hz^{1/2}$ W^{-1}, 0.01 cm^2 detector element, Infrared Associates, Inc., Stuart, FL, U.S.A.). Light coupling from the spectrometer into the fiber optic waveguide and to the detector at the distal end of the fiber was achieved by a custom-built mirror arrangement utilizing one off-axis parabolic mirror (focal length *f*=50.8 mm) at the spectrometer/fiber interface and two similar mirrors (*f*=43 mm) at the fiber/detector interface. SMA compliant connectors for silver halide fiber based optical cables in combination with xyz-positioners ensure rapid and reproducible alignment and connection of the fiber optic probe. For all measurements 100 spectra were averaged with a spectral resolution of 4 cm^{-1} in the spectral range of 4000 cm^{-1} to 400 cm^{-1}. The setting of the aperture after the light source (SiC globar) was open; apodization was a medium Norton-Beer algorithm.

Head Space Gas Chromatography (HS GC)

The HS GC reference analysis was done on a HP 5890 series II GC equipped with flame ionization detection (FID) and electron capture detection (ECD) capability. A Dani HSS 86.50 headspace autosampler (Dani, Milao, Italy) was coupled to the HS GC system. A J&W Scientific Inc. DB-624 capillary column (30 m x 0.250 mm, 1.4 µm stationary phase) was used with nitrogen as carrier and make up gas. Samples were softly shaken for 6 min in the autosampler and injected afterwards using a split/splitless injector kept at 120 °C. The temperature program started at 80 °C for 10 min followed by a ramp to 150

°C at 20 °C/min and finished holding a temperature of 150 °C for 5 min. The detectors were kept at a temperature of 200 °C.

Sensor Head and Fiber Cables

Measurements under field conditions demand careful design of the sensing system and particularly of the sensor head and the fiber optic cables. The mechanical construction of the sensor head has to protect the active sensing region of the fiber at any given time from mechanical damage, while being small enough to be lowered into groundwater monitoring wells with inner diameters of usually 5 cm. Furthermore, unrestricted intimate contact between the active transducer and the probed aqueous phase has to be ensured. Finally, the fiber has to be protected from mechanical damage by strain, squeezing or overbending. Following these requirements a sensor probe was developed and optimized during several field measurement campaigns.

Figure 3.8 Close up of the sensor head, showing the active transducer section of the fiber (1), the o-ring seal lead-troughs (2), the two stainless steel tubes containing the sensor head/fiber cable interface (3) and the fiber cables (4). For the described measurements the semicircular part covering parts of the active transducer was removed.

The prototype sensor head is shown in Figure 3.8 and features a fiber mount made from black Teflon with two O-ring sealed watertight stainless steel lead-throughs accommodating the active sensing zone of the fiber as a loop with a bending radius of 40 mm. Attached to the Teflon mount are two stainless steel tubes forming the interface to the fiber cable and providing water tight sealing and pull relief by O-ring sealed fittings. Each leg of the fiber cable is approx. 3 m long and the silver halide fiber is contrived into dual-layer tubing (inner tubing diameter: 1 mm; outer tube diameter: 8 mm). Connectors attached at the end of the fiber cables follow the SMA standard and hold onto the fiber and the tubing by a watertight O-ring system.

Coating of the sensing zone is performed from a solution of 2.1 % (w/v) E/P-co in hexane/octane (1:1), which was refluxed until fully dissolved. The exposed central section of the fiber was manually dipped twice into the hot E/P-co solution and after

approx. 45 min. once into the same solution at ambient temperature. Afterwards the coating was homogenized with a heat gun set to approx. 100° C for approx. 5 min.

3.2.2.4. Aquifer simulation

The pilot scale aquifer simulation allows studying water flow and analyte dispersion in soil under defined conditions. A stainless steel tank with dimensions of 10 m×1 m×2 m (L×W×H) was filled with soil from the Munich Gravel Plain, a quaternary calcareous and very heterogeneous gravel (permeability coefficient k_f = 7×10^{-3} m/s). A bridge-slot screen on both ends of the tank separates two approx. 10 cm wide compartments from the soil-bed. Water guided into the front compartment (volume V=104 L), i.e. the mixing chamber, will continuously flow through the bridge-slot screen into the soil bed following a 1 % slope. At the rear end of the aquifer system the water is drained through a constant head setup. By varying the height of the run off the hydraulic gradient can be varied between 1 % and 10 %. The water flow $\dot{\mathbf{V}}$ was set to 20 L/min 2.5 weeks before the actual measurements to ensure equilibrium conditions within the soil bed. The water inlet of the front mixing chamber has several nozzles below the water table to homogeneously spread water flowing into the chamber. Additionally, the tubing is filled with a diffuser material (steel wool) further improving the water/analyte mixing process. Analytes are added as highly concentrated methanolic solutions upstream from the nozzle arrangement to the influent water by a peristaltic metering pump. Since the selected model analytes do not dissolve easily in water, they are regarded as dense, nonaqueous phase liquids (DNAPL) and a supplementary solvent has to promote their solubility, as opposed to natural environments where such analytes dissolve at a longer time scale [170].

Furthermore, the content of the mixing chamber is cycled by a submersible pump at a rate of 6 L/min. The distributed injection of the analytes and the circulation via the submersible pump in the mixing chamber closely resemble the concept of a perfectly stirred tank reactor (PSR). The PSR assumes a homogeneous analyte concentration c within the volume of the reactor and its efflux. If the influent concentration c_0 is a function of time t, the generic solution for the concentration $c(t)$ in the reactor is given by

$$c(t) = e^{-t/\tau}\left[C_0 + \frac{1}{\tau}\int_0^t e^{s/\tau} c_0(s)\,ds\right] \qquad (3\text{-}1)$$

where C_0 is the initial concentration within the PSR and τ is the dwell time, respectively [171].

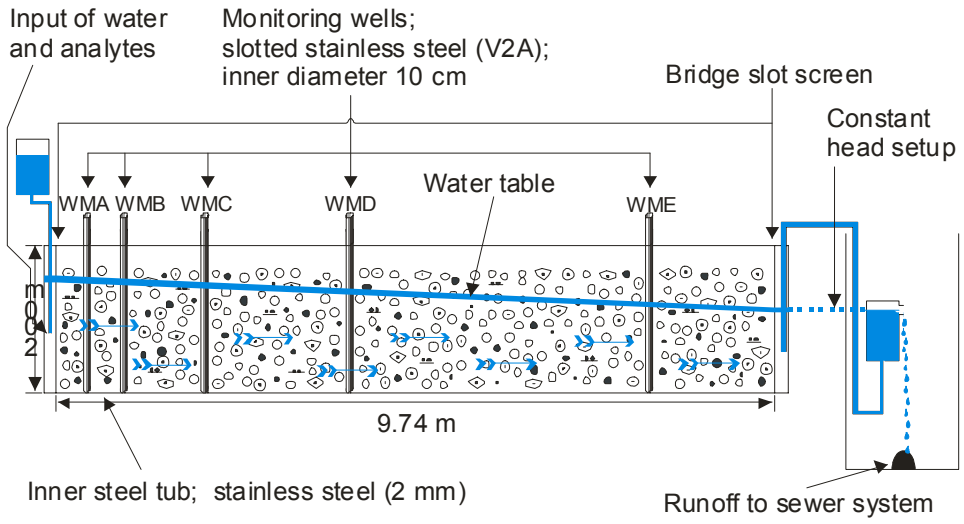


Figure 3.9 Sketch of the aquifer simulation “Munich North” (side view).

Five monitoring wells made from slotted stainless steel tubes are accessible along the aquifer system. With an inner diameter of 150 mm, sensors and sampling lines can be directly introduced into the groundwater flow. All wells cover the entire height of the soil bed allowing measurements in all possible depths (Figure 3.9).

3.2.3. Sensor Calibration and Validation

The aim of the following experiments was to study (a) the sensor response to different analyte concentration profiles, (b) the differences in the response behavior of the sensor readout vs. reference analysis based on off-line HS GC and (c) long-term stability and field readiness of the sensing system. Representative results from several field measurement campaigns are conclusively discussed below demonstrating the feasibility of the proposed mid-infrared chemical sensing concept for the determination of VOCs under field conditions.

Ten hours prior to the first experiment the sensor was installed in monitoring well "WMB", approx. 45 cm below the water surface (water temperature was constant at $\approx$10.2° C) and kept at this location during the whole measurement campaign. Well "WMB" is located 93 cm downstream of the mixing chamber. At an average water flow rate of 20 L/min analytes arrive in well WMB approx. 30 min after their injection to the influent water stream (Figure 3.9). The high water flow rate resulting in a short dwell time τ of 5.2 min only, the multi nozzle array and the pump assisted water recirculation within the mixing chamber lead to a better sample homogenization as observed during previous field experiments, where formation of two phase regions and strong analyte evaporation effects at a flow rate of 6 L/min (τ = 17 min) were evident [95].

Three different analytes, 1.2-dichlorobenzene (DCB), TeCE and TriCE, were selected as relevant model analytes, and their respective characteristic absorption bands in the mid-infrared spectral range were evaluated. At 1036 cm^{-1} and 748 cm^{-1} the aromatic C-Cl stretching vibration and aromatic C-H out of plane vibration of DCB can be observed. TeCE exhibits a strong absorption feature of the C-Cl stretching vibration at 911 cm-1 and TriCE shows two features at 842 cm^{-1} and 932 cm^{-1} respectively. Quantitative information was obtained by conventional evaluation of the respective peak areas as the spectral features appear well separated in the ATR spectrum. Spectra were recorded in 5 min intervals while samples for reference HS GC analysis were collected semi-automatically with a computer-controlled multichannel peristaltic metering pump approx. every 30 min. Sensor and sampling line were placed at the same depth. Hence, the reference samples were collected from the water volume also probed by the IR sensor. Before each sample collection cycle the sampling lines were drained to a waste container for one minute by flushing with water from the monitoring well avoiding adsorption losses to the wall of the tubing and carry-over artifacts.

To study the sensor response to different concentration gradients, three different analyte input functions where defined. In the first run, two rectangular peaks of equal width (120 min) and height (DCB, TeCE and TriCE: 4 mg/L) were injected. Selected concentrations correspond to the order of magnitude of contamination levels found in leachates collected from landfills and contaminated sites. In the second run, two rectangular sample peaks of almost equal width (peak one: 135 min, peak two: 119 min) were injected. The second injected peak had twice the height of the first peak (peak one: DCB and TeCE: 4 mg/L, TriCE: 8 mg/L). In the last experiment analyte concentrations were stepwise increased and decreased subsequently (DCB and TeCE: 4 mg/L, 8 mg/L, 6 mg/L, 4 mg/L; TriCE: double concentration of DCB). Each concentration level was kept for 90 min.

Sensor calibration was performed after the field measurements at laboratory conditions. For the calibration the sensor was immersed into a beaker filled with 1 L of distilled water magnetically stirred for rapid homogenization. Defined amounts of methanolic stock solution containing 10.000 ppm (v/v) of each DCB, TeCE and TriCE were added to the beaker with an adjustable pipette (Transferpette 100-1000, Eppendorf, Hamburg, Germany). Spectra were recorded in 5 min intervals for 40 minutes to monitor sensor response.

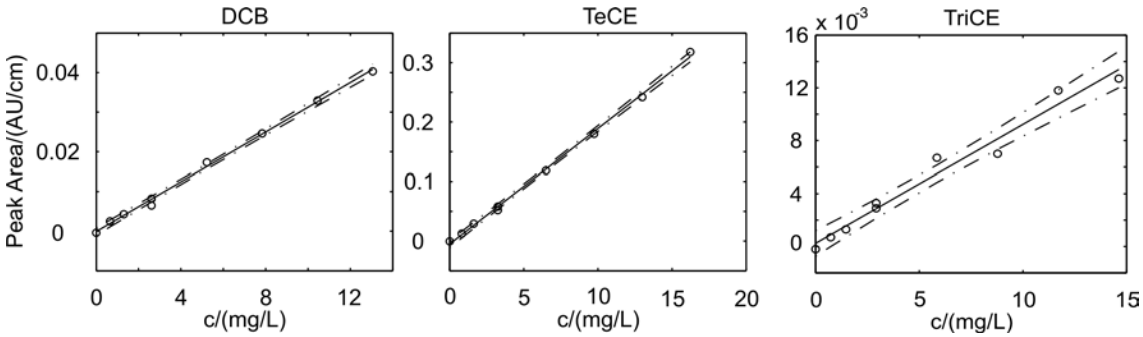


Figure 3.10 Linear regression (solid) and confidence intervals (P=95 %; dash dotted) of the sensor response (circles).

For each concentration step the respective maximum in peak area was used for calculation of the calibration curve. The sensor was calibrated for concentrations in the range between 0 ppm (v/v) and 10 ppm (v/v) (Figure 3.10). The resulting calibration data is given in.

Table 3.2.

Table 3.2 Statistical data of the sensor calibration. Intercept *a*, slope *b*, product-moment correlation coefficient *r*, prediction error SS, standard deviation of slope s_b, intercept s_a and limit of detection LOD. n=9.

Substance	Peak	*a*	s_a	*b*	s_b	*r*	LOD[1)]	SS
	cm^{-1}	AU cm^{-1}	AU cm^{-1}	AU L $mg^{-1}cm^{-1}$	AU L $mg^{-1}cm^{-1}$		mg L^{-1}	mg^2 L^2
DCB	1036	-0.0001	0.0004	0.0031	0.0001	0.9987	0.8	0.4532
TeCE	911	-0.0049	0.0025	0.0194	0.0003	0.9991	0.8	0.4788
TriCE	932	0.0002	0.0004	0.0009	0.0001	0.9866	2.8	5.8846

3.2.4. Results

The experiments demonstrate for the first time that continuous online monitoring of VOCs in groundwater over a period of three days is feasible with the mid-infrared fiber optic sensor probe developed in this study. The aquifer simulation facility grants conditions in an outdoor environment similar to those found at contaminated sites or landfills. In Figure 3.11 the concentration data obtained from the IR sensor during the three experimental series is compared to reference data acquired from simultaneously collected samples analyzed by HS GC.

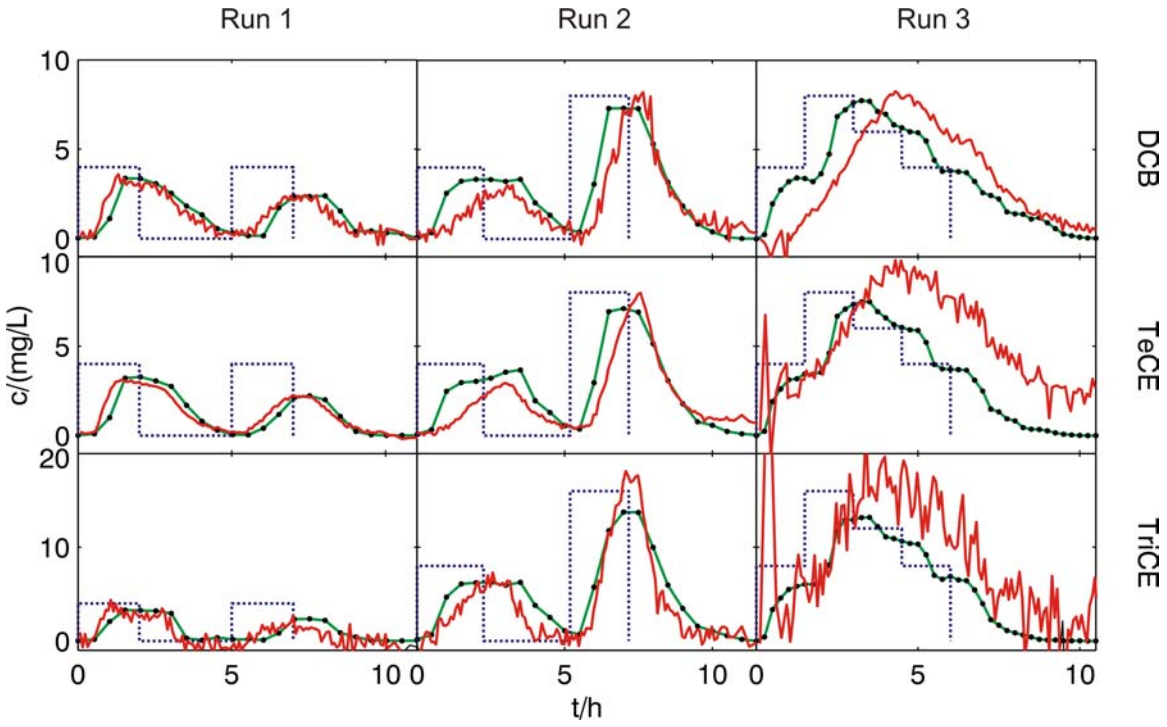


Figure 3.11 Comparison of concentrations measured by IR sensor (red), reference HS GC (green with dots) and analyte concentration added to the water stream (blue). Confidence intervals (P=95%, n=9) at $\bar{y}$ are ± 0.6 mg/L (DCB), ± 0.7 mg/L (TeCE) and ± 2.3 mg/L (TriCE), respectively.

The two rectangular sample peaks injected to the influent during series one were monitored simultaneously by the IR sensor and samples analyzed by HS GC as reference. Although the two peaks were equal in width and height, the recorded HS GC reference concentration profiles differ from each other. The first peak appears broader (peak one 2.5 h, peak two 2.2 h) and higher (peak one 3.4 mg/L, peak two 2.4 mg/L) than the second peak. During the first day the IR sensor response was faster compared to the HS GC reference. The maximum sensor readout was observed 0.2 h (average for the three analytes) before the maximum reference values. This trend was continued in the second peak. In general, reference data shows a high degree of agreement with the sensor data, i.e. 97 % of the DCB, 96 % of the TeCE and 100 % of the TriCE reference values are within the confidence interval of the sensors data.

During the progress of series two agreement between sensor and reference data is not as good as during day one. Again, two peaks have been introduced to the mixing chamber, and the sensor shows much slower response to the rise in concentration for the first peak. Peak concentrations are recorded 0.4 h (DCB), 0.7 h (TeCE) and 0 h (TriCE) after the respective reference values reached their concentration maxima. Also, no distinct concentration plateau as observed in the reference measurements could be determined.

The third day of measurements is characterized by a strong drop in overall light throughput within the fiber optic sensor system (Figure 3.12).

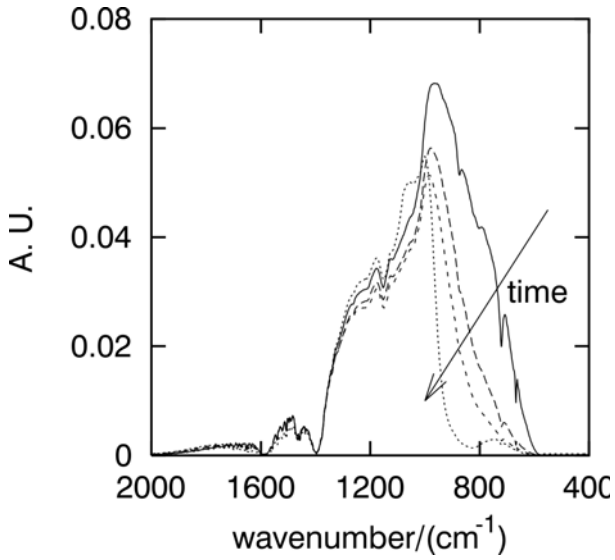


Figure 3.12 During a measurement campaign of three days the single beam spectra changed in shape and overall light throughput as a result of water intrusion. The first spectrum was recorded ten hours before the first experiment (i.e. series 1), the other three at the beginning of each series, respectively. For all spectra the sensor head was immersed in water. Note the strong decrease in light throughput on the third day leading to an impaired signal-to-noise ratio.

Evidently, due to a leak in the sensor head water made contact with the silver halide fiber and thus caused strong absorption losses. This hypothesis is supported by the

observed reversibility of the effect as light throughput and characteristics of the single beam IR spectrum recovered after storage of the sensor in a dry place for several days.

The absorption band of DCB was least effected by this change in light throughput, thus the DCB data shows the best agreement with the reference values during the third day of the field measurements. Sensor and reference show the same maximum concentration values (~7.9 mg/L), however, sensor data appears delayed by 1 h. During stepwise decrease of analyte concentration the time shift between reference values and sensor readout in fact decreases and after 10 h the reference values are again located within the confidence interval of the sensor data.

Despite successful field tests the experiments revealed that during a period of three days changes of the coating absorption behavior are observed affecting the sensor response time. Effectively, this results in a drift of the baseline in the IR absorption spectra and is introducing errors during evaluation of the obtained concentration values. Recently, two novel chemometric methods based on principal component regression (PCR) have been developed in our research group automatically compensating for baseline drifts as integrated algorithm of chemometric data evaluation schemes. Since baseline drifts are broad features compared to analyte absorption peaks the drift contributions can be modeled by polynoms orthogonal to the principal components modeling the concentration result (‘polyPCA’). In a second, more sophisticated approach, drift components are modeled by synthetic pseudo-principal components along with conventional principal components characterizing the analyte peaks (‘pPCA’) [59,60]. Both algorithms have successfully been tested with synthetic spectra and real-world data acquired with mid-infrared chemical sensors and will substantially improve field applicability of chemical sensors in general [60].

Recently, organically modified sol-gels have successfully been tested as novel enrichment matrix in combination with planar infrared waveguides [172-174] and coated onto the surface of silver halide fibers [175].

3.2.5. Conclusions

In this chapter the concept of a fiber optic mid-infrared sensor system for on-line and in-situ monitoring of VOCs in groundwater has been investigated and successful applied during aquifer field studies. During the course of a three day field experiment in an artificial aquifer system the sensor data showed high agreement with the reference data acquired by conventional HS GC analysis of collected water samples with respect to the found concentration maxima and the progression of the analyte concentration levels with time. Due to baseline separated analyte absorption bands in the mid-infrared conventional peak integration methods were used to gain quantitative information. Over time, presumably a delamination process of the E/P-co coating changed the behavior of the sensor system. Response times increased, so that the determined concentration maxima were shifted backwards in time when compared to the reference analysis. Furthermore, decreased light throughput in the relevant portion of the spectral window resulted in a lower signal to noise ratio. The experiments have shown that a better understanding of the long time behavior and ageing of the coating materials is of great importance for successful application of the sensor system. In measurement scenarios where the sensor is deployed and used for autonomous online monitoring, stability of the sensor characteristics are of vital importance.

3.3. Field Conditions – Chlorobenzene in Groundwater

The first tests of ATR based polymer coated sensor systems under real world field conditions were performed at the SAFIRA site (German acronym for ‘‘Remediation Research in Regionally Contaminated Aquifers’’), a remediation pilot plant, in the region of Bitterfeld / Wolfen (Saxonia-Anhalt, Germany) [176,177].

The ground water aquifer in this region has been contaminated over an area of 25 km^2, with a total volume of approximately 200 million m^3, due to activities in open cast lignite mining and related chemical industries for more than 100 years [178]. Since a few years, a local aquifer in the southeast of the city Bitterfeld, contaminated mainly with chlorobenzene (CB), was selected to develop and test new in situ reactive barrier technologies within the German ground water remediation project SAFIRA [179]. The reactive barrier technologies are based on various chemical, physical, and biological processes. The entire on-site pilot plant in Bitterfeld consists of 5 shafts, each with a depth of 23 m and a diameter of 3 m, and a shaft-to-shaft distance of 19 m, housing a total of 20 reactors [177,179].

The following systems are being tested as part of in situ technologies:

- Biodegradation of chlorinated contaminants in an anaerobic/microaerobic system
- Adsorption and simultaneous microbial degradation on activated carbon
- Zeolite-supported palladium catalysts
- Membrane-supported palladium catalysts
- Oxidative solid metal catalysts
- Activated carbon filtration

- Anaerobic microbial degradation of pollutants
- Combination of redox reactors

The subsurface consists predominantly of gravel, which is embedded in lignite and Bitterfeld mica sand. Three aquifers are separated by watertight layers; the reactors of the in situ pilot plant are supplied exclusively with ground water from deeper zones of the quaternary aquifer. In the quaternary aquifer, the contaminants are strongly stratified: groundwater from 5 to 9.5 m in depth is almost non-polluted; in a depth of 9 to 16 m, CB is the dominating contaminant at a concentration of approximately 2 mg/L; in 16 to 22 m of depth, CB concentrations increase to levels up to 51 mg/L. Since the beginning of the measurements in 1997, the hydrochemical parameters and the concentrations of the pollutants in the quaternary aquifer have not changed [176]. The comparatively high contamination levels coinciding with presumably constant concentration levels, along with modern, flexible sampling systems rendered the SAFIRA site highly suitable for first field measurements with the developed IR chemical sensor systems.

3.3.1. Experimental Setup

3.3.1.1. Instrumentation

Sensor Calibration

Data was recorded in a spectral range of 600 cm^{-1} to 1400 cm^{-1} using a Bruker Equinox 55 FT-IR spectrometer (Bruker Optics Inc., Billerica, MA) equipped with a liquid N_2 cooled mercury-cadmium-telluride (MCT) detector (Infrared Associates, Stuart, FL). Flow speed: 4.5 mL/min.

Field Measurements

Data was recorded in a spectral range of 600 cm^{-1} to 1400 cm^{-1} using a Bruker Vector 22 FT-IR spectrometer (Bruker Optik GmbH, Ettlingen, Germany) equipped with a liquid N_2 cooled mercury-cadmium-telluride (MCT) detector (Infrared Associates, Stuart, FL).

Parameters Applicable to Both Scenarios

A total of 100 scans were averaged for each spectrum with a spectral resolution of 4 cm^{-1}. For ATR measurements, a horizontal ATR accessory (Specac, Smyrna, GA) in combination with trapezoidal ZnSe ATR elements (72*10*6mm, 45º; Macrooptica Ltd., Moscow, Russia) and an aluminum flow-cell (custom made, Volume: 2 mL, free contact area to ATR crystal: ~7.2 cm^2) were used. A schematic and a picture of the flow-cell is shown in Figure 3.13. An Alitea C8-Midi peristaltic pump (Watson-Marlow Alitea, Wilmington, MA) was used to ensure continuous flow of the analyte solutions through the ATR cell. In order to minimize adsorption and diffusion losses stainless steel tubing was exclusively used to deliver analyte solutions to the flow cell. A schematic illustration of the setup is shown in Figure 3.14.

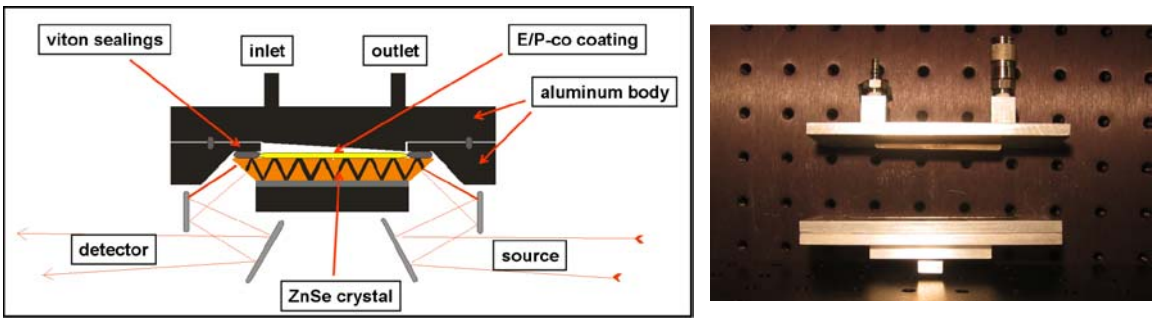


Figure 3.13 Left: Scheme of the custom made flow cell. Right: Picture of the flow cell (disassembled)

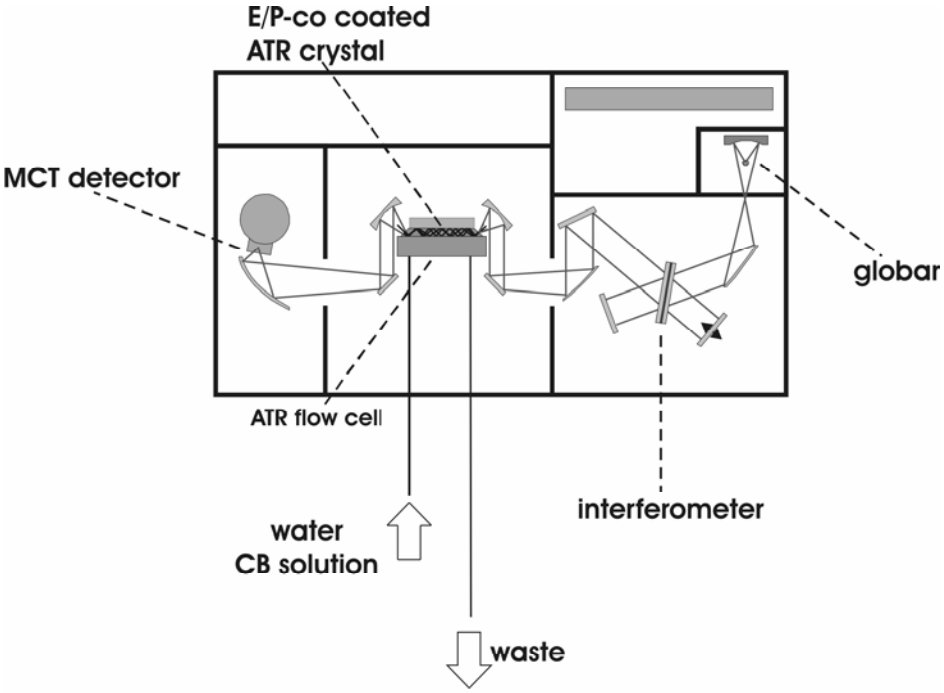


Figure 3.14 Schematic of setup for on-site chlorobenzene measurements.

HS-GC Validation Measurements

The HS GC reference analysis was done on a HP6890 series II GC equipped with flame ionization detection (FID). An Agilent 7694 headspace autosampler was coupled to the HS-GC system. A Chrompack CP-Sil 6B capillary column (30 m x 0.250 mm, 50 µm stationary phase) was used with nitrogen as carrier and make up gas. Samples were softly shaken and extracted at 60°C for 60 min in the autosampler and subsequently injected using a split/splitless injector kept at 250 °C. The temperature program started at 45°C for 5 min followed by a ramp to 200 °C at 20°C/min and finished holding a temperature of 200 °C for 3 min. The detector was kept at a temperature of 280 °C.

3.3.1.2. Preparation of the Extractive Polymer Membrane

A 1 % (w/v) coating solution of E/P-co was prepared by dissolving 0.5 g of granular polymer under reflux in 50 ml n-hexane. Prior to coating, a new ATR crystal was thoroughly rinsed with methanol. About 210 µL of clear, hot solution were applied to the surface of the ATR crystal using an Eppendorf pipette. The crystal was kept at room temperature for at least 2 h ensuring evaporation of most of the solvent. Subsequently, the polymer coating was exposed to hot air treatment with a hot air gun at 150 °C for 5 min to remove remaining traces of solvent and then kept tempered at 80°C in an oven overnight. The thickness of the layer was determined by differential weighing to be 3.2 µm.

3.3.1.3. Sensor System Calibration

Among all VOCs chlorobenzene (CB) is the main pollutant in the groundwater aquifer around the SAFIRA site by several orders of magnitude. According to previously published data on the composition and concentrations of the pollutant cocktail in the groundwater of the Bitterfeld region [176] calibration for CB of the sensor system was conducted in the concentration range from 10 mg/L to 80 mg/L at the ASL laboratory at Georgia Tech. Field measurements were conducted applying a (smaller) Bruker Vektor 22 FT-IR spectrometer. The high agreement of the results obtained under field conditions and the laboratory measurements demonstrate the transferability of the calibration data.

Prior to the calibration measurements the coated sensor element was submersed in water and equilibrated over night. Following, the calibration set consisting of 10 mg/L, 20

mg/L, 30 mg/L, 50 mg/L and 80 mg/L of CB in water was measured regenerating the E/P-co layer after each calibrant. From the MIR absorption spectra of CB, the band with the highest intensity (aromatic C-H out of plane vibration around 740 cm^{-1}) was selected for data evaluation via peak integration.

Figure 3.15 shows the significant part of the CB absorption spectrum for 5 different concentrations of CB in water after partitioning into the E/P-co layer for 24 minutes.

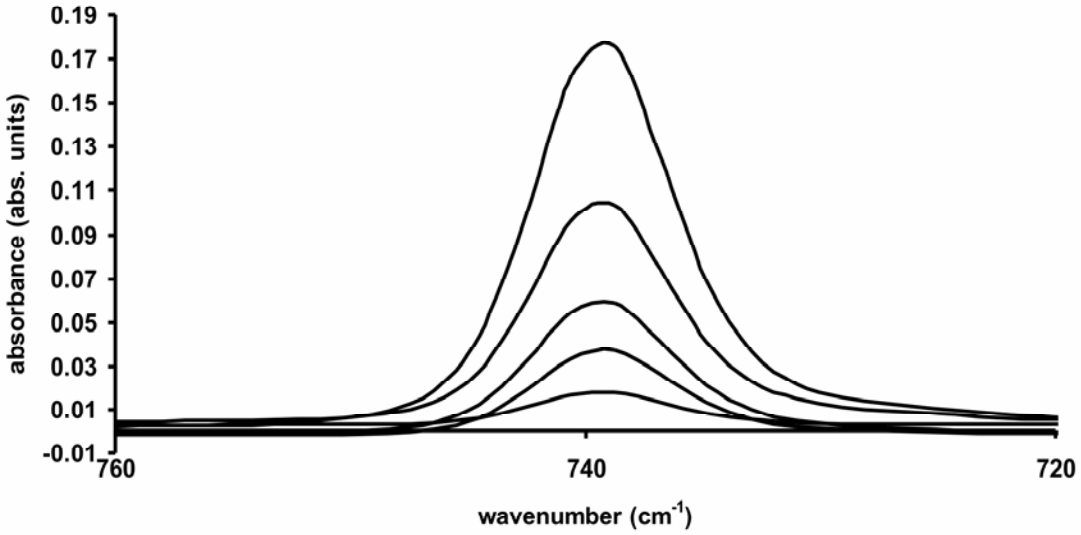


Figure 3.15 The C-H out of plane vibration band of CB for 5 different concentrations after partitioning into the E/P-co layer.

Data evaluation in this case could be performed by simple band integration, as there is only one compound present in the solution. Calibrations have been performed before and after the measurement campaign with approx. 4 weeks of time lapse in between the first and the last calibration set. Figure 3.16 shows a linear calibration derived from band integration for 3 repetitive (1 measured before and 2 after the measurement campaign) runs of the calibration set.

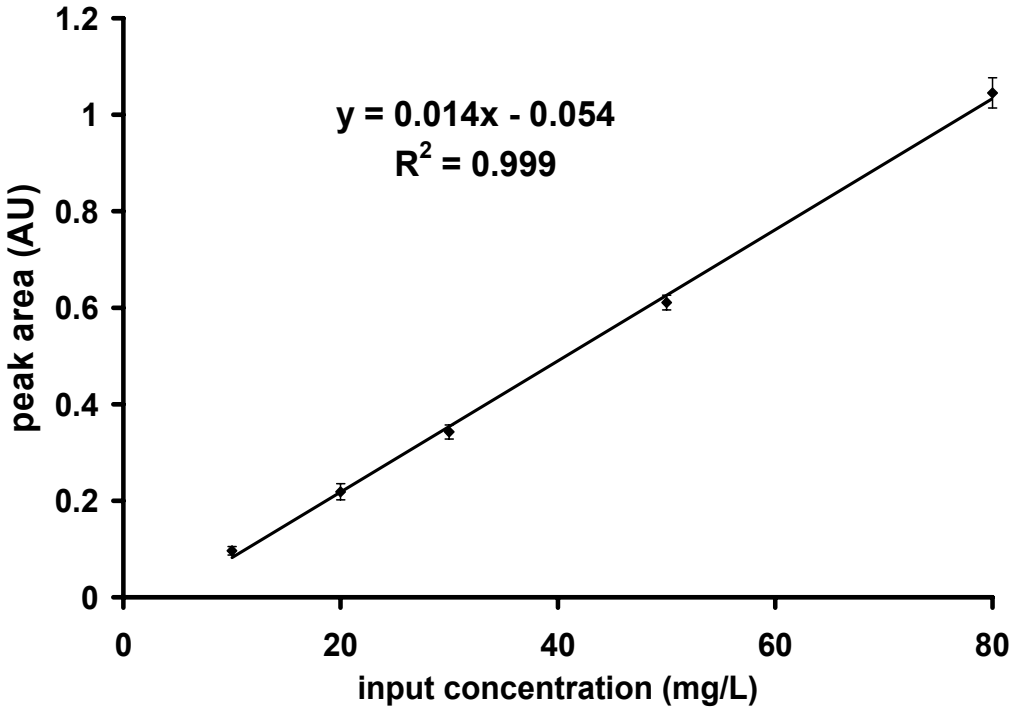


Figure 3.16 Calibration curve for 3 repetitive measurements of the calibration set of CB. Error bars are derived from the standard deviation for each data point.

The acceptable linearity and the small standard deviation proof that the sensor provides superior stability despite being submersed in a highly polluted sample during the measurement campaign, the mechanical stress of being transported and several drying/wetting cycles.

3.3.2. Results

3.3.2.1. SAFIRA Measurement Surrounding Conditions

The SAFIRA site offers an automated sampling system. A partial flow of groundwater is permanently pumped from various depths in the shafts through an array of glass bottles,

which are placed in a cooled storage chamber in the adjacent analytical laboratory and from there back to the reactors for remediation procedures. The same flow configuration is available for water exiting the reactors after the remediation processes. Thus, samples can be conveniently collected from the glass bottles in the laboratory environment rather than descending instrumentation into the shafts. For this first test of an ATR based sensor system for groundwater monitoring it was decided to perform repetitive measurements of water from one shaft for several days in order to verify accuracy and stability of the developed sensor system at field conditions. Concentration levels in the groundwater flow in the Bitterfeld region can be considered constant for the measurement period of several days [176]. Table 3.3 shows the concentration and composition of the groundwater revealing chlorobenzene as the main pollutant by almost 2 orders of magnitude.

Table 3.3 HS-GC validation measurements of groundwater sample from shaft 5 at the SAFIRA site.

sample	*B5-HB*
date	*9/22/2003*
chlorobenzene	***27.92 mg/l***
ethylene	*12.19 µg/l*
vinylchloride	*0.05 mg/l*
1,2-trans-dichloroethylene	*0.04 mg/l*
1,2-cis-dichloroethylene	*0.06 mg/l*
benzene	*0.13 mg/l*
2-chlortoluolene	*0.05 mg/l*
1,4-dichlorobenzene	*0.52 mg/l*
1,2-dichlorobenzene	*0.29 mg/l*

Besides the listed pollutants, the groundwater is characterized by rather high hydrogensulfide contents (up to 5 mg/L) and low concentrations of inorganic pollutants (e.g. heavy metals, arsenic, etc.). Other noteworthy characteristics are high levels of sulfate (up to 1000 mg/l) and chloride (approx. 1300 mg/l).

3.3.2.2. Chlorobenzene Enrichment Behavior

After sensor equilibration with distilled water over night, a stable baseline without noticeable spectral changes due to water diffusion was obtained. Subsequently, groundwater was pumped through the flow cell and spectra were recorded every 2 min until equilibrium was reached. Figure 3.17 shows exemplary spectra of a groundwater sample from shaft 5 and a calibration solution of 50 mg/L CB in water. Besides a concentration related difference in band intensities and the bands related to E/P-co swelling (approx. 780 cm^{-1} – 800 cm^{-1}), both spectra appear to be identical. Hence, we can deduct that all contaminants except chlorobenzene are below the threshold level of detection for this sensor system. The aromatic C-H out of plane vibration of CB around 740 cm^{-1} was selected for data evaluation via peak integration.

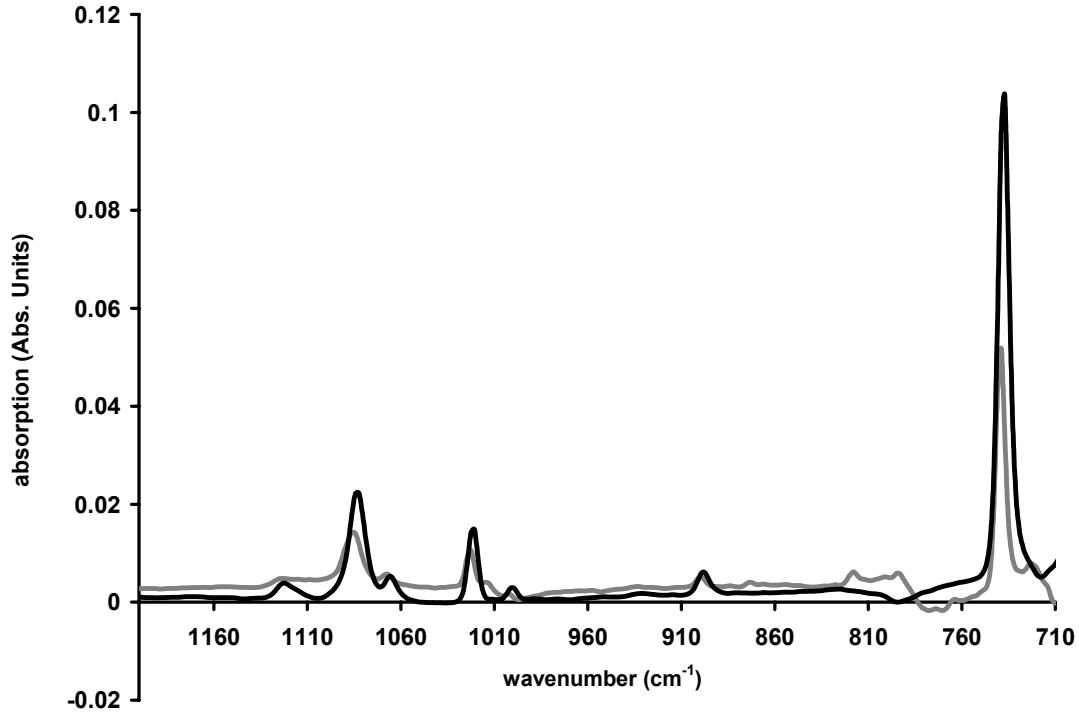


Figure 3.17 Exemplary spectra of a groundwater sample from shaft 5 (grey line) and a calibration solution of 50 mg/L CB in water (black line). Spectra were recorded after 24 min of exposure time to the polymer coated transducer. The peak area of the band at 740 cm^{-1} is used for data evaluation.

A small but noticeable detail can be extracted from the comparison of the spectra in Figure 3.17:

The groundwater spectrum (and all other spectra recorded at the SAFIRA site), is blue shifted by about 3 cm^{-1}. This is explained by the fact that the Vector 22 instrument used on-site had not been calibrated for a number of years and obviously drifted in the wavelength accuracy over that long period of time. This could be compensated for by adapting the spectral region of the band integration during data evaluation in respect to the calibration measurements (performed on the Equinox 55).

Figure 3.18 and Figure 3.19 show the enrichment behavior of CB into the E/P-co layer by plotting the peak area (740 cm^{-1} band) over time for 2 different flow rates of the sample solution (Figure 3.18: 4 mL/min; Figure 3.19: 23 mL/min).

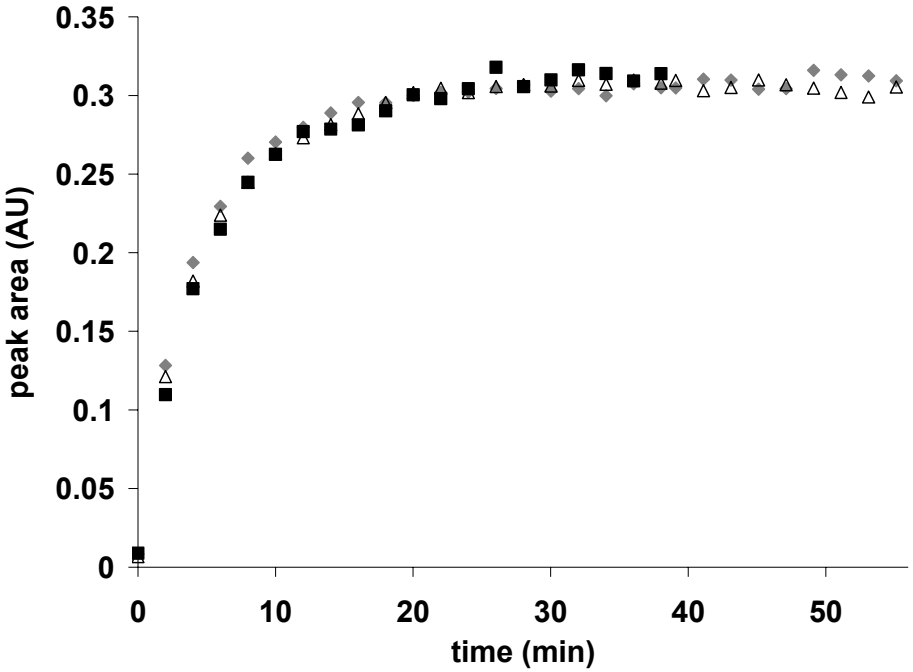


Figure 3.18 Enrichment curves of CB from groundwater at SAFIRA site into the E/P-co layer at a flow rate of 4 mL/min. The 3 measurements were performed at 3 different days.

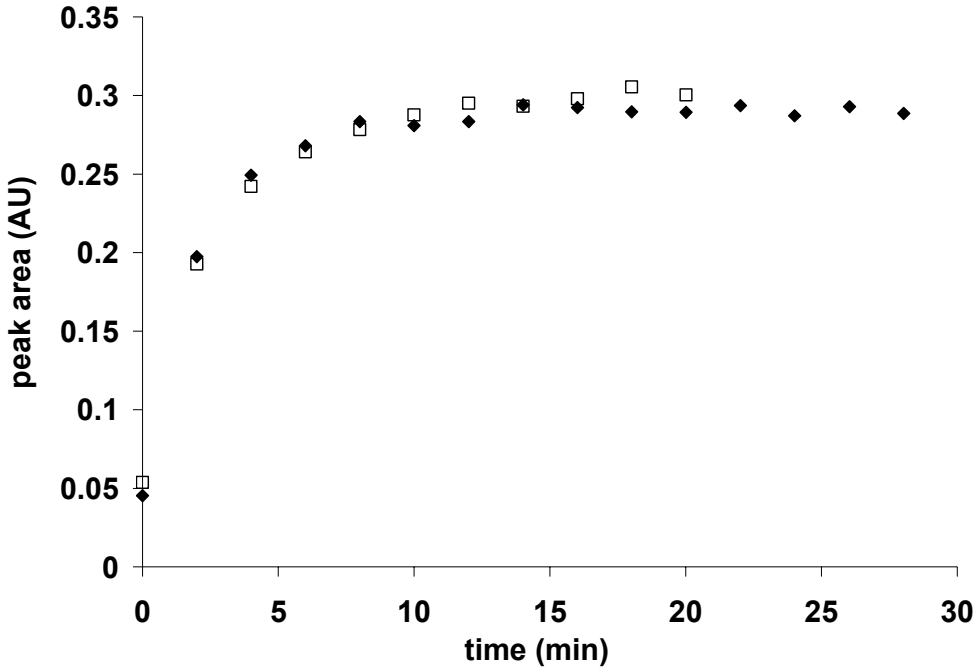


Figure 3.19 Enrichment curves of CB from groundwater at SAFIRA site into the E/P-co layer at a flow rate of 23 mL/min. The 2 measurements were performed at 2 different days.

It can be seen that after reaching equilibrium conditions the peak area remains almost identical for both flow rates, while for the higher flow rate the equilibrium is reached faster. These results indicate that changes in the flow conditions above the polymer layer affect the response time of the sensor. The dependence of the response time on the flow rate or, more specifically, the flow conditions above the extractive membrane has only recently been studied for infrared sensors accounting for results reported from pervaporation and ultrafiltration experiments, which suggested such dependencies [155-157]. Some experimental considerations of these effects were discussed by Roy et al [158]. Extensive CFD simulations have been presented by Phillips et al [43] and Louch [159] on related issues. For a detailed description of the influence of the flow velocity on the equilibration process please refer to chapter 2.6.

A promising result for future IR chemical sensor applications is the obtained reliable and stable performance over a time period of at least 5 days at field conditions. We feel confident to extrapolate this time span significantly considering that the calibration data has been recorded over a period of several weeks prior to and after the field measurement campaign without deviation in performance.

To verify the accuracy of the ATR measurements, the peak areas of the absorption feature at 740 cm^{-1} of all 5 measurement periods have been evaluated with a linear regression function (Figure 3.16) and the values have been compared to the HS-GC validation measurement (Table 3.4).

Table 3.4 Comparison of the ATR measurements to the HS-GC measurements.

	Run 1	*Run 2*	*Run 3*	*Run 4*	*Run 5*
measured conc. CB (mg/L)	**28.71**	**27.54**	**31.05**	**28.01**	**29.05**
HS-GC (mg/L)	27.92	27.92	27.92	27.92	27.92
differnce mg/L	0.79	0.38	3.13	0.09	1.13
error %	2.82	1.36	11.21	0.31	4.03

With an average deviation of only 1.10 mg/L (3.94%) to the validation measurement the ATR sensor system provided surprisingly accurate results. Additionally, these results verify that it is valid to calibrate such a sensor systems at laboratory conditions, still yielding reliable accurate results at field conditions. In this first study the cocktail of pollutants present in the groundwater (although at lower concentrations than CB), the difference in pH level and the measurements at cooled conditions (groundwater samples was at least several degrees Celsius colder than the laboratory calibration samples) did not significantly affect the sensor system performance.

3.3.2.3. Sensor Regeneration

In this series of experiments it has been observed that the enrichment of CB into the E/P-co layer was completely reversible. In Figure 3.20 exemplary spectra of a groundwater sample from shaft 5 and a successively recorded spectrum after regenerating the sensor with distilled water are shown. The enrichment step was performed until equilibrium was reached (approx. 24 min at a flow rate of 4mL/min) and sensor regeneration was performed by rinsing the flow cell with distilled water for the same period of time and at the same flow rate.

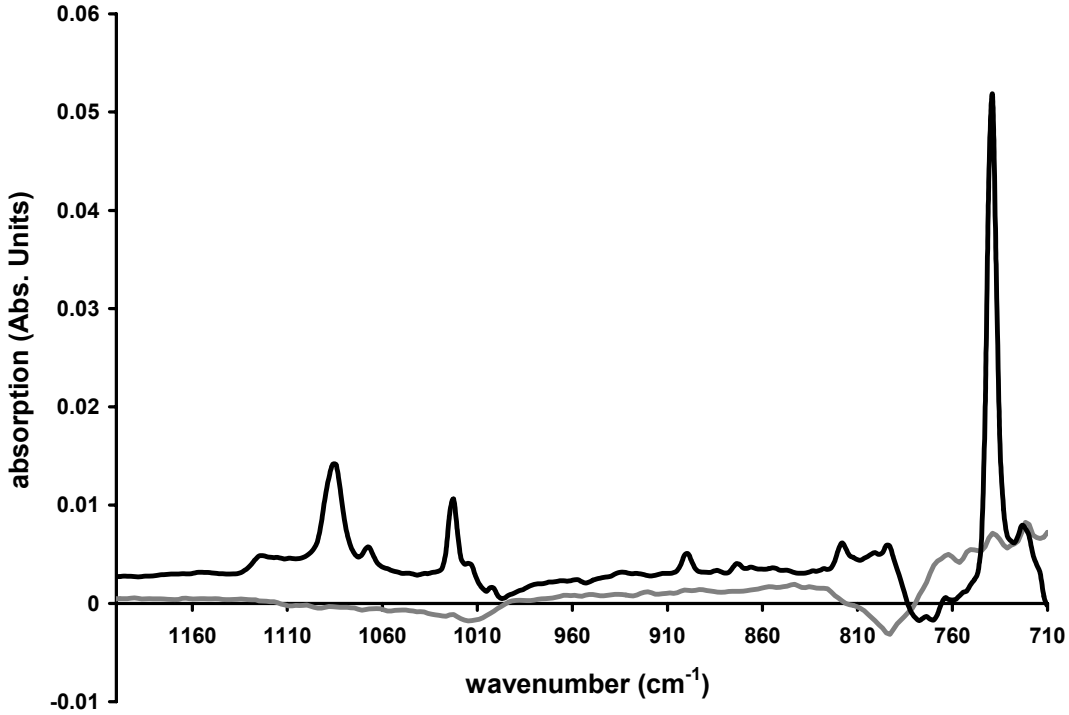


Figure 3.20 Exemplary spectra of a groundwater sample from shaft 5 (black line) and a successively recorded spectrum after regenerating the sensor with distilled water (grey line). Enrichment time and regeneration time were both 24 min with a flow rate of 4 mL/min. The peak area of the band at 740 cm^{-1} is used for data evaluation.

It is evident that an equally long regeneration time is not sufficient for complete CB removal from the E/P-co layer, as a weak absorption feature at 740 cm^{-1} can still be observed.

In Figure 3.21 enrichment of CB into the polymer and sensor regeneration are plotted for 2 flow rates, 4 mL/min and 23 mL/min, respectively. As expected, the curve for the higher flow rate reaches equilibrium conditions faster and shows a more rapid depletion of the analyte within the polymer layer during the regeneration step.

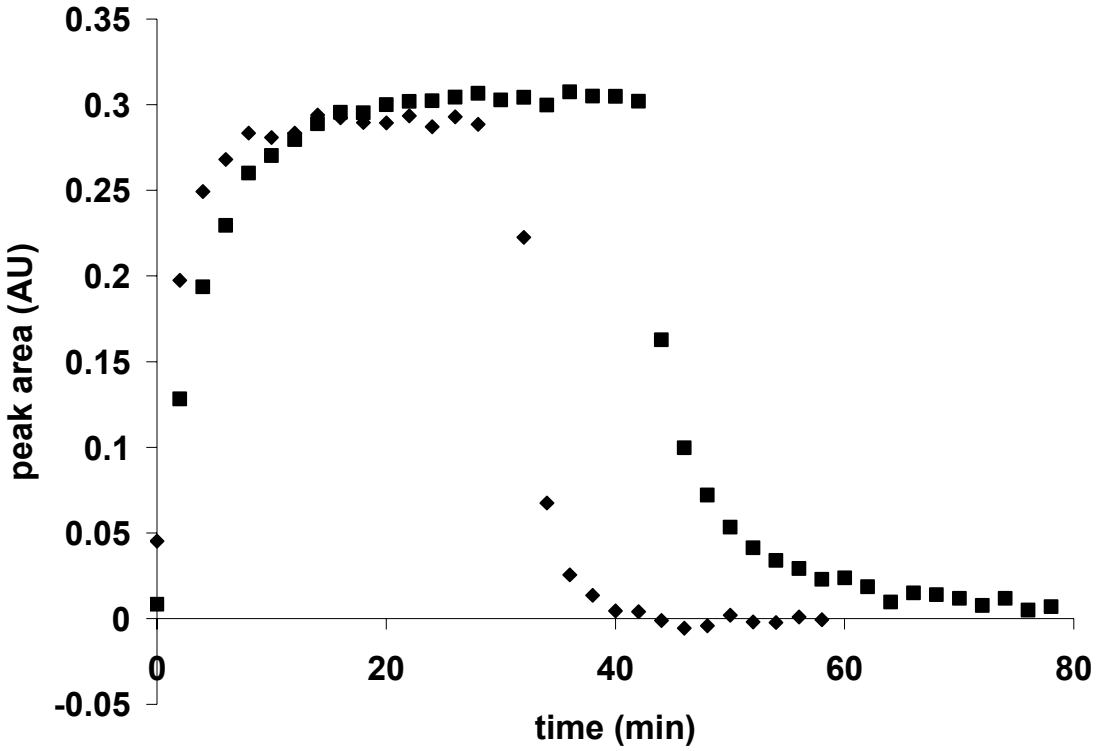


Figure 3.21 Enrichment and regeneration cycle for CB at 4 mL/min (squares) and 23 mL/min (diamonds).

According to these results, the time needed for performing one full enrichment / regeneration cycle for CB with concentrations in the mg/L range at a flow rate of

23 mL/min can be estimated at around 30 min. For lower flow-rate this time easily exceeds 60 min.

However, it can also be concluded that CB completely diffuses out of the polymer layer when regenerated with distilled water (no “memory effect”), potentially enabling numerous measurement cycles with a single sensor system.

3.3.2.4. Long Term Stability

An important figure of merit for the performance of a chemical sensor (system) is its ability to provide accurate readings over a long period of time and without the need for sensor re-calibration or other measures interrupting a continuous monitoring process. In the case of polymer coated ATR sensor systems based on infrared spectroscopy such events may include for example:

- a drifting baseline
- changes in the extraction performances of the polymer layer (biofouling, extensive swelling, etc.)
- degradation of the IRE (e.g. oxidation processes)

As already mentioned, a promising result was the fact that the performance of the sensor system for single measurement procedures was not significantly affected during a time period of several weeks, including the calibration measurements. Nevertheless, for detecting e.g. baseline drifts it is necessary to perform a single continuous measurement over a longer period of time (Figure 3.22).

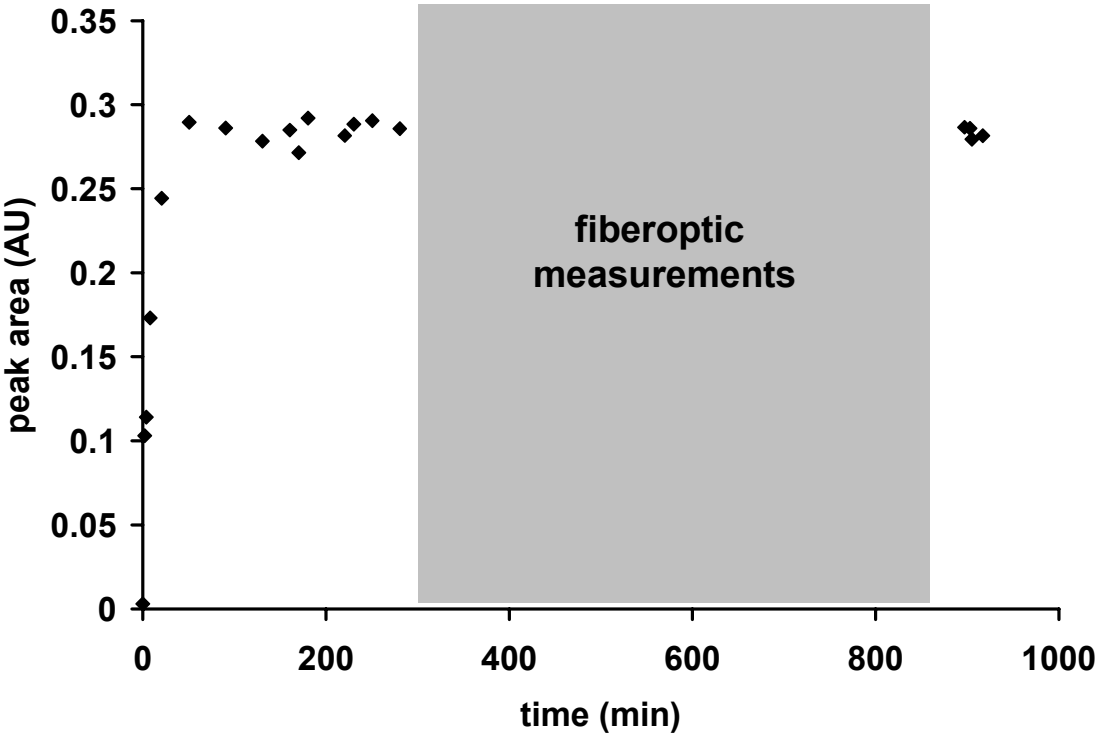


Figure 3.22 Long term stability test for CB measurements in groundwater. The flow rate was set to 4 mL/min. The lack of data in the time period from 400 to 900 min is due to occupancy of the spectrometer by fiberoptic measurements.

This measurement period covered 15 h revealing no significant deviation of the peak area of CB in groundwater once equilibrium was reached. The lack of data in the time period from 400 to 900 min is due to occupancy of the spectrometer by fiberoptic measurements. These data provide a first indication that the developed sensor systems have the potential for delivering reliable results also in case of a continuous monitoring scenario over extended periods of time.

3.3.2.5. Dynamic Sensor Behavior

In the following experiments a change of contamination level was introduced to the sample by adding CB to the groundwater. The resulting sensor behavior provides insight on the dynamic performance of the sensor system and the response to a concentration gradient in the groundwater aquifer, for instance in case of a chemical spill event. For these experiments the sensor system was continuously measuring groundwater and after equilibrium was reached a significant amount of CB was added to the sample (at t=14 min, total CB concentration increased to approx. 100 mg/L).

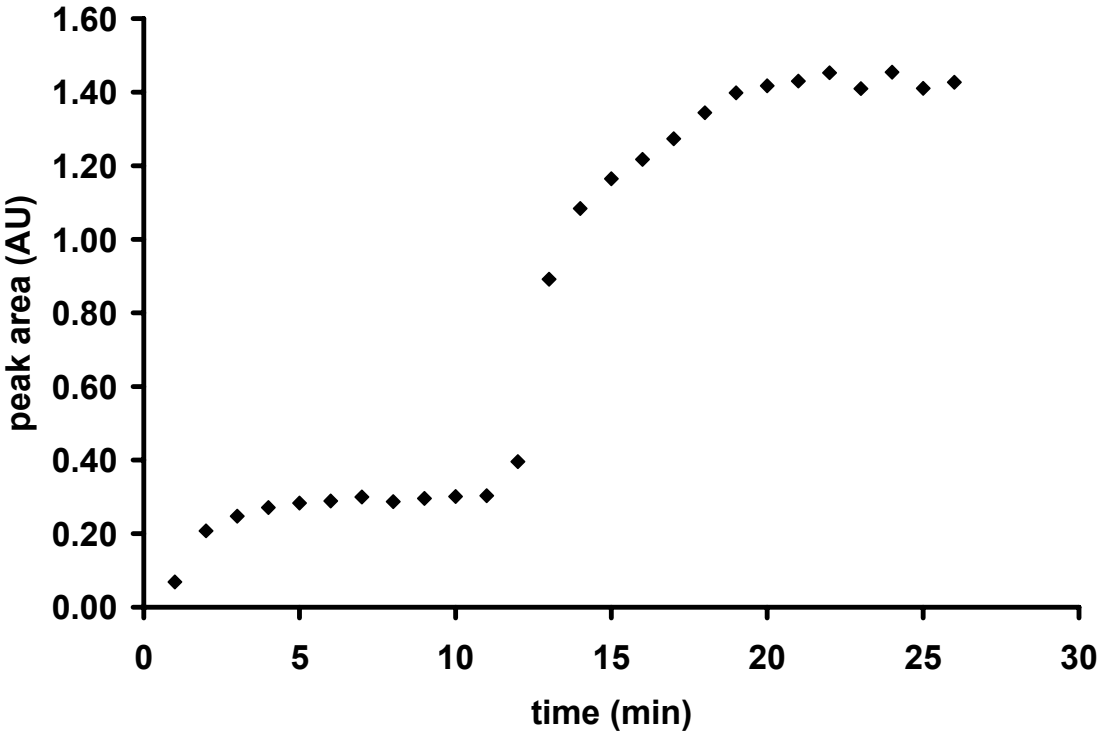


Figure 3.23 Simulation of a chemical spill event, by adding a significant amount of CB to the groundwater sample (at t=14min) after the sensor system was equilibrated with the groundwater sample.

In Figure 3.23 the response of the sensor to a substantial increase of pollutant can be observed as an immediate increase in peak area of the respective band of CB. Approx. 10 min after the spiking event the increase in peak area levels off again approaching equilibrium conditions. Analogous to all other experiments, the response time to a change of analyte concentration occurs in < 2 min, which is sufficient to serve as a chemical spill detector in groundwater streams.

3.3.3. Conclusions

The first measurement campaign deploying a polymer coated IR-ATR sensor system at field conditions has successfully been performed for the determination of chlorobenzene in the groundwater aquifer of a remediation site:

- Performance of the sensor was accurate and stable over a period of time of several weeks.
- Quantitative results were in excellent agreement with HS-GC validation measurements.
- The cocktail of pollutants present in the groundwater did not significantly affect the sensor performance, no cross-interference could be detected.
- Calibration of such sensor systems at laboratory conditions has proven to be valid; the performance was not affected by the pH level and turbidity of the real world sample.
- The effect of changing flow conditions on the equilibrium times has experimentally been confirmed, as suggested by CFD simulations.

- The enrichment of CB into the polymer membrane was completely reversible and there was no indication of an observable memory effect.
- The sensor performed for individual measurements and in continuous monitoring operation, and proved suitable during the simulation of a chemical spill event.
- The minimum measurement repetition time for a complete enrichment and sensor regeneration cycle for the available setup was determined to be approx. 30 min, however, could be improved with higher sample flow rates and the introduction of an optimized flow cell geometry.
- Dynamic sensor behavior has been shown to be < 2 min for increasing and decreasing pollutant concentration in the analyzed sample. A timescale sufficient for remediation processes.

3.4. Modeling the Diffusion Behavior of Chemical Sensors – How Accurate are Existing Models?

It has been confirmed during the studies of this thesis that sensor response times to changes in analyte concentrations usually occur at very short time scales (Figure 3.23). According to these measurements the response time of a detectable change in signal can be estimated to be less than 2 min, resulting in a significant increase of the absorption peaks after analyte introduction. This behavior has been observed for all analytes that have been studied throughout this thesis. For many monitoring applications this will be a satisfactory time resolution for detecting significant changes in the probed

sample. In addition, commonly applied FT-IR instrumentation allows to increase the measurement rate. If needed, time resolutions at the order of a few seconds can thereby be achieved.

In case of accurate quantification, the most common solution is to perform data evaluation as soon as diffusion equilibrium conditions are reached. Depending on the experimental circumstances, this time frame can range from a few minutes up to several hours. For many applications this time frames are not acceptable and considerable efforts are directed toward lowering sensor response evaluation times, usually by evaluating data prior to equilibrium conditions. A good overview on different approaches based strictly on gradient methods of diffusion curves for off-equilibrium data evaluation has been given by Buerck et al. [42]. In this work, it is shown that the gain in evaluation time usually comes along with a loss in sensitivity as a consequence of the evaluation of fewer data points.

Hence, an evaluation algorithm, which could “predict” diffusion curves based on first physical principles, should deliver more accurate results even working with a very limited number of data points. Some approaches have been presented with numerical algorithms exclusively based on Fickian diffusion of analytes in the polymer membrane. The most widely used algorithm nowadays has been introduced by Fieldson and Barbari [185], however, this algorithm – as will be discussed in chapter 3.4.2 - entirely neglects the critical influence of the flow conditions.

Recently performed CFD simulations (chapter 2.6.1) predict a significant influence of the flow conditions in the surrounding solution of the polymer membrane on the sensor response. These predictions could be verified with experimental data in the studies encompassed in this thesis (chapter 3.3.2.2).

In the following chapters the diffusion coefficient of CB in E/P-co will be calculated with two of the commonly applied models. It will be shown that the introduction of a varying

flow rate leads to contradicting results rendering these models insufficient for describing the real world measurement situation in dynamic (flow based) chemical sensor systems. Hence, published diffusion coefficients, which have been determined with such models via data derived from polymer coated IR-ATR sensor systems, are presumably incorrect and should be evaluated with caution. Furthermore, it will be shown that the predicted trends from CFD simulations can be observed in the experimental data, which leads to the conclusion that the consideration of flow conditions with models for diffusion based IR-ATR chemical sensors is inevitable.

3.4.1. First Case Study: A Simple Diffusion Model applied to Experimental Data

Based on a very simplified, but generally accepted method for the determination of the diffusion coefficient of a compound penetrating a polymer membrane, it will be demonstrated that incorrect results will be obtained if the flow conditions are neglected. For this test, a data set from the field measurements with the developed IR-ATR sensor system described in chapter 3.3 is utilized, as it has been obtained with constant experimental parameters, except for a variation of the flow rate.

The data set consists of 5 measurements of the same groundwater sample (refer to chapter 3.3 for a detailed description of the experiment), with a CB concentration of 27 mg/L. The enrichment process of CB into the E/P-co layer (thickness: 3.2 µm) was evaluated via peak integration of the aromatic C-H out of plane vibration (around 740 cm^{-1}). Three runs have been performed at a flow rate of 4 mL/min (Figure 3.24) and two runs at a flow rate of 23 mL/min (Figure 3.25).

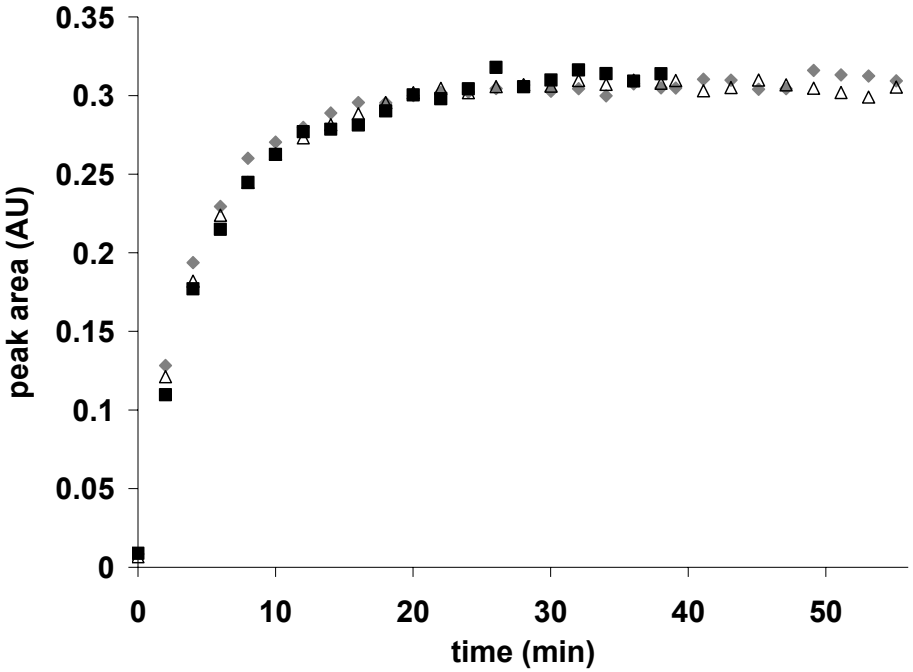


Figure 3.24 Enrichment curves of CB from groundwater at SAFIRA site into an E/P-co layer at a flow rate of 4 mL/min. The 3 measurements were performed at 3 different days.

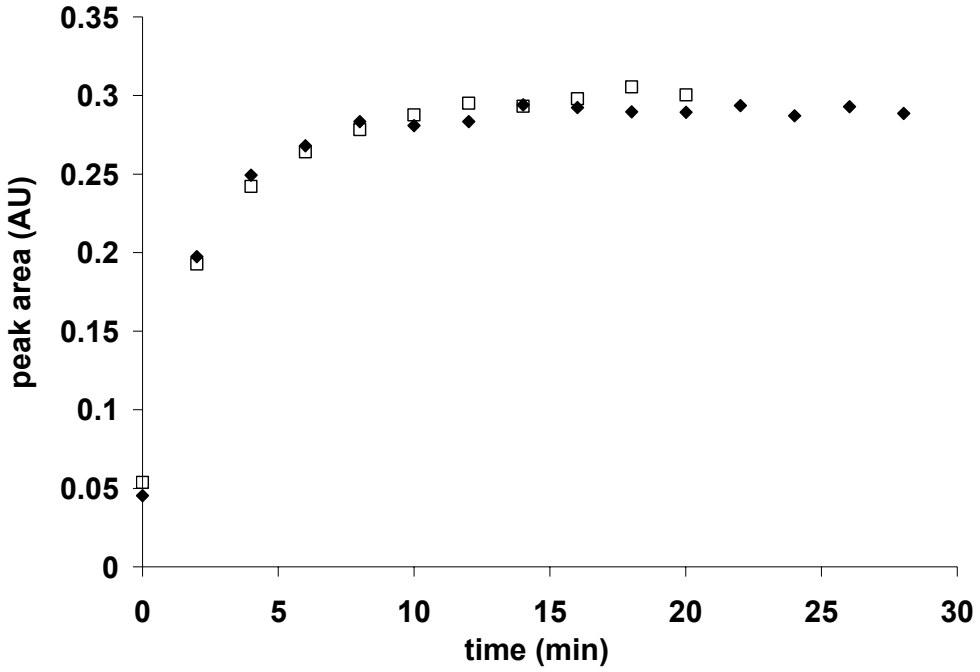


Figure 3.25 Enrichment curves of CB from groundwater at SAFIRA site into the E/P-co layer at a flow rate of 23 mL/min. The 2 measurements were performed at 2 different days.

The difference in flow rate clearly has an impact on the enrichment process. The measurements performed with 23 mL/min more rapidly reach equilibrium conditions. A calculation example will show that neglecting the impact of the flow conditions leads to misleading results, for instance if the diffusion coefficient is derived from such enrichment data. The only published value for a diffusion coefficient of CB in E/P-co was given by Goebel et al [180] as $5 \cdot 10^{-9}$ cm^2/s, derived with stopped-flow experiments using a box model algorithm, which will not be described in more detail in this thesis [180,181]. Generally, the one-dimensional molecular diffusion in a polymer film with a constant diffusion coefficient can be described by the second Fickian law [182].

$$\frac{\partial c}{\partial t} = D \frac{\partial^2 c}{\partial x^2} \qquad (3\text{-}2)$$

where *c* is the concentration of the penetrant, *D* is the diffusion coefficient and *x* the direction normal to the facet of the IRE element.

If a polymer film is placed into contact with a solution containing a diffusant, it has been shown that under certain boundary condition (no diffusion at the edges of the film) the mass transported at the time *t* can be expressed by [182,183]

$$\frac{M_t}{M_{max}} = 4\left(\sqrt{\frac{D}{\pi}}\right)\frac{\sqrt{t}}{d} \qquad (3\text{-}3)$$

where M_{max} is the mass uptake at saturation, M_t is the mass uptake at time *t*, *d* is the film thickness, and *D* the diffusion coefficient. It has been demonstrated that for $M_t/M_{max}<0.5$ the diffusion coefficient of the diffusing species can be derived according to [184]

$$D = \frac{\pi}{16} I_s^2 \qquad (3\text{-}4)$$

where I_s is the initial slope in a plot of M_t/M_{max} versus $t^{0.5}/d$. In a simplifying assumption, M_t and M_{max} can be regarded as absorption intensities Abs_t and Abs_{max} of characteristic peaks of the respective analyte. The resulting graph is shown in Figure 3.26.

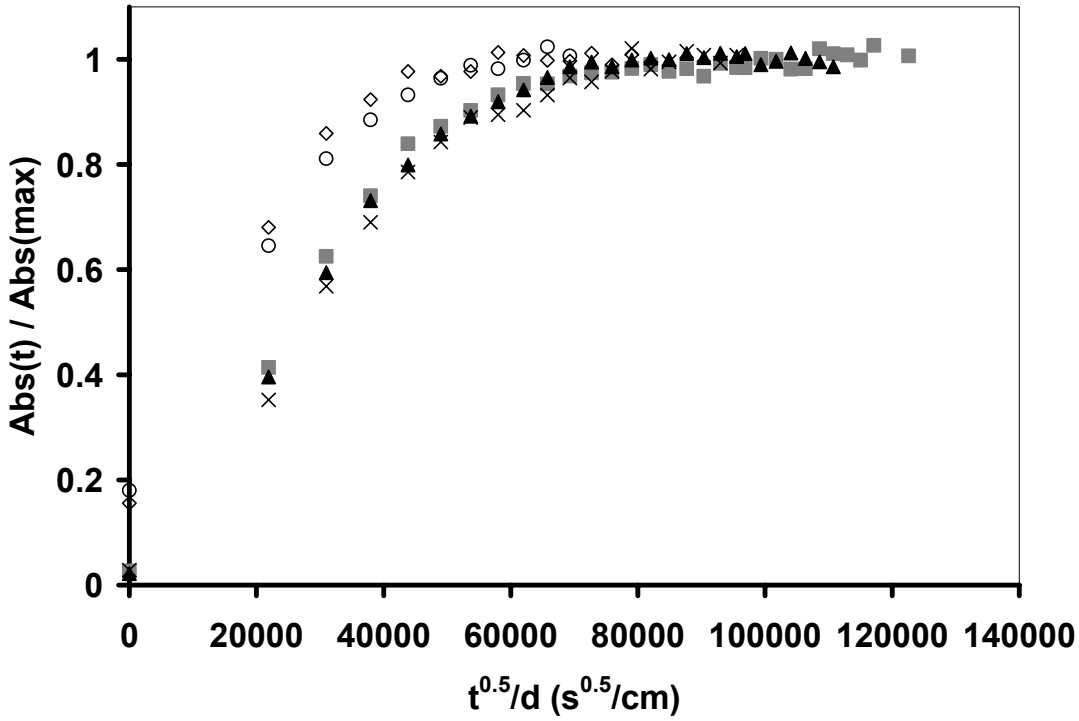


Figure 3.26 $Abs_{(t)} / Abs_{max}$ versus $t^{0.5} / d$ plot for the 5 CB enrichment experiments. The series marked with diamonds and circles refer to the high flow rate (23 mL/min) the other 3 series to the lower flow rate (4 mL/min).

Finally, the diffusion coefficients are derived by extracting the initial slopes of the 5 data series and applying equation 3-4. The results are listed in Table 3.5.

Table 3.5 Calculated diffusion coefficients (cm^2/s) for CB in E/P-co

	flow rate 4 mL/min			*flow rate 23 mL/min*	
initial slope	1.77E-05	1.73E-05	1.71E-05	2.04E-05	2.12E-05
diffusion coefficient (cm^2/s)	6.12E-11	5.86E-11	5.72E-11	8.21E-11	8.86E-11
average diffusion coefficient (cm^2/s)	**$5.90 \cdot 10^{-11} \pm 2.0 \cdot 10^{-12}$**			**$8.53 \cdot 10^{-11} \pm 4.6 \cdot 10^{-12}$**	

The discrepancy of the averaged diffusion coefficients for the two flow rates proves the necessity of considering the flow conditions within the model for the enrichment studies

and render quantitative calculations or modeling based on such deficient algorithms insufficient.

3.4.2. Second Case Study: A Numerical Simulation Model based on Fickian Diffusion applied to Experimental Data

A widely accepted numerical simulation algorithm based on Fickian diffusion has been introduced by Fieldson and Barbari [185] and will be described briefly in the following.

For one-dimensional molecular diffusion in direction x into a polymer film with a constant diffusion coefficient, the basic expression for transient Fickian diffusion is given by Fick's second law (equation 3-2 page 112):

$$\frac{\partial \mathbf{c}}{\partial \mathbf{t}} = \mathbf{D}\frac{\partial^2 \mathbf{c}}{\partial \mathbf{x}^2} \qquad (3\text{-}5)$$

where c is the concentration of the penetrant and D is the diffusion coefficient. The diffusion coefficient defines the flux for a given concentration gradient enabling quantization of the diffusion process for the investigated system conditions. Using initial conditions and boundary conditions for the case of constant surface concentration with no transport of penetrant through the lower polymer interface (i.e. x=0), Fieldson and Barbari [185] combined and integrated the concentration profile with the infrared evanescent field intensity to obtain an analytical solution for the case of a single diffusant, i.e. one-dimensional Fickian diffusion. The derived expression is as follows:

$$\frac{A(t)}{A_\infty} = 1 - \frac{8}{\pi d_p \left(1 - e^{-\frac{2L}{d_p}}\right)} \sum_{n=0}^{n=\infty} \frac{e^g \left[f e^{-\frac{2L}{d_p}} + (-1)^n \left(\frac{2}{d_p}\right) \right]}{(2n+1)\left(\frac{4}{d_p^2} + f^2\right)} \qquad (3\text{-}6)$$

where:

$$f = \frac{(2n-1)\pi}{2L} \qquad (3\text{-}7)$$

$$g = \frac{-D(2n+1)^2 \pi^2 t}{4L^2} \qquad (3\text{-}8)$$

L is the thickness of the polymer membrane, $A(t)$ the absorbance of the measured band at any time (t), A_∞ the absorbance at equilibrium, and d_p the depth of penetration of the evanescent field given by equation 2-3 (page 38). As all other parameters are known, the diffusion coefficient (D) of the penetrant in the polymer may be calculated by regressing experimental absorbance data with equation 3-6. Several groups have and are still applying this algorithm to directly calculate diffusion coefficients from enrichment data obtained with polymer coated evanescent wave sensing systems [186-190] extracting small organic molecules from aqueous solution. Based on the findings described in the previous chapter it is questionable that this algorithm is comprehensive enough to deliver “absolute values” for the diffusion coefficients, as again the effect of the flow conditions is neglected. To support this statement the algorithm was again applied to the experimental data obtained during the measurement campaign in

Bitterfeld, where data at two different flow velocities have been recorded, while other parameters remained constant (Figure 3.24 & Figure 3.25).

The numerical iteration was performed with a visual basic script (see appendix), which was incorporated into an already existing graphical interface for data evaluation developed by our research group. All experimental parameters such as penetration depth, polymer membrane thickness, and measurement time have been included following the actually performed experiments in Bitterfeld (Table 3.6).

Table 3.6 Parameters from the Bitterfeld experiments included in the Friedson and Barbari algorithm.

parameter	**symbol**	**value**
refractive index of crystal	n_1	2.41
refractive index of E/P-co	n_2	1.48
evaluated wavelength	λ (cm)	1.35E-03
angle of incidence	θ	45
polymer layer thickness	L (cm)	3.2E-04
time for iteration algorithm	t (s)	3000
number of iterations	n	100

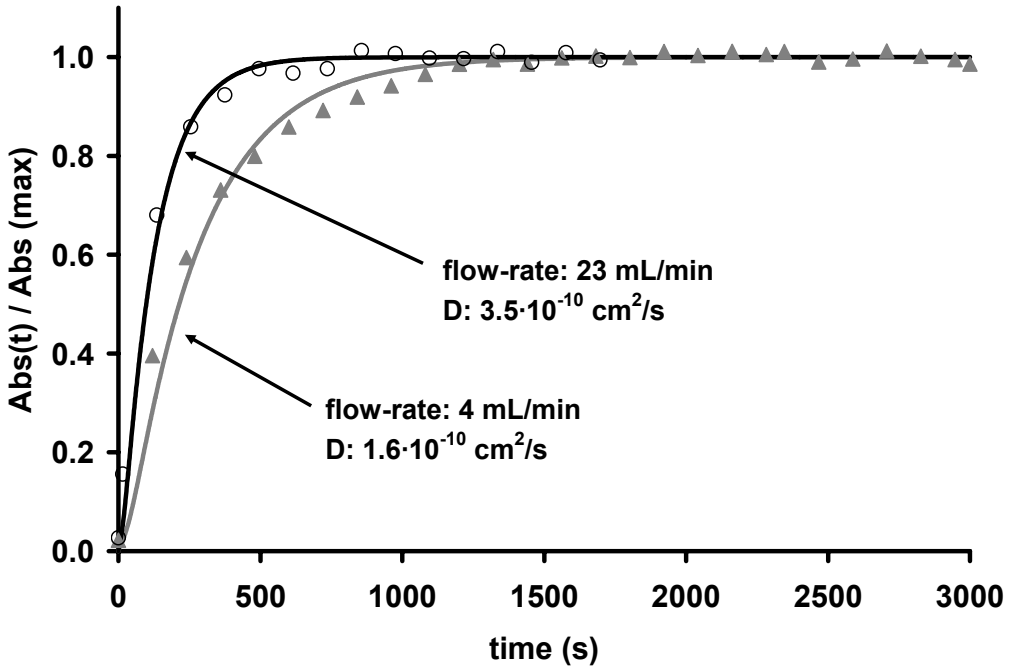


Figure 3.27 Exemplary results from the iterations for one data series obtained with low flow rate (4mL / min, Run3) and one with high flow rate (23 mL/min, Run4). The calculated diffusion coefficients are printed in the graph.

A diffusion curve is simulated utilizing the parameters listed in Table 3.6 and adapted by varying the diffusion coefficient until best fit with the experimental data is achieved.

Exemplary results from the iterations are shown in Figure 3.27 for one data series obtained with low flow rate (4mL / min, Run3) and one with high flow rate (23 mL/min, Run4). The diffusion coefficients derived from that algorithm for all 5 measurement series together with the values obtained with the "initial slope" method (chapter 3.4.1) are listed in Table 2.1.

Table 3.7 Calculated diffusion coefficients for the Bitterfeld field measurements via two different methods

	flow rate 4 mL/min			*flow rate 23 mL/min*	
Fieldson and Barbari	Run1	Run2	Run3	Run4	Run5
diffusion coefficient (cm^2/s)	1.55E-10	1.40E-10	1.60E-10	3.50E-10	3.00E-10
average diffusion coefficient (cm^2/s)	**$1.52 \cdot 10^{-10} \pm 1.0 \cdot 10^{-11}$**			**$3.25 \cdot 10^{-10} \pm 3.5 \cdot 10^{-11}$**	
Initial Slope					
slope	1.77E-05	1.73E-05	1.71E-05	2.04E-05	2.12E-05
diffusion coefficient (cm^2/s)	6.12E-11	5.86E-11	5.72E-11	8.21E-11	8.86E-11
average diffusion coefficient (cm^2/s)	**$5.90 \cdot 10^{-11} \pm 2.0 \cdot 10^{-12}$**			**$8.53 \cdot 10^{-11} \pm 4.6 \cdot 10^{-12}$**	

The results show very clearly the susceptibility of these two methods to variations of the flow conditions, as both algorithms show a significant difference in the value of the calculated diffusion coefficient of up to a factor of 2 when the flow rate is changed from 4 to 23 mL/min. Furthermore, results derived from the Fieldson and Barbari algorithm do not correlate with the "initial slope" method by at least a factor of 3. Goebel et al [180] determined *D* for the diffusion of CB in E/P-co in similar experiments but with a chromatographic box model to be $5 \cdot 10^{-9}$ cm^2/s. This is the only reference value found in literature determined with a similar method, however, its correctness is also questionable as the applied model is again not taking varying flow conditions into account.

However, algorithms such as the Fieldson and Barbari model could still be used for predictive determination of parameters such as the t_{95} value, if experimental parameters are kept constant. However, it is evident that accurate determination of the diffusion coefficient by polymer coated IR-ATR spectroscopic experiments is only possible by extensive CFD simulations.

3.4.3. Third Case Study: CFD Simulations (FLUENT) applied to Experimental Data

In this chapter findings based on the newly developed CFD model are applied to data obtained during the field measurements with the IR-ATR sensor system described in chapter 3.3.

Unfortunately, the software package was not available for simulations during this thesis. However, some interesting consequences can be derived from the initial simulations based on the model flow cell (Figure 2.13) performed by Phillips et al [43 191] in collaboration with our research group.

In this section it will be shown that the results obtained with the measurement setup during the Bitterfeld campaign are in agreement with the results modeled via CFD simulations.

Due to the different parameters used for the CFD simulations and the actual experimental parameters (see Table 3.8) all conclusions at this stage are only qualitative. Nevertheless, they clearly indicate that the predicted trends from the model are experimentally observable.

Table 3.8 Comparison of the parameters of the basic flow cell used in CFD simulations (Figure 2.13 and [43]) and the flow cell used during the Bitterfeld measurements.

	baseline flow-cell parameters	**experimental flow-cell parameters**
flow-channel length	2 cm	7 cm
flow-channel height	1mm	3 mm
polymer layer thickness	5 µm	3.2 µm
flow-speed - slow	0.15 cm/s	0.25 cm/s
flow-speed - fast	1.5 cm/s	1.4 cm/s

CFD simulations using FLUENT have been performed investigating the effect of the flow velocity on the time the sensor system needs to reach diffusion equilibrium conditions (time to steady state). All parameters except the channel height of the experimental flow cell were comparable to the basic flow cell geometry and the results for velocities of 0.15 cm/s and 1.5 cm/s are shown in Figure 3.28.

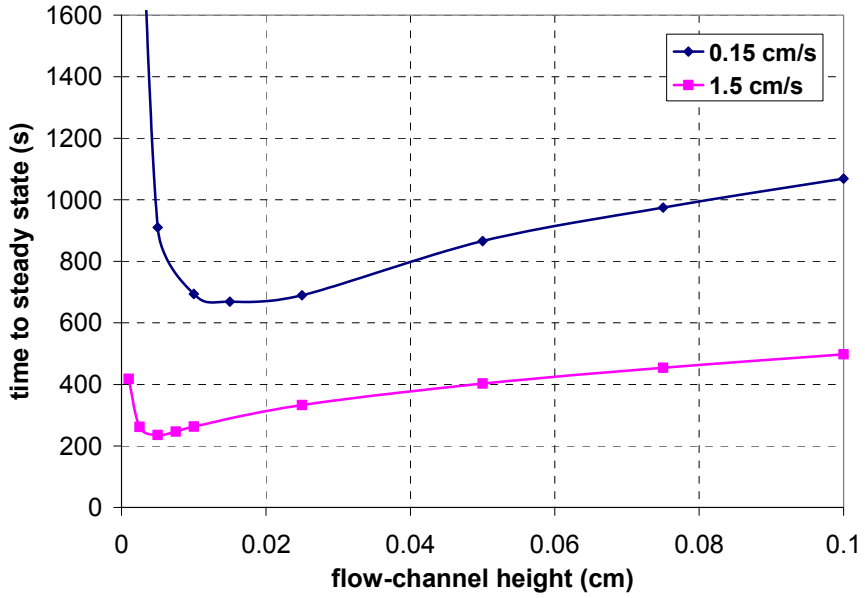


Figure 3.28 Time to steady state (equilibrium) vs. flow channel height for 2 different flow velocities modeled with CFD for the basic flow cell shown in Figure 2.13.

Figure 3.28 shows that the enrichment time is increasing with decreasing velocity. Hence, it takes much longer to saturate the polymer with analyte. The reverse is true for an increase in velocity. The change in optimal height is also due to the thickness of the concentration boundary layer that changes with the flow velocity. The relationship between the channel height and the concentration boundary layer thickness controls the maximum resistance in the flow cell. As the flow velocity increases, the thickness of the concentration boundary layer at the exit of the channel decreases. Thus, the optimal (critical) height at which the advection resistance becomes equivalent to the diffusion

resistance also decreases. Furthermore, if the channel height decreases to very small values, it can be expected that practically every analyte molecule will be partitioning into the polymer layer promoting the trend to longer equilibrium times.

However, the experimental flow cell used in the Bitterfeld campaign had a flow channel height of approx. 3 mm (in contrast to < 0.1 mm of the simulated flow cell), therefore, exceeding the range of the simulation. It can be extrapolated though that the trend of increasing equilibrium times will be continued when further increasing the channel height. This should hold true for both flow velocities, clearly indicating that in general there is a difference in equilibrium times for different flow rates. Evidently, the CFD simulations has have been performed with comparable flow velocities to the experimental data. In the simulation, a change of flow velocity by a factor of 10 changes the equilibration time approx. by a factor of 2, which is of the same order of magnitude as observed in the experimental data at two different flow velocities (see chapter 3.2.4).

In summary, CFD simulations seem to reflect real world measurement situations much more accurately than the currently applied and generally accepted models, which have been discussed and compared in this thesis. Consequently, it should be of substantial interest to implement CFD simulations as a routine tool for simulating the behavior of polymer coated IR-ATR sensors along with advanced flow cell design.

4. Conclusions and Outlook

With steadily increasing use of synthetic and natural organic compounds in industry and agriculture their impact as pollutants in air, water and soil ecosystems is generating an increasing threat to the environment. Amongst other pollutants volatile organic compounds (VOCs) represent a major threat to ground water and surface water, due to

their high abundance in industrial processes. It is a known fact that the significance of such compounds as environmental pollutants benzene, chloroform and trichloroethylene occupy a permanent place among the 20 most relevant priority pollutants in the listings all over the globe.

Contemporary analysis of VOCs in groundwater mostly relies on headspace gas chromatography (HS-GC). HS-GC is a classical off-line technique requiring substantial preparation steps prior to analysis: (i) sampling, (ii) sample transportation and storage, and (iii) depending on analyte concentration and matrix composition sample clean-up and preconcentration procedures. An increased number of processing steps before the actual analysis promotes introduction of errors and is above all time consuming. Consequently, novel analytical technologies should readily assist pollution screening and site-assessment procedures demanding for low cost, rapid and reliable sensor systems capable of quantitative online in situ measurements with high molecular selectivity under field conditions. Mid-IR based spectroscopic techniques represent a feasible and valuable approach for establishing in situ sensing systems with high inherent molecular specificity. In this thesis, the capability of polymer coated ATR-FTIR sensor systems for on site, sensing / monitoring applications in the field could be proved. Based on preliminary experiments the extractive polymer membrane consisted out of a several µm thick layer of E/P-co coated onto either a ZnSe ATR crystal or onto a silver halide fiber. First simultaneous, quantitative determination of environmentally relevant mixtures of BTX in water at trace level concentrations under laboratory conditions showed the suitability of the proposed sensor system for demanding analytical tasks. For the first time LODs for a multi-component measurements, measured with diffusion based evanescent wave sensor systems were determined to be in the low µg/L region for all members of the BTX group. An important part of this improvement in sensitivity can be related to the introduction of the *Mixmaster* an automated mixing system, specially

designed for handling diluted solutions of VOCs. This system can be applied for extensive sensor calibration, hence improving reliability and reproducibility of tedious calibration tasks, by eliminating manual sample preparation as an error source.

Two major measurement campaigns could be fulfilled in the scope of this thesis. The first one at artificial aquifer system, enabling the continuous simultaneous monitoring of three relevant pollutants TriCE, TeCE and DCB could be accomplished. A 6 m long AgX fiber, partially coated with E/P-co acting as both transducer and fiber-optic sensor head enabled the direct determination of the three compounds in a borehole of the aquifer system. Very good agreement with simultaneously performed HS-GC validation measurements showed the enormous potential of this method. However, a steadily increasing degradation process of the physically and chemically vulnerable AgX fiber could be observed throughout the measurement duration of three days. An eminent lack of highly transmitting and chemically inert fiber materials in the MIR is still a common problem, which hopefully will be resolved in the future. It might also be considered to improve AgX stability via ultra protective coatings, similar to the approach as has been shown for ATR crystals recently [192].

The second measurement campaign in the Bitterfelder groundwater at the SAFIRA remediation site showed that evanescent wave spectroscopic setups based on ZnSe ATR crystals show a much better stability, showing no signs of degradation for a period of several weeks. The sensor system, consisting of an E/P-co coated ZnSe crystal mounted into a flow-cell, was able to measure CB in highly contaminated groundwater samples in high agreement with HS-GC validation measurements. Apart from the accuracy of the system other parameters such as long term stability, dynamic sensor behavior and reversibility of the enrichment process have been successfully tested throughout these measurements, the being first ones conducted under field conditions for a polymer coated ATR-FTIR sensor system. Furthermore, it could be experimentally

shown that flow-conditions in the solution in contact with the extractive membrane have a high impact on the signal generation kinetics for the described sensor system, or diffusion based evanescent wave sensor system in general. This has been overseen by the scientific community for years and was predicted by extensive CFD simulations recently. With the calculation of the diffusion coefficient of CB in E/P-co with using different models, it could be demonstrated, that methods like the numerical algorithm introduced by Fieldson and Barbari, which are solely based on Fickian diffusion of the analyte in the polymer layer lead to incorrect results. The CFD simulations seem to reflect real world situations much better than the other models, which have been discussed. Therefore it should be of a high interest to implement such simulations as a regular tool for polymer coated ATR sensor development and evaluation, especially as the preliminary results have shown such tools maybe very helpful in flow-cell design in the future.

PART II

From the Field to the Lab - Investigating IR Signatures for Remote Sensing Applications

A Thesis

Presented to

The Academic Faculty

by

Manfred Karlowatz

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy in Chemistry

Georgia Institute of Technology

May 2004

5. From the Field to the Lab – Investigating IR Signatures for Remote Sensing Applications

6. Introduction

6.1. Landmines – A Global Problem

More than 26,000 people are killed or maimed by mines every year, which is equivalent to one victim every 20 min. For example, in Cambodia one out of every 236 people is a landmine amputee. The casualty ratio rises to one out of every 140 people in Angola, which has more mines than people. In addition to fatal casualties and enormous financial losses, mines ruin large areas of fertile farmland and waterways. In Cambodia, approximately 40% of the rice fields have been mined and abandoned [193]. Most tragically, many victims are children and most mine-afflicted countries are among the poorest countries. Worldwide landmine distribution and its clearance status are summarized in Table 6.1.

Table 6.1 Worldwide landmine distribution and clearance status

Countries	Mines (10^6) UN[a]	USSD[b]	Cleared mines	Mined area (km^2)	Cleare area (km^2)	Casualties[c]
Afghanistan	10	7	158000	550~780	202	300~360/month
Angola	15	15	10000	Unknown	2.4	120~200/month
Bosnia	3	1	49010	300	84	50 ┌ month
Cambodia	6	6	83000	3000	73.3	38786 or 100/month
Croatia	3	0.4	8000	11910	30	677
Egypt	23	22.5	11000000	3910	924	8301
Eritrea	1	1	Unknown	Unknown	2.48	2000
Iran	16	16	200000	40000	0	6000
Iraq	20	10	37000	Unknown	1.25	6715
Laos	NA	NA	251	43098	Unknown	10649
Mozambique	3	1	58000	Unknown	28	1759
Somalia	1	1	32511	Unknown	127	4500
Sudan	1	1	Unknown	800000	0	700000
Vietnam	3.5	3.5	58747	Unknown	65	180 ┌ month

[a]UN Landmine Database 1997 [193]

[b]US State Department Report ‘‘Hidden Killer 1998. The Global Landmine Crisis’’ [194].

[c]Casualty reporting varies drastically among countries; estimates provided by UN or the host government [195]

Because of the potentially catastrophic results of unintentional mine encounters, the process of detecting and removing mines (“demining”) is particularly important. Manual demining is extremely dangerous; one deminer has been killed for every 2,000 mines removed, with even more civilian victims. The cost to purchase and position a typical antipersonnel mine ranges from $3 to $30, while the cost to remove a single mine ranges from $300 to $1000. The European Commission and the United States have invested 138 million dollars for demining activities during last two years [193]. However, these cleared mines are just the tip of the iceberg. In 1994, approximately 200,000 mines were removed, while two million new mines were planted. Many experts believe that it would take more than ten centuries to remove every mine in the world with the

current clearance rate, even if no additional mines were planted. Because mines can be made of both metallic and nonmetallic materials, detection using only conventional metal detectors is not sufficient. Reports also indicate that metal detectors are subject to many false alarms in former battlefields due to the presence of small fragments of munitions. Although manual detection ("probing"), works well for a wide variety of mines, high labor cost and the slow pace involved are encouraging development of alternative techniques. Although some military demining equipment has been developed and used during the Gulf War by the US Army, civilian related demining ("humanitarian demining") is quite different from the military work. The object of humanitarian demining is to find and remove abandoned landmines without any hazard to the environment. The UN requires a probability of 99.96% mine detection accuracy to find a 4 cm radius object at a 10 cm depth, and a localization ability of up to a 0.5m radius [193]. To meet the strict requirements for humanitarian demining, various techniques in the area of sensor physics, signal processing, and robotics have been studied during the last decade. Most mine detection techniques consist of sensor, signal processing, and decision processes. For the sensor part, ground penetration radar (GPR), infrared (IR), and ultrasound (US) sensors are among the most commonly applied techniques nowadays [195] and will be briefly described in the next section.

6.2. Commonly Applied Landmine Detection Methods

6.2.1. Ground Penetration Radar

GPR consists of an active sensor, which emits electromagnetic (EM) waves through a wideband antenna and collects signals reflected from its surroundings. The principle of

GPR is almost the same as in a seismic wave measurement system except for the carrier signal. The commonly used frequency band of the GPR, EM wave is between 100 MHz and 100 GHz [196]. This band is wide enough to carry the necessary information. Reflection occurs when the emitted signal encounters a surface between two electrically different materials. The direction and intensity of the reflection depend on the roughness of the surface and electrical properties of the medium material [196]. A rough surface reflects the incident wave in a diffused manner, while a smooth surface tends to reflect the wave in one direction, where the angle between the surface normal and the reflected wave is the same to the angle between the surface normal and the incident wave. The electrical properties of the medium determine the amount of refraction and absorption of the EM waves and subsequently affect the direction and intensity of the reflection. The penetration depth of the wave into soil usually depends on two factors, the humidity in the soil and the wavelength of the EM wave [7]. The content of water in the soil significantly reduces the depth of penetration of a wave with relatively shorter wavelength. Based on the reflection and penetration properties, GPR works best with low-frequency EM waves in dry sand. Low-frequency signals, however, tend to make low-resolution maps of data, which decreases the accuracy of mine detection. Since the EM waves cannot penetrate water, GPR cannot detect underwater mines, which are common in many countries [194].

6.2.2. Ultrasound

The audio frequency range is between 20 and 20,000 Hz. Ultrasonic waves have the frequency band above this audible range. The principle of ultrasonic sensing systems is

very similar to GPR except that ultrasound uses much lower frequency waves than the GPR system.

The ultrasonic system emits ultrasound signals and collects reflected signals from the surroundings. Note that a sound wave propagates as a mechanical disturbance of molecules in the form of waves [197], while a radar signal makes no physical disturbance in the medium. When a sound wave propagates through a medium, the wave consists of the molecules of the medium oscillating around their equilibrium position.

The speed of sound is dependent on the physical properties, density, and elasticity of the medium. The speed of sound propagation, denoted by c, is given as

$$\mathbf{c} = \mathbf{f} \cdot \lambda \qquad (6\text{-}1)$$

where f represents the wavelength of the wave, and f the frequency. Sometimes c is a material constant. In a uniform homogeneous medium, the ultrasound wave propagates along a straight line and is reflected and refracted when the wave encounters a boundary between two different media. At the boundary, the speed of the wave and the density of the medium affect the behavior of propagation. In mine detection, the frequency of the ultrasound wave decides the penetration depth as is also true for GPR. The lower frequency wave tends to penetrate further than the high frequency wave [195].

Table 6.2 Speed of sound in different media [195]

material	steel	lead	water	soft tissue	bones
speed of sound (m/s)	5000	1300	1460	1500	2500–4900

The ultrasound wave propagates well in humid or underwater conditions, but it is significantly attenuated in air, while the EM wave of GPR behaves oppositely at the same conditions. Table 6.2 summarizes the speed of sound propagation in different materials.

6.2.3. Infrared Sensor

Since visualization is easier than with other sensors, the IR spectral range has widely been applied for mine detection. However, the performance of IR is highly dependent on the environment at the moment of measurement. There are two different methods for sensing IR waves. Passive IR systems sense only natural radiation from the object, while active IR systems need an extra heat source and receive radiation created by that heat source after interaction with the sample surface [198].

Dynamic Thermography

The general concept of using IR thermography for mine detection is based on the fact that mines may have different thermal properties from the surrounding material. If the response is due to an energy flux that varies with time, the objects will follow a temperature curve that will not coincide with the soil. When this contrast occurs by alteration of the heat flow due to the presence of the buried mine, it is called the v*olume effect* [195]. On the other hand, when the contrast results from the disturbed soil layer created by the burying operation, this is called the *surface effect* [195]. The surface effect is detectable only for a limited time after mine burial. During this detectable period the thermal contrast is quite distinctive. Once a sequence of images has been acquired,

various processing techniques can be applied to enhance the contrast between the potential targets and background, which is called *dynamic thermography* [199].

However, at present, the application of thermography in landmine detection mainly focuses on soil surfaces [200] and thin subsurface soil layer detection, e.g. detection of objects with depths less than 5 mm [201] or equal to 8 mm [202], which is still far from satisfying considering UN imposed requirements for buried landmine removal of at least 130 mm of depth penetration [203]. There are two main groups of factors limiting depth sensitivity of thermography: (i) variations and non-homogeneities of mine-neighboring soil [200], and (ii) the diffusive character of the thermal response of buried objects, which is intrinsically linked to heat conduction in the examined region of soil [201].

The influence of these effects is generally seen in randomly distributed changes of projections of buried object boundaries and in the suppression of thermal contrast caused by local difference in thermal characteristics, respectively. These limitations have been known for some time and have caused the focus on research and development for detection of surface laid mines and shallowly buried (flush buried) mines. Hence, detecting deeply buried landmines is a subject rarely investigated [204], which results in the absence of rather important data on buried conditions in realistic mine affected regions.

6.3. Current Developments in Remote Landmine Detection – The Disturbed Soil Approach

In order to overcome some of the eminent problems of remotely detecting buried landmines, considerable efforts were focused on detecting the changed adjacent soil conditions (“disturbed soil”) rather than detecting the buried object itself. The advantage

of that approach is obvious: in order to put a landmine into place, the soil around has to be moved (disturbed), including the soil at the surface. If the disturbed soil exhibits different spectroscopic characteristics than undisturbed (pristine) soil, the location where the mine was buried is spectroscopically "labeled". Hence, these spectroscopic changes would also be observable at the soil surface with the conditions at the surface being responsible for the majority of the emitted and/or reflected spectrum of the soil.

In 1998 Johnson et al. showed that there are significant spectral contrast differences of spectra of pristine and disturbed soils predominantly in the LWIR[†] region of the electromagnetic spectrum [5]. Hence it was proposed that hyperstrectral imaging in the atmospheric window of approx. 8-12 µm is a potentially useful method for determining possible mine spots in future. Hyperspectral imagers operating in this wavelength region have already been presented, with the most prominent example being the Airborne Hyperspectral Imager (AHI) system [2].

The problem of "disturbed soil" and its applicability for reliable detection of buried objects has drawn substantial interest in the remote sensing community. Derived from the few published works on the investigations of the disturbed soil phenomena the following findings can be summarized:

- The difference in spectral contrast is strongest immediately after the disturbing event [1].
- The strength of these effect decays over time, most likely due to weathering processes (wind, rain, erosion, etc.) [1].
- The difference in spectral contrast may be related to a difference in particle size distribution at the surface between disturbed and pristine soil [1,5].

[†] In the remote sensing community the expression LWIR (long wave infrared radiation) is the commonly used equivalent expression for the MIR (mid-IR) region used by spectroscopists.

However, for exploitation of this phenomenon in field applications there are still too many uncertainties that have not yet been investigated. Prior to elucidating the necessary experiments in order to understand the spectroscopic features of disturbed and pristine soils, brief insight into remote sensing techniques is given.

6.3.1. Remote Sensing

Remote sensing is the science of acquiring information about the Earth's surface without actually being in physical contact by sensing and recording reflected or emitted energy and processing, analyzing, and applying that information.

In remote sensing applications, this process involves interaction between incident radiation and the targets of interest. This is exemplified by the use of remote sensor systems involving the following main elements (Figure 6.1):

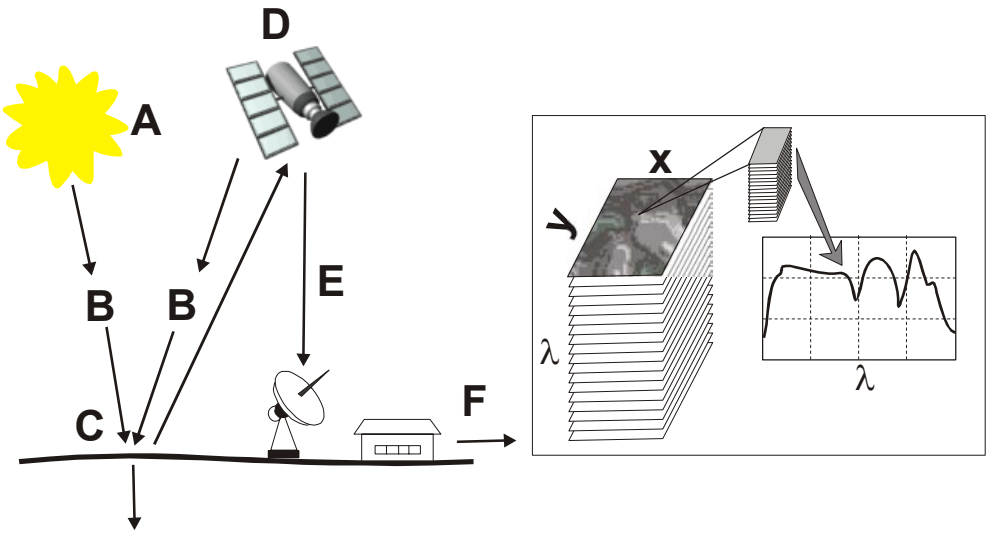


Figure 6.1 Principle of Remote Sensing Techniques

1. **Energy Source or Illumination (A)** - the first requirement for remote sensing is an energy source, which illuminates or provides electromagnetic energy to the target of interest.
2. **Radiation and the Atmosphere (B)** - as energy travels from its source to the target, it will interact with the atmosphere it passes through. This interaction may occur twice as the energy travels from the target to the sensor.
3. **Interaction with the Target (C)** - once energy makes its way to the target through the atmosphere, it interacts with the target depending on the properties of both the target and the radiation.
4. **Recording of Energy by the Sensor (D)** - after energy has been scattered by, or emitted from the target, a sensor is required (remote - not in contact with the target) to efficiently collect and record the electromagnetic radiation.
5. **Transmission, Reception, and Processing (E)** – the signal generated by energy recorded at the sensor has to be transmitted to a receiving and processing station, where the data are processed into an image (hardcopy and/or digital).
6. **Interpretation and Analysis (F)** - the processed image is visually and/or digitally/electronically interpreted to extract information about the target, which was illuminated.

6.3.2. Hyperspectral Imaging

During the last 15 years a new sensor type called imaging spectrometers has been developed [2,205,206,207]. Recently, the term hyperspectral imaging has been established for these systems, which can be considered synonymous with imaging spectrometry. Imaging spectrometry is defined as the acquisition of images in many

(hundreds or more) of contiguous, registered, spectral bands, such that each pixel has a complete attached radiance spectrum. From the radiance spectrum, an apparent reflectance spectrum can be derived by modeling and removing the absorption and scattering effects of the atmosphere. The term “apparent reflectance” is used because the surface irradiance is a function of topographic slope and aspect and cannot be directly derived from the radiance data itself. While the absolute reflectance is unknown, the relative reflectance among spectral channels is correctly derived. The shape of the spectral reflectance curves ultimately contains information on the chemical composition. Image-derived reflectances can be analyzed in the same fashion as laboratory-produced reflectance spectra. Hence, the entire range of chemometric analysis techniques is applicable to data derived from hyperspectral images. However, the process of calibration is much more cumbersome since no pixel is compositionally pure and sampling a typical pixel (nominal size: 20 x 20 m) of a surface for analysis by a primary method is easily subject to error. New techniques such as pixel unmixing using the statistics of the image data themselves are proving valuable to quantitatively derive the composition and relative abundance of individual components making up a pixel [208]. The reflectance data contain information on surface material composition, provided that sufficient spectral resolution is available. Weathered surfaces containing OH-bearing minerals such as clays, have diagnostic overtone combination absorptions that fall within the atmospheric windows in the 1–20 μm region. Hence, minerals have diagnostic spectral features that can be mapped using imaging spectrometry. Because of their molecular structure, solids and liquids do not allow rotational degrees of freedom, and no hyperfine features are observed in the reflectance spectrum. Therefore, sampling the spectrum at 10 nm intervals is sufficient to resolve the most prevalent diagnostic absorption bands found in materials covering the Earth’s surface.

Sensor Technology

In general, there are two engineering implementations of imaging spectrometers: the *whiskbroom* is an opto-mechanical system, the *pushbroom* a solid-state system. The opto-mechanical systems utilizes an oscillating or spinning mirror prior to the foreoptics to scan across the flight direction building up an image as the platform moves downtrack. The entrance slit to the spectrometer is placed in the focal plane of the telescope using line-array detectors. The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), flown aboard the NASA ER-2 at 20 km altitude, is the best-known opto-mechanical hyperspectral imaging system [209].

In the case of the *pushbroom* technique a linear array of detectors located at the focal plane of the image formed by lens systems which are "pushed" along in the flight track direction are applied. Each individual detector measures the energy for a single ground resolution cell and thus the size of the detectors determines the spatial resolution of the system. A separate linear array is required to measure each spectral band or channel. For each scan line, the energy detected by each detector of each linear array is sampled electronically and digitally recorded.

A complete introduction to the complex field of remote sensing and hyperspectral imaging can be either found in published books [210-212] or via permanently available and frequently updated web tutorials [213, 214].

6.4. Scope of this Thesis

Landmine detection via remote sensing techniques is a challenging analytical and spectroscopic task. For example, measurements of disturbed soils have shown different spectral contrast in comparison to undisturbed soils [1-5]. However, these findings are predominantly based on experimental data obtained in real world environments using

hyperspectral imaging systems, where many environmental parameters of influence can easily obscure the results. Evidently, measurements at such uncontrollable conditions do not represent the desired general conditions for principal studies.

Hence, it is of great interest to fundamentally investigate the disturbed and undisturbed soil phenomena in a controlled environment. Based on these measurements, reliable theoretical models could be established leading to improved interpretation of these features during landmine detection scenarios.

In a first step, measurements at controlled laboratory conditions have been performed to investigate individual minerals of the soil matrix and their spectral characteristics at a variety of environmental conditions. Attenuated total reflection (ATR) spectroscopy has been identified as a suitable spectroscopic technique superior to emissivity or reflectance measurements, mainly due to its reproducibility and versatility, while contributing useful data toward fundamental understanding of spectral signatures relevant to remote sensing. Due to the high abundance in natural soils, pure quartz sand (SiO_2) has been selected as the first test matrix.

For the investigation of spectral differences between pristine and disturbed quartz sand, a wetting/drying procedure with subsequent sample aerating has been developed, which in a first approximation represents a sufficient simulation of weathering processes and their impact on related soil disturbances.

In order to support the hypothesis that spectral differences of pristine and disturbed soils mainly result differences in particle size distributions of the probed surface, soda lime glass spheres of different diameters have been investigated in a next step.

It is assumed that the findings and the deductions drawn from these first measurements significantly improve the understanding of the spectroscopy of pristine and disturbed soil samples and are a starting point for extensive series of studies investigating different soil components in scalable laboratory experiments.

7. Background

7.1. Mid-IR Spectroscopy of Minerals

In the following chapter the fundamentals of the MIR spectroscopy of minerals will be discussed with main focus on the interpretation of the vibrational spectrum of quartz, which is the relevant mineral predominantly investigated in this thesis.

The spectral features of minerals in the MIR range considered here are the result of vibrational transition processes. Their number, intensity and shape are dependent on atomic masses, interatomic force fields and, particularly, molecular geometry. One goal of spectroscopic investigations is to quantitatively describe the vibrational process enabling the origin of each absorption band to be traced. Sophisticated calculations have been made consistent with observations at least for some minerals [215], although not necessarily claiming correctness. Even if a vibrational mode is precisely understood, it is virtually impossible to describe such a motion simply and concisely for such complex structures as silicates. Consequently, one must rely on some very general descriptions, such as "Si-O symmetric stretch," to denominate the vibrations predominantly involving the symmetric expansion and contraction of the silicon-oxygen bonds.

Using such simplified visualizations, we can successfully generalize the fundamental aspects of the spectral behavior of minerals.

For example, atoms with low mass vibrate at higher frequencies (shorter wavelengths) than heavier atoms when substituted into the same crystalline structure. However, higher bond strengths also result in higher frequencies of vibrational transitions. This change in bonding within silicates is related to the degree of polymerization of the SiO_4

ion [216]. This results in a systematic change in wavelengths of the fundamental vibration bands of silicates as the framework structure ultimately is based on isolated tetrahedra. Finally, bond-stretching vibrations in covalent structures are located at higher frequencies than bending modes. Such internal molecular vibrations are typically found at higher frequencies than lattice modes [217].

In summary, the most prominent features in infrared spectra of minerals can be understood in the context of the generalized descriptions of the main vibrational features as outlined above and are described below for different types of minerals. The attribution of more complex vibrational features resulting from overtones and combination bands of the internal vibrations and lattice modes is more speculative in nature, even for the simplest minerals.

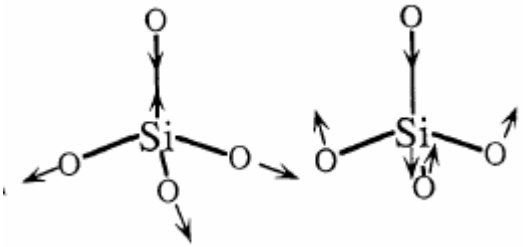


Figure 7.1 Infrared active internal vibrations of quartz. Left: ν_3-asymmetric stretch Right: ν_4-asymmetric bend [218].

For many minerals, the types of vibrational modes may be divided into two main categories: *internal modes* and *lattice modes*. Internal modes are vibrations which can be associated with those of a molecular unit, shifted (and possibly split) by interaction with the crystalline environment into which the molecular unit is bonded: the vibrations of the silica tetrahedron shown in Figure 7.1 are typical examples of these types of motions, which give rise to internal modes. Such internal modes are typically associated

with the most strongly bonded units in a crystal, and thus with the highest frequency vibrations of a given material. Here, it should be mentioned again that even the simple picture of molecular vibrations is often complicated by the presence of interacting molecular units within a crystal. For example, it is difficult to associate different bands in feldspars with stretching vibrations of distinct AlO_4 or SiO_4 tetrahedron due to interlinking tetrahedra. A silica symmetric stretching vibration such as is shown in Figure 7.1 will involve a stretching motion of the adjoining AlO_4 tetrahedron, and vibrations of these two species must be viewed as coupled within such structures.

Lattice modes comprise both a range of (often comparatively low frequency) vibrations not readily describable in terms of molecular units, and so-called external modes. External modes are those involving motions of a molecular unit against its surrounding lattice: for example, displacement of a SiO_4 tetrahedron against the surrounding lattice.

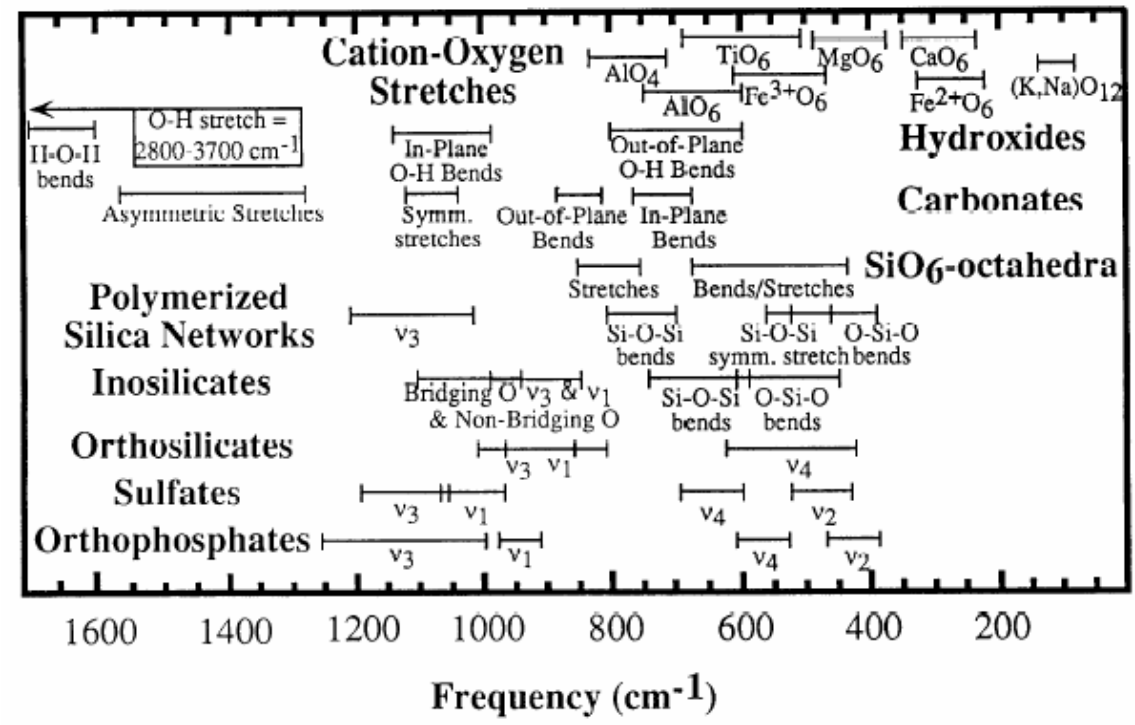


Figure 7.2 Approximate frequency range of common internal vibrations of silicates, oxides and other functional groups within minerals [218].

The most intense spectral features of quartz, occurring between 830 and 1250 cm^{-1} (8 to 12 μm), are generally simplified as fundamental asymmetric Si-O-Si stretching vibrations. The appearance of these features typically changes in reflectance measurements. For example, the weak side band near 1180 cm^{-1} (8.5 μm) in the transmittance spectrum of quartz appears as well-defined lobe of a prominent reflectance doublet between 1000 and 1250 cm^{-1} (8 and 10 μm). The reflectance spectrum of quartz glass displays a much weaker short-wavelength lobe.

Approximate frequency ranges of common internal vibrations of silicates, oxides and other functional groups within minerals are summarized in Figure 7.2 [218]. These most intense features are located within the atmospheric window (700 to 1250 cm^{-1}; 8-14 μm), rendering them most useful for remote sensing of silicates [216].

The second most intense silicate bands are broadly characterized as O-Si-O deformation or bending modes, which occur in the region of 400 to 560 cm^{-1} (18-25 μm). Weaker bands in quartz spectra between 670 and 830 cm^{-1} (12 - 15 μm) have been attributed to symmetric Si-O-Si stretching vibrations [217]. When some of the silicon atoms are replaced by aluminum, as in feldspar minerals, additional Si-O-Al stretching vibrations are added. For example, albite displays eight characteristic bands in its spectrum between 500 to 830 cm^{-1} (12-20 μm). Again, such bands are greatly simplified or eliminated in the spectra of glasses [219]. Additional weak bands are displayed as troughs between 400 to 1450 cm^{-1} (3-7 μm). Such bands in silicate spectra have largely been ignored because they are usually too weak to be observed in transmittance spectra. However, they can be very useful in spectral identification of fine particulate minerals and rocks, where these features appear quite prominent [220]. Since these features have not been assigned with any certainty, such bands are usually referred to as overtone/combination tone bands of internal and lattice modes.

It is well known that glass spectra of mineral compounds show differences in the spectral appearances, which have been attributed generally to broadening of the bands [217]. However, broadening does not sufficiently explain the reduced intensity of the 1180 cm^{-1} (8.5 μm) band in the spectrum of glass compared to that of crystalline quartz. An alternative explanation is that the short-wavelength lobe of the quartz reflectance doublet is not entirely resulting from internal molecular vibrations, but depends to some extent on the long-range order in the crystal [221]. A brief description of the progress in the understanding of the vibrational spectra of glasses is given in the next section.

7.2. Mid-IR Spectra of Glasses

The spectroscopy of glasses has been somewhat neglected in solid state physics over the last century, as compared to the amount and quality of work performed on crystalline structures. The early frontiers of solid state spectroscopy have targeted understanding of periodic materials, for which progressively more refined experiments were performed and supported by elegant theories. Resulting, the structural, mechanical, thermal, electronic, optical, and magnetic properties of perfect crystals are nowadays known in considerable detail. To the extent that defects could be treated as perturbations of otherwise periodic structures, experiments and related theories met with considerable success. This also applies to the vibrational properties of defective and mixed crystals, elaborated by Lifshitz and Maradudin, among others [222].

This situation radically changes considering glasses, owing to the lack of periodicity, which initially led to substantial difficulties in experiments, analytical theories and simulations [223]. This explains why the field of structural glasses remained relatively unexplained for many decades. However, due to significant efforts in this field of research this situation is improving steadily and for a big part the progress in glass analytics is due to optical investigations.

The systematic observation of optical spectra resulting from glasses started in the 1950s using both Raman scattering (RS) and IR spectroscopic techniques [83].

Unfortunately, long-range electric forces, the so-called Coulomb forces have been neglected for the description of the optical behavior of glasses until the late 1970s. The importance of electric forces was first recognized by Galeener and Lucovsky, who reported transverse optical (TO) and longitudinal (LO) TO-LO mode splittings in silica

(and ν-GeO_2) and thereby contributed considerably to the clarification of their optical spectra [224,225].

It was then realized from IR and RS spectra of glasses that these were “distressingly similar to a smeared-out version of the corresponding crystal”, as described by e.g. Gaskell et al. [226,227]. In comparison to the “parent crystal”, quite generally one anticipates three types of contributions in the optical spectra of the corresponding glasses:

a. There must be bands directly related to the bands of the crystals (e.g. to the bands of quartz in the case of silica), which are active in the corresponding glass. These will occur blurred due to the disordered structure.
b. Forbidden bands of the crystals may appear active in the corresponding glass, as the selection rules are likely to be relaxed by disorder.
c. Defects that are absent from the parent crystals may occur in glasses, such as e.g. small rings of repeating Si-O units in the case of silica. If such defects are optically active and sufficiently prevalent, they may contribute to the optical spectrum.

7.3. Mode Splitting in MIR-Spectra of Crystals and Glasses

In recent years, especially the interest in semiconductors and accordingly the demand for improved analysis of thin amorphous SiO_2 (a-SiO_2) layers, promoted a higher output of scientific literature in the field of optical glass analysis.

Thorough investigations on both the structure and the defects present within thin films formed on silicon is of crucial importance in microelectronics. IR spectroscopy is a

powerful tool for such investigations providing information on the structure, thickness, density, carrier concentration and several other important properties of thin films. Unfortunately, some aspects of these methods and the resulting spectroscopic data are still not entirely understood. In particular and despite an enormous amount of work dedicated to analyzing the structure of a-SiO_2 using IR spectroscopy, the interpretation of some spectral features observed at normal and oblique incidence of light, especially in the spectral region of 1000 – 1300 cm^{-1}, is still a matter of substantial debate [228- 254]. Although these spectral features are generally accepted to originate mostly from asymmetric stretching vibrations of Si-O-Si bridging sequences [255], different interpretations are still prevalent in the relevant literature. Galeener [225] attributed the partially resolved pair of peaks in the reflectivity spectrum of bulk a-SiO_2 to an asymmetric stretching vibration of the bridging oxygens parallel to the Si-Si direction plus some Si-cation motion, which could be resolved via Kramers-Kronig analysis of transverse optical (TO) and longitudinal (LO) components arising from long-range Coulomb coupling. The TO component involves nonzero derivatives at the equilibrium inter-nuclear configuration for dipole moment components perpendicular to the propagation vector of the phonon waves, whereas the higher-frequency LO component involves dipole moment changes parallel to the propagation vector.

In the case of thermal SiO_2 films, Boyd [231] suggested the occurrence of some shorter bonds within each SiO_4 tetrahedron in order to explain a slightly asymmetric peak near 1080 cm^{-1}. This feature is typical for IR transmission spectra of thin films, which did not exhibit the high-frequency shoulder of thicker films at approx. 1200 cm^{-1}. In contrast, Huebner et al. [245] used IR transmission spectroscopy at oblique incidence (55° off-normal) to simultaneously detect the TO and LO components, at 1091 and 1260 cm^{-1}, respectively, for 500 nm thick thermal SiO_2 films. Olsen and Shimura [233] used multiple internal reflectance at 60° incidence with linearly polarized light and they were able to

detectthe TO and LO components at 1080 and 1240 cm^{-1} in 3 nm thick SiO_2 films using parallel polarized (p) light. In perpendicular polarized (s) light essentially only the TO mode was detected.

In crystals, these mode splittings arise as a result of long-range Coulomb forces, which are a consequence of the internal electric field created by the motions of the ions during the vibrations. In glasses, there is also some theoretical support for the occurrence of mode splitting effects [239], in particular discussed in the work of de Leeuw and Thorpe [229]. In their work they calculated the optical response of a computer-generated random network with 1536 ions. By introducing long–range Coulomb forces in an exact way LO-TO split vibrational mode frequencies were obtained. In contrast, Phillips [228] has suggested that LO-TO splittings imply a macroscopic polarization effect accompanying the vibrational modes, which is not possible in the continuous random network model of glass structures [243]. Consequently, the vibrational spectra of a-SiO_2 were associated with a para-crystalline model, including a large density of Si=O bonds on internal surfaces, where LO-TO pairs would physically be possible.

Because of the transverse character of electromagnetic radiation, in conventional transmission spectroscopy at normal incidence only TO modes can be detected. It was shown by Berreman [232] that transmission spectra of crystal films at 30° off normal incidence enable the detection of LO modes, which has also been observed for thermal a-SiO_2 thin films, in both transmission [245] and ATR measurements [233]. Berreman's argument substantiates in showing that for crystalline thin films zone-center (long wavelength) phonons have a wave vector perpendicular to the film surface, such that normal incidence radiation can only interact with TO vibrations (parallel to the surface). In contrast, the p-polarized component of oblique incident radiation has sub-components parallel and perpendicular to the film surface, which can excite both TO and LO phonons, respectively. Almeida et al. [246] suggested that this argument can be

extended to the case of bulk samples and the high-frequency LO mode of bulk a-SiO_2, which was detected by diffuse-reflectance FT-IR spectroscopy during variation of the incidence angle between 20° and 70°.

Kirk [235] quantitatively analyzed the IR-absorption spectrum of a-SiO_2 in terms of its TO-LO vibrational modes. Disorder induced mechanical coupling between the asymmetric O-Si-O stretch (AS_1) mode (in-phase motion of adjacent O atoms) and the relatively optically inactive O-Si-O asymmetric stretch (AS_2) mode (out-of-phase motion of adjacent O atoms) was introduced into the oscillator model. Coupled AS_1- and AS_2-mode LO-TO frequency pairs were experimentally observed as peaks at 1076-1256 and 1160-1200 cm^{-1}, respectively, in oblique incidence p-polarized absorption spectra of thin thermally grown a-SiO_2 films. Two other LO-TO mode pairs were observed in these spectra as absorption peaks at approx. 810-820 and 457-507 cm^{-1}. The simplest form of the coupled-mode model consistent with experimental data is one in which the AS_1-mode LO-TO frequency splitting is resulting from the AS_1 transverse effective charge and the AS_2-mode LO-TO splitting relates to the mechanical coupling between these two modes and not to the AS_2 transverse effective charge, which is negligibly small.

However, this assignment has been questioned by Almeida [243] based on his own reflectance measurements and theoretical calculations of transmittance and reflectance spectra performed by Phillips [238] for a-SiO_2 films at different angles of incidence. Later, Gole et al. [242] and Shaganov et al [252] suggested a reassignment of a section of the 1176 cm^{-1} band to the Si=O stretching mode of silanone-based oxyhydrides, based on their quantum chemical calculations. This type of vibration is observed on oxidized porous silicon structures, which have crystalline Si in their core and SiO_x (x=1,2) at their surface.

Discussions about the exact assignments of the spectral features of amorphous and crystalline SiO_2 structures extended into the present literature [247,253,254] and will not

be further discussed within the scope of this thesis, as the most important aspects have been summarized in the section above.

Quantitative band assignments and calculations for glasses naturally become even more complicated when the material is composed of more than one basic unit. One of the wide varieties of mixed glass compositions, soda lime glass spheres have been investigated in this thesis (chapter 9.2), since (a) their main component (~70%) is SiO_2, thus findings presumably can be related to the quartz results and (b) such glass spheres are available in mono-disperse samples in a wide size regime (nm to mm).

Current approaches to the problem of connecting the vibrational spectra with the structures of glasses are still restricted to the qualitative evaluation of experimentally recorded IR or Raman spectra. and subsequently applying semi-quantitative calculations to fit the empirical data [256, 257].

Quantitative analyses usually involve, as the necessary first stage, data treatment intending to converting the recorded spectra into “reduced” spectra, considering only main spectral features [258]. Moreover, the vibrational spectra of inorganic glasses are multiband spectra with overlapping bands virtually always present. Therefore, spectral data evaluation for quantitative data treatment should always include the deconvolution of the multiband spectrum into individual bands. Interpretation of the obtained individual bands remains a challenging task due to a multitude of effects of vast variety of vibrational active structures such as (i) particular types of polyhedra, (ii) R-X-R bridges (R being Si, P, etc. and X being O, F, S, etc.) and $(RX_m)^{n-}$ terminal groups, and (iii) superstructural units (various SiO_4 rings for example), amongst others [237,250,257,258,259]. The problem of reasonable selection of fragments to be separated is related to the problem of changes in a spectrum resulting from an isolated molecule-like monomer vs. a polymeric structure. Atomic displacements in a structural group entering into a polymeric crystal lattice or glass network necessarily cause atomic

displacements in neighboring groups. As a result, the vibrational selection of a polymeric lattice or network significantly differs from those of the normal modes of an isolated group. Due to the lack of translational symmetry in a glass network, it is even more difficult to specify the degree of interaction of vibrations within a network and the actual "dimensions" of a region, which gives rise to a certain band. As a result, current approaches to the formation of vibrational spectra resulting from glasses provide a variety of different answers to these questions.

In summary, a general scheme for quantitative band assignment and a scheme related in more detail to IR reflection spectra are shown in Figure 7.3 and Figure 7.4, respectively [257,259].

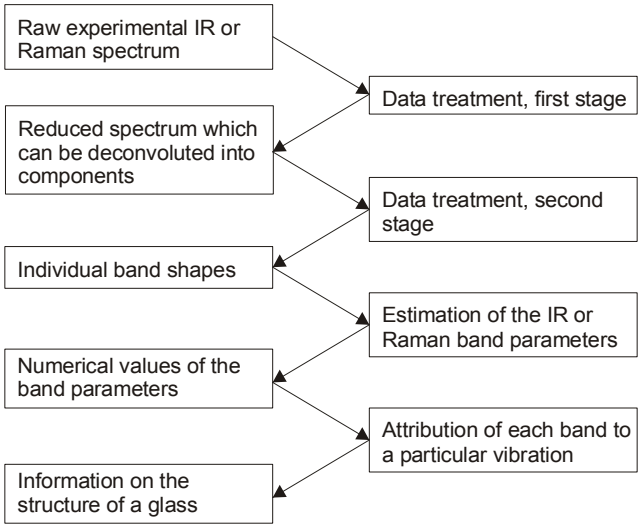


Figure 7.3 General scheme for quantitative IR band assignment for glasses. Reproduced from [257].

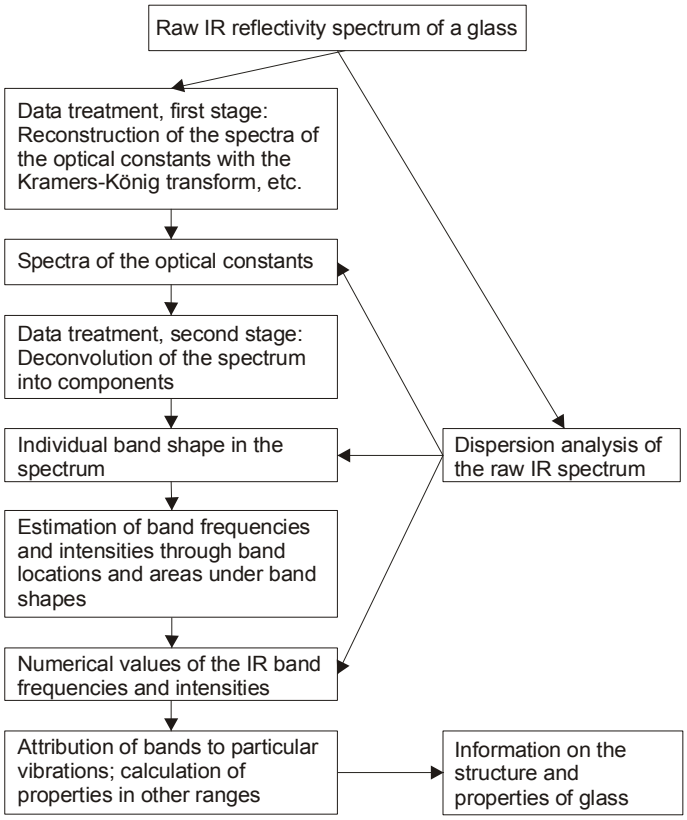


Figure 7.4 Scheme for quantitative IR band assignment for glasses starting from IR reflectance measurements. Reproduced from [257].

7.3.1. Impact on this Thesis

The absence of a comprehensive model providing sufficient explanation of the IR spectra of quartz and amorphous silica solids does not represent a significant drawback the this first study on the exploitation of IR-ATR data for improvements in understanding remote signatures of such materials. Effects such as the disturbed/undisturbed soil problem can still be investigated and obtained results will be traced back to theories such as the Berreman effect and LO-TO splitting. As the main consensus of the works referenced in this thesis [232-272] it can be concluded that in both cases, for quartz and a-SiO_2, the band positions and intensities depend at a basic level on a set of main

parameters, which are defined for the experimental conditions in the presented studies as follows:

- Polarization state of the incident light: measurements will be performed at unpolarized, p-polarized, and s-polarized illumination conditions.
- Angle of incidence: in the presented ATR setup the sample will be interrogated with a broad angular distribution of incident light (~45° +/- 15°). Thus, “bulk” responses of the sample will be obtained.
- Film thickness and/or particle size: natural quartz samples with non-uniform size distribution will be investigated in order to simulate field conditions. Furthermore, mono-disperse soda lime glass spheres will be used as samples to prove the influence of the particle size on the resulting ATR spectra. By using spherical particles signal dependencies on particle shape are effectively eliminated.

The experimental section is divided into experiments with natural quartz samples (chapter 9.1) assessing the general applicability of this measurement approach, and experiments with mono-disperse soda lime glass spheres (chapter 9.2) for a deeper insight into the signal generation with particles of controlled geometry and dimensions.

8. Experimental

8.1. Samples

Natural quartz samples were obtained from Ward’s Natural Science (Rochester, NY). 1 – 3 µm and 4 – 10 µm soda lime glass spheres were purchased from the MO-SCI Corporation (Rolla, MO), all larger spheres were obtained from Whitehouse Scientific Ltd. (Waverton, Chester, UK).

Methanol for cleaning ATR crystals has been purchased from Aldrich (Milwaukee, WI) and was of analytical grade.

8.2. Laboratory Setup

Instrumentation

Data was recorded in a spectral range of 4000 cm^{-1} to 400 cm^{-1} with a Bruker Equinox 55 Fourier transform infrared (FT-IR) spectrometer (Bruker Optics Inc., Billerica, MA) equipped with a liquid N_2 cooled mercury-cadmium-telluride (MCT) detector (FTIR-22-1.0, Infrared Associates, Stuart, FL). A total of 100 scans were averaged for each spectrum with a spectral resolution of 1 cm^{-1}. A complete list of the measurement parameters is given in Table 8.1

Table 8.1 Measurement Parameters for ATR studies

Zero Filling Factor	2	*Instrument Type*	EQUINOX55
Stored Phase Mode	No	*Number of Background Scans*	100
Start Frequency Limit for File	6000 cm^{-1}	*Acquisition Mode*	Double Sided
End Frequency Limit for File	400 cm^{-1}	*Correlation Test Mode*	No
Phase Resolution	8	*Delay Before Measurement*	0
Phase Correction Mode	Mertz	*Stabilization Delay*	0
	Blackman-Harris		
Apodization Function	3-Term	*Wanted High Frequency Limit*	7800 cm^{-1}
High Folding Limit	7900.32 cm^{-1}	*Wanted Low Frequency Limit*	400 cm^{-1}
Low Folding Limit	0 cm^{-1}	*Sample Scans*	100
Sample Spacing Divisor	2	*Resolution*	1 cm^{-1}
Actual Signal Gain	1	*Beamsplitter Setting*	KBr
Switch Gain Position	14070	*Iris Aperture*	2300 µm
Gain Switch Window	300	*Low Pass Filter*	Open
		Scanner Velocity	11 ; 100.0 KHz

A horizontal ATR accessory (Specac, Smyrna, GA) equipped with trapezoidal ZnSe ATR elements (72*10*6 mm, 45º; Macrooptica Ltd., Moscow, Russia) was used. A holographic thallium bromoiodide (KRS-5) polarizer (period: 0.25 µm, Specac, Smyrna, GA), which was mounted in a motorized polarizer rotation unit (#A121, Bruker Optics Inc, Billerica, MA), was applied for measurements at linear polarized light conditions. A schematic of the experimental setup is shown in Figure 8.1

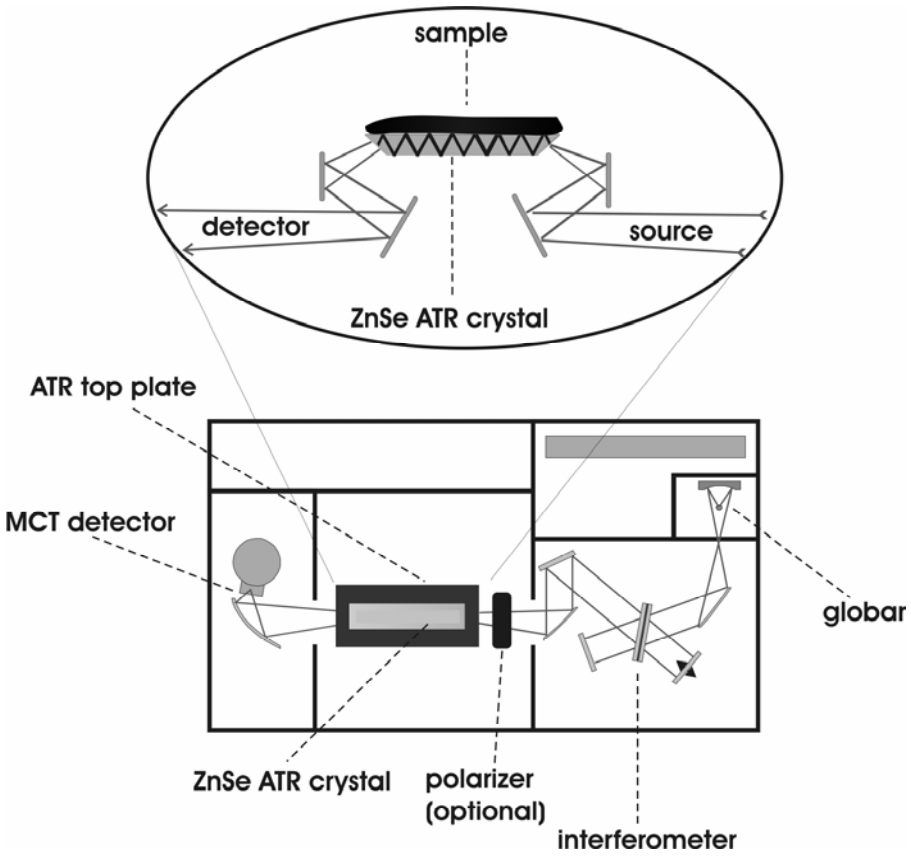


Figure 8.1 Experimental setup for cyclic wetting/drying studies of quartz sand via ATR spectroscopy in the MIR regime.

9. Results

9.1. ATR Spectra of Polydisperse Natural Quartz

9.1.1. Experimental

The ZnSe crystals have been thoroughly cleaned with methanol prior to measurements and reference spectra of bare crystals at unpolarized, p- and s-polarized illumination conditions have been recorded. Approx. 2 – 3 g of quartz sand were applied onto the crystal ensuring complete coverage of the crystal surface with a layer thickness of several millimetres, definitely exceeding the penetration depth of the evanescent field within the investigated spectral range. Following spectral measurements of the “pristine” quartz spectrum, the weathering process was simulated by addition of few droplets of deionized water to form a slurry. Within a timeframe of few hours the majority of the aqueous phase is evaporated, evident by decreasing water absorption bands (e.g. at 1650 cm^{-1}), which were continuously monitored. In this study, spectra of the quartz sample after the wetting/drying cycle will be referred to as “dried” spectra. Finally, a disturbance event was introduced by stirring up the dried quartz sand sample using a plastic spatula. Consequently, spectra recorded after the disturbance event are referred to as “disturbed” spectra. This cyclic procedure has been investigated at unpolarized, p-polarized and s-polarized illumination conditions and related to the corresponding reference spectra. The resulting evanescently recorded absorption spectra have been compared and analyzed at the conditions schematically summarized in Figure 9.1.

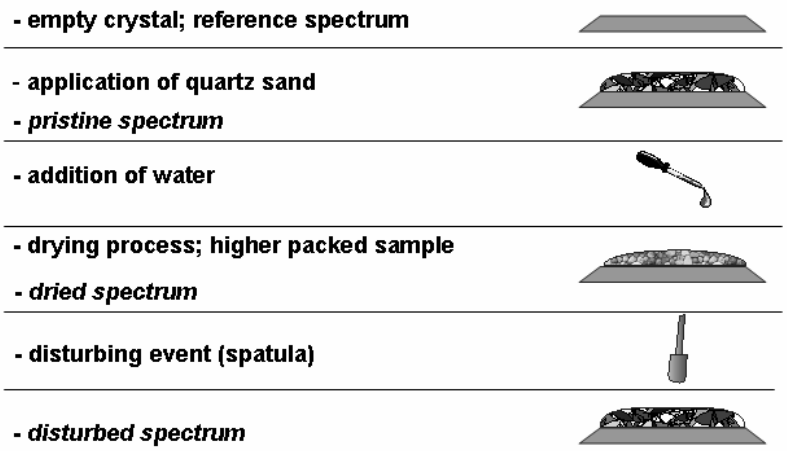


Figure 9.1 Overview of experimental procedure.

9.1.2. Wetting / Drying

As can be seen in Figure 9.2, IR-ATR spectroscopy is a highly suitable, yet comparatively simple method providing infrared spectra of quartz sand. This method allows investigation of a wide variety of samples including other minerals, clays or soil samples at constant and highly reproducible measurement conditions (data not shown). In the IR spectrum of pristine quartz, the broad absorption feature with a maximum around 1090 cm^{-1} is attributed to asymmetric SiO_4 stretching vibrations. The less intense double peak located at around 800 cm^{-1} is due to symmetric stretching and the peak at 690 cm^{-1} is related to Si-O-Si bending transitions.

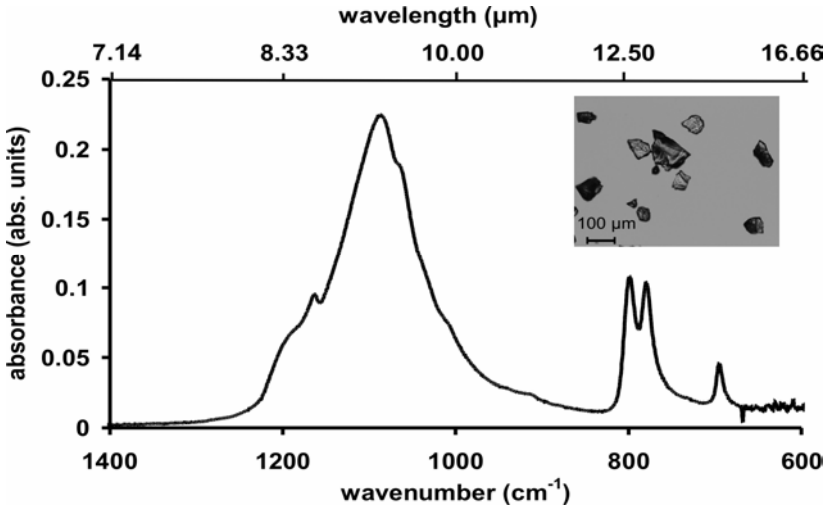


Figure 9.2 Exemplary IR-ATR spectrum of pristine quartz sand (Fluka 83340). The broad absorption feature with a maximum at around 1090 cm^{-1} is attributed to asymmetric stretching vibrations, the double peak at around 800 cm^{-1} relates to a symmetric stretching of the SiO_4 unit cell and the peak at 690 cm^{-1} is related to Si-O-Si bending vibrations. The inset shows an optical microscopy image of the sample.

In Figure 9.3, the comparison between a pristine, dried and disturbed spectrum of quartz sand following the wetting/drying cycle described in the experimental section shows some initially surprising differences. The dried spectrum shows significantly higher absorption features throughout the entire investigated spectral range. This circumstance can be explained by a much higher compactness of the quartz particles resulting from the submersion in water. While the initial (pristine) state of the quartz sand typically shows a high void volume in between the particles mostly due to friction and static forces, the addition of water promotes filling of the interstitial spaces by compacting of the sample leading to an increased density of the sample packed onto the ATR crystal surface. Therefore, more sample material is present within the evanescent field leading to higher intensity of the absorption features. After disturbing the sample by stirring the compacted quartz sand with a spatula, the absorption intensities return to near their initial values.

Most importantly, these findings are in agreement with field and laboratory remote sensing studies, where changes in spectral contrast have been reported as the predominant difference between spectra of pristine and disturbed soils [1,2,5].

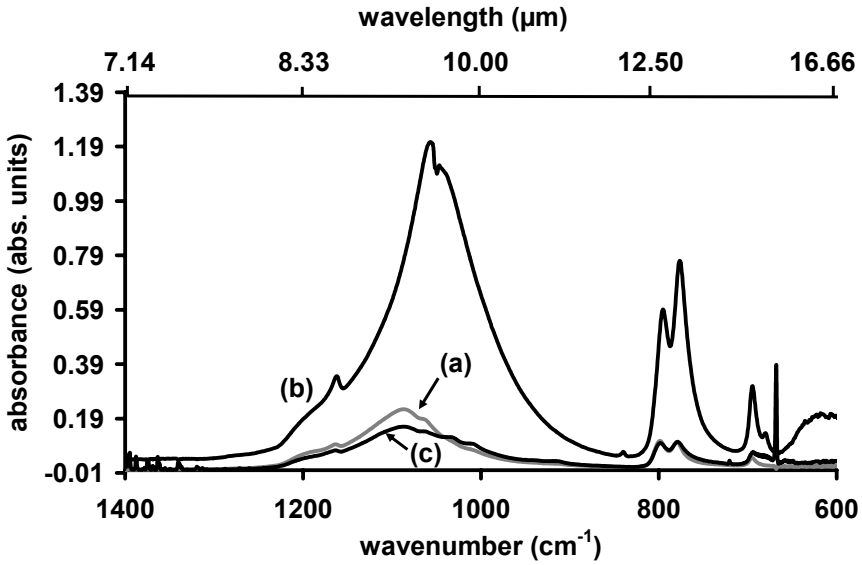


Figure 9.3 Pristine (a), dried (b) and disturbed (c) spectra of quartz sand. The sharp band around 670 cm^{-1} results from atmospheric CO_2 present after opening the sample compartment.

Another noticeable difference is the change of shape and shift of spectral position of the maximum associated with the main absorption feature. We observe a reversible shift of the peak maximum from 1090 cm^{-1} (pristine sample) to 1060 cm^{-1} (dried sample), and back to 1090 cm^{-1} (disturbed sample). This phenomenon was reproducibly observed when the same sample was cycled several times in the order wetting/drying/disturbing (data not shown). This apparently significant and pronounced spectral shift may potentially be a characteristic spectral feature useful to remote detection of disturbed soil sites. In depth investigations of this effect led to the following hypothesis.

The addition of water promotes ultra-fine particles (<10µm), which initially adhere to larger particles, into a suspension state facilitating mobility within interstitial spaces.

Potentially driven by capillary forces, these particles accumulate at or close to the surface of the ZnSe crystal during the drying process. Evidence is derived from removing the majority of the quartz sample layer from the ATR crystal after complete water evaporation and still detecting a layer of ultrafine particles adhering to the surface of the crystal. To further prove this assumption some grams of quartz sample have been suspended in acetone in a vial which was closed severely shaken and a sample has been drawn from the immediately formed sediment, representing the most coarse fraction of the investigated multi-disperse quartz sample (for a comparison of the finest and coarsest fractions of this sample see Figure 9.4).

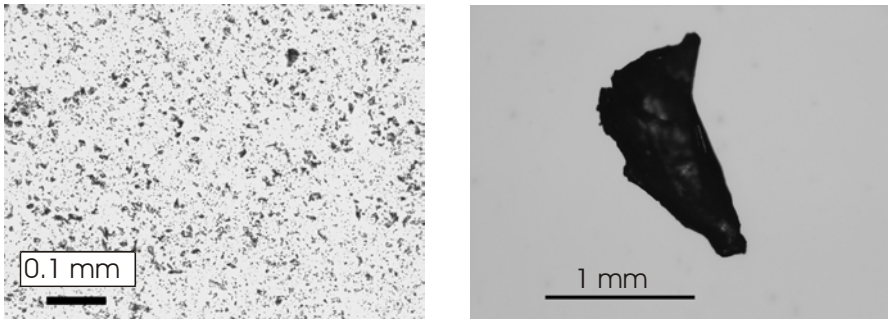


Figure 9.4 Microscopy pictures of the finest fraction (left) and the coarsest section (right) of the investigated quartz sample.

This coarse fraction when measured alone, produced only very weak absorption features, even after a wetting and drying step. In Figure 9.5 the dried spectra of the initial quartz sample, the coarse fraction and the fine fraction, which was obtained as described above are shown.

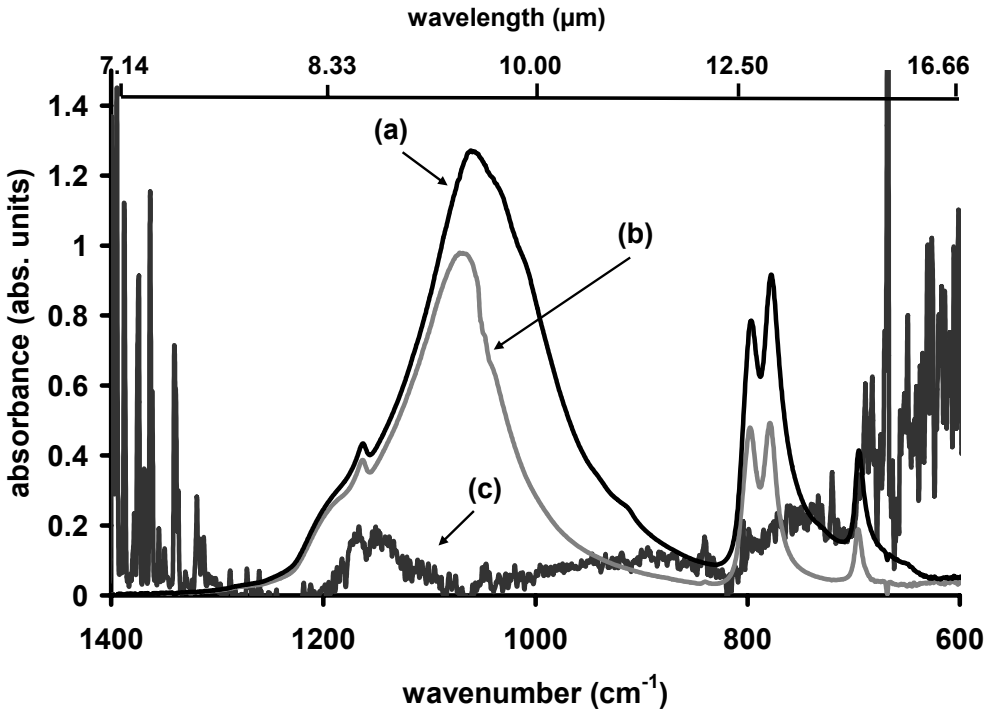


Figure 9.5 Dried Spectra of a mixed (a), fine (b) and coarse (c) fraction of quartz sand. The spectra of the coarse fraction were scaled, due to very weak absorption features.

It can be clearly seen that the spectrum of the mixed (unaltered) quartz resembles the spectrum from the fine fraction to a high extent. The difference in intensities can be related to the fact that during the removing of the major part of the quartz, the ATR element was not completely covered with sample. Consequently, the particle size distribution of the sample is changing throughout the simulated weathering process facilitating migration of ultrafine particles into interstitial spaces of larger grains detected by an increased abundance of material within the analytical volume probed by the evanescent field resulting in dramatically increasing absorption intensities recorded in the spectrum of the dried sample. A schematic of these findings is shown in Figure 9.6.

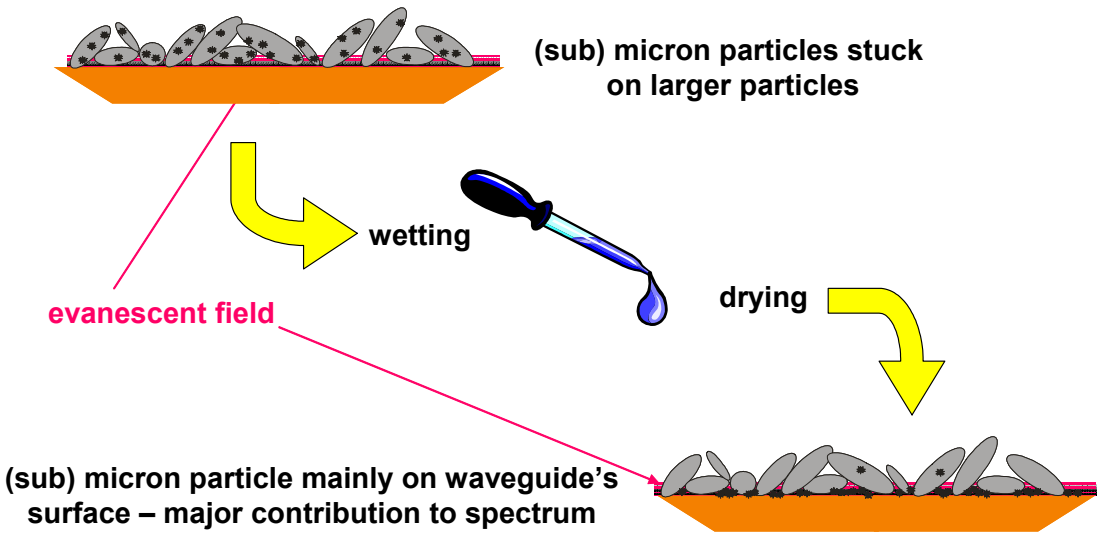


Figure 9.6 Schematic of agglomeration process of ultrafine quartz particles on the crystal surface during the wetting / drying process.

Additionally it is obvious that the peak maximum of the major absorption feature around 1100 cm^{-1} (asymmetric stretch vibrations) is clearly shifted from the coarse to the fine fraction, showing an obvious particle size dependency.

These findings appear plausible and in analogy to previous reports hypothesizing that changes in particle size distribution when investigating undisturbed vs. disturbed soils are a major contribution to the detected differences in the respective spectra [1,5,260].

Judging from these preliminary results the logical next step is to perform studies with polarized light, followed by investigations of mono-disperse samples. Soda lime glass spheres are commercially available in a wide range of particle sizes and therefore have been selected for further investigations (chapter 9.2).

9.1.3. Polarized Light

The influence of using s- and p-polarized radiation to probe quartz sand reveals further interesting spectral aspects of the sample. ATR spectra of a pristine (Figure 9.7) and

dried (Figure 9.8) quartz sand samples have been recorded with unpolarized (a), p-polarized (b) and s-polarized (c) infrared radiation.

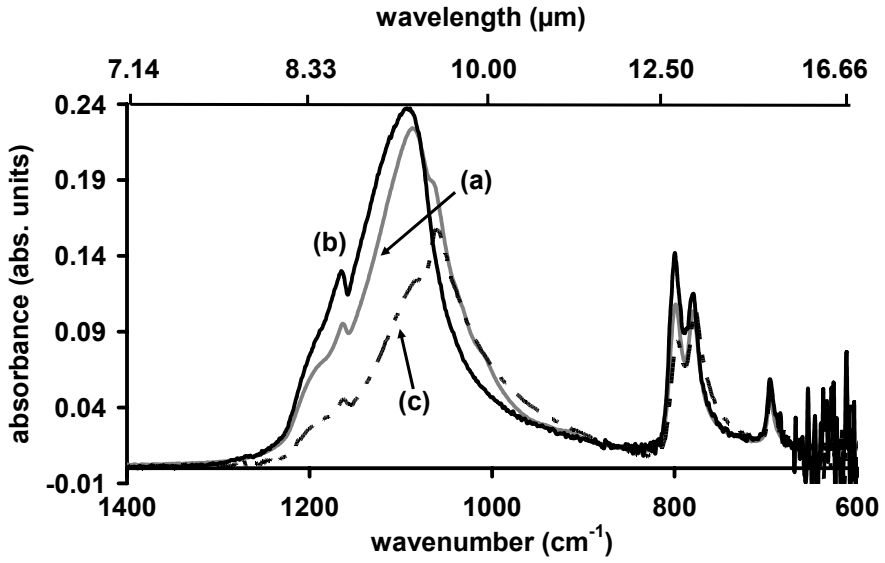


Figure 9.7 ATR spectra of pristine quartz sand samples recorded at different polarization states of infrared radiation: (a) unpolarized light (grey line), (b) p-polarized light (black line), (c) s-polarized light (dotted line).

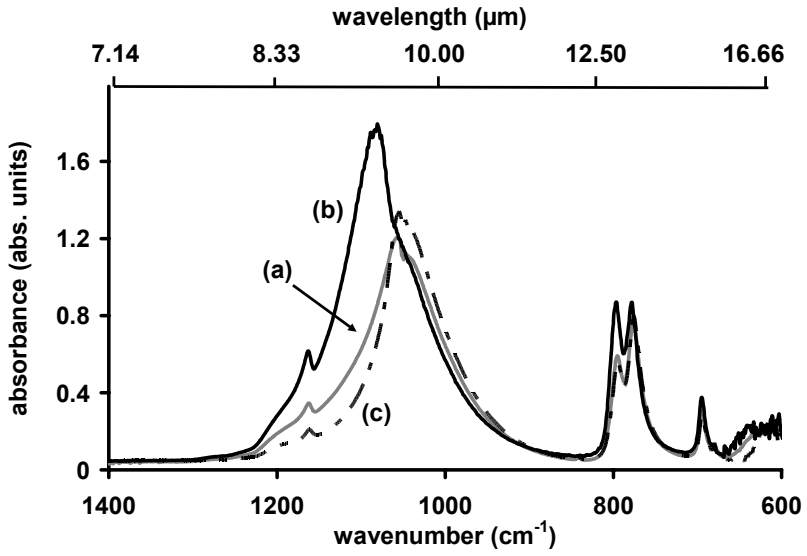


Figure 9.8 ATR spectra of pristine quartz sand samples recorded at different polarization states of infrared radiation: (a) unpolarized light (grey line), (b) p-polarized light (black line), (c) s-polarized light (dotted line).

The pronounced splitting of the dominant absorption feature at 1090 cm^{-1} in both cases is presumably related to a transversal optical (TO) and longitudinal (LO) mode splitting of the asymmetric stretch vibrational mode for SiO_2 as described by Berreman in 1963 [232]. Berreman disapproved of the commonly accepted assumption that IR-spectra of cubic crystals only show vibrational features of TO modes when probed with p-polarized light. He showed that this assumption only holds true for the case of perpendicular light incidence, but was shown to be incorrect for thin films of SiO_2 crystals and oblique incidence angles of light. He related his results in a rather general approach to “special boundary conditions” applicable to thin (semi infinite) films. Harbecke et al. [244] proved that illumination with p-polarized light results in spectral features at the frequencies of TO and LO resonances in reflection and transmission spectra. The LO structure is generated by the surface charges due to the normal component of the electric field. However, it is a prerequisite that the thickness of the film is small compared to the wavelength in vacuum. Furthermore, LO frequencies not only depend on the resonance

frequency of the microscopic oscillator, but also on the dielectric background. Therefore, this effect is related to macroscopic properties of the film. For example, the frequency position of the LO resonance can shift depending on the compactness of the deposited material. From this, the so called Berreman thickness can be derived, which is the film thickness responsible for the maximum effect.

In 1988 Kirk [235] published a contribution for the quantification of mode splitting in case of SiO_2 films. According to his theory, TO-LO splitting occurs for two main reasons: (i) As_1 (asymmetric vibration, O-atoms in phase) mode: LO-TO splitting occurs due to transverse effective (surface) charges, and (ii) As_2 (asymmetric vibration, O-atoms 180 degree out of phase to each other): splitting occurs due to mechanical coupling between the LO and TO mode.

In the same year Piro et al.[236] conducted the first ATR measurements with 2 mm thick α-quartz plates observing LO-TO splitting of the vibrational modes in a thick quartz film and showed that this effect is not limited to ultrathin layers.

From the results shown in Figure 9.7 and Figure 9.8 it can be concluded that the Berreman effect is also observable for particulate materials and is not a unique property of films. This effect has been observed for particulate films for the first time in the course of this study.

9.1.4. Conclusions

It has been shown that IR-ATR spectroscopy in the mid-infrared band provides a reliable methodology for fundamental spectroscopic studies of quartz sand, which potentially benefit interpretation of data provided by the remote sensing community. Besides the already established differences in spectral contrast of disturbed and undisturbed soil, a strong spectral shift of the maximum of the main absorption feature at 1090 cm^{-1} could

be observed. When probed with s- or p-polarized light, the quartz sample showed strong LO-TO mode splitting, which is most likely related to the Berreman effect. These findings advance the variety of spectral characteristics useful to the detection of disturbed soils (i.e. possible landmine sites) with mid-infrared imaging systems. The wetting and drying studies also reveal that the main reason for spectral differences of pristine and disturbed soils eventually relates to changes of the particle size distribution of the sample due to rearrangement of ultrafine particles facilitated by water.

These preliminary results strongly propose the potential of ATR spectroscopic methods for the investigation of signatures derived from remote sensing. Not only the difference in spectral contrast of disturbed and pristine soil could be reproduced, also the assumption of particle size related origin of this phenomenon could be shown. Furthermore, derived from the presented results so far unnoticed spectral shifts in spectra of disturbed and pristine samples was observed clearly, being possibly an exploitable feature for remote disturbed soil detection. In order to render these results useful for remote sensing purposes several experimental studies seem necessary, especially the investigation of mono-disperse samples seems unavoidable, before any quantification and model building for observed effects such as LO-TO mode splitting and absorption intensities can be performed.

Additionally, it is suggested to perform diffuse reflectance measurements applying the same wetting and drying cycles with similar samples in order to ensure that these findings are in coherence with the presented ATR measurements.

9.2. ATR Spectra of Mono-disperse Soda Lime Glass Spheres

In the following section mono-disperse samples (soda lime glass spheres) will be investigated at the same experimental conditions as the quartz samples in the previous

chapter. Resulting, particle size related changes in the spectra during the wetting and drying cycles should be eliminated. Furthermore, the aspect of a possible effect of various particle shapes in case of quartz samples will be suppressed. It is also expected that for at least larger spherical particles the coverage of at the surface of the ATR element will be closely to the most dense packing state and occurring changes in spectral shapes can be associated solely with the different discrete particle size of the sample.

9.2.1. Samples

1–3 µm and 4–10 µm soda lime glass spheres were purchased from the MO-SCI corporation (Rolla, MO), all larger spheres were obtained from Whitehouse Scientific Ltd. (Waverton, Chester, UK).

It should be mentioned that the investigated samples had to be obtained from two different sources in order to cover particle sizes from 1 to >100 µm size regime as the entire dimensional range is not available from one provider. Relevant properties of the materials are listed in Table 9.1. However, the data sheets for both samples mentioned that chemical compositions (and thus the expected absorption spectra) may vary from batch to batch.

Table 9.1 Relevant properties and chemical compositions of the soda lime glass spheres.

Chemical Composition	**MO-SCI Corp.**	**Whitehouse Scientific**
Silica (SiO_2)	65-75%	72.50%
Aluminum oxide (Al_2O_3)	1-5%	0.40%
Calcium oxide (CaO)	9-12%	9.80%
Magnesium oxide (MgO)	1-5%	3.30%
Sodium oxide (Na_2O)	10-20%	13.70%
Iron Oxide (Fe_2O_3)	< 0.3%	0.20%
Physical Properties		
Specific Gravity	2.5 (g/cm^3)	2.49 (g/cm^3)
Index of Refraction	1.51	1.51
Softening Temperature	650°C	740°C
Coefficient of Thermal Expansion	$9x10^{-6}$/°C (30-300°C)	$7.75\ x10^{-6}$/°C (30-300°C)
Diameter	1 - 3 µm 4 - 10 µm	25 - 32 µm 112 - 125 µm 400- 425 µm

In the following sections spheres with diameters of “1-3 µm“ will be named “1 µm“, and “4–10 µm“ will be named “4 µm”, etc..

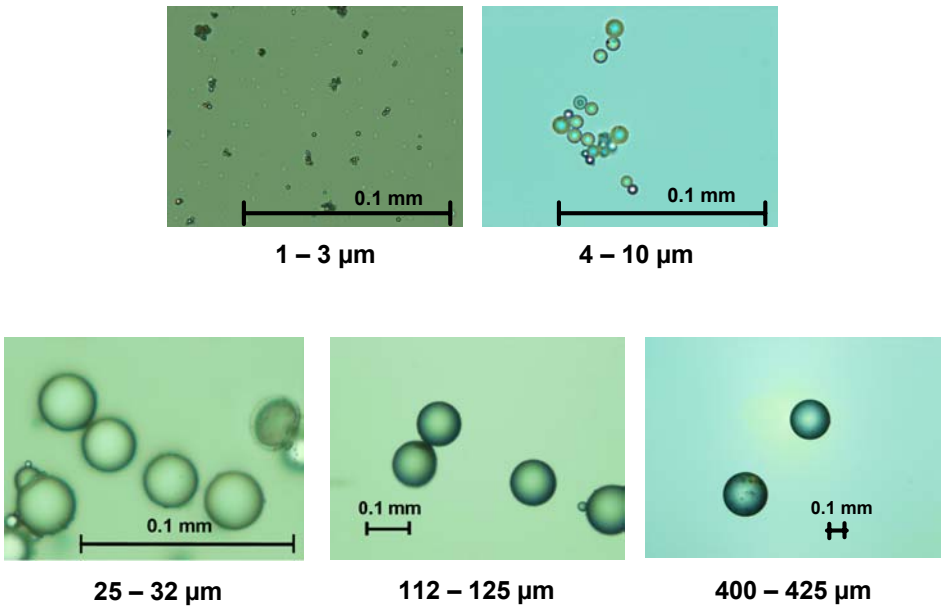


Figure 9.9 Optical microscopy images of soda lime glass spheres. The two smallest size fractions were obtained from the MO-SCI corporation (images on top). The larger size fractions were obtained from Whitehouse Scientific (images at the bottom).

Optical microscopy images of the different glass sphere batches show high quality of the samples except for rare defects, as can be seen in the lower left image of Figure 9.9. Judging from various images of each batch, the number of shape defects and size outliers are insignificantly small and should not significantly influence the obtained IR-ATR spectra.

9.2.2. Experimental

Setup and experimental procedures are similar to the study on quartz samples. Refer to chapter 8.2 and chapter 9.1.1 for details.

9.2.3. Wetting / Drying

In Figure 9.10 the ATR spectrum of the 112 µm spheres are shown illuminated with unpolarized light. Band assignments are given in Table 9.2 [261-263].

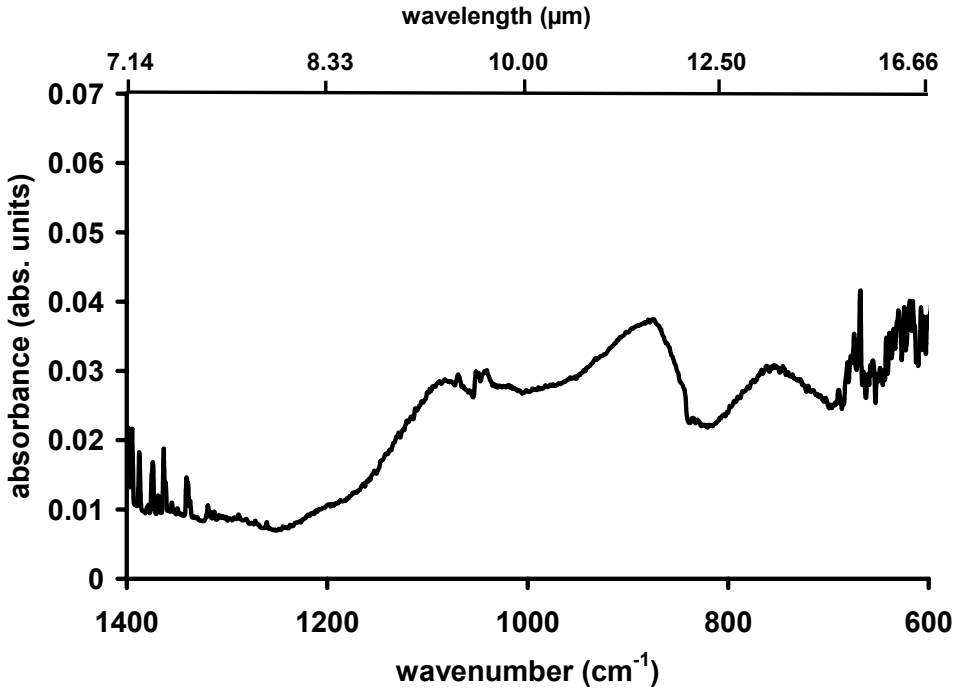


Figure 9.10 The ATR spectrum of 112 µm soda lime glass spheres.

It is clearly observable that the glass sphere spectra show band broadening in comparison to the crystalline quartz spectra, as predicted by theory [218,221].

Table 9.2 Band assignments for soda lime glasses with similar composition [261,262,263]

Peak position (cm^{-1})	Appearance	Assignment
460–480	very sharp	bending vibrations of Si–O–Si linkages
640–680	shoulder	Si–O–Si and O–Si–O bending modes
775–800	sharp	symmetric stretching vibration of [O–Si–O] bonds
960	shoulder	vibration of nonbridging oxygens
1050–1060	broad and very sharp	antisymmetric stretching + vibrations of bridging oxygens
1120	shoulder	Si–O–Si antisymmetric vibrations of bridging oxygens

However, these band assignments have to be considered with caution, as the only obtainable reference of glass with exactly the same composition (Abo-Naf et al. [[261]) shows relating absorption spectra, which resemble the spectra of this thesis only

remotely. A comparison of a transmission spectrum of soda lime glass particles of relatable size (150 to 250 µm) recorded with the KBr pellet technique from the work Abo-Naf with ATR results for soda lime glass spheres is shown in Figure 9.11.

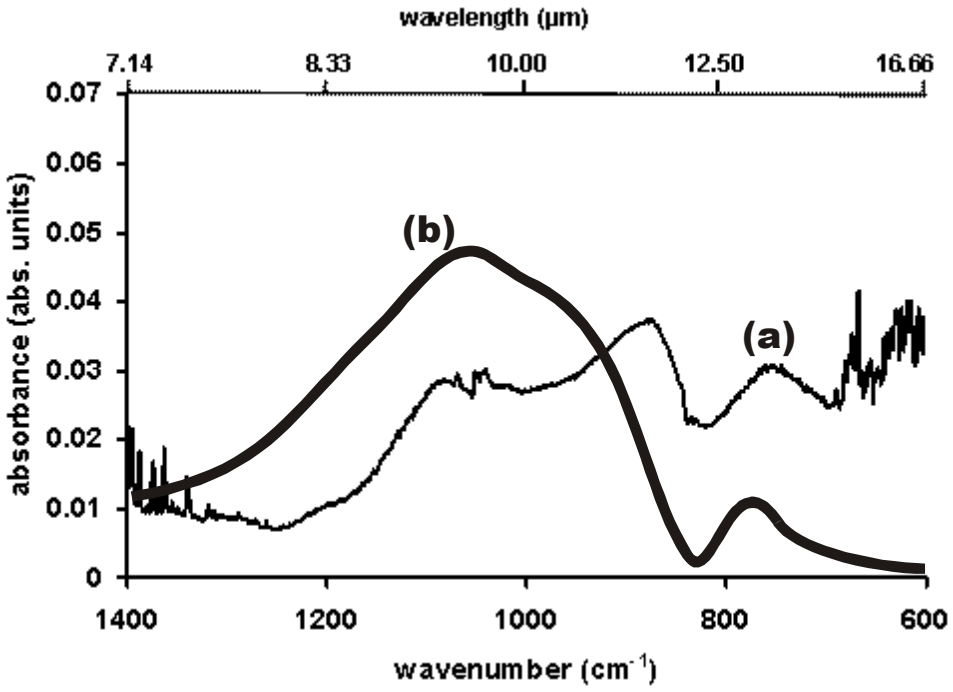


Figure 9.11 The ATR spectrum of 112 µm soda lime glass spheres (a) compared to a transmission spectrum (KBr pellet) of a 150 to 250 µm size fraction of soda lime glass with a very similar composition (b) (spectrum reproduced from [261]).

Especially the band at 960 cm^{-1} (assigned to vibrations of non-bridging oxygens) is of much higher relative intensity than in the cited reference (small shoulder). Another interesting fact is that this reference indicates no contribution of NaO_2 related vibrations in the band assignments, despite it's the rather high content in the sample.

Efimov [257,258] contradicts these band assignments and describes a semi-empirical model, which complements the soda lime glass spectra recoded in this study in a much more satisfactory way. Within the spectra of alkali disilicate crystals, the range of the Si-

O-Si asymmetric stretches can extend down to 900 cm^{-1}, whereas the range of non-bridging oxygen vibrations never extends to frequencies lower than 1000 cm^{-1}. Therefore, band assignments as given in Table 9.2 should be reversed for these two vibrational modes. Furthermore, it is illustrated that the TO mode of the asymmetric stretch vibrations of Si-O-Si is increasingly red-shifted with an increase of Na_2O content. In [257] Efimov presents experimental data and calculations to determine the band position of this asymmetric Si-O-Si vibration to be located around 960 cm^{-1} for a Na_2O-$2SiO_2$ glass providing the strongest oscillator of all vibrational modes at this frequency position. This change of band assignment together with the band positions from Table 9.2 seem to be in better agreement with the measured spectra in this section. Furthermore, as this band corresponds to a strong TO mode, it should be strongly pronounced in measurements with s-polarized light (see chapter 9.2.5). An overview of the revised band assignments is given in Table 9.3.

Table 9.3 Revised band assignments for soda lime glass as suggested by Efimov [257].. The revised bands have been shaded for clarity.

Peak position (cm^{-1})	**Assignment**
460–480	bending vibrations of Si–O–Si linkages
640–680	Si–O–Si and O–Si–O bending modes
775–800	symmetric stretching vibration of [O–Si–O] bonds
900-1000	antisymmetric stretching + vibrations of bridging oxygens
1000-1050	vibration of nonbridging oxygens
1120	Si–O–Si antisymmetric vibrations of bridging oxygens

Similar to the quartz experiments, wetting/drying cycles were investigated with the pristine and dried spectra of the soda lime glass spheres shown in Figure 9.12.

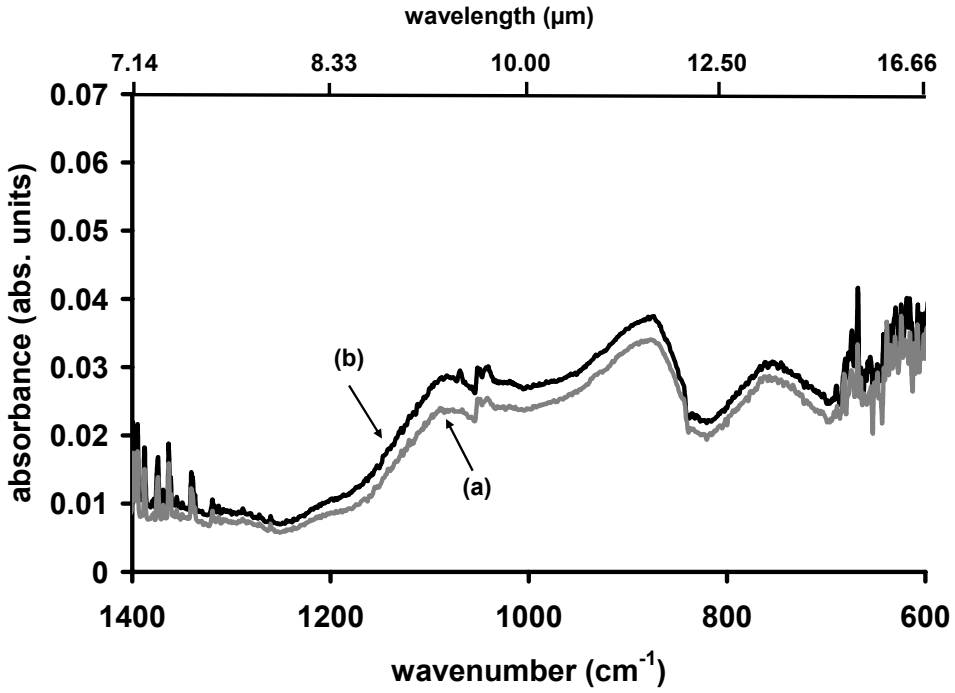


Figure 9.12 Pristine (a), dried (b) spectra of the 112 µm soda lime glass spheres.

The only observable difference in the pristine and dried spectra of the sample is expressed by a small change in spectral intensity, which is related to minor re-arrangements of the spheres during the wetting step. This result is expected given the monodispersity of the spheres already densely packed at the ATR crystal surface and strongly corroborates the assumption that spectral shifts upon wetting and drying are solely related to particle sizes if a fraction of significantly smaller particles is present.

9.2.4. Particle Size Dependence of Absorption Features of Soda Lime Glass Spheres in ATR Spectra

The particle size dependent of ATR spectra of soda lime glass spheres for 3 different diameters (25 µm, 112 µm and 400 µm) are shown in Figure 9.14. The most prominent trend observable is that smaller spheres produce spectra with higher absorption

intensities. This is expected due to the much lower void volume in between the particles on the surface. Upon deposition onto the ATR element surface, these samples seemed to form a film with highly dense uniform surface coverage, which could be visually observed for the larger spheres. In such cases, band intensities derived via ATR techniques might be used as a tool for particle size determination as has been shown for instance by Yoshidome et al. [264], providing (close to) complete coverage of the ATR element.

Apart from the expected change in spectral intensities, there is also a significant blue-shift of the absorption feature around 960 cm^{-1} (asymmetric Si-O-Si stretch vibration) with increasing particle size. In the work of Yoshidome et al. [264], where different sized silica spheres (sizes: 0.81 to 5.2 µm) were investigated, shifting of peaks was not reported. Although, when visually evaluating the published data it appears that shifts are also present in their study, however, apparently have been overlooked (Figure 9.13).

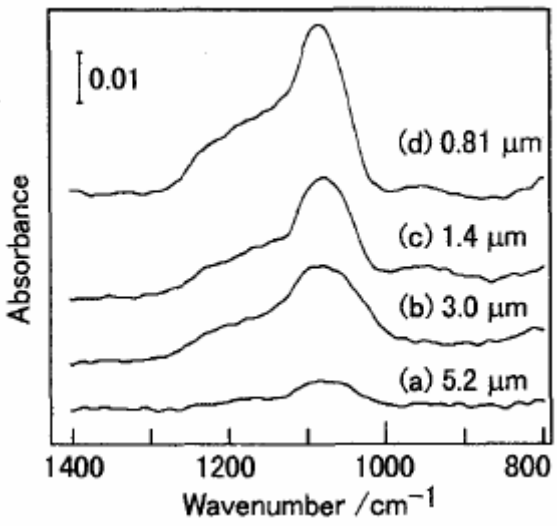


Figure 9.13 ATR spectra of silica-gel particles with various diameters [264].

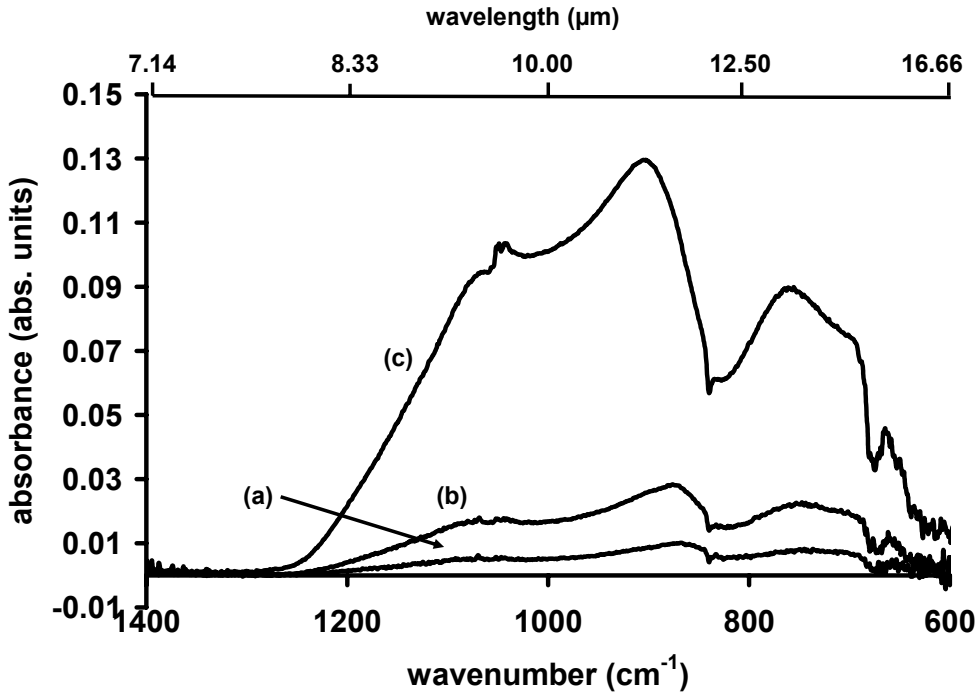


Figure 9.14 ATR spectra of soda lime glass spheres with different diameters: 400 µm (a), 112 µm (b), 25 µm (c).

It is a known problem for ATR spectroscopy of powders that if particles are sufficiently small, electrostatic forces will produce particle conglomerates and distribution by simply applying the powder onto the ATR surface will not ensure complete coverage.

In case of the glass spheres used in this study this was almost certainly the case for the two smallest fractions (1 and 4 µm). A possible solution to this problem is to suspend these samples in a volatile liquid (e.g. chloroform), apply the suspension onto the crystal and wait for solvent evaporation leaving a generally rather homogeneous film on the substrate. However, in case of strongly absorbing materials the amount of solids has to be optimized so that the resulting layer thickness does not produce spectra where total absorption takes place. This happened in several unsuccessful tries with different amounts of deposited suspension. Hence, an alternative approach for the two smallest sphere sizes was developed. In order to obtain useful spectra the powders were gently

pressed onto the ATR element via an aluminum block throughout the entire measurements, which finally yielded satisfactory results.

ATR spectra of all soda lime spheres in the range from 1 to 400 µm are shown in Figure 9.15.

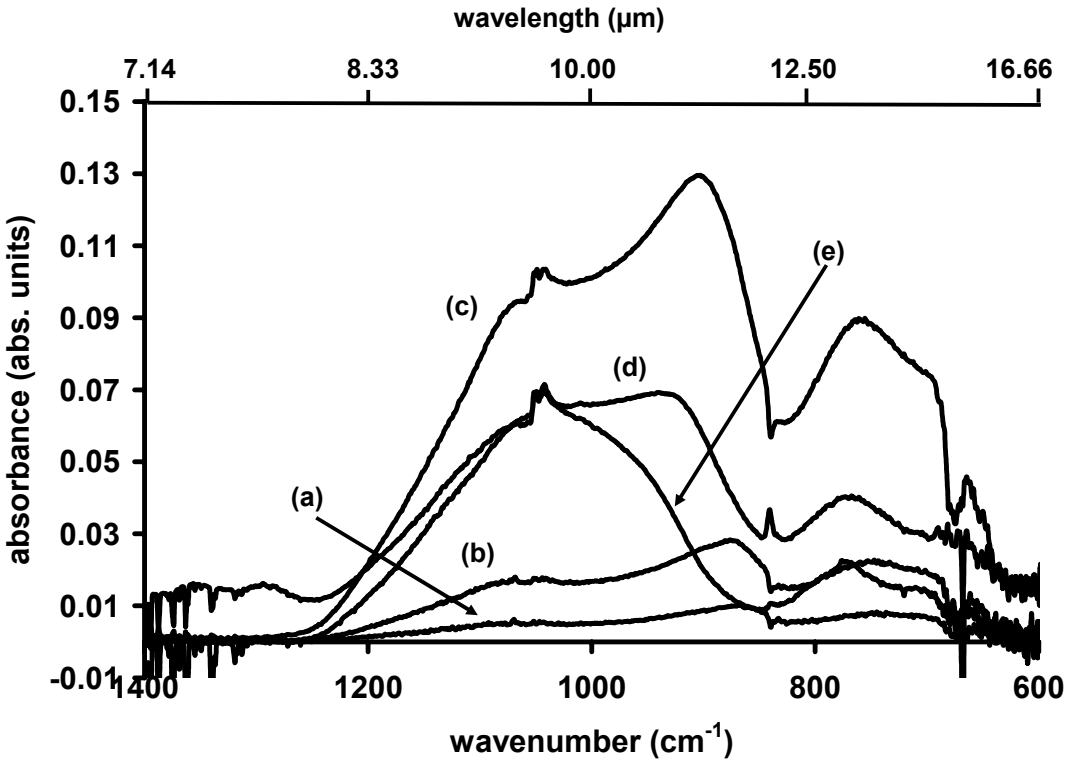


Figure 9.15 ATR spectra of soda lime glass spheres with different diameters: 400 µm (a), 112 µm (b), 25 µm (c), 4 µm (d) and 1 µm (e).

The uncertainty of complete coverage of the two smallest fractions is confirmed by the fact that they apparently do not follow the logical trend showing the most intense absorption features. However, a very interesting size-related effect can be observed: the initially most intense band at 960 cm^{-1} is continuously decreasing in intensity and is practically vanished in the spectrum of the 1 µm spheres. In order to be able to follow this trend more precisely, the spectra have been normalized at 1040 cm^{-1}, as this band

appears to be a spectral region where only minor changes occur throughout the variation of sphere diameters (Figure 9.16) .

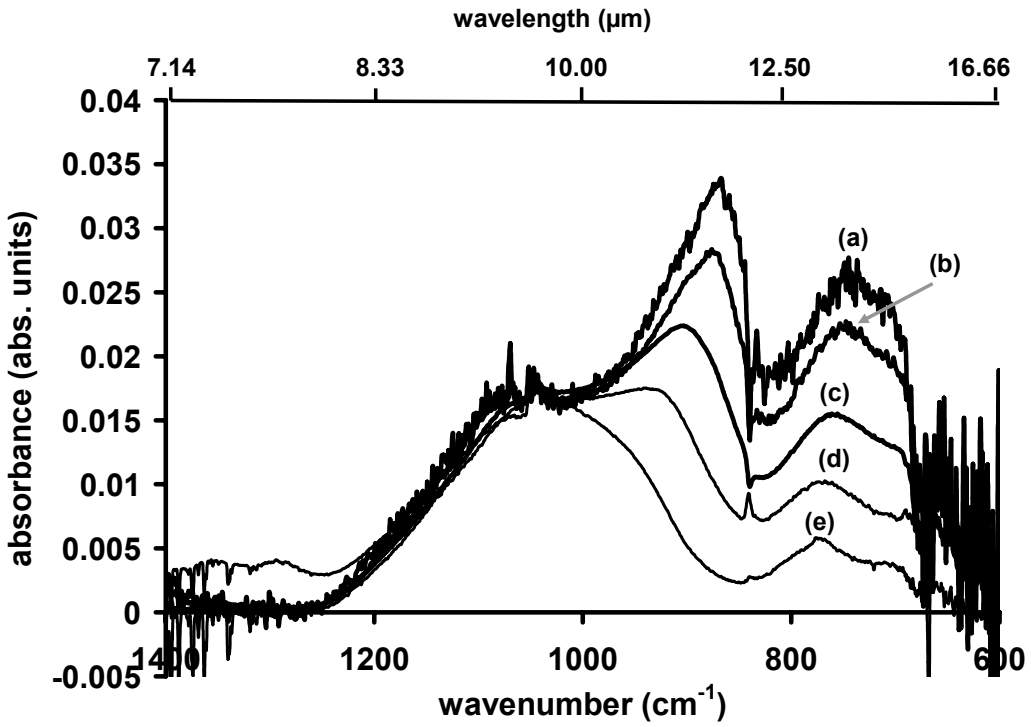


Figure 9.16 Normalized (at 1040 cm^{-1}) ATR spectra of soda lime glass spheres with diameters of: 400 µm (a), 112 µm (b), 25 µm (c), 4 µm (d) and 1 µm (e).

It appears that normalization at a frequency of 1040 cm^{-1} is a valid operation, as the higher frequency component of the recorded spectra appears very similar for all particle diameters.

Assuming band assignments (Table 9.3) are correct, a monotonously decreasing intensity of the TO mode of the non bridging oxygen (NBO) vibrational band (~870 cm^{-1} for the 400 µm spheres) with decreasing particle size is observable. It should be mentioned that due to the normalization of the spectra this decrease is to be recognized as a relative change in absorption in relation to the other absorption features. The initial strong intensity of NBO vibrational band can be explained by the high content of cations

the glass composition (Na^{+}>10%, Ca^{2+}~10%, Fe^{3+}~0.5%), which strongly promote the abundance of NBO sites [265-267]. The cations provoke the disruption of the amorphous network due to the breaking of some of the Si–O–Si bonds leading to the formation of nonbridging oxygen groups (Si–O–NBO). According to the spectra shown in Figure 9.16, the number of NBO sites is increasing together with particle size, which is expressed by the increasing intensity of the respective absorption band in respect to the other spectral features. In their work, Serra et al. [267] show that similar effects can be observed due the influence of cation / SiO_2 ratio in glasses as shown in Figure 9.17.

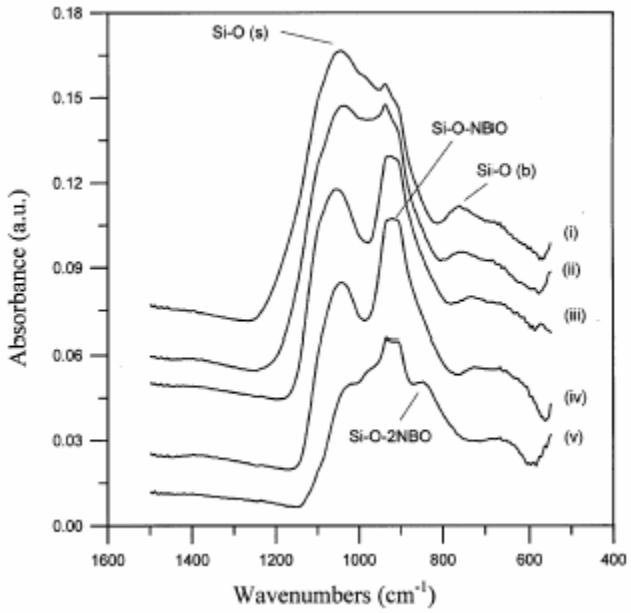


Figure 9.17 FTIR spectra of Na_2O–CaO–P_2O_5–K_2O–MgO–B_2O_3–SiO_2 glasses with different SiO_2 content: (i) 66%, (ii) 59%, (iii) 55%, (iv) 50% and (v) 42% [267].

It can be seen that a decreasing SiO_2 content goes along with a strong rise of the Si-O-NBO feature around 900 cm^{-1}.This indicates that the results shown in Figure 9.16 do not necessarily prove a particle size related effect, but could also arise from a systematic - particle size related - change in composition of the glass spheres during the

manufacturing process. In order to clarify such assumptions a chemical analysis of the soda lime glass spheres is suggested for future studies.

9.2.5. Polarized Light

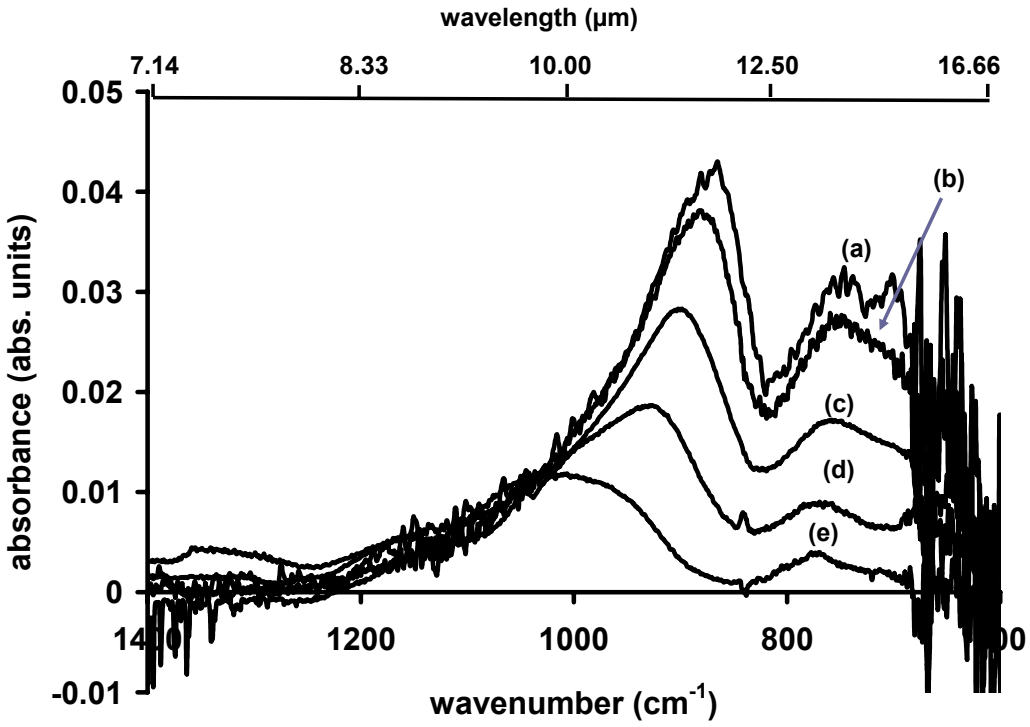


Figure 9.18 S-polarized ATR spectra of soda lime glass spheres with different diameters: 400 μm (a), 112 μm (b), 25 μm (c), 4 μm (d) and 1 μm (e). Data has been normalized in intensity.

Figure 9.18 shows the ATR spectra of all sizes of the glass spheres under s-polarized illumination. Theoretically, only TO modes should be observable. At least three modes are immediately noticeable:

- ~750 cm^{-1} (symmetric Si-O-Si stretch vibration)
- ~ 880 – 940 cm^{-1} (shifting, decaying NBO stretch vibration)

- ~1050 cm^{-1} (asymmetric Si-O-Si stretch vibration)

These band assignments are based on the considerations discussed earlier in chapter 9.2.3 (page 169).

The most obvious trend in these spectra can be attributed to the strong decrease of the intensity of the NBO stretch vibrational band with decreasing particle size. This mode expresses a strong apparent blue shift as well, exclusively based on the evaluation of the shift of the maximum peak position shifting from ~ 880 to 940 cm^{-1} for the examined particle sizes. However, the substantial shift of the peak maximum over almost 80 wavenumbers is at least partly promoted due to the strong overlaps of the absorption features in the spectra.

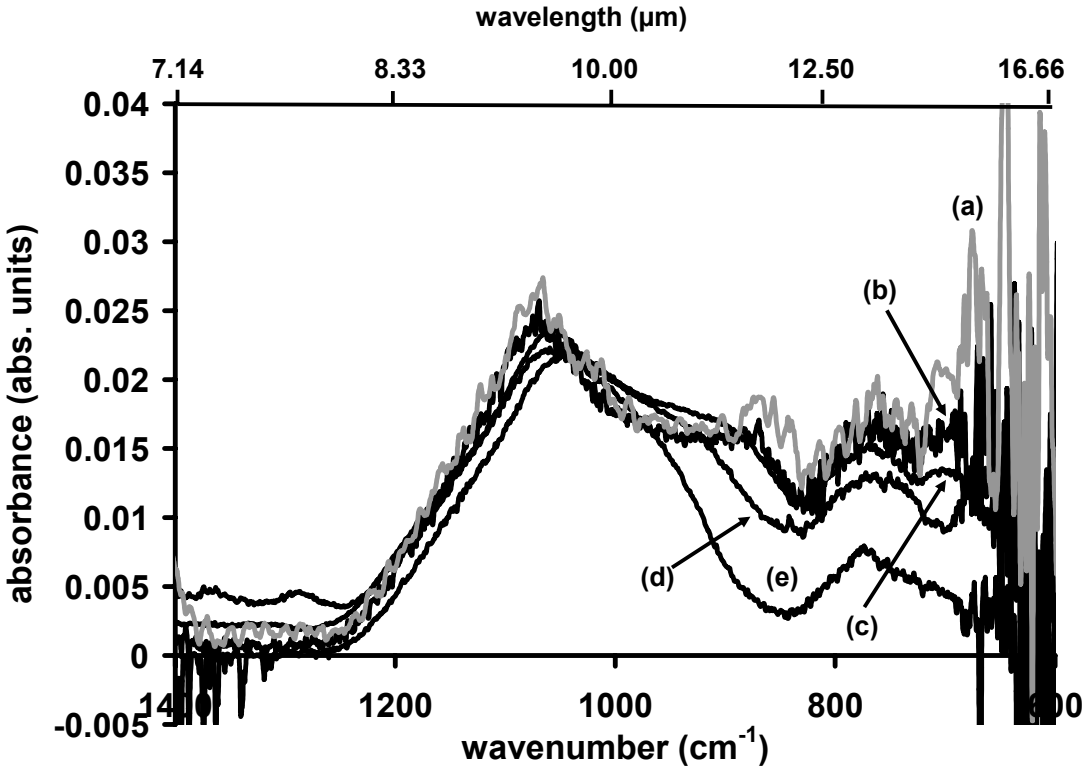


Figure 9.19 P-polarized ATR spectra of soda lime glass spheres with different diameters: 400 µm (a), 112 µm (b), 25 µm (c), 4 µm (d) and 1 µm (e).

Figure 9.19 shows the ATR spectra of all sizes of the glass spheres under p-polarized illumination. Theoretically, LO modes should be strongly expressed and TO modes should be visible to a lower extent. It is known that LO modes only express minor dependency on peak position and peak width upon varying layer thickness [see e.g. 254,257]. The experimental data corroborate these assumptions: the major absorption feature shifts by approx. 15 cm^{-1} from ~1050 to 1065 cm^{-1} with increasing particle diameter. The changes in the spectra in the longer wavelength region are attributed to also expressed TO modes of the glass spheres. Usually, band intensity and peak widths of LO modes show strong dependency on the angle of incidence e.g. [252**Error! Bookmark not defined.**]. However, with the present ATR setup this parameter is in principle fixed and cannot be reproduced in the course of these experiments.

9.2.6. Conclusions

ATR measurements of mono-disperse soda lime glass spheres lead to the following conclusions

- Wetting and drying studies of a sample consisting of glass spheres of only one particle size showed no differences in the pristine and dried spectrum. This circumstance is another indication that the different spectral properties of disturbed and pristine soils are a particle size related effect.
- Particle size dependent ATR spectra of the soda lime glass spheres showed significant changes in the relative band intensities of the absorption features. After band assignment it could be concluded that the intensity of the non bridgin oxygen stretch vibration band (~860 cm^{-1}) decreased with decreasing sphere diameters in respect to the other major absorption features.

- Experiments under s- and p-polarized illuminations corroborated the proposed band assignments and effects.
- The present setup with a broad angle distribution of the incident light in respect to the ATR element surface is suitable for showing general trends, but represents a problem for detailed evaluation of the spectra as LO-TO mode splitting effects are strongly dependent on the angle of incidence.
- For quantitative results it is recommended to modify the sample illumination technique in order to be able to address different, defined angles of incidence.

10. Conclusion and Outlook

10.1. Are ATR spectroscopic Studies suitable as supporting Method for Remote Sensing?

It has been shown that Mid-IR-ATR spectroscopy provides a reliable methodology for fundamental spectroscopic studies of quartz sand, which potentially benefit interpretation of data provided by the remote sensing community. Besides the already established differences in spectral contrast of disturbed and undisturbed soil, a strong spectral shift for quartz samples of the maximum of the main absorption feature at 1090 cm^{-1} could be observed. When probed with s- or p-polarized light, the sample showed strong LO-TO mode splitting, which is most likely related to the Berreman effect. These findings advance the variety of spectral characteristics useful to the detection of disturbed soils (i.e. possible landmine sites) with mid-infrared imaging systems. The wetting and drying

studies also reveal that the main reason for spectral differences of pristine and disturbed soils eventually relates to changes of the particle size distribution of the sample due to rearrangement of ultrafine particles facilitated by water.

These preliminary results strongly propose the potential of ATR spectroscopic methods for the investigation of signatures derived from remote sensing. Not only the difference in spectral contrast of disturbed and pristine soil could be reproduced, also the assumption of particle size related origin of this phenomenon could be shown. Furthermore, derived from the presented results so far unnoticed spectral shifts in spectra of disturbed and pristine samples was observed clearly, being possibly an exploitable feature for remote disturbed soil detection.

A Mid-IR-ATR study of mono-disperse soda lime glass spheres with diameters in the range from 1 to 400 µm subsidized the findings of the quartz measurements and due to the more defined sample led to deeper insight of the reasons for spectral changes in relation for the disturbed and undisturbed soil problem.

Wetting and drying studies of a sample consisting of glass spheres of only one particle size showed no differences in the pristine and dried spectrum., which is in high agreement with the assumption that different spectral properties of disturbed and pristine soils can is related to particle size. Furthermore, in the comparison of ATR spectra of different mono-disperse glass spheres, relative band intensity shifts were observed, another potentially interesting finding for remote sensing of disturbed soils. The results showed a relative intensity change of the TO non bridging oxygen stretch vibrational band in respect to other major bands in the spectra. This vibrational mode becomes less pronounced with decreasing sphere diameters. Measurements performed under s- and p-polarized illumination of the sample corroborated these findings. However, in order to perform quantitative data evaluation, a modified setup is recommended where the angle of incidence of the IR radiation can be chosen.

In order to render these results useful for remote sensing purposes fundamental experimental studies seem necessary, especially the investigation of mono-disperse quartz samples are needed, before profound quantification and model building for observed effects such as LO-TO mode splitting and absorption intensities can be performed.

For further studies an environmental chamber has been developed, which is compatible with the laboratory based ATR setup. It allows to control relevant parameters such as temperature and humidity, which potentially influence the spectral behavior of samples in the field (see Appendix, Figure 10.1).

Additionally, it is suggested to perform diffuse reflectance or emissivity measurements applying the same wetting and drying cycles with similar samples in order to ensure that the presented ATR measurements are in coherence with data derived from remote sensing.

APPENDIX

Instruments and Major Components

Instruments, components	Brand, specification etc.	Company
FT-IR spectrometer	Bruker Vektor 22 Bruker EQUINOX 55	Bruker Optics Inc (Billerica, MA, USA)
Vertical ATR Module	Standard Mirror Bench	Specac Inc. (Smyrna, GA, USA)
Horizontal ATR Module	Standard Mirror Bench	Specac Inc. (Smyrna, GA, USA)
MCT detector	D316-type	Infrared Associates (Stuart, FL, USA)
ZnSe ATR elements	50*20*2 mm, 45°, trapezoid	Macrooptica Ltd. (Moscow, Russia)
ZnSe ATR elements	72*10*6 mm, 45°, trapezoid	Macrooptica Ltd. (Moscow, Russia)
Breadboard	Aluminium, 30*60*1.27 cm	Thorlabs (North Newton, NJ, USA)
High precision piston pump 10-port selection valve 6-port injection valve	Cavro XL3000, vol. 25,000 µL Valco C25Z-3180EMH Valco C22Z-3186EH	Global FIA Inc (Fox Island, WA, USA)
Profilometer	Dektak^3	Veeco/Sloan Technology (Santa Barbara, CA, USA)
Spin-coater	WS-400A-6NPP-LITE	Laurell Technologies Corporation (North Wales, PA, USA)

Visual Basic Script for Modeling Diffusion via Fieldson and Barbari (Chapter 3.4.2)

```
Private Function CalcF(n As Long, _
l As Double) As Double
CalcF = ((2 * n + 1) * PI) / (2 * l)
End Function

Private Function CalcG(D As Double, _
n As Long, _
l As Double, _
t As Long) As Double

CalcG = (-D * (2 * n + 1) ^ 2 * PI ^ 2 * t) / (4 * (l ^ 2))

End Function

Private Function CalculateA(n As Long, t As Long) As Double

Dim dSum As Double
Dim l As Long

For l = 0 To n
dSum = dSum + (Exp(CalcG(mdD, l, mdL, t)) * (CalcF(l, mdL) * Exp((-2 * mdL) / mdDp) + (-1) ^ l *
(2 / mdDp))) / ((2 * l + 1) * (4 / (mdDp ^ 2) + (CalcF(l, mdL)) ^ 2))

Next

CalculateA = 1 - (8 / (PI * mdDp * (1 - Exp((-2 * mdL) / mdDp)))) * dSum

End Function
```

The Environmental Chamber

For investigating the effects of varying temperature and humidity on ATR spectra of quartz (and other minerals) a small environmental chamber (volume approximately 500mL) has been developed for future use (Figure 10.1).

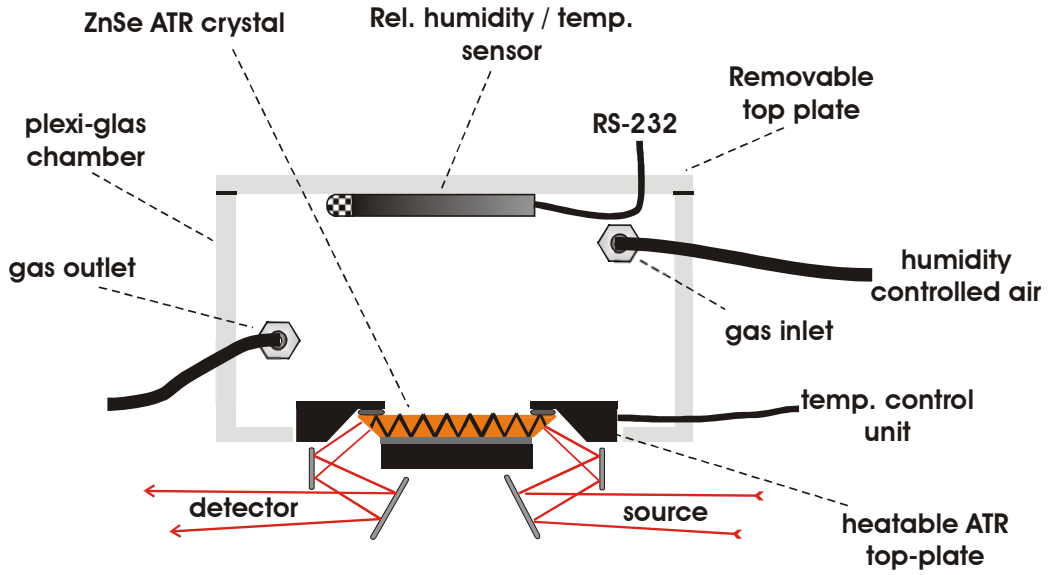


Figure 10.1 Schematic of the environmental chamber developed for temperature and humidity studies on ATR measurements for quartz and other minerals.

Major parts are labeled in the schematic picture, a brief description of the components is as follows"

- Humidity and temperature validation: Omega, OM-CP-RHTEMP101 is a miniaturized RH/T sensor that can be used as a datalogger
- Heatable ATR-setup: Specac Inc. 11155, allows controlled heating (constant temperatures and temperature ramps) of the ZnSe crystal from room temperature up to 150 °C via an external temperature controller.

- Humidity controlled air: with the help of a relative humidity controlling system from (Sable Syste,s, DG-1) allows to introduce an airflow of controlled humidity (working range: approx. 10 to 90% RH, accuracy +/- 1%) or drive humidity ramps / cycles.
- housing: this first generation environmental chamber consists of an easy accessible (sample introduction, disturbing...) plexi-glass material, which is glued together and provides sufficient sealing to the outside.

The environmental chamber is fully assembled, is being tested for stability and general performance.

REFERENCES

1 J. E. McFee, H. T. Ripley *Detection of buried landmines using a casi hyperspectral image*" Detection and Remediation Technologies for Mines and Minelike Targets II, ed A. C. Dubey, SPIE, 1997, Vol **2079**, 738-749.

2 P. G. Lucey, T. J. Williams, M. Winter, *Two years of operations of AHI: a LWIR hyperspectral imager*, Infrared Imaging Systems: Design, Modeling, and Testing XI, ed. G. Holst, SPIE, 2000, Vol. **4030**, 31-40

3 C. A. DiMarzio, T. Vo-Dinh, H. E. Scott, *Some Approaches to Infrared Spectroscopy for Detection of Buried Objects*, Detection and Remediation Technologies for Mines and Minelike Targets III, eds A. C. Dubey, J. T. Broach, SPIE, 1998, Vol **3392**, 158-166.

4 H. T. Haskett, R. Rupp, T. Moore *Quantitative performance of buried mine hyperspectral reflectance signatures (.35-2.5µm) in various soils* Detection and Remediation Technologies for Mines and Minelike Targets V, eds A. C. Dubey, J. F. Harvey, J. T. Broach, R. E. Dugan, SPIE, 2000, Vol **4038**, 886-899.

5 J. R. Johnson, P. G. Lucey, K. A. Horton, E. M. Winter, "Infrared Measurements of Pristine and Disturbed Soils 1. Spectral Contras Differences between Field and Laboratory Data", Remote Sensing of the Environment, 1998, **64**, 34–46.

6 L. Ritter, K. Solomon, P. Sibley, K. Hal, P. Keen, G. Mattu, B. Linton, *Sources, Pathways, and relative risks of contaminants in surface water and groundwater: a perpective prepared for the walkerton inquiry*, Journal of Toxicology and Environmental Health Part A, 2002, **65(1)**, 1-142

7 P. Williams, L. Benton, J. Warmerdam, P. Sheehan, *Comparative risk analysis of six volatile orga*nic *compounds in california drinking water*, Environmental Science and Technology*,* 2002, **36(22)**, 4721–4728.

8 P. T. C. Harrison, *Health effects of environmental chemicals,* R. M. Harrison (Editor) Pollution: Causes, Effects and Control, (4th edition), 2001, 500-523.

9 J. H. Sevigny, M. J. Tindal, G. L. Robins, W. Staudt, L. Serbin, *Importance of different volatile petroleum hydrocarbon fractions in human risk assessment*, Human and Ecological Risk Assesment 2003, **9 (4),** 987-1001.

10 B. C. J. Zoeteman, K. Harmsen, J. B. H. J. Linders, C. F. H. Morra, W. Slooff, *Persistent organic pollutants in river water and groundwater of the Netherlands,* Chemosphere 1980, **9(4)**, 231-249.

11 W. Giger, C. Schaffner, *Ground water pollution by volatile organic chemicals Studies in Environmnental*, Science, 1981, **17**, 517-522.

12 S. Ramamoorthy, S. Ramamoorthy, *Chlorinated Organic Compounds in the Environment,* Lewis, 1997, Boca Raton, Fla.

13 T. Trouwbourst, *Groundwater pollution by volatile halogenated hydrocarbons; sources of pollution and methods to estimate their relevance*, Science Total Environment 1981, **21**, 41-46.

14 P. J. Squillace, M. J. Moran, W. W. Lapham, *Volatile organic compounds in untreated ambient groundwater of the united states, 1985-1995*, Environmental Science and Technology, 1999, **33(23)**, 4176-4187.

15 P. J. Squillace, J. C. Scott, M. J. Moran, B. T. Nolan, D. W. Kolpin, *VOCs, pesticides, nitrate, and their mixtures in groundwater used for drinking water in the united states*, Environmental Science and Technology 2002, **36(9)**, 1923-1930.

16 Permanent Website: http://www.atsdr.cdc.gov/clist.html.

17 EPA 811-F-95-004-T, *Contaminant Specific Fact Sheets - Volatile Organic Chemicals*, Oct 1995.

18 EPA 816-R-99-006, *A Review of Contaminant Occurrence in Public Water Systems*, Nov 1999.

19 R. Krska, E. Rosenberg, K. Taga, R. Kellner, A. Messica, A. Katzir, *Polymer coated silver halide infrared fibers as sensing devices for chlorinated hydrocarbons in water,* Appl. Phys. Lett., 1992, **61(15)**, 1778-1780.

20 E. Martinez, S. Lacorte, I. Lobet, P. Viana, D. Barcelo, *Multicomponent analysis of volatile organic compounds in water by automated purge and trap coupled to gas chromatography-mass spectrometry*, Journal of Chromatography A, 2002, **959(1-2)**, 181–190.

21 M. Rosell, S. Lacorte, A. Ginebreda, D. Barcelo, *Simultaneous determination of methyl tert.-butyl ether and its degradation products, other gasoline oxygenates and benzene, toluene, ethylbenzene and xylenes in Catalonian groundwater by purge-and-trap-gas chromatography-mass spectrometry*, Journal of Chromatography A, 2003 ,**1007(1-2)**, 209-210 .

22 M. C. Hennion, *Solid-phase extraction: method development, sorbents, and coupling with liquid chromatography*, Journal of Chromatography A, 1999, **856(1-2)**, 3–54

23 I. Liska, *Fifty years of solid-phase extraction in water analysis - historical development and overview*, Journal of Chromatography A, 2000, **885(1-2)**, 3–16.

24 A. M. Dietrich, W. F. daCosta, *Measurement and monitoring of pollutants. Chemical species,* Water Environment Research, 1997, **69(4)**, 391-403.

25 H. C. Hennion, V. Pichon, D. Barcelo, *Surface water analysis (trace-organic contaminants) and EC regulations*, Trends in Analytical Chemistry, 1994,**13(9)**, 361-372.

26 W. Goepel, J. Hesse, J. N. Zemel (Editors), *Sensors*, VCH, Weinheim, Germany 1995.

27 R. F. Taylor, J. S. Schultz (Editors) *Handbook of Chemical and Biological Sensors,* Institute of Physics, Bristol, UK, 1996.

28 B. R.Eggins, Chemical Sensors and Biosensors, John Wiley & Sons, West Sussex, UK 2002.

29 J. Janata; A. Bezegh *Chemical sensors*. Analytical Chemistry, 1988, **60(12)**, 62R-74R

30 J. Janata, *Chemical sensors*, Analytical Chemistry, 1990, **62(12)**, 33R-44R

31 J. Janata, *Chemical sensors*, Analytical Chemistry, 1992, **64(12)**, 196R-219R

32 J. Janata, M. Josowicz, D. M. DeVaney, *Chemical sensors*, Analytical Chemistry, 1994, **66(12)**, 207R-228R

33 J. Janata, M. Josowicz, P. Vanysek, D. M. DeVaney, *Chemical Sensors*, Analytical Chemistry, 1998, **70(12)**, 179R – 208R.

34 J. Janata, *Centennial retrospective on chemical sensors*, Analytical Chemistry, 2001, **73(5)**, 150A-153A

35 O. S. Wolfbeis (Editor), *Fiber Optic Chemical Sensors and Biosensors*, Vol. 1, CRC Press, Boca Raton, Fl, 1991.

36 G. Holst, B. Mizaikoff, *Optical fiber sensors for environmental applications. Handbook of fiber optic sensing technology: principles and application*, J. M. Lopez-Higuerra (Editor), John Wiley & Sons Ltd, 2002, 729-755.

37 R. Krska, K. Taga, R. Kellner, *New IR fiber-optic chemical sensor for in-situ measurements of chlorinated hydrocarbons in water*, Applied Spectroscopy, 1993, **47(9)**, 1484-1487.

38 B. Mizaikoff, R. Goebel, R. Krska, K. Taga, R. Kellner, M. Tacke, A. Katzir, *Infrared fiber-optical chemical sensors with reactive surface coatings*, Sensors and Actuators B, 1995, **29(1-3)**, 58-63.

39 J. E. Walsh, B. D. MacCraith, M. Meaney, J. G. Vos, F. Regan, A. Lancia, S. Artjushenko, *Sensing of chlorinated hydrocarbons and pesticides in water using polymer coated mid-infrared optical* fibers, Analyst, 1996, **121(6)**, 789-792.

40 G. Schwotzer, I. Latka, H. Lehmann, R. Willsch, *Optical sensing of hydrocarbons in air or in water using UV absorption in the evanescent field of fibers*, Sensors and Actuators B, 1997, **38(1-3)**, 150-153.

41 J. Buerck, S. Roth, K. Kraemer, M. Scholz, S. Klaas, *Application of a fiber-optic NIR-EFA sensor system for in-situ monitoring of aromatic hydrocarbons in contaminated groundwater*, Journal of Hazardous Materials, 2001, **83(1-2)**, 11-28.

42 J. Buerck, M. Schlagenhof, S. Roth, H. Mahieu, *Kinetic evaluation method of SPME-NIR measurements of analytes with long equilibration time Field*, Analytical Chemistry and Technology, 2001, **5(3)**, 131-142.

43 C. Phillips, M. Jakusch, H. Steiner, B. Mizaikoff, A. G. Fedorov, *Model-based optimal design of polymer-coated chemical sensors*, Analytical Chemistry, 2003, **75(5)**, 1106-1115.

44 O. S. Wolfbeis, *Fiber-optic chemical sensors and biose*nsors, Analytical Chemistry 2000, **72(12)**, 81R-89R.

45 O. S. Wolfbeis, *Fiber-optic chemical sensors and biosensors*, Analytical Chemistry, 2002, **74(12)**, 2663-2678.

46 B. Mizaikoff, *Mid-IR fiber-optic sensors*, Analytical Chemistry, 2003, **75(11)**, 258A-267A.

47 F. Baldini, S. Bracci, *Polymers for optical fiber sensors Polymer*, Sensors and Actuators, 2000, 91-107.

48 V. Acha, M. Meurens, H. Naveau, S. N. Agathos, *ATR-FTIR sensor development for continous on-line monitoring of chlorinated aliphatic hydrocarbons in a fixed-bed bioreactor*, Biotechnology and Bioengineering, 2000, **68(5)**, 473-487.

49 J. Buerck, M. Mensch, K. Kraemer, *Field Experiments with a portable fiber-optic sensor system for monitoring hydrocarbons in water Field*, Analytical Chemistry and Technology, 1998, **2(4)**, 205-219.

50 J. Buerck, S. Roth, K. Kraemer, H. Mathieu, *OTDR fiber-optical chemical sensor system for detection and location of hydrocarbon leakage*, Journal of Hazardous Materials, 2003, **102(1)**, 13-28.

51 B. L. Wittkamp, D. C. Tilotta, *Determination of BTEX compounds in water by solid-phase microextraction and Raman spectroscopy*, Analytical Chemistry, 1995, **67(3)** 600-605.

52 B. L. Wittkamp, S. B. Hawthorne, *Determination of aromatic compounds in water by solid phase microextraction and ultraviolet absorption spectroscopy. 1. Methodology*, Analytical Chemistry 1997, **69(6)**, 1197-1203.

53 S. A. Merschman, S. H. Lubbad, D. C. Tillota, *Poly(dimethylsiloxane) films as sorbents for solid-phase microextraction coupled with infrared spectroscopy*, Journal of Chromatography A, 1998, **829(1+2)**, 377-384.

54 D. C. Stahl, D. C. Tilotta, *Partition infrared method for total gasoline range organics in water based on solid phase microex*traction, Environmental Science and Technology, 1999, **33(5)**, 814-819.

55 M. Lamotte, P. Fornier de Violet, P. Garrigues, M. Hardy, *Evaluation of the possibility of detecting benzenic pollutants by direct spectrophotometry on PDMS solid absorbent*, Analytical and Bioanalytical Chemistry, 2002, **372(1)**, 169-173.

56 J. Pawliszyn, *Solid phase microextraction – Theory and practice*, (1st ed.) Wiley, New york, NY, 1997.

57 J. Olsson, G. Traegardh, F. Lipnizki, *The influence of permeant and membrane properties on mass transfer in pervoration of volatile organic compounds from dilute aqueous solutions Separation*, Sciences and Technology, 2002, **37(6)**, 1199-1223.

58 D. M. Haaland, D. K. Melgaard, *New prediction-augmented classical least-squares (PACLS) methods: application to unmodeled interferents*, Applied Specroscopy, 2000, **54(9)**, 1303-1312.

59 F. Vogt, B. Mizaikoff, *Secured PCR (sPCR) for detection and correction of PCR calibration model failures induced by uncalibrated spectral features, Journal of Chemometrics*, 2003, **17(4)**, 225-236.

60 F. Vogt, B. Mizaikoff, *Introduction and application of secured principal component regression for analysis of uncalibrated spectral features in optical spectroscopy and chemical sensing*, Analytical Chemistry, 2003, **75(13)**, 3050-3058.

61 J. Einax, H. W. Zwanziger, S. Geiss, *Chemometrics in Environmental Analysis*, VCH, Weinheim Germany, 1997.

62 K. R. Beebe, R. J. Pell, M. B. Seasholtz, *Chemometrics: A practical guide*, Wiley, New York, N.Y., 1998.

63 R. G. Brereton, *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*, John Wiley & Sons, New York, N.Y. 2003.

64 D. M. Haaland, E. V. Thomas, *Partial least-squares methods for spectral analysis. 1. Relation to other quantitative calibration methods and the extraction of qualitative information*, Analytical Chemistry 1988, **60 (11)**, 1193-1202.

65 E. V. Thomas, D. M. Haaland, *Comparison of multivariate calibration methods for quantitative spectral analysis*, Analytical Chemistry, 1990, **62 (10)**, 1091-1099.

66 D. M. Haaland, L. Han, T. M. Niemczyk, *Use of CLS to understand PLS IR calibration for trace detection of organic molecules in water*, Applied Spectroscopy 1999, **53(4)**, 390-395.

67 M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, J. Pages, *NIR calibration in nonlinear systems: different PLS approaches and artificial neural networks*, Chemometrics and Intelligent Laboratory Systems, 2000, **50(1)**, 75-82.

68 P. C. Jurs, G. A. Bakken, H. E. McClelland, *Computational methods for the analysis of chemical sensor array data from volatile analytes*, Chemical Reviews, 2000, **100(7)**, 2649-2678.

69 D. M. Haaland, *Synthetic multivariate models to accommodate unmodeled interfering spectral components during qualitative spectral analysis*, Applied Spectroscopy, 2000, **54(2)**, 246-254.

70 R. A. S. Lapa, J. L. F. C. Lima, B. F. Reis, J. L. M. Santos, E. A. G. Zagatto, *Multi-pumping in flow analysis: concepts, instrumentation, potentialities*, Analytica Chimica Acta, 2002, **466(1**), 125-132.

71 F. Albertus, B. Horstkotte, A. Cladera, V. Cerda, *A robust multisyringe system for process flow analysis. Part I. On-line dilution and single point titration of protolytes*, Analyst, 1999, **124(9)**, 1373-1381.

72 E. Richards, C. Bessant, S. Saini, *A liquid handling system for the automated acquisition of data for training, validating and testing calibration models*, Sensors and Actuators B 2003, **88(2)**, 149-154.

73 F. Vogt, M. Karlowatz, M. Jakusch, B. Mizaikoff, *The automated sample preparation system MixMaster for investigation of volatile organic compounds with mid-infrared evanescent wave spectroscopy*, Analyst, 2003, **128(4)**, 397-403.

74 F. R. P. Rocha, B. F. Reis, E. A. G. Zagatto, J. L. F. C. Lima, R. A. S. Lapa, J. L. M. Santos, *Multicommunication in flow analysis: concepts, applications and trends*, Analytica Chimica Acta, 2002, **468(1)**, 119-131.

75 R. M. Silverstein, G. C. Bassler, T. C. Morill, *Spectrometric Identification of Organic Compounds*, John Wiley and Sons, New York, NY, 1981.

76 F. Vogt, M. Tacke, M. Jakusch, B. Mizaikoff, *A UV spectroscopic method for monitoring aromatic hydrocarbons dissolved in water*, Analytica Chimica Acta, 2000, **422(2)**, 187-198.

77 F. Vogt, M. Tacke, W. Bohmer, *Selective measurement of aromatic hydrocarbons in water by use of UV derivative spectroscopy*, Spectroscopy Europe, 1999, **11(5)**, 12,14,16,18.

78 E. Huber, M. Frost, *Light scattering by small particles Aqua*, 1988, **47(2)**, 87-94.

79 K. Sturmwohrer, N. Matsche, S. Winkler, *Influence of changes of the wastewater composition on the applicability of UV-absorption measurements at combined sewer overflows Water*, Science and Technology, 2003, **47(2)**, 73-78.

80 G. Langergraber, N. Fleischmann, F. Hofstadter, *A multivariate calibration procedure for UV/VIS spectrometric quantification of organic matter and nitrate in wastewater*, Water Science and Technology, 2003, **47(2)**, 63-71.

81 M. D. DeGrandpre, L. W. Burgess, *Long path fiber-optic sensor for evanescent field absorbance measurements*, Analytical Chemistry, 1988, **60(23)**, 2582-2586.

82 S. A. Mershman, D. C. Tilotta, *Fiber-optic sensor for aromatic compounds in water based on solid-phase microextraction and ultraviolet evanescent wave absorption spectroscopy*, Applied Spectroscopy, 1998, **52(1)**, 106-111.

83 T. Miya, Y. Terunuma, T. Hosaka, T. Miyashita, *Ultimate low-loss single-mode fiber at 1.55 μm*, Electronics Letters 1979, **15(4)**, 106-108.

84 B. Mizaikoff, *Infrared optical sensors for water quality monitoring Water*, Science and Technology, 2003, **47(2)**, 35-42.

85 M. D. DeGrandpre, L. W. Burgess, *A fiber-optic FT-NIR evanescent field absorbance sensor*, Applied Spectroscopy, 1990, **44(2)**, 273-279.

86 J. Buerck, J.-P. Conzen, B. Beckhaus, H.-J. Ache, *Fiber-optic evanescent wave sensor for in-situ determination of non-polar organic compounds in water*, Sensors and Actuators B, 1994, **18(1-3)**, 291-295.

87 E. Sensfelder, J. Buerck, H.-J. Ache, *Determination of hydrocarbons in water by evanescent wave absorption spectroscopy in the near-infrared region*, Fresenius' Journal of Analytical Chemistry, 1996, **354(7-8)**, 848-851.

88 D. S. Blair, L. W. Burgess, A. M. Brodsky, *Evanescent fiber-optic chemical sensor for monitoring volatile organic compounds in water*, Analytical Chemistry, 1997, **69(13)**, 2238-2246.

89 D. S. Blair, J. Bando, *Quantitative monitoring of volatile organic compounds in water using an evanescent fiber optic chemical sensor*, Environmental Science and Technology, 1998, **32(2)**, 294-298.

90 G. L. Klunder, R. E. Russo, *Core-based intrinsic fiber-optic absorption sensor for the detection of volatile organic compounds*, Applied Spectroscopy 1995, **49(3)**, 379-385.

91 J. S. Sanghera, I. D. Aggarwal, *Infrared fiber optics*. Boca Raton: CRC Press, 1998.

92 B. Mizaikoff, B., Lendl, *Handbook of Vibrational Spectroscopy*, J.M. Chalmers, P.R. Griffiths (Editors), John Wiley & Sons, Ltd, 2002,1560-1573.

93 M. Jakusch, B. Mizaikoff, R. Kellner, A. Katzir, *Towards a remote IR fiber-optic sensor system for the determination of chlorinated hydrocarbons in water*, Sensors and Actuators B, 1997, **38(1-3)**, 83-87.

94 F. Regan, B. D. MacCraith, J. E. Walsh, K. O'Dwyer, J. G. Vos, M. Meaney, *Novel Teflon-coated optical fibers for TCE determination using FTIR spectroscopy*, Vibrational Spectroscopy, 1997, **14(2)**, 239-246.

95 H. Steiner, M. Jakusch, M. Kraft, M. Karlowatz, T. Baumann, R. Niessner, W. Konz, A. Brandenburg, K. Michel, C. Boussard-Pledel, B. Bureau, J.Lucas, Y. Reichlin, A. Katzir, N. Fleischmann, K. Staubmann, R. Allabashi, J. M. Bayona, B. Mizaikoff, *In-situ sensing of volatile organic compounds in groundwater: First field tests of a mid-infrared fiber-optic sensing system*, Applied Spectroscopy, 2003, **57(6)**, 607-613.

96 M. Kraft, B. Mizaikoff, *A MID-infrared sensor for monitoring of chlorinated hydrocarbons in the marine environment*, International Journal Environmental Analytical Chemistry, 2000, **78(3-4),** 367–383.

97 M. C. Ertan-Lamontagne, S. R. Lowry, W. R. Seitz, S. A. Tomellini, *Polymer-coated tapered cylindrical ATR elements for sensitive detection of organic solutes in water*, Applied Spectroscopy, 1995, **49(8)**, 1170-1173.

98 R. Howley, B. D. MacCraith, K. Dwyer, H. Masterson, P. Kirwan, P. McLoughlin, *Determination of hydrocarbons using sapphire fibers coated with poly(dimethylsiloxane)*, Applied Spectroscopy, 2003, **57(4)**, 400-406.

99 V. Acha, M. Meurens, H. Naveau, S. N. Agathos, *Detoxification of a mixture of aliphatic chlorinated hydrocarbons in a fixed-bed bioreactor: continuous on-line monitoring via an attenuated total reflection-Fourier transform infrared sensor,* Water Science and Technology, 1999, **40(8)**, 41-47.

100 B. Mizaikoff, *Mid-infrared evanescent wave sensors – a novel approach for subsea monitoring*, Measurement Science and Technology, 1999, **10(12)**, 1185-1194.

101 M. Kraft, M.Jakusch, M. Karlowatz, A. Katzir, B. Mizaikoff, *New frontiers for mid-infrared sensors: Towards deep sea monitoring with a submarine FT-IR sensor system*, Applied Spectroscopy, 2003, **57(6)**, 591-599.

102 T. Beyer, P. Hahn, S. Hartwig, W. Konz, S. Scharring, A. Katzir, H. Steiner, M. Jakusch, M. Kraft, B. Mizaikoff, *Mini spectrometer with silver halide sensor fiber for in-situ detection of chlorinated hydrocarbons*, Sensors and Actuators B, 2003, **90(1-3)**, 319-323.

103 D. L. Heglund, D. C. Tilotta, *Determination of volatile organic compounds in water by solid phase microextraction and infrared spectroscopy*, Environmental Science and Technology, 1996, **30(4)**, 1212-1219.

104 R. Krska, R. Kellner, U. Schiessl, M. Tacke, A. Katzir, *Fiber optic sensor for chlorinated hydrocarbons in water based on infrared fibers and tunable diode lasers*, Applied Physics Letters, 1993, **63(14)**, 1868-1870.

105 J. F. Kastner, M. Tacke, A. Katzir, B. Edl-Mizaikoff, R. Goebel, R. Kellner, *Optimizing the modulation for evanescent-wave analysis with laser diodes (EWALD) for monitoring chlorinated hydrocarbons in water*, Sensors and Actuators, B: Chemical, 1997, **38(1-3)**, 163-170.

106 R. K. Chang, T. E. Furtak (Editors), *Surface-enhanced Raman scattering*, Plenum New York, 1982.

107 J. Corset, J. Aubard, *Special Issue--Surface Enhanced Raman Scattering: New Trends and Applications,* Journal of Raman Spectroscopy, 1998, 29(8), 1-113.

108 S. P. Mulvaney, C. D. Keating, *Raman Spectroscopy*, Analytical Chemistry, 2000, **72(12)**, 145-157. (title missing)

109 D. L. Jeanmaire, R. P. Van Duyne, *Surface Raman spectroelectrochemistry. Part I. Heterocyclic, aromatic, and aliphatic amines adsorbed on the anodized silver electrode*, Journal of Electroanalytical Chemistry and Interfacial Electrochemistry, 1977, **84(1)**, 1-20.

110 T. L. Williams, T. W. Collette, *Environmental applications of Raman spectroscopy to aqueous systems*, Practical Spectroscopy, 2001, **28**, 683-731.

111 T. W. Collette, T. L. Williams, *The role of Raman spectroscopy in the analytical chemistry of potable water*, Journal of Environmental Monitoring, 2002, **4(1)**, 27-34.

112 G. E. Walrafen, J. Stone, *Intensification of spontaneous Raman spectra by use of liquid core optical fibers*, Applied Spectroscopy, 1972, **26(6)**, 585-589.

113 R. Altkorn, E. Koev, A. Gottlieb, *Waveguide capillary cell for low-refractive-index liquids*, Applied Spectroscopy, 1997, **51(10)**, 1554-1558.

114 M. Holtz, P. K. Dasgupta, G. Zhang, *Small-volume Raman spectroscopy with a liquid core waveguide*, Analytical Chemistry, 1999, **71(14)**, 2934-2938.

115 C. M. Stellman, K. J. Ewing, F. Bucholtz, I. D. Aggarwal, *Determination of BTEX contaminants in water via a long-pathlength fiber-optic Raman 'dip-stick'*, Sensors and Actuators B, 1998, **53(3)**, 173-178.

116 S. Tanikkul, J. Jakmunee, M. Rayanakorn, K. Grudpan, B. J. Marquardt, G. M. Gross, B. J. Prazen, L. W. Burgess, G. D. Christian, R. E. Synovec, *Characterization and use of a Raman liquid-core waveguide sensor using preconcentration principles*, Talanta, 2003, **59(4)**, 809-816.

117 K. Carron, L. Peitersen, M. Lewis. *Octadecylthiol-modified surface-enhanced Raman spectroscopy substrates: a new method for the detection of aromatic compounds*, Environmental Science and Technology, 1992, **26(10)**, 1950-1954.

118 K. Mullen, K. Carron, *Adsorption of Chlorinated Ethylenes at 1-Octadecanethiol-Modified Silver Surfaces*, Analytical Chemistry, 1994, **66(4)**, 478-483.

119 B. J. Kennedy, R. Milofsky, K. T. Carron, *Development of a cascade flow cell for dynamic aqueous phase detection using modified SERS substrates*, Analytical Chemistry, 1997, **69(22)**,4708-4715.

120 T. Murphy, H. Schmidt, H.-D. Kronfeldt, *Use of sol-gel techniques in the development of surface-enhanced Raman scattering (SERS) substrates suitable for in-situ detection of chemicals in seawater*, Applied Physics B: Lasers and Optics, 1999, **69(2)**, 147-150.

121 T. Murphy, S. Lucht, H. Schmidt, H.-D. Kronfeldt, *Surface-enhanced Raman scattering (SERS) system for continuous measurements of chemicals in sea-water*, Journal of Raman Spectroscopy, 2000, **31(10)**, 943-948.

122 S. Lucht, T. Murphy, H. Schmidt, H.-D. Kronfeldt, *Optimized recipe for sol-gel-based SERS substrates*, Journal of Raman Spectroscopy, 2000, **31(11)**, 1017-1022.

123 J. H. Richardson, M. E. Ando, *Sub-part-per-trillion detection of polycyclic aromatic hydrocarbons by laser induced molecular fluorescence*, Analytical Chemistry 1977, 49(7), 955-959.

124 S. H. Lieberman, *Direct-push, fluorescene-based sensor systems for in-situ measurement of petroleum hydrocarbons in soils Field*, Analytical Chemistry and Technology, 1998, **2(2)**, 63-73.

125 J. J. S. Rodriguez, C. P. Sanz, *Fluorescence techniques for the determination of polycyclic aromatic hydrocarbons in marine environment: an overview*, Analusis, 2000, **28(8)**, 710-717.

126 D. Patra, *Applications and new developments in fluorescence spectroscopic techniques for the analysis of polycyclic aromatic hydrocarbons*, Applied Spectroscopy Reviews, 2003, **38(2)**, 155-185.

127 D. A. Cremers, L. J. Radziemski, *Laser Plasmas for Chemical Analysis, in Laser Spectroscopy and Its Applications*, L. J. Radziemski, R. W. Solarz, J. A. Paisner (Editors) Marcel Dekker, New York, NY, 1987.

128 J. B. Jeffries, G. A. Raiche, L. E. Jusinski, *Detection of chlorinated hydrocarbons via laser-atomization/laser-induced fluorescence*, Applied Physics B, 1992, **55(1)**, 76-83.

129 L. M. Berman, P. J. Wolf, *Laser-induced breakdown spectroscopy of liquids: aqueous solutions of nickel and chlorinated hydrocarbons*, Applied Spectroscopy 1998, **52(3)**, 438-443.

130 G. C. Guilbault, *Practical Fluorescence*, Marcel Dekker, New York, NY,1973.

131 J. R. Lakowicz (Editor), *Principles of Fluorescence Spectroscopy*, Plenum Pr (2nd edition), 1999.

132 P. Karlitschek, F. Lewitzka, U. Buenting, M. Niederkrueger, G. Marowsky, *Detection of aromatic pollutants in the environment by using UV-laser-induced fluorescence*, Applied Physics B: Lasers and Optics, 1998, **67(4)**, 497-504.

133 J. Bloch, B. Johnson, N. Newbury, J. Germaine, H. Hemond, J. Sinfield, *Field test of a novel microlaser-based probe for in-situ fluorescence sensing of soil contamination*, Applied Spectroscopy, 1998, **52(10)**, 1299-1304.

134 J. Wu Pepper, A. O. Wright, J. E. Kenny, *In-situ measurements of subsurface contaminants with a multi- channel laser-induced fluorescence system*, Spectrochimica Acta, Part A: Molecular and Biomolecular Spectroscopy, 2002, **58(2)**, 317-331.

135 N. J. Harrick, *Internal reflection spectroscopy*, Wiley, New York, 1967.

136 J. Fahrenfort, *Attenuated total reflection. A new principle for the production of useful infrared reflection spectra of organic compounds*, Spectrochim. Acta, 1961, **17**, 698-709.

137 A. W. Snyder, J.D. Love, *Optical Waveguide Theory*, Chapman and Hall, 1983.

138 M. Kraft, *Mid-infrared spectroscopy using optical waveguides in the marine environment*, Dissertation, Vienna University of Technology, Austria, 2000.

139 F. Regan, Meaney, M.; Vos, J. G.; MacCraith, B. D.; Walsh, J. E., *Determination of pesticides in water using ATR-FTIR spectroscopy on PVC/chloroparaffin coatings*, Anal. Chim. Acta, 1996, **334(1-2)**, 85-92.

140 M. Kraft, M. Karlowatz, B. Mizaikoff, R. Stück, M. Steden, M. Ulex, H. Amann, *Sensor head development for mid-infrared fiber-optic underwater sensors,* Measurement Science and Technology, 2002, **13(8)**, 1294-1303.

141 J. Yang, J.-W. Her, *Gas-Assisted IR-ATR Probe for Detection of Volatile Compounds in Aqueous Solutions*, Analytical Chemistry, 1999, **71(9)**, 1773–1779.

142 J. Yang, S.-S. Tsai, *Cooled internal reflection element for infrared chemical sensing of volatile to semi-volatile organic compounds in the headspace of aqueous solutions*, Analytica Chimica Acta, 2002, **462(2)**, 235–244.

143 F. O. Libnau, O. M. Kvalheim, A. A. Christy, J. Toft, *Spectra of water in the near- and mid-infrared region*, Vibrational Spectroscopy, 1994, **7(3)**, 243-254.

144 Z. Zhang M. J. Yang, J. Pawliszyn, *Solid-Phase Microextraction. A Solvent-Free Alternative for Sample Preparation*, Analytical Chemistry, 1994, **66(17)**, 844A-852A.

145 M. Jakusch, B. Mizaikoff, *Selective polymer materials: absolute determination of their sorption properties*, Proc. SPIE Int. Soc. Opt. Eng., 2001, **4205**(Advanced Environmental and Chemical Sensing Technology), 93-98.

146 X. Zhou, M. A. Arnold, *Response Characteristics and Mathematical Modeling for a Nitric Oxide Fiber-Optic Chemical Sensor*, Analytical Chemistry, 1996, **68(10)**, 1748-1754.

147 F. Meriaudeau, T. Downey , A. Wig , A. Passian, M. Buncick, T.L. Ferrell, *Fiber optic sensor based on gold island plasmon resonance*, Sensors and Actuators B, 1999 , **54**, 106 – 117.

148 Y. Chen, T.-C. Tan, *Modeling and experimental study of the transient behavior of plant tissue sensors in sensing dopamine*, Chemical Engineering Science, 1996, **51(7)**, 1027-1042.

149 S. Astié, A. M. Gué, E. Scheid, L. Lescouzères, A. Cassagnes, *Optimization of an integrated SnO_2 gas sensor using a FEM simulator*, Sensors and Actuators A,1998, **69(3)**, 205-211.

150 T. Erbey-Gruz, *Transport Phenomena in Aqueous Solutions*, Wiley: New York, 1974.

151 J. Brandup, E. H. Immergut, *Polymer Handbook*, 3rd ed.; John Wiley & Sons: New York, 1989.

152 V. G. Levich, *Physicochemical Hydrodynamics*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1962.

153 R. Krska, K. Taga, R. Kellner, A. Messica, A. Katzir, *A new fiber optic physico-chemical sensor for chlorinated hydrocarbons in water*, Fresenius Journal of Analytical Chemistry, 1992, **342(1-2)**, .

154 R. Krska, K. Taga, R. Kellner, *Simultaneous in-situ trace analysis of several chlorinated hydrocarbons in water with an IR fiber optical system*, Journal of Molecular Structure, 1993, **294**, 1–4.

155 H. H. Nijhuis, M.H.V. Mulder, C. A. Smolders, *Removal of trace organics from aqueous solutions. Effect of membrane thickness*, Journal of Membrane Science, 1991, **61**, 99–111.

156 R. Psaume, P. Aptel, Y. Aurelle, J. C. Mora, J. L. Bersillon, *Pervaporation: importance of concentration polarization in the extraction of trace organics from water*, Journal of Membrane Science, 1988, **36**, 373–384.

157 B. Raghunath, S. T. Hwang, *Effect of boundary layer mass transfer resistance in the pervaporation of dilute organics*, Journal of Membrane Science, 1992, **65**, 147–61.

158 G. Roy, J. A. Mielczarski, *Infrared detection of chlorinated hydrocarbons in water at ppb levels of concentration*, Water Research, 2002, **36**, 1902-1908.

159 D. Louch, S. Motlagh, J. Pawliszyn, *Dynamics of Organic Compound Extraction from Water Using Liquid-Coated Fused Silica Fibers*, Analytical Chemistry, 1992, **64(10)**, 1187-1199.

160 M. Jakusch, *Polymer coated infrared evanescent wave sensors - approaches for their systematic development*, Dissertation, Vienna University of Technology, Austria, 2000.

161 S. Patankar, *Numerical Heat Transfer and Fluid Flow*, Hemisphere, Washington,DC, 1980.

162 J. Ruzicka, G. D. Marshall, *Sequential injection: a new concept for chemical sensors, process analysis and laboratory assays,* Analytica Chimica Acta, 1990 **237(2)**, 329-343.

163 M. Valcárcel, M.D. Luque de Castro, *Flow Injection Analysis, principles and applications*, Ellis Horwood, Chicester, 1987.

164 J. Ruzicka, *The second coming of flow-injection analysis*, Analytica Chimica Acta, 1992, **261(1-2)**, 3-10.

165 R. Goebel, R. Krska, R. Kellner, R. Seitz, S. Tomellini, *Investigation of different polymers as coating materials for IR/ATR spectroscopic trace analysis for chlorinated hydrocarbons in water*, Applied Spectroscopy, 1994, **48(6)**, 678–683.

166 M. Jakusch, B. Mizaikoff, R. Kellner, *Optimized Sensitive Coatings for MIR Fiber Optic Sensors*, Proc. SPIE Vol. 1997, **3105**, 283-288.

167 A. Mohacsi, Z. Bozoki, R. Niessner, Direct diffusion sampling-based photoacoustic cell for in-situ and on-line monitoring of benzene and toluene concentrations in water, Sensors and Actuators B, 2001, **79(2-3)**, 127-131.

168 D. Bunimovich, A. Katzir, *Dielectric properties of silver halide and potassium halide crystals*, Applied Optics, 1993, **32(12)**, 2045-2048.

169 S. Shalem, A. German, N. Barkay, F. Moser, A. Katzir, *Fiber and Integrated Optics*, 1997, **16(1)**, 27-54.

170 M. Maimone, Coping with groundwater contamination, Chemical Engineering, 1997, **104(10)**, 195-202.

171 E. Müller-Erlwein, *Chemische Reaktionstechnik*, B.G. Teubner (Editor), Stuttgart, 1998.

172 L. Han, T. M. Niemczyk, Y. Lu, G. P. Lopez, Chemical sensors based on surface-modified sol-gel-coated infrared waveguides, Applied Spectroscopy, 1998, **52(1)**, 119-122.

173 L. Han, T. M. Niemczyk, D. M. Haaland, G. P. Lopez, Enhancing IR detection limits for trace polar organics in aqueous solutions with surface-modified sol-gel-coated ATR sensors, Applied Spectroscopy, 1999, **53(4)**, 381-389.

174 M. Janotta, M. Karlowatz, F. Vogt, B. Mizaikoff, Sol-gel based mid-infrared evanescent wave sensors for detection of organophosphate pesticides in aqueous solution, Analytica Chimica Acta, 2003, **496(1-2)**, 339-348.

175 M. Janotta, A. Katzir, B. Mizaikoff, Sol-gel-coated mid-infrared fiber-optic sensors**,** Applied Spectroscopy, 2003, **57(7)**, 823-828.

176 C. Vogt, A. Alfreider, H. Lorbeer, J. Ahlheim, B. Feist, O. Boehme, H. Weiss, W. Babel, L. Wuensche, *Two pilot plant reactors designed for the in situ bioremediation of chlorobenzene-contaminated ground water: hydrogeological and chemical characteristics and bacterial consortia*, Water, Air, & Soil Pollution: Focus, 2002, **2(3)**, 161-170.

177 P. Wycisk, H. Weiss, A. Kaschl, S. Heidrich, K. Sommerwerk, *Groundwater pollution and remediation options for multi-source contaminated aquifers (Bitterfeld/Wolfen, Germany),* Toxicology Letters, 2003, **140-141**, 343-351.

178 H. Peter, J. Grossmann, G. Schulz-Terfloth, *Rahmensanierungskonzept des Großprojektes Bitterfeld/Wolfen*, H.-P. Luehr (Editor), Grundwassersanierung 1995, IWS Schriftenreihe 23, Erich Schmidt-Verlag, Berlin, Germany, 1995, pp. 123–128.

179 P. Merkel, G. Teutsch, H. Weiss, H. H. M. Rijnaarts, *Innovative Reactive Barrier Technologies for Regional Contaminated Groundwater*, Proceedings of the Seventh International FZK/TNO Conference on Contaminated Soil (ConSoil 2000), Vol. 1, Thomas Telford publishing, London, 2000, pp. 532–540.

180 R. Goebel, R. W. Seitz, S. A. Tomellini, R. Krska, R. Kellner, *Infrared attenuated total reflection spectroscopic investigations of the diffusion behaviour of chlorinated hydrocarbons into polymer membranes*, Vibrational Spectroscopy, 1995, **8**, 141-149.

181 E. Rosenberg, R. Kellner, *IR-ATR spectroscopic determination of the diffusion coefficients of carbohydrates in polymer matrixes*, Proc.SPIE,1992, vol. 1575, 346-347.

182 J. Comyn, *Polymer Permeability*, Elsevier Applied Science, New York, 1985.

183 J. M. Vanguard, *Liquid Transport Processes in Polymeric Materials: Modeling and Industrial Applications*, Prentice-Hall, Engelwood Cliffs, NJ, 1991.

184 U. Mukai, K. K. Gleason, A. S. Argon, R. E. Cohen*, Poly(Dimethylsiloxane)/Nylon-6 Block-Copolymers - Molecular Mobility At the Interface*. Macromolecules 1995, **28(14)**, 4899-4903.

185 G. T. Fieldson, T. A. Barbari, *The use of attenuated total reflection FTIR spectroscopy to characterize penetrant diffusion in polymers*, Polymer, 1993, **34(6)** 1146-1153.

186 C. M. Balik, W. H. Siemendinger, *An attenuated total reflectance cell for analysis of small moclecule diffusion in polymer thin films with Fourier-transform infrareed spectroscopy*, 1998, **39(20)**, 4723-4728.

187 P. Peterkova, L'. Lapcik Jr*., Determination of the diffusion coefficient of water into atelocollagen type I thin films by attenuated total reflection Fourier transform infared spectroscopy*, Colloid Polymer Science, 2000, **278(10)**, 1014-1016.

188 R. Howley, B.D. MacCraith, K. O'Dwyer, P. Kirwan, P. McLoughlin, *A study of the factors affecting the diffusion of chlorinated hydrocarbons into polyisobutylene and*

polyethylene-co-propylene for evanescent wave sensing, Vibrational Spectroscopy 2003, **31(2)**, 271-278.

189 B. Murphy, P. Kirwan, P. McLoughlin, *Study of the impact of penetrant characteristics upon diffusion into Teflon membranes to further assess the performance of an ATR/FTIR sensor*, Analytical and Bioanalytical Chemistry, 2003, **377(1)**, 195-202.

190 Y. A. Elabd, M. G. Baschetti, T. A. Barbari, *Time-Resolved Fourier Transform Infrared/Attenuatetd Total Reflection Spectroscopy for the Measurement of Molecular Diffusion in Polymers*, Journal of Polymer Science Part B, 2003, **41(22)**, 2794-2807.

191 C. Phillips, *Model-based design optimization of heterogeneous micro-reactors and chemical*, MS Thesis Georgia Institute of Technology, 2003.

192 M. Janotta, F. Vogt, H.-S. Voraberger, W. Waldhauser, J.-M. Lackner, C. Stotter, M. Beutl, B. Mizaikoff, *Direct Analysis of Oxidizing Agents in Aqueous Solution with Attenuated Total Reflectance Mid-Infrared Spectroscopy and Diamond-like Carbon Protected Waveguides***,** Analytical Chemistry, 2004, **76(2)**, 384-391.

193 Permanent website: The United Nations Mine Action Services, http://www.un.org/Depts/dpko/mine.

194 U.S. Department of State, *Hidden killers 1998: The global landmine crisis*, Bureau, of Political–Military Affairs, Office of Humanitarian Demining Programs. (permanent website): http://www.state.gov/www/global/arms/rpt_9809_demine_toc.html, 1998.

195 J. Paik, C. P. Lee, M. A. Abidi, *Image processing-based mine detection techniques: A review*, Subsurface Sensing Technologies and Applications, 2002, **3(3)**, 153-202.

196 P. Machler, *Detection technologies for anti-personnel mines*, Proceedings Symposium on Autonomous Vehicles in Mine Countermeasures, 1995, **6**, 150–154.

197 J. Jr. Stepanic, H. Wustenberg, V. Krstelj, H. Mrasek, *Contribution to classification of buried objects based on acoustic impedance matching*, Ultrasonics, 2003, **41(2)**, 115-123.

198 L. Kempen, A. Katarzin, Y, Pizurion, C. Corneli, H. Sahli, *Digital signal image processing for mine detection, Part 2: Ground based approach*, Proceedings Euro Conference on Sensor Systems and Signal Processing Techniques applied to the Detection of Mines and Unexploded Ordnance, 1999, 54–59.

199 K. Russell, J. McFee, W. Sirovyak, *Remote performance prediction for infrared imaging of buried mines*, Proceedings SPIE Detection and Remediation Technologies for Mines and Minelike Targets II, 1997, **3079**, 762–769.

200 J. Stepanic, M. Malinovec, S. Svaic, V. Krstelj, *Parameterisation of non-homogeneities in buried object detection by means of thermography*, Infrared Physics & Technology, 2004, **45(3)**, 201-208.

201 M. Lundberg, I.Y.H. Gu, *A 3-D matched filter for detection of land mines using spatio-temporal thermal modelling*, Detection and Remediation Technologies for Mines and Minelike Targets V, SPIE Proceedings, 2000, **4038**, 179–188.

202 I. Boras, M. Malinovec, J. Stepanic, S. Svaic, *Modelling of buried object detection using thermography*, QIRT 2000, Reims, Proceedings, 2002, 176–181.

203 UN Mine Action Service, International Mine Action Standards, Chapter 9–Clearance requirements. Permament website: http://www.mineactionstandards.org/IMAS_archive/Final/09.10.pdf.

204 P. Pregowski, W. Swiderski, R. Walczak, *Surface and volume effects in thermal signatures of buried mines: experiment and modeling*, QIRT 1998, Lodz, Proceedings, 1998, 233–238.

205 A. F. H. Goetz, G. Vane, J. Solomon, B. N. Rock, *Imaging Spectrometry for Earth Remote Sensing*, Science, 1985*,* **228,** 1147–1153.

206 A. F. H. Goetz, "Imaging Spectrometry for Remote Sensing: Vision to Reality in 15 Years" Proceedings SPIE Aerosense Symposium, 1995, **2480**, 2–13.

207 A. F. H. Goetz, B. Curtiss, *Hyperspectral Imaging of the Earth: Remote Analytical Chemistry in an Uncontrolled Environment,* Field Analytical Chemistry and Technology, 1996, **1(2)**, 67-76.

208 J. W. Boardman, *Automated Spectral Unmixing of AVIRIS Data Using Convex Geometry Concepts*, Summaries of the Fourth Annual JPL Airborne Geoscience Workshop, JPL Pub. 1993, **93**, 53–56.

209 G. Vane, R. O. Green, T. G. Chrien, H. T. Enmark, E. G. Hansen, and W. M. Porter, *The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS*), Remote Sensing of Environment, 1993, **44**, 127–144.

210 R. A. Schowengerdt, *Remote Sensing* Academic Press, 2nd edition, San Diego, USA, 1997.

211 W. L. Wolfe, *Introduction to Imaging Spectrometers,* D. C. O'Shea (Editors), SPIE, Bellingham, Washington USA, 1997.

212 T. M. Lillesand, R. W. Kiefer, J. W. Chipman, *Remote Sensing and Image Interpretation,* Wiley, New York, USA, 2003.

213 Permanent website: http://rst.gsfc.nasa.gov/Homepage/Homepage.html.

214 Permanent website:
http://www.ccrs.nrcan.gc.ca/ccrs/learn/tutorials/fundam/fundam_e.html.

215 M. M. Elcombe, *Some aspects of the lattice dynamics of quartz*, Proceedings of the Physics Society of London, 1967, **91**, 947-958.

216 L. S. Walter, J. W. Salisbury, *Spectral characterization of igneous rocks in the 8 to 12 micrometers region,* Journal of Geophysical Research, 1989, **94**, 9203-9213.

217 V. C. Farmer (Editor), *The Infrared Spectra of Minerals* Monograph No. 4, Mineralogical Society, London, UK, 1974.

218 Q. Williams, *Infrared, Raman and Optical Spectroscopy of Earth Materials,* T. J. Ahrens (Editor), *Mineral physics and crystallography : a handbook of physical constants*, AGU, Washington, DC, USA, 1995.

219 D. B. Nash, J. W. Salisbury, *Infrared reflectance spectra of plagioclase feldspars*: Geophysical Research Letters, 1991, **18**, 1151-1154.

220 J. W. Salisbury, D. M. D'Aria, E. J. Jarosewich, *Midinfrared (2.5-13.5 micrometers) reflectance spectra of powdered stony meteorites,* Icarus, 1991, **92**, 280-297.

221 J. W. Salisbury, L. S. Walter, N. Vergo, Dana M. D'Aria, *Infrared (2.1 – 25 micrometers) Spectra of Minerals*, Johns Hopkins University Press, Baltimore, MD , USA, 1991.

222 F. Seitz, D. Turnbull (Editors), *Solid State Physics*, Academic Press, New York, 1966, **18**, 274–420. (see also, 1966, **19**, 1-134).

223 R. J. Elliott, J. A. Krumhansl, P. L. Leath, *Theory and properties of randomly disordered crystals and related physical systems*, Reviews of Modern Physics, 1974, **46(3)**, 465-543.

224 F. L. Galeener, G. Lucovsky, *Longitudinal optical vibrations in glasses: germania and silica*, Physical Review Letters, 1976, **37(22)**, 1474-1478.

225 F. L. Galeener, *Band limits and the vibrational spectra of tetrahedral glasses*, Physical Review B, 1979, **19(8)**, 4292-4297.

226 P. H. Gaskell, D. W. Johnson, *The optical constants of quartz, vitreous silica and neutron-irradiated vitreous silica (I)*, Journal of Non-Crystalline Solids, 1976, **20(2)**, 153-169.

227 P. H. Gaskell, D. W. Johnson, *The optical constants of quartz, vitreous silica and neutron-irradiated vitreous silica. (II). Analysis of the infrared spectrum of vitreous silica*, Journal of Non-Crystalline Solids, 1976, **20(2)**, 171-191.

228 J. C. Phillips, *Spectroscopic and morphological structure of tetrahedral oxide glasses*, Solid State Physics, 1982, **37**, 93-171.

229 S. W. De Leeuw, M. F. Thorpe, *Coulomb splittings in glasses*, Physical Review Letters, 1985, **55(26)**, 2879-2882.

230 F. Ruiz, J. R. Martinez, J. Gonzalez-Hernandez, *A simple model to analyze vibrationally decoupled modes on SiO2 glasses,* Journal of Molecular Structure, 2002, **641(2-3)**, 243-250.

231 I. W. Boyd, *Deconvolution of the infrared absorption peak of the vibrational stretching mode of silicon dioxide: evidence for structural order?*, Applied Physics Letters, 1987, **51(6)**, 418-420.

232 D. W. Berreman, *Infrared Absorption at Longitudinal Optic Frequency in Cubic Crystal Films*, Physical Review, 1963, **130(6)**, 2193–2198.

233 J. E. Olsen, F. Shimura, *Infrared reflection spectroscopy of the silica-silicon interface*, Journal of Applied Physics, 1989, **66(3)**, 1353-1358.

234 I. W. Boyd, J. I. B. Wilson, *A study of thin silicon dioxide films using infrared absorption techniques,* Journal of Applied Physics, 1982, **53(6)**, 4166-4172.

235 C. T. Kirk, *Quantitative analysis of the effect of disorder-induced mode coupling on the infrared absorption in siclica*, Physical Review B, 1988, **28(2)**, 1255-1273.

236 O. E. Piro, E. E. Castellano, S. R. Gonzalez, *Attenuated total-reflectance spectra of strongly absorbing anisotropic single cystals: Trigonal α-quartz*, Physical Review B, 1988, **38(12)**, 8437–8443.

237 R. M. Almeida, C. G. Pantano, *Structural investigation of silica gel films by infrared spectroscopy*, Journal of Applied Physics, 1990, **68(8)**, 4225-4232.

238 H. R. Philips, *The infrared optical properties of SiO2 and SiO2 layers on silicon*, Journal of Applied Physics, 1979, **50(2)**, 1053-1057.

239 M. C. Payne, J. C. Inkson, *Longitudinal-optic-transverse-optic vibrational mode splittings in tetrahedral network glasses*, Journal of Non-Crystalline Solids, 1984, **68(2-3)**, 351-360.

240 A. Lehmann, L. Schumann, K. Huebner, *Optical phonons in amorphous silicon oxides. II. Calculation of phonon spectra and interpretation of the IR transmission of silicon oxide (SiOx)*, Physica Status Solidi B: Basic Research, 1984, **121(2)**, 505-511.

241 J. Wang, B. Zou, M. El-Sayed, *A Comparison between the polarized Fourier-transform infrared spectra of aged porous silicon and amorphous silicon dioxide films on Si (100) surface*, Journal of Molecular Structure, 1999, **508(1-3)**, 87-96.

242 J. L. Gole, F. P. Dudel, D. Grantier, D. A. Dixon, *Origin of porous silicon photoluminescence: Evidence for a surface bound oxyhydride-like emitter*, Physical Review B: Condensed Matter, 1997, **56(4)**, 2137-2153.

243 R. M. Almeida, *Detection of LO modes in glass by infrared reflection spectroscopy at oblique incidence*, Physical Review B, 1992, **45(1)**, 161-170.

244 B. Harbecke, B. Heinz, P.Grosse, *Optical Properties of Thin Films and the Berreman Effect*, Applied Physics A, 1985, **38**, 263-267.

245 K. Huebner, L. Schumann, A. Lehmann, H. H. Vajen, G. Zuther, *Detection of LO and TO phonons in amorphous silicon dioxide films by oblique incidence of IR light*, Physica Status Solidi B: Basic Research, 1981, **104(1)**, K1-K5.

246 R. M. Almeida, T. A. Guiton, C. G. Pantano, *Detection of longitudinal optical mode in vitreous silica by infrared diffuse reflectance spectroscopy*, Journal of Non-Crystalline Solids, 1990, **119(2)**, 238-241.

247 P. Innocenzi, P. Falcaro, D. Grosso, F. Babonneau, *Order-Disorder Transitions and Evolution of Silica Structure in Self-Assembled Mesostructured Silica Films Studied through FTIR Spectroscopy*, Journal of Physical Chemistry B, 2003, **107(20)**, 4711-4717.

248 J. A. Moreno, B. Garrido, J. Samitier, J. R. Morante, *Analysis of geometrical effects on the behavior of transverse and longitudinal modes of amorphous silicon compounds,* Journal of Applied Physics, 1997, **81(4)**, 1933-1942.

249 C. Z. Tan, J. Arndt, *Interaction of longitudinal and transverse optic modes in silica glass*, Journal of Chemical Physics, 2000, **112(13)**, 5970-5974.

250 E. Courtens, M. Foret, B. Hehlen, R. Vacher, *The vibrational modes of glasses*, Solid State Communications, 2001, **117**, 187-200.

251 N. Rochat, A. Chabli, F. Bertin, C. Vergnaud, P. Mur, S. Petitdidier, P. Besson; *Infrared analysis of thin layers by attenuated total reflection spectroscopy*, Material Science Eng., 2003, **B102**, 16–21.

252 I. I. Shaganov, T. S. Perova, R. A. Moore, K. Berwick, *Spectroscopic characteristics of SiO and SiO_2 solid films: Assignment and local field effect influence*, Journal of Materials Science: Materials in Electronics, 2001, **12**, 351-355.

253 J. Sarnthein, A. Pasquarello, R. Car, *Origin of the high-frequency doublet in the vibrational spectrum of vitreous SiO_2* , Science, 1997, **275 (5308),** 1925-1928.

254 B. C. Trasferetti, C. U. Davanzo, *s- and p-polarized infrared specular reflectance of vitreous silica at oblique incidences: detection of LO modes*. Applied Spectroscopy, 2000, **54(4)**, 502-507.

255 R. J. Bell, P. Dean, *Properties of vitreous silica. Analysis of random network models*, Nature, 1966, **212(5068)**, 1354-1356.

256 *The key problem for calculating individual* ed. M. Garbuny, Optical Physics, Academic Press, NY, 1965

257 A. M. Efimov, *Vibrational spectra, related properties, and structure of inorganic glasses*, Journal of Non-Crystalline Solids, 1999, **253**, 95-118.

258 A. M. Efimov, *Quantitative IR spectroscopy Applications to studying glass structure and properties,* Journal of Non-Crystalline Solids, 1996, **203**, 1-11.

259 T. G. Mayerhoefer, H. H. Dunken, R. Keding, C. Ruessel, *Interpretation and modeling of IR-reflectance spectra of glasses considering medium range order*, Journal of Non-Crystalline Solids, 2004, **333**, 172-181.

260 J. W. Salisbury, A. Wald, D. M. D'Aria, *Thermal infrared remote sensing and Kirchhoff's law 1. Laboratory measurements*, Journal Geophysical Research,1994, **99(11)**, 897–911.

261 S. M.Abo-Naf, F. H. El Batal, M. A. Azooz, *Characterization of some glasses in the system SiO_2-Na_2O-RO by infrared spectroscopy*, Materials Chemistry and Physics, 2003, **77(3)**, 846-852.

262 C. I. Merzbacher, W. B. White, *The structure of alkaline earth aluminosilicate glasses as determined by vibrational spectroscopy*, Journal of Non-Crystalline Solids, 1991, **130(1)**, 18-34.

263 A. M. Efimov, *IR fundamental spectra and structure of pyrophosphate glasses along the 2ZnO-P_2O_5-2Me_2O-P_2O_5 join (Me being Na and Li), Journal of Non-Crystalline Solids*, 1997, **209(3)**, 209-226.

264 T. Yoshidome, H. Kusumoto, O. Kuroki, S. Kamata, *Application of the attenuated-total-reflection technique to the measurement of silica-gel particle size*, Chemistry Letters, 1998, **(8)**, 747-748.

265 D. Sanders, W. B. Person, L. L. Hench, *Quantitatiive analysis of glass structure with the use of infrared reflection spectra* Applied Spectroscopy, 1974, **28(3)**, 247-255

266 W. J. Dell, P. J. Bray, S. Z. Xiao, *Boron-11 NMR studies and structural modeling of sodium borosilicate (Na2O-B2O3-SiO2) glasses of high soda content,* Journal of Non-Crystalline Solids, 1983, **58(1)**, 1-16.

267 J. Serra, P. Gonzalez, S. Liste, C. Serra, S. Chiussi, B. Leon, M. Perez-Amor, H. O. Ylanen, M. Hupa, *FTIR and XPS studies of bioactive silicate-based glasses,* Journal of Non-Crystalline Solids 2003, **332(1-3)**, 20-27.

ABSTRACT

Title of Document: PHYSICS-BASED DETECTION OF SUBPIXEL TARGETS IN HYPERSPECTRAL IMAGERY

Joshua Bret Broadwater
Doctor of Philosophy, 2007

Directed By: Professor Ramalingam Chellappa
Department of Electrical and Computer Engineering

Hyperspectral imagery provides the ability to detect targets that are smaller than the size of a pixel. They provide this ability by measuring the reflection and absorption of light at different wavelengths creating a spectral signature for each pixel in the image. This spectral signature contains information about the different materials within the pixel; therefore, the challenge in subpixel target detection lies in separating the target's spectral signature from competing background signatures. Most research has approached this problem in a purely statistical manner. Our approach fuses statistical signal processing techniques with the physics of reflectance spectroscopy and radiative transfer theory. Using this approach, we provide novel algorithms for all aspects of subpixel detection from parameter estimation to threshold determination.

Characterization of the target and background spectral signatures is a key part of subpixel detection. We develop an algorithm to generate target signatures based on radiative transfer theory using only the image and a reference signature without the need for calibration, weather information, or source-target-receiver geometries. For background signatures, our work identifies that even slight estimation errors in the number of background signatures can severely degrade detection performance. To this end, we present a new method to estimate the number of background signatures specifically for subpixel target detection.

At the core of the dissertation is the development of two hybrid detectors which fuse spectroscopy with statistical hypothesis testing. Our results show that the hybrid detectors provide improved performance in three different ways: insensitivity to the number of background signatures, improved detection performance, and consistent performance across multiple images leading to improved receiver operating characteristic curves.

Lastly, we present a novel adaptive threshold estimate via extreme value theory. The method can be used on any detector type – not just those that are constant false alarm rate (CFAR) detectors. Even on CFAR detectors our proposed method can estimate thresholds that are better than theoretical predictions due to the inherent mismatch between the CFAR model assumptions and real data. Additionally, our method works in the presence of target detections while still estimating an accurate threshold for a desired false alarm rate.

PHYSICS-BASED DETECTION OF SUBPIXEL TARGETS IN HYPERSPECTRAL IMAGERY

By

Joshua Bret Broadwater

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2007

Advisory Committee:
Professor Ramalingam Chellappa, Chair
Professor Eyad Abed
Professor Larry S. Davis
Professor Adrian Papamarcou
Professor Min Wu

© Copyright by
Joshua Bret Broadwater
2007

Preface

The data used in this dissertation comes from the RDECOM CERDEC Night Vision & Electronic Sensors Directorate (NVESD) of the U.S. Army. The data was collected at significant expense by NVESD and therefore they reserved the right to approve all publications containing their data. Because the NVESD data contains some of the best examples of subpixel target images available, the NVESD imagery is used throughout this dissertation. In order to use their imagery, we had to receive approval from NVESD to publish this dissertation – a ten week process. To help minimize the approval process which dictates that any publication changes must be approved by NVESD, we rewrote the dissertation such that it contains a data chapter. NVESD only requires that this data chapter be approved per e-mail of Mr. David Hicks (NVESD). Fortunately, the addition of this data chapter has provided the added benefit of providing a good explanation of hyperspectral imagery and its idiosyncrasies to motivate the rest of the dissertation.

Dedication

To Andra

Acknowledgements

First, I want to thank my advisor, Prof. Rama Chellappa, for his patience, support, and encouragement while working on this dissertation. He provided me the opportunities to learn, publish, and develop as a scholar and I will always fondly remember my years at the University of Maryland because of those experiences.

Second, I want to thank Dr. Amit Banerjee, Dr. Marc Kolodner, Dr. Reuven Meth, and Dr. Patricia Murphy for their many helpful discussions both on hyperspectral image analysis and their own Ph.D. experiences.

Third, I must thank Ms. Miranda Schatten and Mr. David Hicks of the U.S. Army RDECOM CERDEC NVESD for providing the data used in this dissertation. Hyperspectral data with subpixel targets and detailed ground truth is difficult to find. Without the NVESD data, many of the developments introduced in this dissertation would not have been possible.

Most importantly, I dedicate this dissertation to my wonderful and loving wife, Andra. Without her patience and support, this dissertation would not exist. I also want to thank her for giving me the two best gifts in the world: Sarah and Zachary. I love and appreciate you all more than I can ever express in words.

Table of Contents

PREFACE II
DEDICATION III
ACKNOWLEDGEMENTS IV
TABLE OF CONTENTS V
LIST OF TABLES VII
LIST OF FIGURES VIII
LIST OF ABBREVIATIONS IX
CHAPTER 1: INTRODUCTION 1
1.1. A Brief History of Imaging Spectroscopy 1
1.2. Subpixel Detection 4
1.3. Thesis 5
CHAPTER 2: HYPERSPECTRAL DATA 10
2.1. AVIRIS 10
2.1.1. Sensor Details 10
2.1.2. Imagery 10
2.2. Sensor X 11
2.2.1. Sensor Details 11
2.2.2. Imagery 12
2.2.3. Spectral Signatures 14
2.2.4. Ground Truth 16
CHAPTER 3: TARGET SIGNATURE CHARACTERIZATION 19
3.1. A Review of Radiometry 22
3.1.1. Sun Light 24
3.1.2. Sky Light 28
3.1.3. Upwelled Radiance 30
3.1.4. Atmospheric Transfer Function 31
3.2. Current Target Characterization Algorithms 31
3.2.1. Model-Based Methods 32
3.2.2. In-Scene Methods 34
3.3. Average Relative Radiance Transform 38
3.4. Experimental Results 47
3.4.1. Comparison of Target Radiance Signatures 48
3.4.2. Comparison of Target Signatures for Subpixel Detection 51
3.5. Summary 59
CHAPTER 4: BACKGROUND SIGNATURE CHARACTERIZATION 61
4.1. A Review of Endmember Extraction Methods 62
4.2. Selected Endmember Extraction Techniques 65
4.3. Dimensionality of Hyperspectral Imagery 66
4.3.1. Intrinsic Dimensionality Metrics 67
4.3.2. Virtual Dimensionality Metrics 69
4.4. Experimental Results 73
4.4.1. Individual Image Results 74
4.4.2. ROC Results 78

4.4.3. Conclusions 82
4.5. Summary 82
CHAPTER 5: PHYSICS-BASED HYBRID DETECTORS 84
5.1. Current Subpixel Algorithms 87
5.1.1. Fully Constrained Least Squares (FCLS) 87
5.1.2. Adaptive Matched Subspace Detector (AMSD) 90
5.1.3. Adaptive Cosine/Coherent Detector 92
5.2. Hybrid Detectors 95
5.2.1. Hybrid Structured Detector 95
5.2.2. Hybrid Unstructured Detector 97
5.3. Experimental Results 98
5.3.1. Experimental Design 99
5.3.2. Endmember Sensitivity Analysis 101
5.3.3. Separability Analysis 105
5.3.4. Receiver Operating Characteristics 110
5.3.5. Conclusions 113
5.4. Summary 115
CHAPTER 6: ADAPTIVE DETECTION THRESHOLDS VIA EXTREME VALUE THEORY 117
6.1. Extreme Value Theory 121
6.1.1. The Fisher-Tippett Theorem 121
6.1.2. EVT for the Exponential Class 122
6.1.3. Generalized Pareto Distribution 124
6.2. EVT Adaptive Threshold Algorithm 126
6.3. Experimental Results 132
6.3.1. Experiments with Known Distributions 133
6.3.2. Experiments on Subpixel Target Detectors 135
6.3.3. Conclusions 142
6.4. Summary 143
CHAPTER 7: SUMMARY 144
7.1. Cumulative Performance Results 144
7.2. Future Work 147
7.3. Contributions 150
BIBLIOGRAPHY 152

List of Tables

Table 1: Description of Sensor X Imagery 14
Table 2: Description of Targets 15
Table 3: Target Ground Truth 17
Table 4: Quantitative Comparison of Atmospheric Compensation Algorithms 50
Table 5: Comparison of Dimensionality Estimates for Target 1 75
Table 6: Comparison of Dimensionality Estimates for Target 2 75
Table 7: Comparison of Dimensionality Estimates for Target 3 76
Table 8: Comparison of Dimensionality Estimates for Target 4 77
Table 9: Subpixel Experiment Details 99
Table 10: Endmember Sensitivity Results 104
Table 11: Comparison of MC and GPD on Known Distributions 134
Table 12: Comparison of Threshold Estimates for ACE Results 136
Table 13: Comparison of P_d Estimates for ACE Results 137
Table 14: Comparison of False Alarms for ACE Results 138
Table 15: Comparison of Threshold Estimates for HSD Results 140
Table 16: Comparison of P_d Estimates for HSD Results 141
Table 17: Comparison of False Alarm Rates for HSD Results 142

List of Figures

Figure 1: Hyperspectral Signatures of Common Materials 3
Figure 2: Subpixel Detection Block Diagram..... 7
Figure 3: AVIRIS Image of Cuprite, Nevada 11
Figure 4: Sensor X 1200m Imagery..... 13
Figure 5: Sensor X 300m Imagery..... 14
Figure 6: Target Reflectance Signatures..... 16
Figure 7: Target 3 Radiance Signatures in Image 7..... 18
Figure 8: Target 4 Radiance Signatures in Image 7..... 18
Figure 9: The five sources of light in the reflective wavelengths..... 23
Figure 10: Source-Target-Receiver Geometry..... 25
Figure 11: The Solar Spectrum..... 26
Figure 12: Comparison of Mean Radiance and Reflectance Estimates Using ARRT 42
Figure 13: ARRT Block Diagram..... 46
Figure 14: Comparison of Atmospheric Compensation Algorithms for Target 3 49
Figure 15: Comparison of Atmospheric Compensation Algorithms for Target 4 49
Figure 16: ACE Results for Image 7..... 54
Figure 17: ROC Comparison of Target 1 Signatures..... 57
Figure 18: ROC Comparison of Target 2 Signatures..... 57
Figure 19: ROC Comparison of Target 3 Signatures..... 58
Figure 20: ROC Comparison of Target 4 Signatures..... 59
Figure 21: Comparison of Background Dimension Estimates for Target 1 79
Figure 22: Comparison of Background Dimension Estimates for Target 2 80
Figure 23: Comparison of Background Dimension Estimates for Target 3 81
Figure 24: Comparison of Background Dimension Estimates for Target 4 81
Figure 25: Graphical Comparison of Endmember Sensitivity..... 103
Figure 26: Separability Analysis for Target 1..... 106
Figure 27: Separability Analysis for Target 2..... 107
Figure 28: Separability Analysis for Target 3..... 108
Figure 29: Separability Analysis for Target 4..... 109
Figure 30: Subpixel Detection ROC Curves for Target 1..... 111
Figure 31: Subpixel Detection ROC Curves for Target 2..... 111
Figure 32: Subpixel Detection ROC Curves for Target 3..... 112
Figure 33: Subpixel Detection ROC Curves for Target 4..... 112
Figure 34: Comparison of the GPD to the Empirical CDF for Example 1..... 130
Figure 35: Comparison of the GPD to the Empirical CDF for Example 2..... 131
Figure 36: Comparison of Corrected Samples..... 131
Figure 37: Block Diagram of the EVT Adaptive Threshold Algorithm..... 132
Figure 38: Proposed Subpixel Detection Block Diagram..... 145
Figure 39: Subpixel Detection System ROC Curves..... 146

List of Abbreviations

ACE ..... Adaptive Coherent/Cosine Estimate
AIC ..... Akaike Information Criterion
AMEE ..... Automated Morphological Endmember Extraction
AMSD ..... Adaptive Matched Subspace Detector
ARRT ..... Average Relative Radiance Transform
AVIRIS ..... Airborne Visible Infrared Imaging Spectrometer
BIC ..... Bayesian Information Criterion
BRDF ..... Bidirectional Reflectance Distribution Function
CDF ..... Cumulative Distribution Function
CEM ..... Constrained Energy Minimization
CFAR ..... Constant False Alarm Rate
CSD ..... Constrained Signal Detector
EIF ..... Empirical Indicator Function
ELM ..... Empirical Line Method
EVT ..... Extreme Value Theory
FCLS ..... Fully Constrained Least Squares
GEVT ..... Generalized Extreme Value Theory
GLRT ..... Generalized Likelihood Ratio Test
GPD ..... Generalized Pareto Distribution
GPS ..... Global Positioning System
HIS ..... Hyperspectral Imagery
HSD ..... Hybrid Structured Detector
HUD ..... Hybrid Unstructured Detector
IARR ..... Internal Average Relative Reflectance
IEA ..... Iterative Error Analysis
IS ..... Importance Sampling
JHU/APL ..... The Johns Hopkins University Applied Physics Laboratory
JPL ..... Jet Propulsion Laboratory
LSE ..... Least Squares Error
LVQ ..... Learning Vector Quantization
LWIR ..... Long Wave Infrared
MC ..... Monte Carlo
MDL ..... Minimum Description Length
MEI ..... Morphological Eccentricity Index
MLE ..... Maximum Likelihood Estimate
MNF ..... Maximum Noise Fraction
MODTRAN ..... Moderate Transmission model
MSMA ..... Modified Spectral Mixture Analysis
MWIR ..... Mid-Wave Infrared
NASA ..... National Aeronautics and Space Administration
NIR ..... Near Infrared
NRL ..... United States Naval Research Laboratory
NSP ..... Noise Subspace Projection

NVESD ..... Night Vision & Electronic Sensors Directorate
ORASIS ..... Optical Real-Time Spectral Identification System
OSP ..... Orthogonal Subspace Projection
PCA ..... Principal Component Analysis
PDF ..... Probability Density Function
PPI ..... Pixel Purity Index
ROC ..... Receiver Operating Characteristic
SAA ..... Simulated Annealing Algorithm
SAM ..... Spectral Angle Mapper
SINR ..... Signal to Interference plus Noise Ratio
SOLCD ..... Spectral Object Level Change Detection
SMM ..... Stochastic Mixing Model
SVD ..... Singular Value Decomposition
SWIR ..... Short Wave Infrared
USGS ..... United States Geological Survey
VIS ..... Visible

Chapter 1: Introduction

1.1. A Brief History of Imaging Spectroscopy

The study of a material's spectral properties grew out of the field of reflectance spectroscopy introduced in the 1920s. Reflectance spectroscopy identified the component chemicals in a sample by studying the reflective properties of the material [40]. By the 1930s and 1940s, spectrophotometers were introduced and the field of spectroscopy grew more popular. This work led to radiative transfer theory that was able to measure the reflective properties of a sample and identify the underlying physical mechanisms in such measurements. Radiative transfer theory ultimately led to the development of spectral imagers in the early 1970s [54].

Spectral imagery is, however, not a new concept. Color imagery is the most basic and widely recognized spectral imagery. In spectral imagery, each spatial point or pixel is represented by multiple measurements of different wavelengths in the electromagnetic spectrum. In the case of color imagery, each pixel contains information for the red, green, and blue wavelengths in the visible portion of the electromagnetic spectrum. This idea of measuring the energy in different wavelengths of the spectrum along with radiative transfer theory led to the development of multispectral imagery.

In July 1972, the first space-based multispectral imager was launched under the LANDSAT program [63]. The imager contained four bands across the visible (VIS) to near-infrared (NIR) wavelengths. The LANDSAT program was so successful that the program continues today utilizing new multispectral sensors that are capable of measuring seven bands of the electromagnetic spectrum. The success

of these multispectral sensors led to the development of the hyperspectral sensor in the mid-1980s and its corresponding field of imaging spectroscopy.

Hyperspectral imagery (HSI) differs from its earlier counterpart, multispectral imagery, in two key ways. The first difference is the number of spectral bands collected by hyperspectral sensors. Multispectral sensors typically collect less than ten bands of spectral information per pixel. Hyperspectral imagery contains hundreds of bands of spectral information per pixel. The second difference is that multispectral imagery having so few bands, selects wavelengths that are considered the most informative for a particular application; thus, the bands are non-contiguous. Hyperspectral sensors sample the spectrum creating hundreds of contiguous spectral bands. The result is a spectral signature at every pixel location that can be used to identify the materials imaged within the pixel. The spectral signature can also be decomposed to identify different materials present in the same pixel.

For this dissertation, we focus on hyperspectral sensors that measure energy in the reflectance wavelengths of the electromagnetic spectrum. Reflectance is defined as "the ratio of reflected radiance to incident irradiance" [93]. Simply, reflectance is a measure of the energy reflected from the surface of an object. Therefore, hyperspectral sensors in the reflective wavelengths are passive instruments measuring the light reflected in a scene – typically sunlight. The reflectance wavelengths in the electromagnetic spectrum are composed of three spectral bands: the Visible (VIS) from 400 nm to 700 nm, the Near Infrared (NIR) from 700 nm to 1100 nm, and the Short Wave Infrared (SWIR) from 1100 nm to 2500 nm. Figure 1 displays these three spectral bands and provides three typical materials in a hyperspectral image: road,

soil, and vegetation. This shows figure demonstrates the spectral resolution available in hyperspectral imagery.

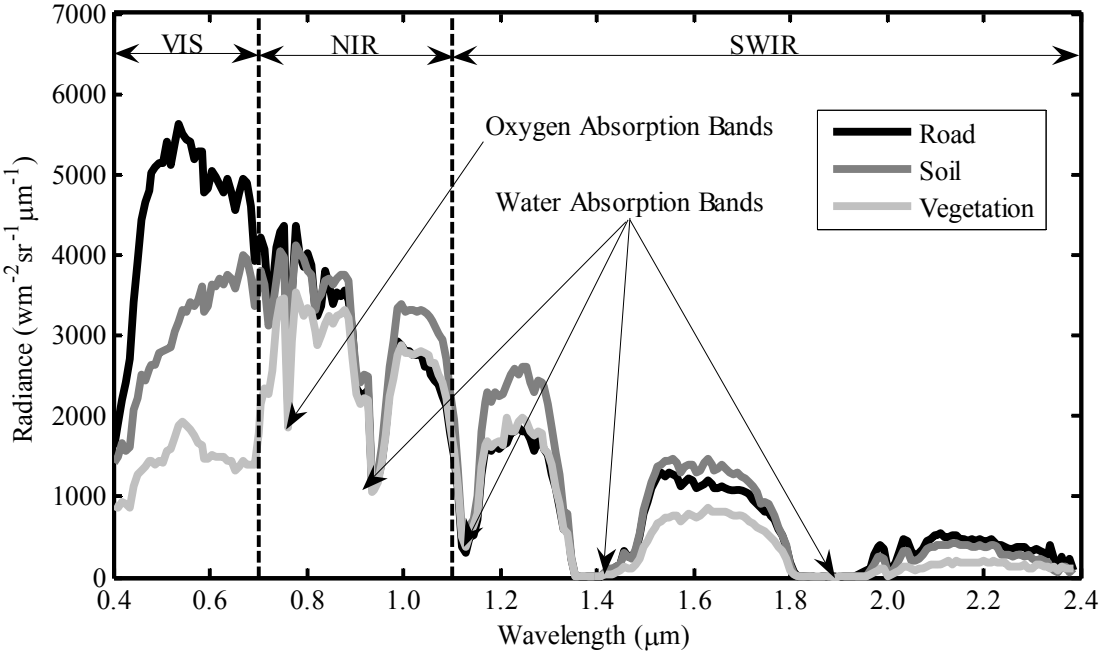


Figure 1: Hyperspectral Signatures of Common Materials

Figure 1 also displays a few of the effects caused by light passing through the atmosphere. Therefore, hyperspectral sensors do not directly measure the reflectance properties of a material. Instead, hyperspectral sensors measure the radiance at each wavelength. Radiance is defined as "radiant flux per unit area per unit solid angle per unit wavelength" [93]. The radiance values not only contain the reflectance properties of the object being imaged, but also contain all of the environmental effects that arise between the imager and the object being imaged. Thus, the hyperspectral sensor not only records the materials in the pixel, but also the spectral signatures due to sunlight and the atmosphere such as the absorption bands shown in Figure 1.

Despite the effects of the atmosphere masking the true reflective signatures of the materials being imaged, a number of applications have been developed to utilize

hyperspectral imagery such as mineral identification [76][77], land cover classification [34], vegetation studies [66], and atmospheric studies [72]. This dissertation focuses on target detection applications – specifically, subpixel detection where the target is literally smaller than the area imaged by a single pixel. This field of study has broad reaching applications from obvious military applications to search and rescue operations [106] to forensic investigations for the space shuttle Columbia incident [78]. The last application is perhaps the most well known use of hyperspectral sensors to perform broad-area searches and find parts of the Columbia that were only one inch long from an altitude of 2000 ft.

1.2. Subpixel Detection

Detection can be considered a special two class case of pattern recognition; however, it differs from classification in a number of ways [69]. In classification, the objective is to minimize the total error across all classes of data [24]. In detection, we only want to identify our desired target class amongst a larger background class. This reasoning fundamentally assumes that the target class is rare and that most pixels are from the background class. Thus, if we minimized the total error as in classification, we could simply identify every pixel as background. Of course, we are interested in maximizing the detection of targets while minimizing Type I errors – identifying background pixels as targets (false alarms) [18]. This maximization of target detection and minimization of false alarms is the fundamental difference between detection and standard pattern recognition.

Spectral subpixel detection in hyperspectral image (HSI) data aims to identify a target smaller than the size of a pixel using only spectral information [71]. Thus, the challenge in detecting subpixel targets lies in separating the target's spectral signature

from other competing signatures within the pixel. To accomplish this "unmixing" of signatures, the field of reflectance spectroscopy provides a model of how these multiple spectra interact with one another [40]. The most common model assumes that the spectra are represented by unique spatially non-overlapping materials. This model is called the linear mixing model and it is the cornerstone for most subpixel detection algorithms.

The linear mixing model assumes that a pixel is made up of endmembers, each with its own abundance. Endmembers are the spectra representing the unique materials in a given image. For instance, in an image that contains soil, vegetation, and road, the endmembers would be the corresponding unique spectral signatures for each of these materials as shown in Figure 1. Abundances are the percentage of each material within a given pixel. Mathematically, the linear mixing model is written as

$$\mathbf{x} = \mathbf{E}\mathbf{a}, \quad a_i \geq 0, \quad \sum_{i=1}^{M} a_i = 1 \tag{1}$$

where $\mathbf{x}$ is an $L \times 1$ vector that represents the spectral signature of the current pixel, M is the number of endmembers within the image, $\mathbf{E}$ is an $L \times M$ matrix where each column represents the i^{th} endmember, and $\mathbf{a}$ is an $M \times 1$ vector where the i^{th} entry represents the abundance value a_i. Note that the linear mixing model includes two constraints on the abundance values: non-negativity and sum-to-one. These constraints place physical limitations on the abundances making sure they represent the percentage of each material present in the pixel.

1.3. Thesis

The interesting part of subpixel detection is not the linear mixing model itself, but the parameters of the linear mixing model. These parameters have been

historically treated only in a statistical sense. The parameters are typically found using maximum likelihood estimates (MLE). This is, of course, a natural way to proceed in solving detection problems since such estimates are guaranteed to be consistent and asymptotically efficient [18]. However, Prof. David Landgrebe, a pioneer in remote sensing, argues in his paper that the improvement in hyperspectral image analysis will not be made by using different statistical algorithms, but by properly modeling the physics of the problem [64]. Instead of using statistical estimates of the parameters, we could use physics-based estimates of the parameters within statistical hypothesis tests to improve subpixel detection.

Some research has already been devoted to this type of physics-based detection approach. The most notable is from Thai and Healey [109]. They present an algorithm that creates a subpixel detector that is invariant to atmospheric effects. They project the desired target reflectance signature to radiance signatures for thousands of different atmospheric profiles using the computational physics model MODTRAN (MODerate TRANsmission) [3]. From these thousands of possible target radiance signatures, they use singular value decomposition (SVD) to extract a set of target singular vectors that minimize atmospheric and illumination effects; however, they only use physics to derive the target signature. The background signatures and detector are still estimated using purely statistical arguments. This has the negative effect of generating abundances that cannot meet the linear mixing model constraints.

Schott [94] and Lee [65] take a slightly different approach to physics-based subpixel detection. From the thousands of different target radiance signatures generated with MODTRAN, Lee uses a simplex method to identify the target

signatures that span the space of all possible target signatures generated. These target "endmembers" are concatenated to the image data and a simplex method such as N-FINDR is used to extract the endmembers [115][116] – some of which they argue will be target signatures. This has the result of creating both target and background endmembers that are physically meaningful. Unfortunately, they too use least squares estimates of the abundances even though physically meaningful abundances could be estimated from their endmember signatures.

Our physics-based subpixel detection approach uses physically meaningful estimates of both the endmembers and their abundances. We show this approach leads to not only improved detection performance over previous approaches, but also provides a level of insensitivity to estimation errors and provides contextual information not obtainable with other methods. Additionally, we propose new algorithms for nearly all facets of subpixel detection (shown in Figure 2) from parameter characterization to threshold estimation.

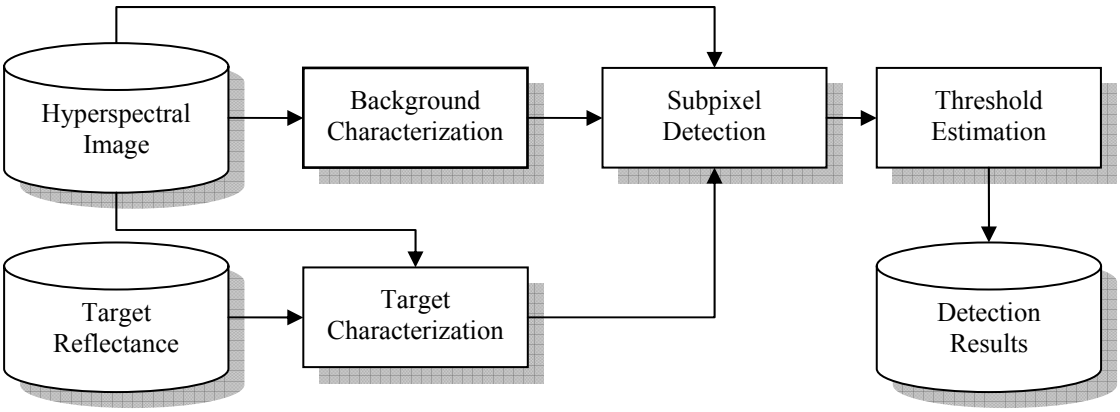


Figure 2: Subpixel Detection Block Diagram

In Chapter 3, we present a novel way to estimate target radiance signatures from reflectance measurements using only the target reflectance signature and the

hyperspectral image. This chapter provides an overview of radiative transfer theory and how MODTRAN and other methods use this theory to estimate radiance signatures from reflectance measurements. We explain how MODTRAN can be used with proper weather, topographic, and geometric data to generate a target signature for a specific hyperspectral image. From this, we develop a new in-scene algorithm that performs similarly to MODTRAN, but uses only a target and reference reflectance signature along with the hyperspectral image to estimate a target radiance signature for subpixel detection.

In Chapter 4, we present a new method to estimate the number of endmembers that maximize subpixel detection performance. The chapter gives a brief overview of endmember extraction techniques and identifies the algorithms we use in this dissertation to obtain physically meaningful endmembers. The chapter documents the sensitivity of subpixel target detection to the number of endmembers showing how slight errors in estimating the number of endmembers can cause severe losses in performance. From this result, we compare a number of different algorithms to estimate the number of endmembers and compare them to our proposed methods relative to subpixel detection performance.

In Chapter 5, we present our physics-based hybrid subpixel detectors [12]. Unlike the subpixel detectors proposed by [41], [49], [58], and [71], we develop a detector that uses all of the linear mixing model constraints including the non-negativity and sum-to-one constraints of the abundances. Our work differs from previous work because of how it models the data. The assumption in the literature is that the error between the linear mixing model and HSI data can be modeled by zero-

mean noise with a covariance matrix of $\sigma^2 I$. This has been shown to be erroneous in [71]. Using this result, we model the remaining noise using a full covariance matrix to account for sensor artifacts and nonlinear mixing effects not represented by the linear mixing model. This results in a subpixel detector that has improved performance and is partially insensitive to the number of background endmembers used.

In Chapter 6, we present a new algorithm to estimate a detection threshold for a desired false alarm rate for any detector. One of the disadvantages of the hybrid subpixel detectors is the use of the non-negativity constraints of the linear mixing model. These constraints disallow a closed-form solution for the detector making derivation of the target and background conditional distributions difficult at best. To overcome this shortfall, we develop an adaptive threshold technique based on Extreme Value Theory (EVT). We show the proposed technique outperforms both theoretical estimates for Constant False Alarm Rate (CFAR) detectors as well as non-parametric methods such as Monte Carlo estimates – especially when targets are present in the imagery.

In Chapter 7, we summarize our work and present an example of the proposed algorithms working together in a subpixel detection process. Besides providing excellent detection of subpixel targets, the result shows the ability of these methods to provide near real-time results using a minimal amount of ancillary information. This result is important to transitioning hyperspectral subpixel detection algorithms from research to practice.

Chapter 2: Hyperspectral Data

In this dissertation, we use hyperspectral imagery from two sensors: the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) and the U.S. Army RDECOM CERDEC Night Vision & Electronic Sensors Directorate (NVESD) Sensor X. The chapter is therefore broken into two sections. Each section contains information about the hyperspectral sensor, its images, available target reflectance signatures, and corresponding ground truth information.

2.1. AVIRIS

2.1.1. Sensor Details

The AVIRIS imagery comes from the National Aeronautics and Space Administration (NASA) Jet Propulsion Laboratory (JPL) at the California Institute of Technology [111]. This sensor collects 224 contiguous spectral bands spanning the wavelengths from 400 to 2500 nm. The sensor was primarily designed for environmental remote sensing applications; therefore, the imagery collected has not been focused on subpixel detection applications. Nevertheless, the AVIRIS sensor has been well calibrated and does not contain any low SNR bands allowing us to use all 224 spectral bands for processing.

2.1.2. Imagery

We chose one image to use from the AVIRIS data sets: the Cuprite, Nevada image [107]. From the Cuprite data set, we chose a sub-image containing a small town shown in Figure 3. The image itself covers a 10.4 km by 5.1 km swath of area with each pixel measuring 17 m per side. While the AVIRIS imagery has not been focused on subpixel detection applications, it can be useful to demonstrate the

atmospheric compensation techniques in Chapter 3. AVIRIS images are delivered as two images: the original radiance image collected by the sensor and another image which is an estimate of the reflectance signatures at each pixel in the image using known ground materials. These reflectance estimates will be used to identify how well our proposed target characterization method identifies radiance signatures generated from flat reflectance signatures.

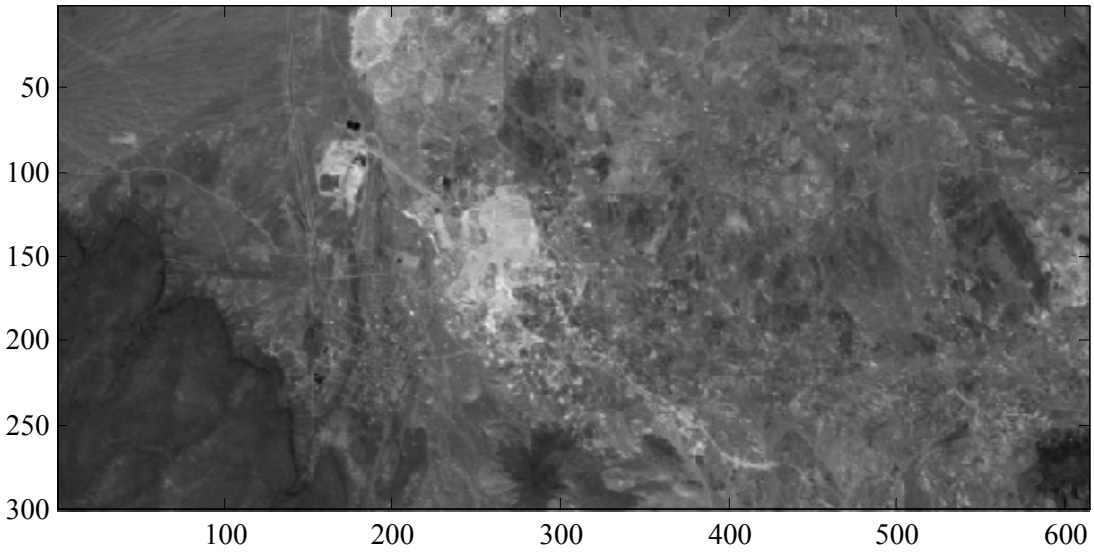


Figure 3: AVIRIS Image of Cuprite, Nevada

2.2. Sensor X

2.2.1. Sensor Details

The Sensor X imagery comes from the U. S. Army RDECOM CERDEC Night Vision & Electronic Sensors Directorate (NVESD). The sensor collects 256 contiguous spectral bands spanning the wavelengths from 400 to 2500 nm. Along with the sensor specifications, we received a spreadsheet containing information about the sensor's spectral bands. For example, the absorption bands for oxygen, carbon dioxide, and water were well documented. The spreadsheet also identified low

SNR bands in the imagery due to sensor artifacts. For our target detection application, these bands are non-informative and only serve to increase processing time without providing any benefits. Because of this, we did not use these bands as is typically done in target detection applications [41],[70],[71]. After removing these bands, we are left with 169 spectral bands for our subpixel detection experiments.

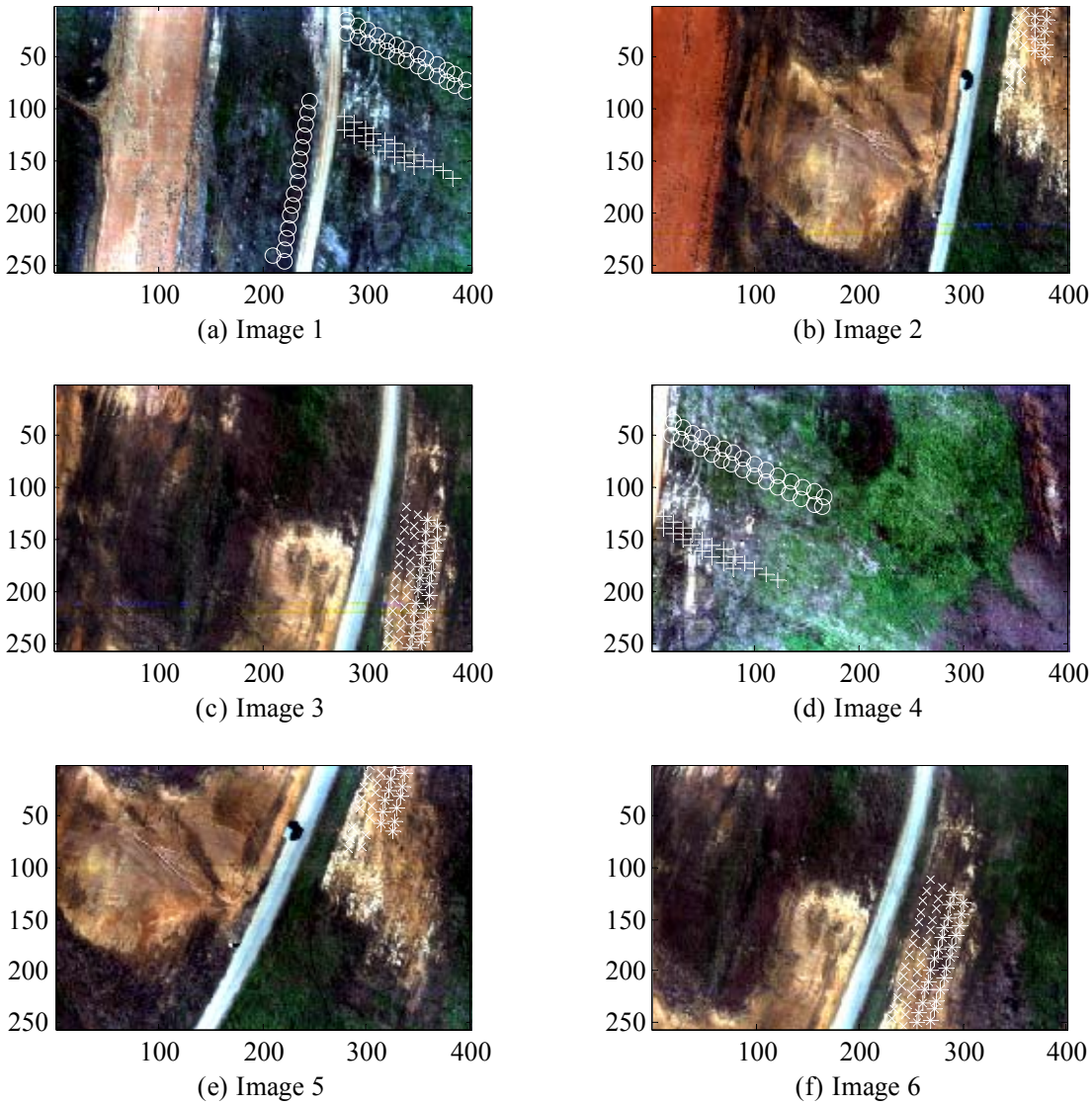
2.2.2. Imagery

We chose seven images to use in this dissertation. The first six images were chosen because of their small fill factors (e.g., percentage of a pixel that is comprised of target) and the difficult background in which the targets lie. The most difficult of these areas is the tall grass site. At this site, the grass is high enough to partially obscure the target causing the pixel fill factors to be smaller than expected. The other two areas are easier since the targets are not obscured. Figure 4 shows the six images with corresponding target locations.

The seventh image is shown in Figure 5. This image was chosen because the targets were full or multi-pixel. This image was selected because the true target radiance signatures could be extracted from the image. These signatures can be compared to the target radiance estimates described in Chapter 3.. Without this image, we would not know how well the target characterization algorithms were performing. The image is only used for Chapter 3. Table 1 identifies each of the images, the type of area imaged, the amount of area imaged, and the spatial resolution of an individual pixel.

Unfortunately, the imagery we received was collected with an uncalibrated sensor. This posed a significant problem. Some of the algorithms within this dissertation use the physics-based model MODTRAN that calculates the radiance of

an object from its corresponding reflectance signature. The radiance signature generated by the model assumes the sensor is calibrated. When the sensor is not calibrated, the model will predict signatures that will not match those in the imagery. This mismatch is severe enough to render a target detection algorithm useless.



(a) Image 1 (b) Image 2

(c) Image 3 (d) Image 4

(e) Image 5 (f) Image 6

Figure 4: Sensor X 1200m Imagery
(Target 1 '+', Target 2, 'o', Target 3 'x', Target 4 '*')

To overcome this problem, we worked with Dr. Marc Kolodner of the Johns Hopkins University Applied Physics Laboratory (JHU/APL). Using MODTRAN, we generated radiance signatures for known background materials in the imagery. We

compared the model-based signatures to the known signatures in the imagery. From these comparisons, an offset and gain vector was created. This offset and gain was applied to each image to vicariously calibrate the image. These new vicariously calibrated images were then used for the experiments in this dissertation.

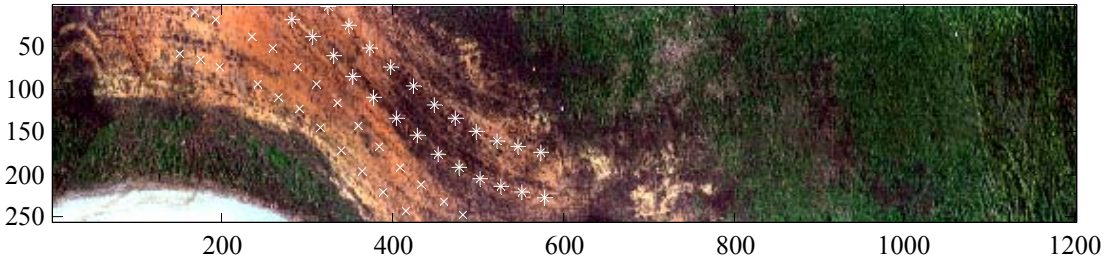


Figure 5: Sensor X 300m Imagery
(Target 3 'x', Target 4 '*')

Table 1: Description of Sensor X Imagery

Image	Background	Clutter Density	Altitude (m)	Area (m^2)	Pixel Size (m^2)
1	Short & Tall Grass	High	1220	18811	0.1823
2	Sparse Grass	Medium	1220	18811	0.1823
3	Sparse Grass	Medium	1220	19464	0.1823
4	Short Grass	Medium	1216	18815	0.1815
5	Sparse Grass	Medium	1215	18542	0.1806
6	Sparse Grass	Medium	1213	19097	0.1806
7	Sparse Grass	Medium	313	7400	0.0241

2.2.3. Spectral Signatures

Besides the imagery, we received spectral libraries containing reflectance signatures for both the targets and background materials. All signatures were collected using hand-held spectrometers in the field. Due to this in-field data capture, multiple signatures were created for each target and background material. These signatures were averaged to form a signature for each material. This method was chosen because the averaged spectral signature reduced variations that occurred when measuring with the hand-held spectrometer.

For the background, numerous signatures were collected. These ranged from different types of vegetation to fiducial markers placed in the field for spatial registration purposes. This information is typically not available in real-world applications, but allows us to vicariously calibrate the images. The signatures are also used as reference signatures to help estimate the amplitude of the target signature as explained in Chapter 3.

From the target signatures, we chose four different targets. The targets were chosen to provide a wide variety of spectral signatures. The targets are typically pieces of metal or plastic small enough to achieve subpixel sizes at 1200m altitudes. Additionally, the targets have different paints which cause the reflectance signatures to vary from very strong (Target 1) to very weak (Target 4) as shown in Figure 6. Table 2 provides a description of each target's geometry, size, material, color, and symbol used in figures throughout the dissertation.

Table 2: Description of Targets

Target	Geometry	Size (m^2)	Material	Color	Symbol
1	Circle	0.0182	Plastic	White	+
2	Circle	0.0869	Metal	Green	o
3	Square	0.1090	Plastic	Green	x
4	Circle	0.0869	Metal	Dark Green	*

Target 3 was an interesting case as that particular target had two spectral signatures. The two signatures existed because it was discovered later that the targets were made of slightly different plastics. The difference was very slight as can be seen in Figure 6, but was significant enough that it was decided two signatures should be used. We chose to use this target because it is the only case where we have multiple target signatures for a single target type.

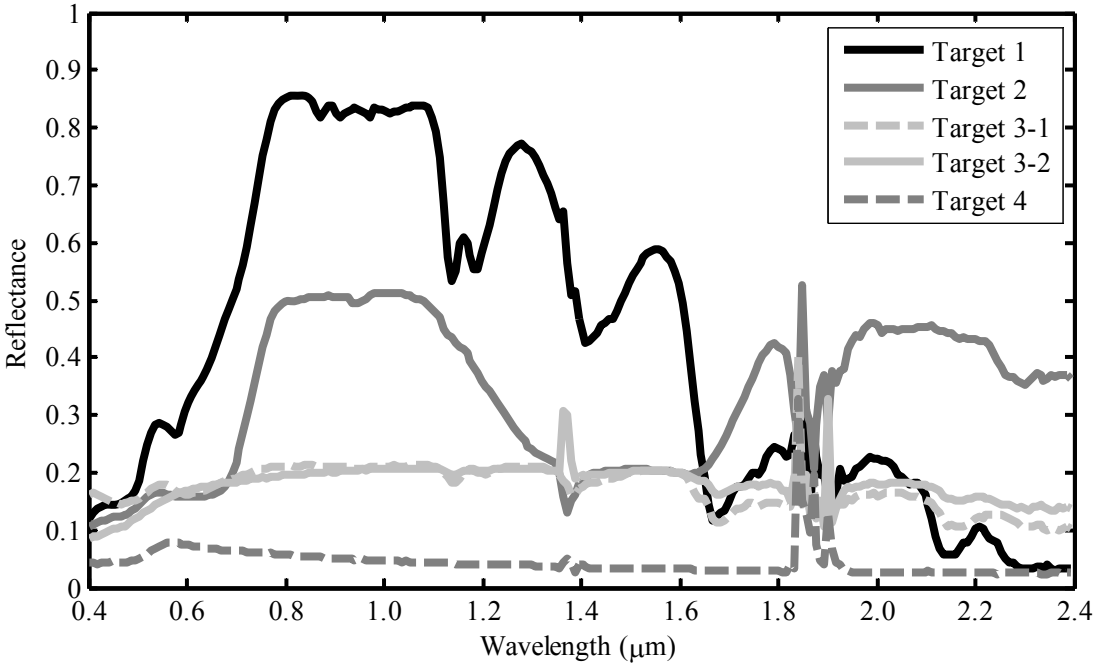


Figure 6: Target Reflectance Signatures

2.2.4. Ground Truth

Along with the imagery and signatures we received from NVESD, we received ground truth information identifying the target locations in the imagery. The ground truth data contained object-level location information. Unlike pixel-level truth which identifies the location of the targets for each pixel and their corresponding abundances, object-level truth specifies an area in the image where the targets are located. Therefore, the ground truth identifies the center of the target even though it may span multiple pixels. Note that this statement is true even with subpixel targets as the target could be located on pixel borders. Table 3 details how many targets are in the seven images arranged by target type and image. The locations of each target in the Sensor X imagery can be seen in Figure 4 and Figure 5.

Given object-level ground truth, we had to cluster the detector outputs to form objects as pixel level analysis was not possible. To obtain these objects, a clustering

threshold is applied to each image. This clustering threshold refers to a threshold that combines adjacent pixels together to form an object which will be classified as either target or clutter. Typically this threshold is chosen to include no more than 1% to 5% of the pixels in the image depending on the application. In our analysis, we chose 1% as we knew the number of targets was far less than 1% of the pixels in any one image. Each cluster is assigned the maximum detection score from all the pixels that make up the cluster. Along with the maximum detection score, each cluster is identified as either target or clutter based on their location relative to the object-level ground truth. This information can then be used to identify how well a detector performs.

Table 3: Target Ground Truth

Image	Target 1	Target 2	Target 3	Target 4	All
1	20	42	0	0	62
2	0	0	12	9	21
3	0	0	25	23	48
4	20	30	0	0	50
5	0	0	15	12	27
6	0	0	28	25	53
7	0	0	24	24	48
All	40	72	104	93	309

From the ground truth information, we were able to extract target radiance signatures from Image 7 due to the targets spanning multiple pixels. These “true” target radiance signatures will be used in Chapter 3 to compare the estimated target radiance signatures with the ones shown in Figure 7 andFigure 8. Each figure contains all of the target radiance signatures found in the image (in gray) and their spectral average (in black). Note the wide variability of target signatures in either case. Despite our best efforts, some background signatures leaked into our “true” target signatures. This occurred because even with four pixels on target, some small

amounts of background signatures may still be present. This is especially the case for Target 4 where the targets spanned on average 3.6 pixels.

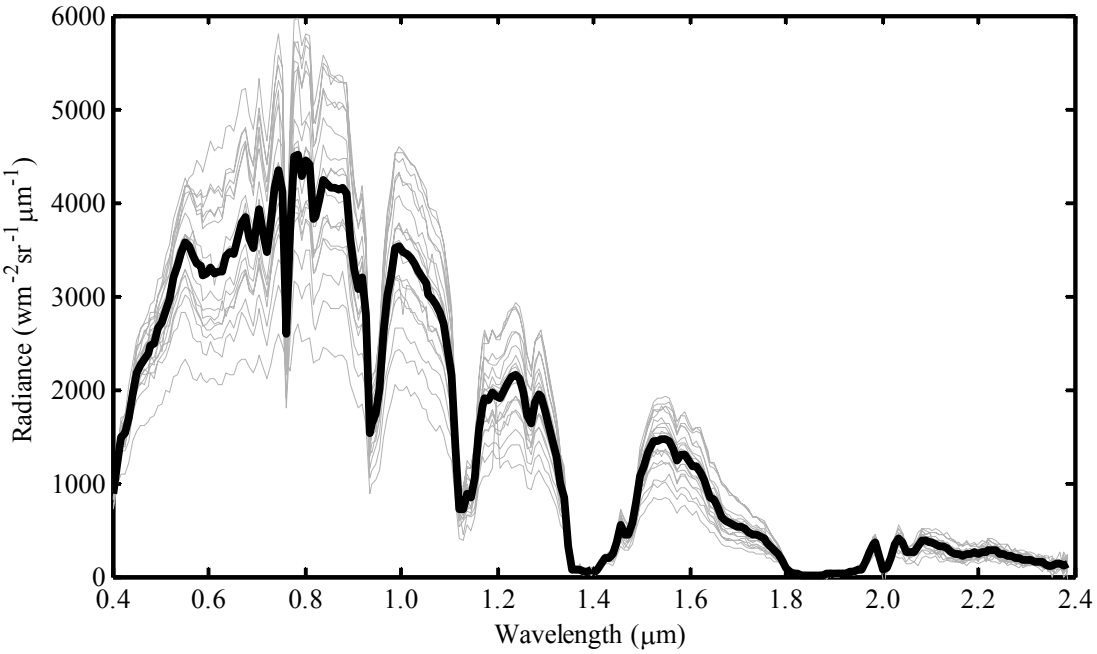


Figure 7: Target 3 Radiance Signatures in Image 7
(Gray lines represent individual targets and black line represents the mean)

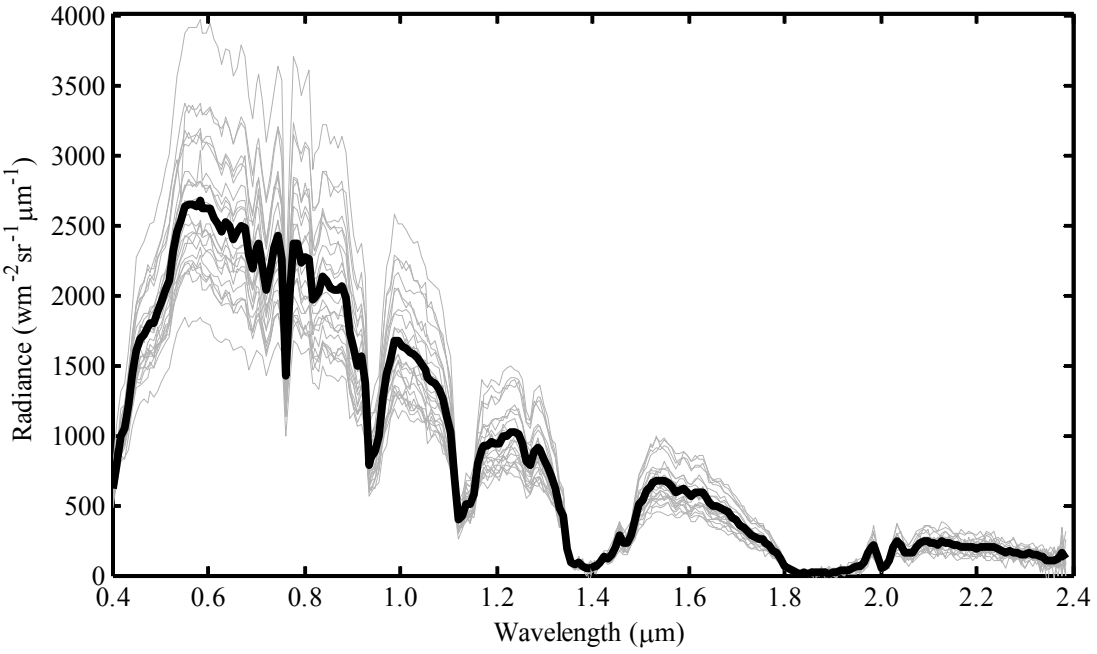


Figure 8: Target 4 Radiance Signatures in Image 7
(Gray lines represent individual targets and black line represents the mean)

Chapter 3: Target Signature Characterization

An important part of subpixel detection is the correct characterization of the target signature. As explained in Chapter 1, target characterization is especially important for hyperspectral detection because the images are collected in terms of radiance while the target signatures are measured in terms of reflectance. The reason for this mismatch is due to the fact that target signatures are typically measured in laboratories or in the field with hand-held spectrometers that are at most a few inches from the target surface. Hyperspectral images, however, are collected hundred to thousands of meters away from the target and have significant atmospheric effects present. Therefore, a transfer function between radiance and reflectance must be obtained. This transfer function is known as atmospheric compensation.

A number of algorithms have been developed to compensate for atmospheric effects. The algorithms can be classified into two primary types: radiance inversion methods and radiance projection methods. Radiance inversion methods were first developed for spectral analysis purposes. Originally, hyperspectral imagery was used to classify images into different natural phenomenon for applications such as mineral mapping [59],[98],[107]. In order to accomplish this type of classification, the logical path was to invert the image from radiance to reflectance and compare the resulting corrected image to known spectral reflectance libraries. The idea in these programs was not to identify a certain material, but to identify the constituent materials in the image for mapping purposes. One such algorithm is FLAASH [3].

While this may be ideal for image analysts wanting to investigate spectral signatures, it is not the best method for detecting subpixel targets. First, the

algorithms process every pixel in the image requiring significant processing time. Second, the algorithms have to make simplifying assumptions to perform the inversion because it is intrinsically an ill-posed problem [75]. So, while these programs have enjoyed some success in target detection applications, they are better suited for spectral analysis by operators that can make informed judgments.

The other class of atmospheric compensation algorithms is based on radiance projection methods. These methods project a reflectance signature into a radiance signature for a particular hyperspectral image. Murphy and Kolodner have one of the most direct approaches: calculate the radiance of a target signature at the sensor using real-time weather predictions and the known source-target-receiver geometry [75]. This type of atmospheric compensation algorithm makes good use of computational physics using the MODTRAN atmospheric model [3]. It also provides different shading conditions so targets can be modeled in both full sun and full shade (such as in the shade of a tree or cloud). Although this approach is the most direct and computationally simple, it also requires the most ancillary information to work properly. Weather data must be timely and the source-target-receiver geometry known precisely. For new data collections, this is usually not hard information to obtain; however, for past data collections, this method typically cannot be used

Healey and Slater simultaneously developed another forward projection model that was designed to be atmospheric invariant [45]. Based on Healey's earlier work with color imagery, they developed an algorithm that projected a target reflectance signature into approximately 17,000 different environments. From these 17,000 radiance signatures, they used SVD to create a nine-dimensional subspace that could

be used in any environment. Results show that this method works well, but requires a significant amount of pre-processing to create the invariant subspace.

A final set of methods use in-scene information to calculate the target radiance signature. These approaches directly estimate atmospheric effects by using information present in the imagery. The most popular of these is the Empirical Line Method [26]. This method uses an adaptive background estimator to find any vegetation in the imagery. Vegetation is used because it is typically ubiquitous and has a well-known reflectance signature. Using the estimated vegetation signature from the image and the known vegetation reflectance signature allows a direct calculation of the transfer function without MODTRAN or any other physical modeling technique. The only issue with such an approach is that certain environments may not have vegetation in the image such as urban environments, winter scenes, or desert scenes.

This chapter presents our work and analysis of model-based and in-scene based radiance projection methods. To begin, we describe in some detail the atmospheric transfer function and the simplifying assumptions made for estimation purposes. We next describe two current methods for atmospheric compensation: an in-scene method developed by Piech and Walker [80] and a model-based method using MODTRAN with radiosonde information. . We then present our own in-scene method for target characterization called Average Relative Radiance Transform (ARRT). The final sections of the chapter compare ARRT to MODTRAN. It will be these two methods which we will use throughout the dissertation for target signature characterization.

3.1. A Review of Radiometry

Radiometry is the measurement of electromagnetic fields typically in the visible and infrared wavelengths [93]. To understand the measurements at an optical sensor, radiometry (or radiative transfer theory) has produced a model of how photons (light) propagate from the sun and through the atmosphere. By understanding this model, we can understand which parts of the radiance signature measured at the sensor are produced by the target of interest and which are produced by the surrounding environment. We can also understand which parts of the model are more critical than others for target characterization.

For this dissertation, we only cover the most basic radiometric principles; however, there are two excellent books available by Schott [93] and Hapke [40] that provide greater details about this interesting theory. Schott's book is meant primarily for the general scientist and engineer interested in remote sensing. Hapke's book provides a more thorough analysis of the governing equations of light. Both are excellent resources and much of the material in this section is derived from both of these texts.

For this dissertation, we are concerned only with those photons that can be collected by a hyperspectral sensor in the reflectance domain. The reflectance domain identifies a range of electromagnetic wavelengths from 400 nm to 2500 nm where light is primarily reflected from objects. As the wavelengths increase, the dominant effect becomes self emittance of photons (such as heat). While this is an interesting regime, our data is all collected in the reflectance wavelengths and as such, we will restrict our analysis to these wavelengths.

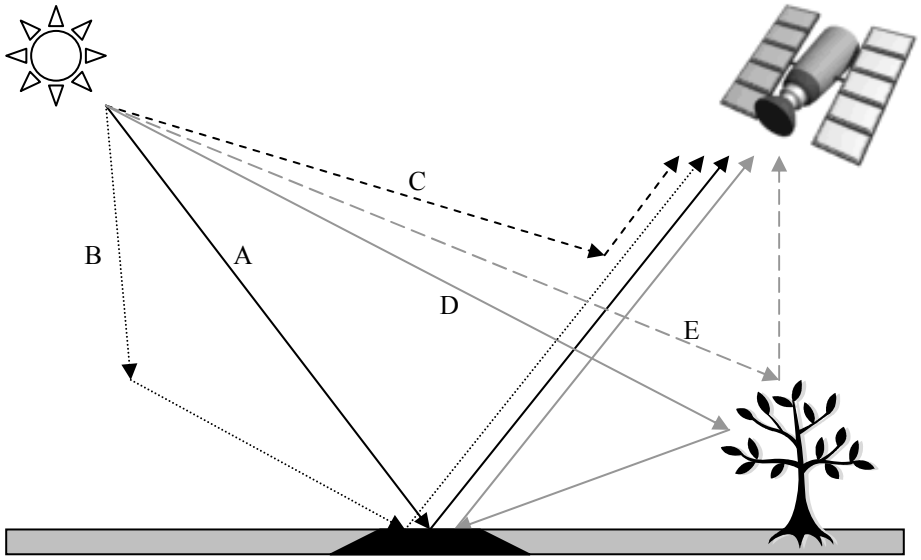


Figure 9: The five sources of light in the reflective wavelengths
(A: Direct Sunlight, B: Sky Light, C: Upwelled Radiance, D: Multipath Effect, E: Adjacency Effect)

In the reflectance domain, there are five main sources of light collected by a sensor: direct sun light, sky light, upwelled radiance, multipath effect, and the adjacency effect. These multiple sources of light are shown in Figure 9. Sun light is the light generated by the sun that passes through the atmosphere, reflects off the area being imaged, and is collected at the sensor. Sky light is the light that is scattered in the atmosphere which reflects off the area being imaged and back to the sensor. Upwelled radiance is the light that is scattered in the atmosphere that never reaches the area being imaged. Instead, this light is scattered directly into the optical path of the sensor. Multipath effects are due to light that reflects off of multiple objects in a scene before arriving at the sensor. The adjacency effect occurs when light scatters off of other background objects near the area being imaged into the optical path of the sensor [52]. The last two sources of light are very small compared to the first three

and are typically not computed in most models. Because of these reasons, only the first three light sources will be treated in greater detail.

3.1.1. Sun Light

The most obvious source of light is the sun. Photons are generated at the sun and pass through the atmosphere onto the object being imaged and back to the sensor. Along the way, the spectral properties of the light are changed as the photons are absorbed and scattered through the atmosphere. These effects can be mathematically modeled as

$$L_{sun}(x,y,\lambda) = KT_u(z_g, z_u, \vartheta_v, \phi_v, \lambda)R(x,y,\lambda)T_d(z_g, \vartheta_0, \phi_0, \lambda)E_0(\lambda)\cos\vartheta_0 \qquad (1)$$

where L_{sun} is the radiance seen at the sensor generated from sun light, K is the amount of energy at the top of the atmosphere, T_u is the upward atmospheric transmittance, R is the reflectance of the object being imaged, T_d is the downward atmospheric transmittance, and E_0 is the exoatmospheric spectral signature of the sun. All of these quantities are a function of the spectral wavelength λ and most of the quantities are based on the geometry of the source (sun), target (object being imaged), and receiver (camera) geometry as shown in Figure 10. The geometries are based on cylindrical coordinates where z_g is the elevation of the sun, z_u is the elevation of the camera, θ_v is the declination of the camera from a normal vector to the surface, θ_0 is the declination of the sun from the same normal vector, ϕ_0 is the azimuth of the sun and ϕ_v is the azimuth of the camera.

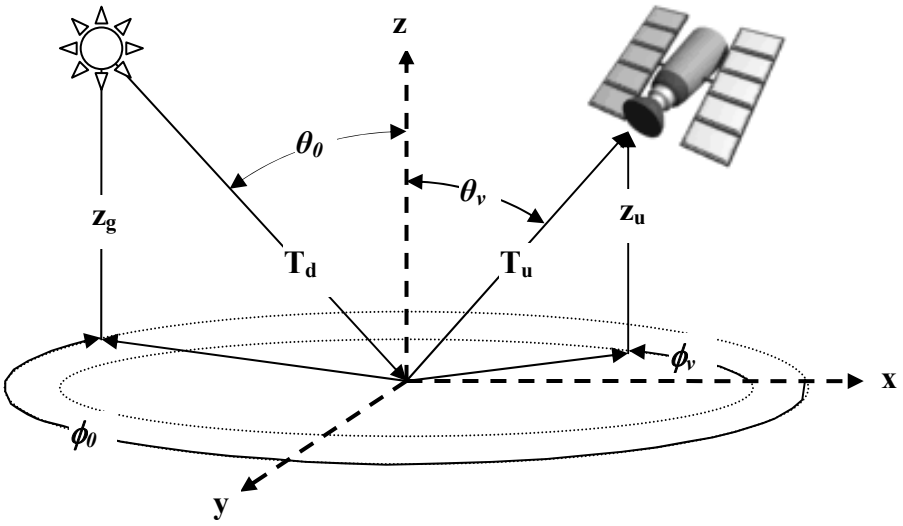


Figure 10: Source-Target-Receiver Geometry

3.1.1.1. Solar Spectral Signature E_0

For light to reach the sensor, light must first be generated. Ideally, the light source should be spectrally flat equally distributing the energy across all wavelengths. This can be accomplished in a laboratory setting, but in hyperspectral applications, the light source is typically the sun which has its own spectral signature. The sun's atmosphere is made of 73.46% hydrogen, 24.85% helium (by-product of the fusion of hydrogen atoms), and a fraction of other naturally occurring elements. These gases absorb certain wavelengths of light causing the documented Fraunhofer Absorption Lines [55]. Additionally, the fusion reaction produces more energy in the visible wavelengths. When these two effects are combined, it produces the typical solar spectrum seen in Figure 11. Thus, all images are colored with this solar spectrum.

The amount of sun light that reaches an object is a function of the sun declination angle and the downward atmospheric transmittance. The declination angle determines how much sun light directly hits an object. For example, when the sun is

directly overhead, the declination angle is zero and all the sun light reaches the object (cos(0°) = 1). When the declination angle is 60°, the amount of energy is only half of the energy when the sun is directly overhead. The interesting result of this effect is that the declination angle can be caused by either the sun being lower in the sky or the object sitting on a non-level surface. Thus, besides the angle of the sun relative to the horizon, even minor changes in topography can change the overall amount of sun light an object receives.

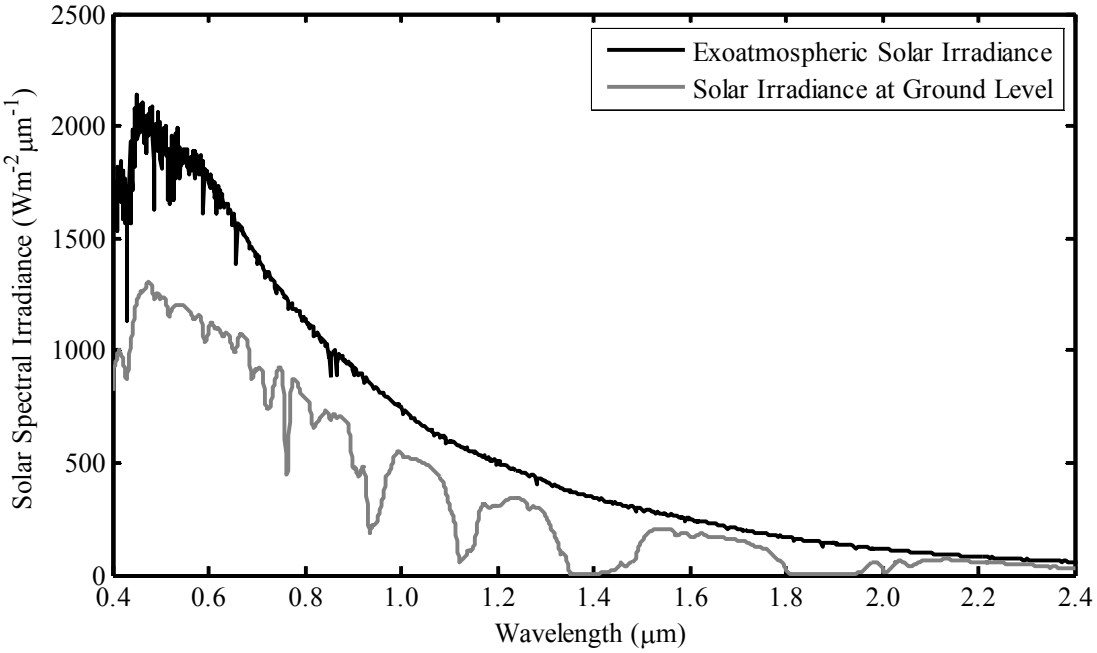


Figure 11: The Solar Spectrum

3.1.1.2. Downwelled Atmospheric Transmittance T_d

The other effect that reduces the sun light reaching an object is the downwelled atmospheric transmittance. The downwelled atmospheric transmittance quantifies the scattering and absorption effects that occur as light passes through the atmosphere. Scattering disperses the photons out of the direct path of the object thereby reducing the amount of light reaching the ground. The other dominant effect

is absorption which reduces the energy in certain wavelengths due to such molecules as water and carbon dioxide. By the time the light reaches the object being imaged, it has both the spectral properties of the sun and the intervening atmosphere as shown in Figure 11.

We can model how the atmosphere affects the sun light using a number of cylinders stacked on top of one another representing different altitudes. Each of these cylinders has a certain temperature, pressure, and humidity. These measurements dictate the amount of absorption and scattering that occurs within each cylinder and at each wavelength. Near the top of the atmosphere, there are very few particles and hence the three measurements are not as critical as near the bottom of the atmosphere. Thus, the cylinders are tall at the top of the atmosphere and become smaller as they reach the surface. This occurs because the dense atmosphere is located near the surface and causes a significant portion of the transmittance effects. This dense atmosphere is also the most variable as weather changes occur mostly in this region making signatures vary from one location to another.

3.1.1.3. Reflectance R

Once the sun light reaches the object, the reflectance of the object dictates which wavelengths of light are absorbed and which are reflected in various directions. The spatial reflectance attributes of a material are described by its bidirectional reflectance distribution function (BRDF). This function measures the reflectance for all wavelengths and input-output angles. A full BRDF characterization of a material is rare; so, materials are typically classified into gross categories ranging from specular reflectors to diffuse reflectors (also known as Lambertian). Specular materials reflect light in one direction such as mirrors. Diffuse reflectors reflect light

in all directions equally such as flat paint. Most materials fall between these two categories, but tend to be more diffuse then specular. Because BRDF characterizations are rare and most materials can be treated as diffuse, we assume diffuse reflectors for the remainder of this dissertation.

3.1.1.4. Upwelled Atmospheric Transmittance T_u

After the light has been reflected from the object being imaged, it passes back through the atmosphere to the sensor. The upwelled atmospheric transmittance quantifies these atmospheric effects. Upwelled atmospheric transmittance is very similar to downwelled atmospheric transmittance. The real difference between the two transmittances is upwelled transmittance only affects light between the object and the sensor. Therefore for low altitudes (e.g. 300m), this effect is minimized. On the other hand, the sensor could be space-borne in which case the light passes through the entire atmosphere. Either way, T_u is modeled the same way as T_d using cylinders of the atmosphere along the light path to quantify the scattering and absorption effects. As described in (1), the light reaches the sensor after being affected by the solar spectral signature, downwelled atmospheric effects, reflectance of the object being imaged, and upwelled atmospheric effects.

3.1.2. Sky Light

In the previous sections about atmospheric transmittance, scattering played an important part of how the spectral signature of the sun light was changed. This scattering of light has another side effect causing a secondary light source called sky light. Sky light can be mathematically modeled as

$$L_{sky}(x,y,\lambda) = R(x,y,\lambda)T_u(z_g, z_u, \vartheta_v, \phi_v, \lambda)\int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi/2} E_s(\theta,\phi,\lambda)\cos\vartheta\sin\theta d\theta d\phi \qquad (2)$$

where L_{sky} is the sky light radiance at the sensor, R is the reflectance of the object being imaged, T_u is the upwelled atmospheric transmittance, and E_s is the amount of energy scattered by the atmosphere.

Sky light takes a very similar path to sun light. Once the light reaches the object being imaged, it reflects the same as the sun light (assuming a diffuse material), and is reflected back up through the atmosphere to the sensor along the same path as the sun light. The main difference between sky light and sun light is the source of sky light is the scattering of photons in the atmosphere. These scattered photons arrive at the object being imaged from all directions. Therefore, these different patches of sky light are integrated over the hemisphere above the object being imaged. This produces the two integrals seen in (2) replacing the $T_d(z_g, \vartheta_0, \phi_0, \lambda)E_0(\lambda)\cos\vartheta_0$ term in (1).

There are three types of scattering that take place. The most well known scattering effect is Rayleigh scattering as explained by Lord Rayleigh to answer why the sky was blue [67]. Rayleigh scattering occurs when light interacts with the very small molecules that make up the atmosphere. The scattering occurs mostly in the blue wavelengths while other wavelengths are absorbed creating the blue color of the sky.

The other well known scattering effect is Mie Scattering [105]. This type of scattering occurs when photons interact with particles that are roughly the same size. These particles are typically composed of aerosols, combustible by-products, and small dust particles. This effect causes the scattered light around cities to be much different from the light scattered in rural areas.

The final effect is called non-selective scattering. This type of scattering occurs when the particles are much larger than the photons of light. Examples of such particles are water droplets and ice crystals that are due to cloud formations. Thus, scattered light can be affected by the amount and types of cloud cover in the image. Theses different scattering effects explain why images taken of rural areas on cloudless days can be very different from images taken of cities on partially cloudy days.

3.1.3. Upwelled Radiance

While some light is scattered so that it illuminates the object, other light is scattered directly towards the sensor. Unlike all the previous sources of light, upwelled radiance, $L_{up,}$ never reaches the object being imaged. This light is scattered directly into the sensor's optical path from the atmosphere. Like sky light, it undergoes the same three scattering processes making it vary based on location and weather conditions. This has two effects on the imagery. The first effect reduces the overall contrast of the image. The second effect causes a blue shift (an increase in energy at the blue wavelengths) as the upwelled radiance term is typically dominated by Rayleigh scattering.

A good example of upwelled radiance is fog. As fog settles in, our eyes cannot see objects far away because they are obscured by the scattering of light towards our eyes from the water vapor particles (Mie and non-selective scattering). The effect is those objects disappear in a haze of gray. This effect is always present except it typically scatters such a small amount of photons relative to sun and sky light to make it undetectable in most situations.

The same can be said about the upwelled radiance reaching a sensor. In normal environmental conditions, upwelled radiance has a very small effect relative to the other sources of light. However, as the sensor is placed higher in altitude, the scattering effect becomes more predominant and can start to reduce the contrast of the image at the sensor. This occurs because there are more particles and thus more opportunities for scattering to occur.

3.1.4. Atmospheric Transfer Function

We can now mathematically define the radiance L reaching a sensor from an object with reflectance R as

$$\begin{aligned} L(x,y,\lambda) = {} & R(x,y,\lambda)T_u(z_g,z_u,\vartheta_v,\phi_v,\lambda)T_d(z_g,\vartheta_0,\phi_0,\lambda)KE_0(\lambda)\cos\vartheta_0 \\ & + R(x,y,\lambda)T_u(z_g,z_u,\vartheta_v,\phi_v,\lambda)\int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi/2}E_s(\theta,\phi,\lambda)\cos\vartheta\sin\theta d\theta d\phi \\ & + L_{up}(z_g,z_u,\vartheta_v,\phi_v,\lambda). \end{aligned} \tag{3}$$

The radiance equation in (3) states that the radiance at the sensor is a linear combination of the sun light, sky light, and upwelled radiance contributions. Although the final equation is a linear combination, the previous sub-sections detail how complex the atmospheric transfer function is to compute. Detailed weather information, source-target-receiver geometries, topography, and BDRFs are required to solve all the necessary functions. Typically, all of this information is not available and algorithms have to make simplifying assumptions. What assumptions are made depends on the type of algorithm.

3.2. Current Target Characterization Algorithms

Nearly all algorithms that convert reflectance to radiance or vice-versa are based on (3). The difference between these algorithms is the simplifying assumptions

they make and how they estimate each of the light sources. These algorithms can be broken down into two general methods: model-based methods and in-scene based methods.

3.2.1. Model-Based Methods

Model-based methods attempt to solve (3) directly. This type of solution requires a wealth of ancillary information besides the image. From Figure 10, the exact locations of the source, target, and receiver are required. This information is easy to obtain from the Global Positioning System (GPS). The location of the sun relative to a ground location is also well understood and can easily be found on the internet for a given location and time.

The information that is not as easy to obtain is weather data. In the modeling of atmospheric transmittance, the temperature, humidity, and pressure at varying levels of altitude need to be measured (i.e, the cylinders of the atmosphere). Typically, this is done using radiosondes. Radiosondes are weather sensors attached to balloons that measure all the needed weather information. Unfortunately, radiosonde information is not always available or applicable. For example, radiosondes are collected at certain locations which may be too far from the area being imaged to be applicable. If radiosonde data is available, the information is typically collected only twice a day and may describe the atmospheric profile that occurred hours in the past.

Murphy and Kolodner developed another way to get the requisite weather data [75]. If radiosonde data is not present or is inaccurate due to the aforementioned issues, weather maps generated from weather stations can be used. These weather maps produce an atmospheric profile that can be estimated via interpolation between

weather stations. This information is fused with satellite imagery to produce an accurate atmospheric profile at any location on the planet. This information is then used as the model inputs.

Once the ancillary information has been collected, a computational model can calculate the radiance for a given reflectance at any angle, source-target-receiver geometry, and wavelength via (3). MODTRAN is arguably the most used computational model [3]. It produces an estimate for every function in (3) and can make estimates for large declination angles as well as areas with variable topography. For most of the functions, it performs a direct calculation, but for the atmospheric transmittance functions, it has to make a simplifying assumption.

The scattering and absorption is not only a function of humidity, temperature, and pressure, but also of the constituent particles in the atmosphere. To model these particles in the atmosphere, MODTRAN uses one of many atmospheric profiles for urban, desert, or rain forest areas to name a few. Each profile uses a lookup table to provide an estimate of how light is scattered based on the types of particles found above each area type. Unfortunately, real world situations can vary significantly from the atmospheric profiles included with MODTRAN. While this may not greatly effect the radiance estimate, such assumptions can be very important when estimating weak target signatures such as Target 4.

Model-based methods have become the standard for atmospheric compensation techniques. They can make estimates for every parameter and function in the atmospheric transfer function. These estimates can take into account any type of topography and source-target-receiver geometry – even when the sensor may be on

or near the ground. To accomplish this calculation, they require a significant amount of ancillary information about source-target-receiver geometry, weather, and atmospheric profile type.

3.2.2. In-Scene Methods

The problem with model-based methods is that we sometimes lack all of the necessary ancillary information (or any estimate thereof). This is especially true with images collected in the past where such information was simply not collected. Because the information is either inaccurate or not available, another way to estimate the atmospheric transfer function was created using only the image data. These methods are called in-scene methods.

In-scene methods have to make a number of simplifying assumptions as well. The first assumption is that the area being imaged is small enough that the atmospheric profile (azimuths, altitudes, declination angles, etc.) is the same for all pixels even though this may not be true in a number of cases (e.g. water vapor [32]). The second assumption is that the pixels being used to estimate the atmospheric transfer function have Lambertian scattering properties. This assumption again is not necessarily true [89], but materials can be found that have near Lambertian properties that are acceptable for in-scene methods. Third, pixels that contain only one material (pure pixels) must exist in the image. Thus, in-scene methods are best for aerial images that cover a small amount of ground area.

3.2.2.1. Piech and Walker Shadow Method

One of the earliest and most accurate in-scene methods was developed by Piech and Walker [80]. They noted that shadow regions could be used to estimate the three main light sources in the atmospheric transfer function. Instead of estimating

detailed functions such as atmospheric transmittance, the atmospheric transfer function was simplified to

$$L(\lambda) = R(\lambda)L_{sun}(\lambda) + R(\lambda)FL_{sky}(\lambda) + L_{up}(\lambda) \tag{4}$$

where F is the fraction of the sky above the area being imaged (i.e., in shadow zones the amount of sky not blocked by the object creating the shadow). All x,y coordinates have been removed since we assume Lambertian scattering with equal amounts of light at each pixel.

The key to this method is realizing that in shadow zones, (4) becomes

$$L_{shade}(\lambda) = R(\lambda)FL_{sky}(\lambda) + L_{up}(\lambda) \tag{5}$$

since the sun light term has been reduced to zero. The algorithm therefore requires a material that is in both direct sun and shade conditions. When this occurs, the sunlight term can be easily calculated by taking the difference between (4) and (5) and solving for the sun light term to obtain

$$L_{sun}(\lambda) = \frac{L(\lambda) - L_{shade}(\lambda)}{R(\lambda)}. \tag{6}$$

To isolate the upwelled radiance term, equations (4) through (6) can be combined so the total radiance term is a linear regression of the shade radiance term as

$$\begin{aligned} L(\lambda) &= \frac{L_{sun}(\lambda) + FL_{sky}(\lambda)}{FL_{sky}(\lambda)} L_{shade}(\lambda) - \frac{L_{sun}(\lambda) + FL_{sky}(\lambda)}{FL_{sky}(\lambda)} L_{up}(\lambda) + L_{up}(\lambda) \\ &= m(\lambda)L_{shade}(\lambda) + b(\lambda). \end{aligned} \tag{7}$$

Using multiple materials with varying reflectance signatures, (7) can be solved to obtain the m and b terms at each wavelength. Rearranging these terms provides the upwelled radiance estimate

$$L_{up}(\lambda) = \frac{b(\lambda)}{1 - m(\lambda)}. \quad (8)$$

Equations (6) and (8) provide a way to establish the last light source such that

$$L_{sky}(\lambda) = \frac{L(\lambda) - R(\lambda)L_{sun}(\lambda) - L_{up}(\lambda)}{R(\lambda)F}. \quad (9)$$

This algorithm provides estimates of each light source within the atmospheric transfer function. The algorithm requires a shadow area which contains numerous pixels of the same material in both full sun and full shade conditions. Additionally, the algorithm requires multiple materials to be identified (historically by hand) to make estimates of the upwelled radiance term. In cases where these constraints cannot be met, we must rely on other methods.

3.2.2.2. Empirical Line Method

The empirical line method (ELM) is simpler than the shadow method and does not require any shadows in the imagery. ELM also does not estimate all of the light sources in the atmospheric transfer function. Instead, ELM makes the following simplification

$$L(\lambda) = R(\lambda)L_{sun+sky}(\lambda) + L_{up}(\lambda) \quad (10)$$

where the $L_{sun+sky}$ term combines the sun light and sky light into a single term assuming $F = 1$ due to the lack of shadows. Equation (10) identifies that the total radiance term is a linear combination of the upwelled radiance, the combined sun and sky light terms, and the reflectance. Thus, a linear relationship could be established by identifying a material with known reflectance in the scene. From this knowledge, the combined sun and sky light and upwelled radiance terms could be calculated for

each wavelength via linear regression. The linear regression is performed to estimate reflectance signatures from the radiance measurements in the image.

Various papers have identified numerous ways ELM can be implemented. All perform linear regression, but vary the number of materials required to estimate the parameters. The simplest implementations use one material and assume zero reflectance objects have zero radiance [26],[73]. This, of course, is not true as it assumes the upwelled radiance term simply does not exist. Not surprisingly, studies show errors of up to 20% in the predicted reflectance when compared to the true reflectance signature. Further studies used multiple known materials [26],[83] which show that four materials make the best estimates varying only a few percent from the actual reflectance signature.

While ELM has removed the need to have shadows, it does still require a significant number of known materials exist in the image. In cases where the study area is well documented or panels of known reflectance are placed in the scene, ELM performs very well. However, in images where only one material is well known, another method called dark object subtraction may be more applicable.

3.2.2.3. Dark Object Subtraction

Dark object subtraction is very simple. The idea is to find the minimum radiance values for each band in the image. These minimum values should represent the upwelled radiance assuming that the dark pixels have near-zero reflectivity. Using this dark object estimate as the upwelled radiance term allows the linear regression in ELM to take place without needing more than one known material.

This assumption holds in the NIR and SWIR bands, but the visible bands can have significant errors. The errors are especially troublesome when working with

subpixel targets which have low reflectance signatures. These low reflectance values from the targets inadvertently become part of the estimated upwelled radiance estimate. The overall effect in such cases is a corruption of the atmospheric transfer function and thus it is not well suited for subpixel detection.

3.3. Average Relative Radiance Transform

Another way to estimate the atmospheric transfer function is to use detection theory. There are a few reasons for approaching target characterization in this manner. First, the imagery does not have all the necessary ancillary information required by model-based methods. Second, the in-scene methods require user interaction to identify the materials with known reflectance in the image. This can be a time consuming process requiring a person with significant knowledge of remote sensing. Third, the simpler in-scene methods requiring the least amount of information are the most variable making them inappropriate for subpixel detection. Fourth, both in-scene and model-based methods were developed for analysis purposes. The idea was to map the radiances measured in the image back to reflectance values for comparison against spectral libraries for environmental research such as land class mapping and deforestation studies.

These reasons led us to develop a new atmospheric compensation algorithm for subpixel detection applications. To make subpixel detection applications accessible to a wide variety of users, the target characterization algorithm should automatically generate a target signature that can be used by a detector with little or no user intervention. The method should also use as little ancillary information as possible because this data may not always be available (e.g. historical image collections or analysis of areas for which information is not available). Finally, the

target characterization algorithm needs to provide enough fidelity that a detector can identify the target even among materials with similar spectral signatures.

The aforementioned constraints led us to develop the Average Relative Radiance Transform (ARRT). ARRT has a number of advantages. First, the algorithm is computationally efficient. Instead of projecting thousands (possibly millions) of pixels from radiance to reflectance, ARRT projects a few target reflectance signatures to radiance – a thousand or more so improvement in processing time. Second, ARRT is an in-scene atmospheric compensation technique requiring very little ancillary information. The algorithm only requires the image, the desired target reflectance signature, and a reference background reflectance signature. Source-target-receiver geometries and detailed weather information are not required. Third, ARRT is fully automated requiring only the aforementioned input signatures and image. Fourth, since ARRT is an in-scene method, the sensor need not be calibrated. As long as the errors in the sensor are uniform across the image, ARRT will account for the calibration errors where model-based methods cannot.

The original ARRT idea is based on the Internal Average Relative Reflectance algorithm (IARR) [59]. The IARR algorithm uses the spectral mean of an image as the atmospheric transfer function (ignoring upwelled radiance effects). The fundamental idea assumes that the image is comprised of many different underlying reflectance signatures that cancel one another when averaged together. The end result is the average spectral signature has a flat reflectance with some unknown multiplying factor K. Our early work demonstrated that applying IARR to generate target radiance signatures could work for subpixel detection algorithms [15]. The

drawback of the method is the assumption that the reflectance signatures cancel one another. Typically, the spectral mean still contains some of the reflectance characteristics of the dominant material. For example, if vegetation dominates the image, the spectral mean will have characteristics of the vegetation making it ineffective for certain targets.

This drawback led us to an updated ARRT algorithm that uses a two-pass detection method. The first detection pass identifies pixels with radiance values that most likely contain flat reflectances. This is very much like the underlying idea in IARR; however, ARRT directly detects these radiance signatures in the image instead of relying on the spectral mean.

To detect these highly probable flat reflectance materials in the image, a band ratio technique is employed. Band ratio techniques have been used in other analyses to identify vegetation, soil types, and other materials [48],[88]. For this application, we use a ratio between bands located on either side of the red-edge wavelength (700 nm). The red-edge effect causes a significant increase in reflectivity near 700 nm that corresponds to chlorophyll content (Figure 1) [90]. For radiance signatures generated from flat reflectance materials, the radiance drops slightly from 550 nm to 730 nm causing a band ratio less then one. Empirically, we found the value 0.8 to work best at identifying flat radiance signatures using both real-world HSI data and flat reflectance signatures generated by MODTRAN. Using this band ratio, radiance signatures with highly probable flat reflectances are found in the image and averaged together. As with IARR, the average reduces material and sensor variability to provide a better estimate of the flat reflectance than any single pixel found in the image.

To demonstrate the band ratio technique, we use the AVIRIS image in Chapter 2. For this data, we have two images with one being the true radiance measurement at the AVIRIS sensor and the other image being the estimated reflectance signatures for each pixel. The reflectance signatures were generated using model-based atmospheric compensation techniques validated by ground measurements of the scene [21]. Therefore, we will assume the reflectance estimates are accurate.

Figure 12 shows the results of the first stage of the ARRT algorithm on the AVIRIS data. In the top sub-figure, the mean spectrum of the radiance signatures chosen by ARRT to have highly likely flat reflectances is plotted. Using those pixel locations, we calculate the mean reflectance signature from the AVIRIS data in the second sub-figure. The reflectance is nearly flat across the spectrum except for some slight nonlinear effects near the lowest wavelengths. This slight decrease in reflectance is most likely an artifact of the AVIRIS reflectance estimation model. For example throughout the entire AVIRIS image, no one signature has a flat reflectance despite the presence of concrete in the image – a material with a known flat reflectance. Nevertheless, ARRT is finding radiance signatures that have a nearly flat reflectance signature.

The result of the first detection pass determines the spectral shape, but not amplitude. The average flat radiance signature is mathematically expressed as

$$L_{flat}(\lambda) = RL_{sun+sky}(\lambda) + L_{up}(\lambda) \tag{11}$$

where L_{flat} is the flat radiance signature estimated from the image. Because we assume the reflectance is flat, the reflectance term R should be constant for all

wavelengths. Additionally, L_{flat} includes the upwelled radiance term which causes a blue shift and loss of contrast as detailed in Section 3.1.3. Nevertheless, the L_{flat} term contains most of the spectral shape characteristics. Therefore, multiplying a reflectance signature by L_{flat} obtains a good representation of the spectral shape of the target material; however, the amplitude is still unknown as we do not have an estimate for R.

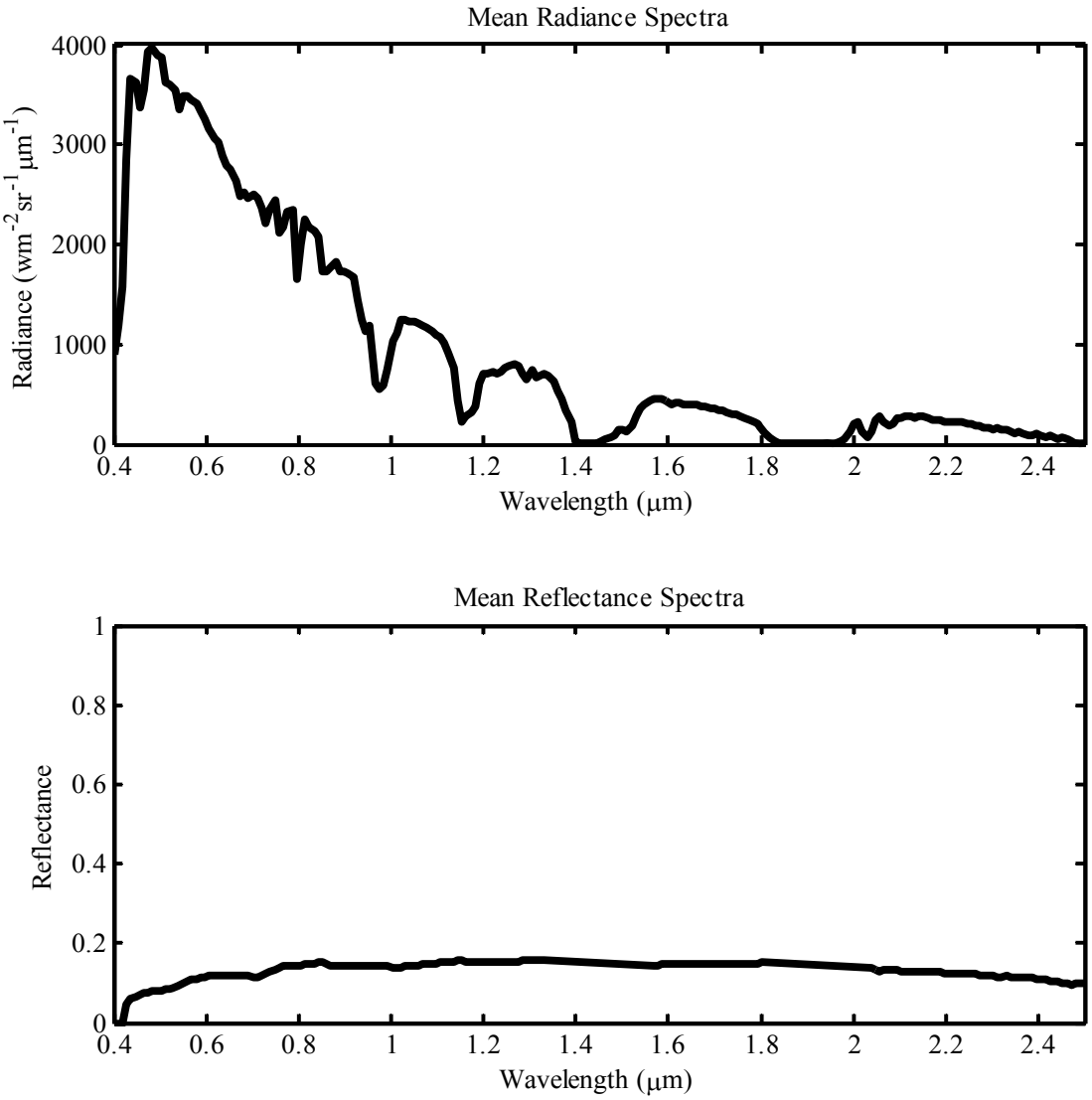


Figure 12: Comparison of Mean Radiance and Reflectance Estimates Using ARRT

It has been proposed that the amplitude mismatch is not problematic for detection applications. This statement is true in full pixel detection algorithms which use a replacement model (i.e., the pixel is either background or target, but not both). For full pixel target detection, the detectors normalize the pixels and desired target signature by their L^2 norm (see Spectral Angle Mapper [54],[95]). The result of such a normalization procedure makes the shape of the spectral signature the important determining factor as opposed to the amplitude. For replacement models, this is a desired result.

In subpixel target detection, the model is additive (i.e., the pixel is background or background plus target). To understand what happens if we divide a pixel by its L^2 norm, we describe a pixel using the linear mixing model introduced in Chapter 1:

$$x_{norm} = \frac{x}{\|x\|_2} = \frac{\sum_{i=1}^{M} \alpha_i e_i}{\left\| \sum_{i=1}^{M} \alpha_i e_i \right\|_2} = \frac{\sum_{i=1}^{M} \alpha_i e_i}{\left(\sum_{j=1}^{M} \sum_{i=1}^{M} \alpha_j \alpha_i e_j^T e_i \right)^{1/2}} \neq \frac{\sum_{i=1}^{M} \alpha_i e_i}{\sum_{i=1}^{M} \alpha_i \|e_i\|_2}. \quad (12)$$

Unlike full pixel targets, subpixel targets contain a number of background endmembers that are not a simple linear combination of their norms (i.e., cross terms exist in the solution). Therefore, normalizing the pixel, the background endmembers, and target spectra independently does not achieve the same result as full pixel target detection.

Because of this result, subpixel target detection requires a signature that is correct both in shape and amplitude. To estimate the amplitude, a second detection pass is required with a known reference material. Known reference materials refer to signatures within the image for which their reflectance signature is known. For

example, some ELM implementations use vegetation as a reference material. ARRT has no restriction on the reference material except that it has a moderate to strong reflectance signature and occurs as a pure pixel in the image.

A number of methods exist to choose a proper reference material. For example, reference signatures can be found based on the geographic region where the image was collected. If the image was collected over a desert region, sand would be an excellent reference signature while in forests, certain deciduous tree varieties would be a better match. All of these signatures are freely available from the United States Geological Survey (USGS) website (http://speclab.cr.usgs.gov/). Additionally, the USGS and other organizations have land class databases that describe the natural attributes of any area on the planet. From these two sources, a reference material for any image can be found.

Once a reference material and its corresponding spectral signature have been identified for the image of interest, the ARRT algorithm uses the Spectral Angle Mapper (SAM) algorithm to find the corresponding reference radiance signatures in the image [54]. Those pixels that pass a detection threshold are then ranked by their detection score. The top N detection scores are averaged to obtain the corresponding reference radiance signature for the image. Note we do not use the top N detection scores directly; instead we use the top N detection scores above a detection threshold. The reasoning behind this decision is that a given reference signature may not actually be within the image and the algorithm should not blindly use detection scores that fail to pass a minimum threshold. If there are no detections found in the image,

ARRT will inform the user and ask for another reference signature that better matches what is available in the image.

If a reference radiance signature is found, it is used to calculate the unknown reflectance R value in (11). The solution is

$$\hat{R} = \frac{\left\| L_{flat} R_{ref} \right\|_2}{\left\| L_{ref} \right\|_2} \tag{13}$$

where R_{ref} is the reflectance signature of the reference material and L_{ref} is the radiance signature estimated from the image for the reference material. R can be estimated assuming the reference signature has a high reflectance signature thus minimizing the effect of the upwelled radiance term.

An estimate of the upwelled radiance term can also be calculated as

$$\hat{L}_{up}(\lambda) = \begin{cases} \hat{R} L_{ref}(\lambda) - L_{flat}(\lambda) R_{ref}(\lambda), & \lambda < 700nm \\ 0, & \lambda \geq 700nm \end{cases}. \tag{14}$$

The estimated upwelled radiance term is the difference between the estimated radiance signature and the detected radiance signature of the reference material in the visible wavelengths. In the near infrared and short-wave infrared wavelengths, errors due to noise dominate the signature. In the visible wavelengths, the Rayleigh and Mie scattering effects dominate, being significantly stronger than the error terms; thus, we clip the estimated upwelled radiance to only affect the visible wavelengths.

The final estimated target radiance signature can be calculated as

$$T(\lambda) = \left(\frac{L_{flat}(\lambda)}{\hat{R}} - \hat{L}_{up}(\lambda) \right) R_T(\lambda) + \hat{L}_{up}(\lambda) \tag{15}$$

where R_T is the reflectance signature of the desired target. To help clarify the ARRT algorithm, Figure 13 provides a block diagram describing the two-pass detection process and what inputs are necessary at each stage to arrive at (15).

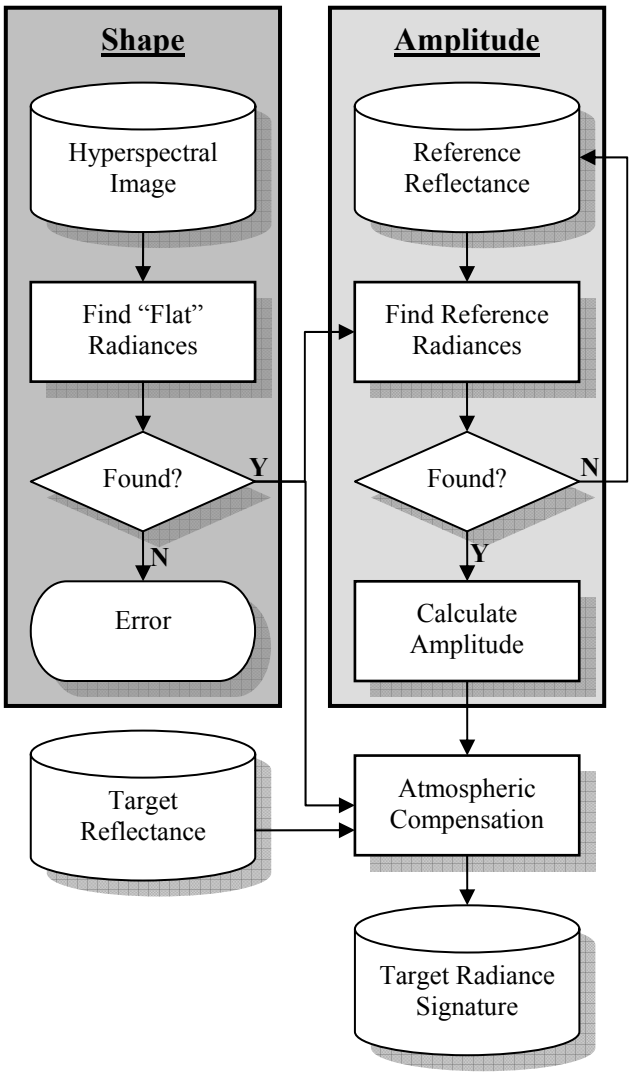


Figure 13: ARRT Block Diagram

Similar to other in-scene atmospheric compensation techniques, ARRT is only valid for certain conditions. First, ARRT was designed for aerial imagery where the upwelled radiance terms are small compared to the sun light and sky light terms. Second, ARRT requires a reference signature that has moderate to high reflectivity and has at least one pure pixel in the image. Currently, ARRT does not handle

shadow zones, but this can be addressed in another version that merges these techniques with Piech and Walker's work [80]. This will be discussed in more detail in Chapter 7.

3.4. Experimental Results

As with any atmospheric compensation algorithm, certain assumptions had to be made with ARRT. To validate whether these assumptions are valid and allow ARRT to produce useful target radiance signatures, we have designed two experiments. The first experiment uses Image 7 from Sensor X to directly compare target signatures generated by MODTRAN and ARRT to known target radiance signatures in the image. The second experiment compares target radiance signatures estimated using MODTRAN and ARRT relative to subpixel target detection performance.

Besides the imagery used for these experiments, a wealth of ancillary data was also collected. Radiosonde information was available from a nearby airport; however, this data was six hours old by the time the imagery was collected. Source-target-receiver geometry was also well documented as GPS was used on the airplane carrying the sensor. Numerous hand-held spectrometers were used on the ground to measure the reflectance of both target and background materials. While the sensor was not calibrated, the soil reflectance and radiance signatures were measured to correct for calibration errors via vicarious calibration as explained in Chapter 2. All of this ancillary data makes the following comparisons between MODTRAN and ARRT possible.

3.4.1. Comparison of Target Radiance Signatures

This experiment was used to validate the ARRT algorithm produces target signatures that match the actual target radiance signatures in an image. Image 7 from Sensor X was used for this experiment. The image was flown at 313m altitude so that each pixel imaged 0.0241 m^2 of area. The image contains Targets 3 and 4 with areas of 0.1090 m^2 and 0.0869 m^2 respectively. Targets thus spanned on average 4.5 and 3.6 pixels respectively.

Because the targets are multi-pixel, using the ground truth we received with the image, we were able to extract the true target radiance signatures from the image as shown in Figure 7 andFigure 8. These figures show the spectral variability of each target and their corresponding mean spectra. For Target 3, the mean spectrum is used in this experiment. For Target 4 however, we used only one signature pulled from a pixel that contained pure target spectra. Unfortunately, the smaller Target 4 only covers 3.6 pixels and thus has some background signature that "bleeds" into the target area as explained in Chapter 2. This minor corruption of the target signatures can be very serious when dealing with low reflectance targets. When the mean spectrum for Target 4 was used to test the subpixel detectors, it provided the worst detection performance supporting the hypothesis that many of the "true target" signatures were corrupted by background.

ARRT and MODTRAN were used to estimate Target 3 and Target 4 radiance signatures for Image 7. In the case of ARRT, two variants were used: one version estimated the upwelled radiance term while the other did not. The three estimated radiances were plotted against the known Target 3 and 4 radiance signatures in Figure 14 andFigure 15 respectively.

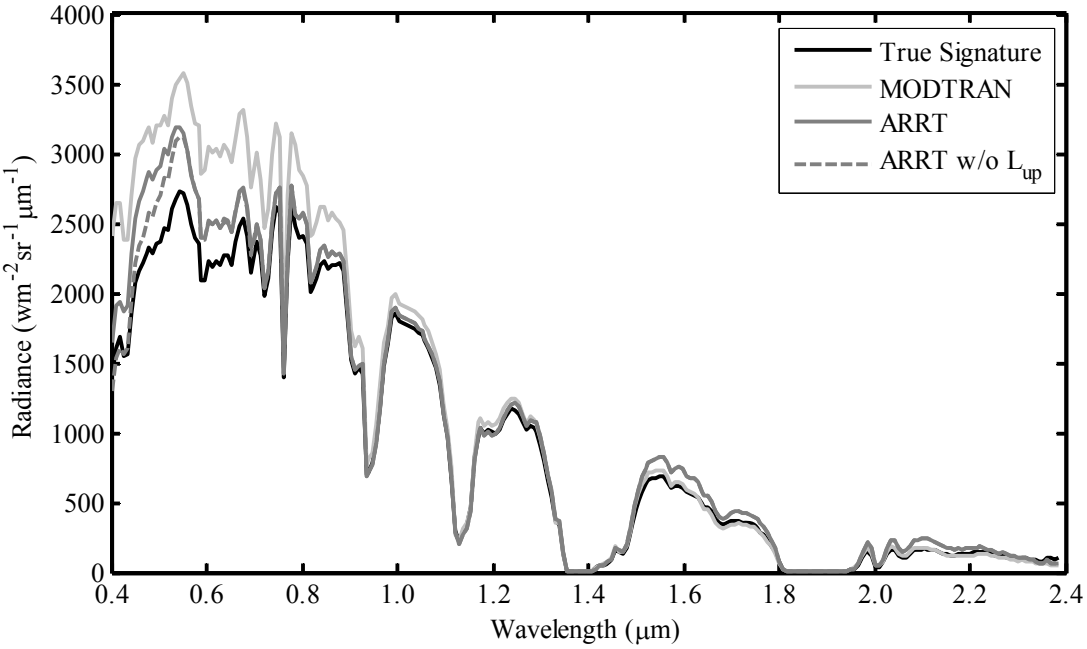


Figure 14: Comparison of Atmospheric Compensation Algorithms for Target 3

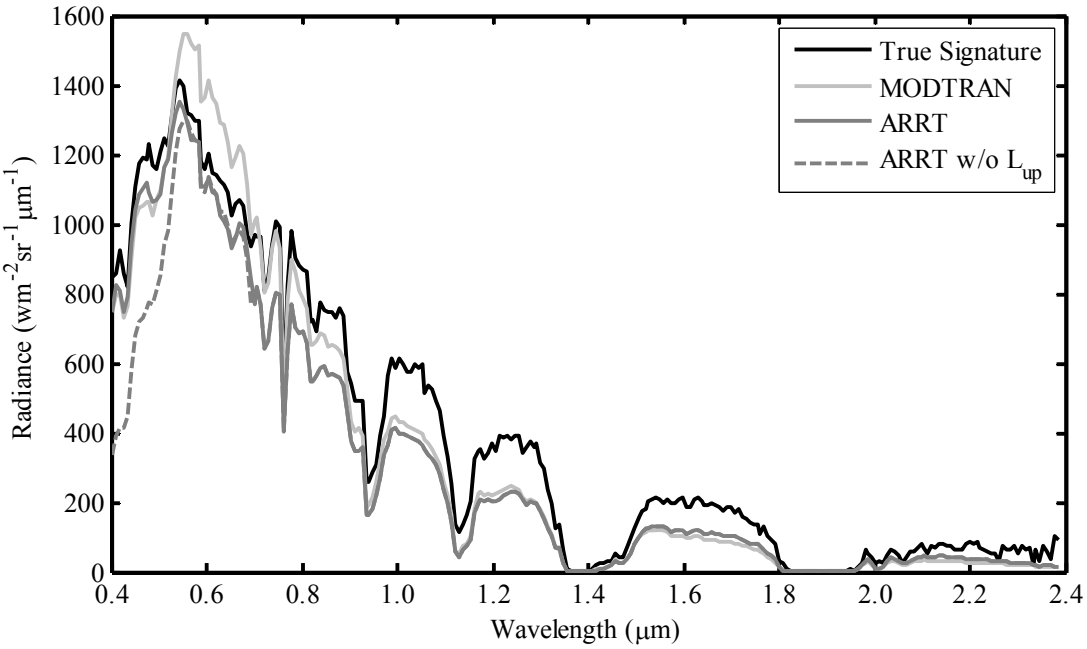


Figure 15: Comparison of Atmospheric Compensation Algorithms for Target 4

In addition, quantitative measurements are presented in Table 4. For each algorithm and target, two metrics were created measuring the similarity in amplitude

and similarity in shape to the true target signature. The metric for measuring the amplitude similarity is

$$\alpha = \left\| S - \hat{S} \right\|_2 \tag{16}$$

The metric for measuring the shape similarity is the angle between the spectral signatures

$$\vartheta = \cos^{-1}\left(\frac{S^T \hat{S}}{\|S\|_2 \left\|\hat{S}\right\|_2} \right). \tag{17}$$

The estimated target radiance signature that minimizes the above metrics provides a better match to the true target radiance signature.

Table 4: Quantitative Comparison of Atmospheric Compensation Algorithms

Target	Metric	MODTRAN	ARRT	ARRT (No L_{up})
3	α	5664	2547	1888
	θ	5.86°	4.00°	2.98°
4	α	1515	1648	2342
	θ	9.81°	7.83°	10.49°

Comparing the signatures using Figure 14, Figure 15, and Table 4, ARRT estimates the target radiance signatures well. For Target 3, ARRT outperforms MODTRAN in matching the true target signature. The shape and amplitude is a better match and as such we expect to have better detection performance using the ARRT signature. Interestingly, the ARRT version without an upwelled radiance is marginally better than the standard ARRT algorithm.

For Target 4, the results are mixed. MODTRAN estimates the amplitude very well, but does not do as well estimating the overall shape of the signature. The ARRT algorithm estimates the shape better than MODTRAN, but underestimates the

amplitude. The ARRT algorithm without the upwelled radiance term performs the worst of all the variants. All algorithms however underestimate the shape and amplitude of the SWIR bands including MODTRAN. In the next section we show that this underestimation will lead to poor detection performance. Thus, Target 4 is an interesting case for further research into ways to improve all atmospheric compensation techniques.

Overall the ARRT algorithm performs as well as MODTRAN using only the target reflectance signature, reference signature, and imagery. MODTRAN requires radiosonde information, vicarious calibration, and GPS information to produce signatures that are at best only slightly better than ARRT. Considering the amount of time necessary to collect all this information and process it through MODTRAN, ARRT provides similar target estimates with significantly less ancillary data and in a fraction of the time.

3.4.2. Comparison of Target Signatures for Subpixel Detection

While comparing the estimated radiance signatures to their true counterparts is important, it does not answer whether the estimated targets are a good match for subpixel target detection applications. This set of experiments was designed to answer the aforementioned question using the well known Adaptive Cosine Estimate (ACE) algorithm [58]. This detector is one of the better detectors available for subpixel detection in HSI data. Another reason for using this detector is the background is modeled entirely by a multivariate normal distribution; thus, no background endmembers are required. The algorithm's performance is based solely on the image and the target signature. Thus, ACE makes an ideal algorithm to use for experiments

comparing algorithms that generate target radiance signatures. More information on the ACE algorithm is documented in Chapter 5.

For all of these experiments, the ACE algorithm was processed in the following manner. Besides the target signature, a mean and covariance had to be estimated. There are two ways to estimate these parameters: globally or locally. We chose the global method for these experiments as this provided both the best performance and the fastest implementation. Typically, the SAM algorithm is used to detect obvious target detections and remove them from the image before calculating the global mean and covariance as was done for Image 7. In Images 1 through 6 however, the targets are so small, they are not detected by the SAM algorithm and hence were not removed. While this may slightly degrade performance [27], it provides the most honest performance results as real-world applications will not have knowledge of the ground truth a-priori.

Once the ACE detector was run, a detection image was generated. As mentioned in Chapter 2, the ground truth for Sensor X was for object level detection. To obtain objects from our detection images, a clustering threshold is applied. This clustering threshold refers to a threshold that combines adjacent pixels together to form an object which will be classified as either target or clutter. Typically this threshold is chosen to include no more than 1% to 5% of the pixels in the image depending on the application. In our analysis, we chose 1% as we knew the number of targets was far less than 1% of the pixel in any one image. Each cluster is assigned the maximum detection score from all the pixels that make up the cluster. Along with the maximum detection score, each cluster is identified as either target or clutter

based on their location relative to the object-level ground truth. This information can then be used to identify how well a detector performs.

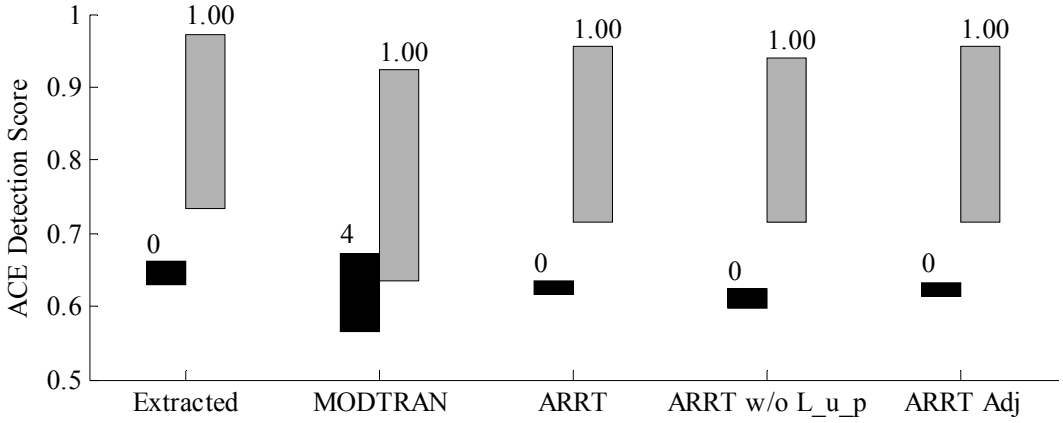
3.4.2.1. Comparison of Full Pixel Detection Performance

The first experiment applies ACE to Targets 3 and 4 in Image 7 from Sensor X using the target signatures generated in the previous set of experiments. For this experiment we use the MODTRAN algorithm and three variants of ARRT: the standard ARRT algorithm described in the previous sections, the ARRT algorithm without the upwelled radiance estimate (ARRT w/o L_{up}), and an adjusted ARRT algorithm where the amplitude has been matched perfectly to the extracted target signatures (ARRT Adj). The ARRT variants were added to identify the benefits of estimating the upwelled radiance term and to test the importance of obtaining a correct estimate of amplitude.

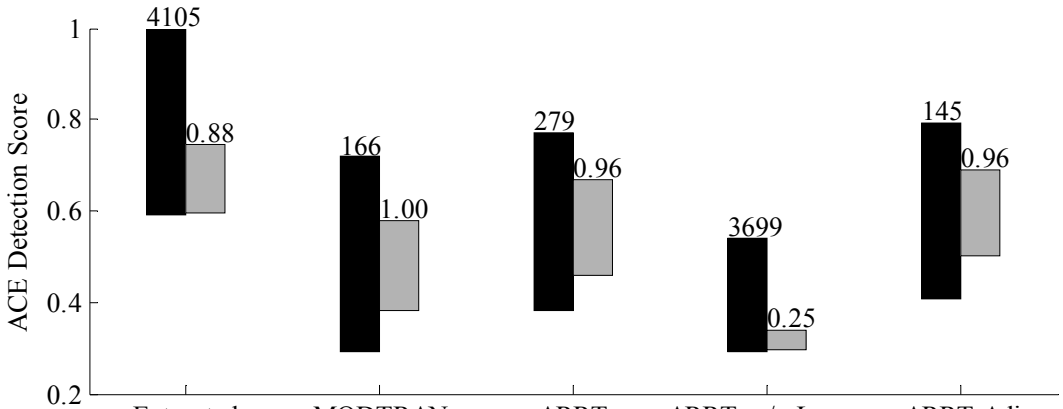
Figure 16 shows the ACE detector results for the estimated target signatures. Each figure contains black and gray vertical bars. The black bars show the range of detection values for the background. The gray bars show the range of detection values for the targets. Ideally, these bars should not overlap indicating the targets are completely separable from the background. Above the black bar, a number is posted identifying how many false alarms occur above the minimum target detection score (i.e., the number of false alarms that are in or above the range of target detection scores). Above the gray bar, a number is posted indicating the percentage of target detected in the image.

Results for Target 3 show all the target estimates are well matched to the targets in the image. The ARRT estimates achieve the ideal case separating the target from the background easily. The MODTRAN signature generated 4 false alarms, but

this was to be expected as it was not as accurate in both shape and amplitude as the ARRT signatures. Even with 4 false alarms, the performance is only marginally worse than using the ARRT signatures.



(a) Target 3 Results



(b) Target 4 Results

Figure 16: ACE Results for Image 7
for (a) Target 3 and (b) Target 4

Results for Target 4 are much more interesting. First, Target 4 is a difficult target to detect because of its low reflectance signature. Not surprisingly, the false alarm counts are significantly higher with this target than with Target 3. The MODTRAN signature provides the best performance outperforming the "true" signature estimated from the mean of the target detections in the image. ARRT provides good detection performance, but has 68% more false alarms. As expected,

the ARRT estimate without the upwelled radiance term performs the worst giving an abysmal 25% P_d providing evidence that the upwelled radiance term is important to subpixel detection applications. Another interesting result is the last set of bars. These results were generated using an ARRT signature that was corrected to have the same amplitude as the target signature taken from the image. The results for this signature rival the performance achieved with MODTRAN. Thus, amplitude plays a considerable role in achieving good subpixel detection performance.

On a final note, the true target signature for Target 4 does not perform as well as most of the target radiance estimates. This is not surprising however given the size of Target 4 in Image 7. Since targets span only 3.6 pixels, most likely some "target" pixels were identified that contained some background materials as well. Thus, the "real" target signature is compromised and this leads to the degraded performance. Another result from this experiment is that even with multi-pixel targets that contain few pixels; atmospheric compensation algorithms may provide a better estimate of the target than can be drawn from the image with known ground truth.

3.4.2.2. Comparison of Subpixel Detection Performance

Image 7 provided us the opportunity to compare target signatures generated using atmospheric compensation algorithms to their true signatures in an image. Unfortunately, the analysis could not provide performance estimates for actual subpixel targets. To provide this type of analysis, we compare the MODTRAN, ARRT, and ARRT without L_{up} on Images 1 through 6 from Sensor X. These images were collected at an altitude of 1220 m so that each pixel imaged approximately 0.1820 m^2 of area. The result of the higher altitude is that the targets have fill factors

(percent of the pixel occupied by target material) ranging from at most 60% to as low as 11%.

As was done in the previous experiment, ACE was applied to the data for the various target types and target radiance estimates. A clustering threshold of 1% was used to form the objects that were identified as either target or clutter using the provided ground truth. Some target did span multiple pixels, but did so with smaller fill factors (e.g., Target 3 has a 60% fill factor that can be split across two pixels as 20% and 40%).

Instead of bar graphs to analyze performance, receiver operating characteristic (ROC) curves were used. These ROC curves were generated across all images so enough targets would be available to make a meaningful ROC curve. As is typical, the y-axis measures the P_d normalized to 1. The x-axis, however, is a measure of false alarm density. This metric is the number of false alarms divided by the total area imaged. Curves for detectors that achieve false alarm densities of 10^{-3} or lower with 50% Pd are considered good performers.

Figure 17,Figure 18, and Figure 19 display the ROC curves for Targets 1 through 3 respectively. In all cases, ARRT performs as well as MODTRAN. This shows that an in-scene technique can perform as well as a complicated model-based technique for subpixel detection performance. This result is expected given the good results seen on Target 3 in the earlier experiments. Additionally, Targets 1 through 3 have moderate to strong reflective signatures as shown in Figure 6. Because the signatures have good reflectance, the algorithms are less prone to small errors and provide good radiance estimates in all cases.

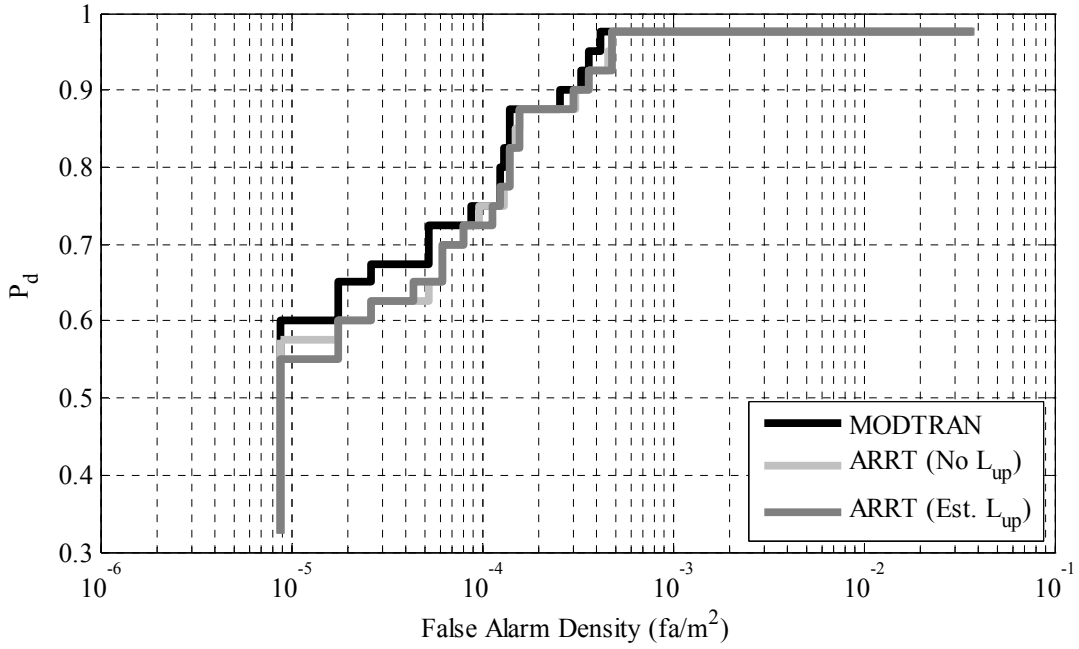


Figure 17: ROC Comparison of Target 1 Signatures

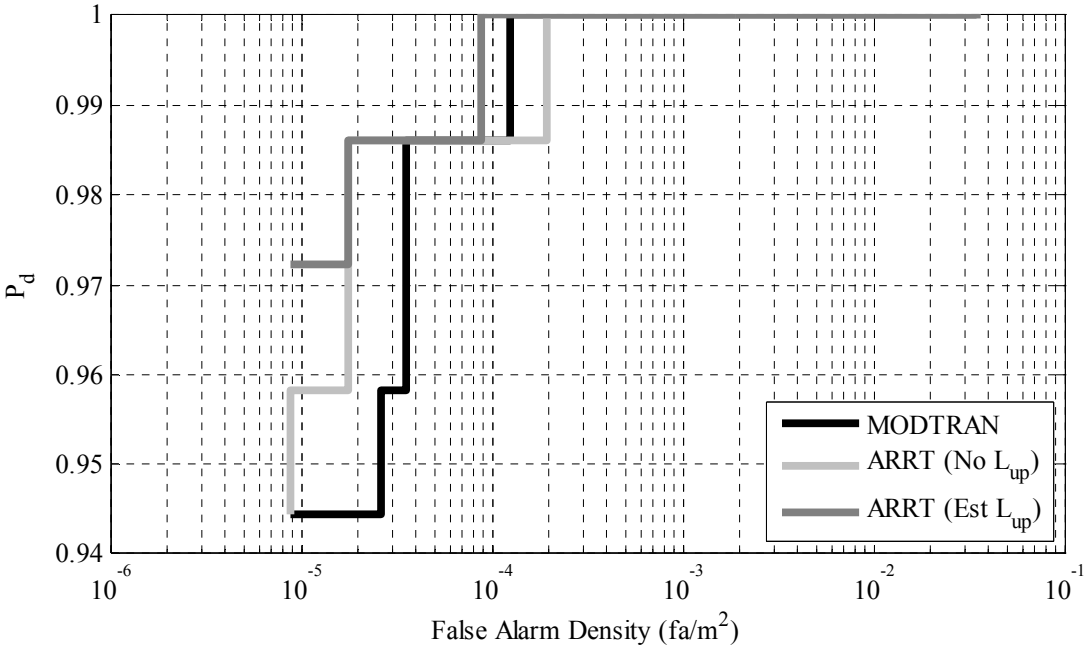


Figure 18: ROC Comparison of Target 2 Signatures

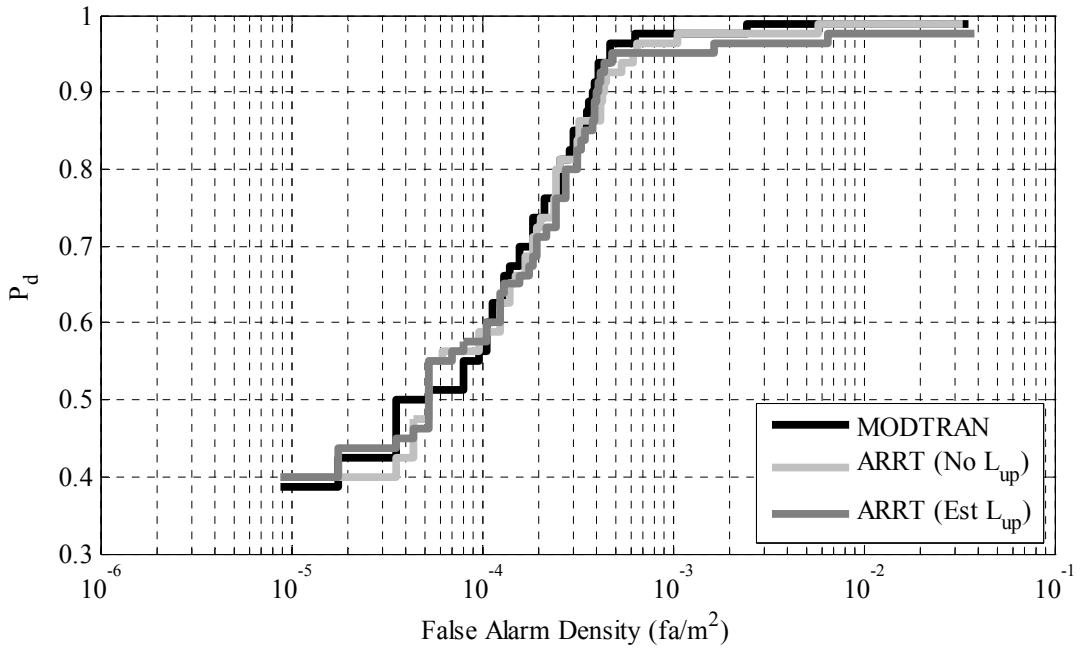


Figure 19: ROC Comparison of Target 3 Signatures

Target 4 is the difficult target. As mentioned in the previous section, the target has a weak reflectance signature making it hard to detect at an altitude of 313m. At 1220m altitudes, the target becomes very difficult to detect. None of the detectors with any target estimate perform well although MODTRAN performs the best as expected. Model-based methods are somewhat immune to sensor collection errors and tend to perform better with low reflectance targets [93]. In-scene methods tend to degrade with such targets as even small errors can seriously affect the shape and amplitude of the estimated target signature which leads to degraded detection performance. Therefore when dealing with weak target signatures, model-based methods still have an advantage over in-scene methods as has been previously documented [93]. This statement holds true for ARRT as well.

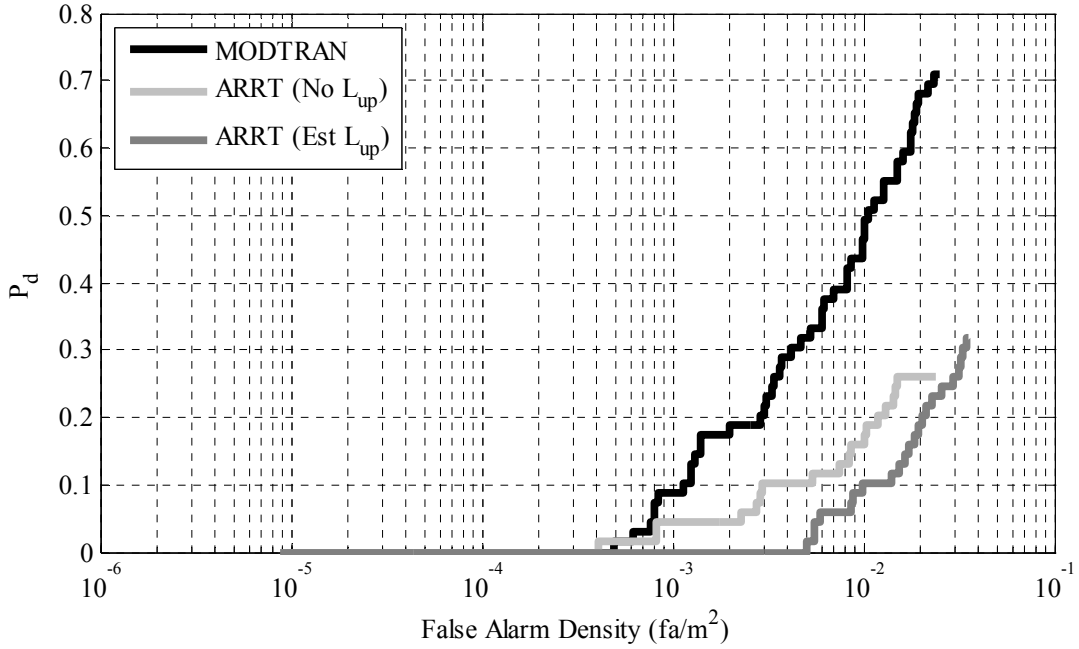


Figure 20: ROC Comparison of Target 4 Signatures

3.5. Summary

Characterization of the target radiance signature is a key part of subpixel detection. Many ways have been developed over the years to estimate the atmospheric transfer function at the heart of target characterization. This work presents a new in-scene algorithm ARRT for characterizing target radiance signatures using only the image and a reference reflectance signature. The algorithm uses detection theory and radiative transfer theory to project a target reflectance signature into the radiance seen at the sensor.

The ARRT algorithm provides a number of advantages over other methods. First, ARRT provides radiance signatures in a fraction of the time of model-based methods since ancillary information such as weather and source-target-receiver geometry are not used. Second, ARRT generates signatures that rival model-based methods. Third, the signatures generated by ARRT have been shown to provide good

subpixel detection performance over a variety of targets. Finally, sensor calibration issues which are problematic for model-based methods pose no problem for in-scene methods such as ARRT. These traits make ARRT very attractive for applications where a model is simply not feasible and or the ancillary information cannot be obtained.

While ARRT does have the aforementioned attractive properties, it also has its limitations. ARRT is meant for aerial imagery as opposed to satellite data or images taken at extreme oblique angles. Additionally, the imagery must contain pure background pixels with moderate to high reflectance signature to estimate the amplitude of the target radiance signature. As expected, ARRT like other in-scene methods has difficulty estimating signatures with low reflectance. However, even in this extreme case, model-based methods perform only marginally better.

Chapter 4: Background Signature Characterization

Target characterization is an important aspect of any detection algorithm. In subpixel detection, however, characterization of the competing background signatures within the pixel is just as important. Unlike conventional full-pixel detection where the pixel contains target or background signatures, subpixel targets are a combination of the target and the competing background signatures as described by the linear mixing model in Chapter 1. Having developed a way to characterize the target signature, we must now focus our attention on characterizing the background signatures.

Unlike target characterization where we have a known target signature, we do not know a-priori all of the background materials in an image. Instead, these background materials must be estimated. While one could use land class maps to identify the main background components in any area, these maps are typically coarse and cannot capture the material variability that may be in the scene. Thus, most subpixel detectors rely on endmember extraction methods which adaptively estimate the background endmembers from the image.

This chapter begins by providing an overview of endmember extraction techniques. The first section describes some of the many endmember extraction techniques available to the community today. While this is not an exhaustive list, it does provide examples of the fundamentally different ways background endmembers can be estimated. From this list, we identify the two endmember extraction techniques we use for the remainder of the dissertation and motivate why we selected them.

In the following sections of the chapter, we discuss the importance of estimating the correct number of endmembers for subpixel target detection purposes. We argue that this topic has been largely ignored by the community based on the different ways researchers have estimated the number of endmembers. We introduce the various state-of-the-art methods from intrinsic dimensionality to virtual dimensionality statistics. We present two of our own proposed methods for estimating the number of endmembers arguing that the estimate should be based on both the endmember extraction algorithm and the desired target signature. We compare our methods to the current state-of-the-art methods showing appreciable gains in a number of experiments. Through these comparisons, we also show how important correct estimation of the number of endmembers is to subpixel detection performance.

4.1. A Review of Endmember Extraction Methods

A number of algorithms have been developed to adaptively estimate the endmembers in an image. A review of the literature shows how many different algorithms exist including Pixel Purity Index (PPI) [9], N-FINDR [116], the Simulated Annealing Algorithm (SAA) [7], Optical Real-Time Spectral Identification System (ORASIS) [37], Iterative Error Analysis (IEA) [77], and Automated Morphological Endmember Extraction (AMEE) [81] to name just a few. A good review of various endmember extraction algorithms can be found in [82]. The intent of this section is to simply and quickly describe the different ways endmembers can be extracted from HSI data.

The PPI, N-FINDR, and SAA algorithms are geometry-based methods. These algorithms project the HSI data into a smaller dimension d using methods such as the

Maximum Noise Fraction (MNF) transform [36]. After the transformation, the algorithms have slightly different approaches. PPI generates random lines onto which the transformed data is projected. The outliers on each line are counted and the process is repeated many times identifying those pixels that continue to be outliers as endmembers. An operator takes this result and uses a *d*-dimensional visualization tool to identify the final number of endmembers. N-FINDR finds the endmembers as the *d+1* vertices of the simplex that contains the maximum amount of the transformed data. N-FINDR is computationally efficient and can be performed in near real-time without operator intervention. SAA is very similar to N-FINDR in that it also identifies endmembers as the vertices of a simplex enclosing the transformed data. Unlike N-FINDR though, SAA creates "virtual endmembers" when no pure pixels are present in the image. This generation of virtual endmembers using a simulated annealing algorithm guarantees endmembers that are pure material spectra. This is also an automatic extraction technique, but is more computationally expensive than N-FINDR due to the simulated annealing.

ORASIS is both a vector quantization method and geometric method. This algorithm developed by the U.S. Naval Research Laboratory (NRL) operates in real-time using a two step process. The first pass reduces the volume of the HSI data using a learning vector quantization (LVQ) process [10]. Using LVQ, exemplar signatures are adaptively found from the image using a distance metric (typically the SAM metric [54]). Once the exemplars are found, a modified Gram-Schmidt process called salient selection is used to project the exemplars onto a smaller dimensional orthogonal subspace. The algorithm identifies the endmembers as those that make up

the vertices of the simplex that encloses the projected data similar to N-FINDR and SAA; however because of the LVQ preprocessing step, this algorithm can run in real-time.

AMEE is a joint spatial and spectral morphological approach to endmember extraction. In this method, no subspace projection is necessary. Instead, the image is iteratively processed using spatial morphological kernels of various sizes. At each pixel location, the spectrally purest and spectrally most mixed pixels are found. The morphological eccentricity index (MEI) is calculated as the angles between these pure and mixed pixels. This is repeated for multiple kernel sizes until an MEI image is created. Segmentation takes place on the MEI image and the endmembers are those chosen from the image after a spatial and spectral growing procedure occurs which removes variability within each spectral class.

The IEA algorithm extracts physically meaningful endmembers that are based on minimizing the mean squared error between the actual image and an unmixed image. The algorithm begins with the target signature and unmixes the image (estimates the endmembers and corresponding abundances) using the Fully Constrained Least Squares algorithm [46] (further details can be found in Chapter 5). An error image is created between the original image and the unmixed image generated using (1). The mean of the pixels that contain the largest mean squared error are chosen as the next endmember. Extraction continues until N number of endmembers is found.

There is another class of endmember extraction methods based on statistical models. Parametric statistical models include the stochastic mixing models (SMM)

[104] based on expectation maximization methods and the Modified Spectral Mixture Analysis (MSMA) which is an approach similar to the SAA algorithm [110]. Non-parametric statistical algorithms have also been used to extract endmembers such as K-Means clustering [29].

4.2. Selected Endmember Extraction Techniques

To characterize the background for subpixel target detection, we are interested in finding an endmember extraction technique that 1) performs well, 2) produces physically meaningful endmembers, and 3) is fully automatic. Using the research from [77] and [82], we decided on a variant of the IEA algorithm for multiple reasons. First, the IEA algorithm produces physically meaningful endmembers that are well matched to the FCLS algorithm – an abundance estimation algorithm that will be used in our subpixel detectors described in Chapter 5. Second, the algorithm provides endmembers that are significantly different from the target signature minimizing the change of background signatures "bleeding" into the target subspace. Third, the algorithm runs quickly taking only a few minutes to extract 30 endmembers. Fourth, the IEA algorithm was identified as one of the best performing endmember extraction techniques in [82]. Since the IEA algorithm is also fully automatic, it meets all of our criteria.

We use another technique defined by the popular Adaptive Matched Subspace Detector (AMSD) – a baseline subpixel detector used in Chapter 5. We use this method because the AMSD algorithm specifically identifies this method be applied in its detector [71][109]. This technique does not extract physical endmembers. Instead it performs an eigenvector decomposition of the image correlation matrix. The resulting eigenvectors comprise the endmembers for the background. Note that while

these endmembers are not physically meaningful, they do minimize the mean squared error when used with the AMSD algorithm. We only use this method for the AMSD algorithm as it does not provide physically meaningful endmembers for our physics-based approach.

4.3. Dimensionality of Hyperspectral Imagery

In addition to the extraction of endmembers, a significant amount of research has gone into identifying the correct number of endmembers for a scene. Most algorithms have focused on what has been termed "intrinsic" dimensionality [19]. These dimensionality measures focus on identifying the unique spectral signatures in an image. For classification purposes, it is important to estimate the intrinsic dimensionality. For target detection applications, intrinsic dimensionality may not be the best measure.

In target detection, the background must be characterized such that the probability of detecting the target is maximized while the probability of detecting a false alarm is minimized. In such cases, the number of endmembers required to characterize the background may be significantly more than the intrinsic dimensionality. The reasons are varied, but can be quickly summarized as the additional endmembers may be signatures due to shadowing effects, sensor artifacts, and finer material identification (e.g. coarse sand vs. fine sand). This has been noted in [19] where the best number of endmembers varied for different applications. This measure of dimensionality relative to detection performance has been termed virtual dimensionality [19].

The next two sections describe the different metrics used to select the "best" number of endmembers from a scene. The intrinsic dimensionality measures are

energy, Akaike Information Criterion (AIC), Minimum Description Length (MDL), and Empirical Indicator Function (EIF). The virtual dimensionality measures are based on work by Chang and Du [19], Thai and Healey [109], and two we propose for subpixel detection applications.

4.3.1. Intrinsic Dimensionality Metrics

4.3.1.1. Energy Metric

This metric is used by Manolakis, Siracusa, and Shaw for the AMSD algorithm [71]. In this paper, they characterize the background as the eigenvalue decomposition of the image correlation matrix. The resulting eigenvalues are sorted in decreasing order. The number of endmembers used is calculated using the sorted eigenvalues such that

$$\hat{m} = \min_{m} \left(\frac{\sum_{i=1}^{m} \lambda_i}{\sum_{j=1}^{M} \lambda_i} \geq 0.99 \right) \tag{18}$$

where M is the total number of endmember extracted and λ_i is the i^{th} ordered eigenvalue.

4.3.1.2. MDL Metric

A set of metrics was developed to estimate the order of a statistical model. One of the first was the AIC published by Akaike in 1974 [2]. The AIC statistic was found to be inconsistent [51] and this led to other works by Rissanen using an information-theoretic criterion [87] and by Kashyap [50] and Schwartz [96] using a Bayesian framework. The researchers independently came to the same result: the Minimum Description Length (or Bayesian Information Criterion as Schwartz identified it). The criterion is

$$\hat{m} = \min_{m}\left(-\log L(x,\alpha_m) + \tfrac{1}{2}k\log N\right) \tag{19}$$

where $L(x,\alpha_m)$ is the statistical likelihood function parameterized by α_m, k is the number of free parameters that must be estimated, N is the number of samples used to estimate the likelihood and its associated parameters, and m is the dimension of the parameters.

Chang and Du used Wax and Kailath's MDL criterion in their research [113]. The results showed poor performance because of two reasons. First, the Wax and Kailath work was designed for time series data where each sample came from an iid zero-mean Gaussian distribution; therefore, the combined likelihood could be expressed entirely in terms of the data covariance matrix. HSI data does not fit this assumption as mentioned in [71] and [103]. Second, Chang and Du used the equation directly from Wax and Kailath [113] which was designed for complex data. HSI data is real-valued and hence the equation they used was inappropriate. Instead, the equation should have been

$$\hat{m} = \min_{m}\left(-\log\left(\frac{\sum_{i=1}^{L}\lambda_i^{1/(L-m+1)}}{\frac{1}{L-m+1}\sum_{i=1}^{L}\lambda_i}\right)^{(L-m+1)N} + \tfrac{1}{4}m(1+2L-m)\log N\right) \tag{20}$$

where λ_i are the eigenvalues of the image covariance matrix, L is the number of spectral bands, and N is the number of pixels in the image. Nevertheless, in all of our experiments, the Wax/Kailath implementation never achieved a minimum (using Wax and Kailath's original equation or (20)).

4.3.1.3. EIF Metric

Malinowski created a metric specifically designed to estimate the number of unique spectra in chemical spectroscopy studies [68]. Using empirical studies based on chemical factor analysis, he created an empirical indicator function (EIF) such that

$$\hat{m} = \min_{m} \left(\frac{\left(\sum_{i=1}^{M} \lambda_i \right)^{1/2}}{N^{1/2} (L - m + 1)^{3/2}} \right) \tag{21}$$

where λ_i are the eigenvalues of the $L \times M$ endmember matrix, M is the total number of endmembers, L is the number of spectral bands, and N is the number of pixels in the image.

4.3.2. Virtual Dimensionality Metrics

4.3.2.1. NSP Metric

The term "virtual dimensionality" was coined by Chang and Du [19]. In this paper, they presented a new way to assess the dimensionality of HSI data relative to classification and detection performance. Interestingly, the Noise Subspace Projection (NSP) metric they developed uses no information about the target or the detector. They do, however, form a binary hypothesis test based on the eigenvalues of the whitened image covariance matrix.

The algorithm begins by estimating the image covariance matrix from the data. The inverse of the covariance matrix is decomposed such that

$$\mathbf{C}^{-1} = \mathbf{DED} \tag{22}$$

where **D** is a diagonal matrix created from the square root of the diagonal elements of $\mathbf{C}^{-1}$ and **E** is a matrix of correlation coefficients of $\mathbf{C}^{-1}$. Using this decomposition, the whitening matrix is defined as

$$\mathbf{W} = \mathbf{D}^{-1}. \tag{23}$$

Using (23), the image covariance matrix is whitened such that

$$\mathbf{C}_{\mathbf{W}} = \mathbf{WCW}. \tag{24}$$

The whitening is performed to reduce the correlations inherent between spectral bands. The whitened matrix is analyzed using Principal Component Analysis (PCA) to extract the eigenvalues for the binary hypothesis test. The hypotheses are

$$\begin{aligned} H_0 &: \lambda_i = 1 \\ H_1 &: \lambda_i > 1 \end{aligned} \tag{25}$$

for each i^{th} eigenvalue. The likelihood function for the null hypothesis is simplied to

$$p_o(\lambda_i) \cong N\left(1, \tfrac{2}{N}\right). \tag{26}$$

Using (26), a threshold can be calculated for a given false alarm probability. Because (26) is independent of the index *i*, the same threshold can be applied to all eigenvalues. Using this information, the number of endmembers can be found using

$$\hat{m} = \max_m \left(\frac{(\lambda_m - 1)\sqrt{N}}{\sqrt{2}} \geq \Phi^{-1}(1-p) \right) \tag{27}$$

where $\Phi^{-1}(1\text{-}p)$ is the inverse of the standard normal cumulative density function (cdf) evaluated at probability *1-p*. From [19], they recommend a value of 0.001 for *p*.

4.3.2.2. Thai/Healey Metric

This metric was developed as an aside in Thai and Healey's invariant subpixel detection paper [109]. The paper is another variant of the AMSD algorithm where the

target subspace is created using Healey and Slater's invariant method [45]. Thai applied this invariant method to subpixel detection and independently derived the AMSD algorithm [71]. Unlike Manolakis, Siracusa, and Shaw [71]who depended on the energy estimate described earlier, Thai and Healey designed a new metric to choose the dimension of their background subspace.

The basic idea is to find the number of endmembers that maximize target detection while minimizing the background. To accomplish this, they created a ratio of AMSD statistics such that

$$\hat{m} = \max_{m \in \mathbf{M}} \left(\frac{\delta_{AMSD}(\overline{\mathbf{S}})}{\delta_{AMSD}(\boldsymbol{\mu})} \right) \tag{28}$$

where $\delta_{AMSD}(\mathbf{x})$ is the AMSD statistic given in Chapter 5, $\overline{\mathbf{S}}$ is the mean of the target signatures, and $\boldsymbol{\mu}$ is the adjusted spectral mean of the image. The adjusted spectral mean is calculated from all the pixels in the image except those whose matched filter score is near one. The set of $\mathbf{M}$ restricts the values of m based on the mean squared error between the original image and the PCA decomposition of the image. Thus, at least m_1 eigenvectors are always used, but not more than m_2 eigenvectors. The reasoning is that the number of eigenvectors that make up the background must be numerous enough to minimize the mean squared error, but not so numerous that the eigenvectors are pure "noise." No discussion is provided on how to derive these limits, or the threshold used in the matched filter.

4.3.2.3. AMSD MDL Metric

The first proposed metric fuses the ideas from the AMSD detector and MDL criterion. The original MDL equation in (19) can be formed for any likelihood. The method used by Du and Chang in their paper assumed that the HSI data could be fully

modeled by the image covariance matrix following Wax and Kailath's work. Unfortunately, this approach is not applicable to hyperspectral analysis as previously discussed in [19]

Instead of the image covariance matrix, we propose using the AMSD likelihood directly in the MDL criterion. This would match the criterion to the specific detector and all of its implicit assumptions. However, with any detector, there are two likelihoods: one for the null hypothesis and one for the alternate hypothesis. For this criterion, we use the alternate hypothesis which includes the target signature(s). The reasoning is the alternate hypothesis includes information about the target signature as well as the detector. Therefore, combining the MDL criterion with the alternate AMSD likelihood is

$$\hat{m} = \min_{m}\left(L\sum_{i=1}^{N}\left(\mathbf{x}_i^T(\mathbf{I}-\mathbf{E}_m(\mathbf{E}_m^T\mathbf{E}_m)^{-1}\mathbf{E}_m^T)\mathbf{x}_i\right)+m(N+1)\log N\right) \tag{29}$$

where $\mathbf{E}_m$ is the concatenation of the target and m background signatures, L is the number of spectral bands, and N is the number of pixels in the image.

4.3.2.4. Subpixel Dimensionality Metric

In the MDL AMSD criterion, the idea was to identify the number of endmembers that minimized the likelihood of the denominator. This is only part of the optimization problem however. Ideally, the number of endmembers should also maximize the numerator. Interestingly, this is the same optimization done in detection theory; so, we can use the detector directly to estimate the number of endmembers. Following the approach of Thai and Healey [109], the subpixel dimensionality metric is

$$\hat{m} = \max_{m \in \mathrm{M}} \left(\delta_{AMSD} \left(a\overline{\mathbf{S}} + (1-a)\boldsymbol{\mu} \right) \right) \tag{30}$$

where the dimensionality is chosen by maximizing the AMSD detection score for a simulated pixel that is a linear combination of the desired target spectra and spectral mean from the image. The abundance *a* is calculated a-priori given the size of the target and the size of the pixels based on the altitude and the sensor's field of view parameters.

This approach has a number of advantages. First, like Thai and Healey's method, the metric can be quickly calculated for all numbers of endmembers. Second, the statistic directly uses the detector accounting for application dependencies unlike the intrinsic dimensionality metrics. Third, the metric chooses the number of endmembers based on both the predicted size and spectral characteristics of the target.

4.4. Experimental Results

Endmember extraction algorithms have been compared in a number of papers [81],[82],[116], but little experimentation has been performed on the impact of background dimensionality on subpixel target detection performance. This section compares the different methods of background dimensionality estimation and their impact on subpixel target detection. The goal is to identify which methods provide good dimensionality estimates for subpixel detection and under what conditions.

The experiments are broken into two parts: individual image results and ROC results. The individual image results present P_d and P_{fa} results for each image and target type. The ROC performance provides results across all images including those that do not contain targets. All the experiments use the Sensor X data for Images 1

through 6 and Targets 1 through 4. As mentioned in Chapter 2, Targets 1 and 2 are relatively easy to identify. Target 3 is more difficult because of the inherent variability in the spectral signature. Target 4 is very difficult to detect due to its low reflectance.

The detector used for these experiments is the AMSD detector described in Chapter 5. This is a standard structured subpixel detector in the literature that uses the eigenvectors of the image correlation matrix as the background endmembers. This type of detector allows us to apply all of the background dimensionality estimates on similar background information (image covariance or image correlation matrix).

4.4.1. Individual Image Results

Tables 5 through 8 provide the results of the individual image experiments for Targets 1 through 4 respectively. In each table, the number of endmembers (m), the P_d, and the number of false alarms (FA) are provided for each of the background dimension estimates described in Section 4.3. The ideal case is also provided in the last column. This case was found using the known ground truth to find the number of endmembers providing the highest P_d while minimizing the number of false alarms. Each table includes only the images in which targets are present.

The results show some intriguing results. First, the energy metric does not perform well as expected. For this implementation, we required 99.9% of the energy be obtained leading to background estimates of 2 to 3 endmembers. Unfortunately in radiance space, these first few eigenvectors comprise most of the environmental effects. This has the effect of providing little separation between target and background for all target types. The P_d is typically low with high false alarm rates.

The impact of this finding shows that papers using this metric [60],[71] are biasing their results against the AMSD detector.

Table 5: Comparison of Dimensionality Estimates for Target 1

Metric	Image	Energy	EIF	NSP	Thai/ Healey	AMSD MDL	SDD	Ideal
m	1	3	103	97	12	11	121	68
	4	2	105	97	39	11	131	125
P_d	1	0.15	1.00	1.00	0.95	1.00	1.00	1.00
	4	0.05	1.00	1.00	1.00	1.00	1.00	1.00
FA	1	36	0	0	280	334	2	0
	4	66	0	0	10	383	0	0

Table 6: Comparison of Dimensionality Estimates for Target 2

Metric	Image	Energy	EIF	NSP	Thai/ Healey	AMSD MDL	SDD	Ideal
m	1	3	103	97	5	11	111	11
	4	2	105	97	6	9	133	7
P_d	1	0.74	1.00	1.00	0.71	1.00	1.00	1.00
	4	0.90	1.00	1.00	1.00	1.00	1.00	1.00
FA	1	175	253	130	212	27	360	27
	4	271	34	17	5	43	92	0

The Thai/Healey and AMSD MDL metrics perform poorly as well. This is a surprising result as these metrics use knowledge of the target signature and detector type to estimate the background dimension. The Thai/Healey metric degrades significantly as the targets become more difficult to detect. Even on the simpler targets, the P_d is less than the other methods with higher false alarm densities. The reason this occurs is because the targets are truly subpixel, but the metric assumes a full pixel target. This causes a mismatch between what is being estimated and what is present in the data. We would expect the metric to perform well on full-pixel targets even though it was developed for subpixel target applications.

Table 7: Comparison of Dimensionality Estimates for Target 3

Metric	Image	Energy	EIF	NSP	Thai/ Healey	AMSD MDL	SDD	Ideal
m	2	3	102	89	18	12	61	31
	3	3	104	89	1	4	82	108
	5	3	106	99	6	11	104	79
	6	3	106	98	1	11	103	117
P_d	2	0.00	1.00	1.00	0.92	0.75	0.92	1.00
	3	0.56	1.00	1.00	0.12	0.68	1.00	1.00
	5	0.20	0.93	0.93	0.67	0.73	0.93	1.00
	6	0.57	0.96	0.93	0.07	0.96	0.96	1.00
FA	2	332	108	711	58	340	248	21
	3	492	3	9	44	112	17	0
	5	13	28	107	35	107	237	339
	6	139	6	5	55	81	4	0

The AMSD MDL criterion also degrades significantly as the targets become more difficult to detect. For Target 1, the criterion does find all the targets, but also provides the highest false alarm numbers. For Target 2, AMSD MDL outperforms the other metrics obtaining estimates close to the ideal. On the last two targets, the AMSD MDL estimate degrades losing significant P_d and obtaining large false alarm densities. The estimates vary because the metric is only treating the denominator of the AMSD statistic without reference to the effect of the numerator. In Target 2, this is not a significant problem, but for all other target types, the numerator decreases as quickly as the denominator causing the metric to erroneously pick the wrong number of endmembers.

The last three metrics (EIF, NSP, and SDD) perform well. The EIF criterion does well without any information about detector type or target signature. This is an interesting result as the other two methods are virtual dimensionality statistics. However, the EIF criterion was developed for identifying the number of spectral

signatures in chemical spectroscopy. This idea seems to have merit when applied to optical spectroscopy even with lack of target and detector knowledge.

Table 8: Comparison of Dimensionality Estimates for Target 4

Metric	Image	Energy	EIF	NSP	Thai/ Healey	AMSD MDL	SDD	Ideal
m	2	3	102	89	4	8	88	6
	3	3	104	89	3	13	83	6
	5	3	106	99	6	10	87	134
	6	3	106	98	1	11	85	70
P_d	2	0.33	0.00	0.00	0.11	1.00	0.00	1.00
	3	0.00	0.22	0.43	0.00	0.00	0.09	0.52
	5	0.42	0.25	0.17	0.00	0.17	0.00	0.50
	6	0.04	0.04	0.12	0.00	0.04	0.00	0.28
FA	2	279	786	766	186	234	721	230
	3	444	789	581	444	230	677	472
	5	109	649	623	429	450	661	898
	6	310	516	493	20	210	755	673

The NSP algorithm which was developed for HSI data performs well. The estimate provides some of the lowest false alarm densities for Targets 1 and 2 while maintaining 100% P_d. As with the other methods, NSP breaks down as the targets become more difficult; however, it does not degrade as fast as energy or AMSD MDL. NSP, in fact, maintains the highest number of target detections on Target 4.

The final algorithm is the proposed SDD metric. This metric performs similarly to the EIF and NSP metrics. As expected, the performance of this metric is directly linked to the detector performance. When the targets become more difficult for the detector to find, this metric degrades as well. The unfortunate outcome of this result is that is provides some of the worst performance on Target 4, but some of the best performance on Target 1.

The overall results of these experiments are mixed. The energy metric is not desirable due to its poor performance across images. The AMSD MDL metric is not desirable due to its variable performance that is uncorrelated with the difficulty of the target type. Thai and Healey's metric which does not account for the subpixel nature of the target provides poor estimates as well. The EIF, NSP, and SDD metrics perform well and degrade gracefully as the target becomes more difficult to detect.

4.4.2. ROC Results

The results from the first experiment show the EIF, NSP, and SDD estimates perform similarly well when applied to images with targets. To see if any separation exists between these methods, it is interesting to look at cases where images that do not contain targets are used. To measure the effect of such images, we use ROC curves.

ROC curves show the average performance of the detector across all images. Ideally, the number of endmembers used should help suppress the background pixels into the same range of detection scores. This allows the ROC curve to apply the same threshold across each image and get similar results. When the background is not confined to the same range of detection scores, the background detection scores on one image may actually be higher than the target detection scores on another image. In such cases, the inconsistency of the detection scores will negatively impact ROC performance. Thus, the impact of images without targets can be assessed on the overall detector performance.

Figures 21 through 24 provide the ROC curves for Targets 1 through 4 respectively. In each figure, there are seven curves. The first curve represents the ideal based on ground truth information obtained with the imagery. The other six

curves represent the estimation algorithms defined in Section 4.3. Similar to the previous experiment, the ideal number of endmembers was selected as those that maximized P_d while minimizing the number of false alarms. In the images without targets, the ideal was chosen as the number of endmembers that suppressed the detection scores into ranges that were similar to the other images.

As expected given the earlier experimental results, the energy, Thai/Healey, and AMSD MDL criterions did not perform well. While these results do not provide good performance, they do highlight the need for good background dimension estimates. The interesting exception to this rule is the AMSD MDL curve for Target 4. For this target, the AMSD MDL curve is one of the best, but this is most likely a coincidence as the estimate simply favors lower numbers of endmembers.

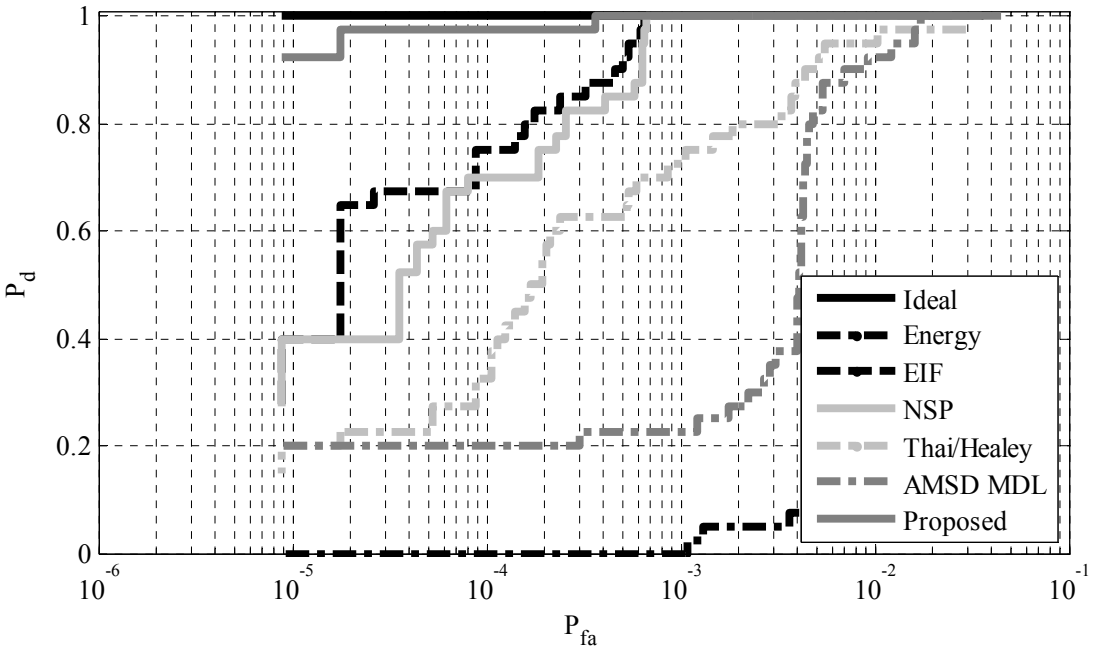


Figure 21: Comparison of Background Dimension Estimates for Target 1

The interesting results occur with the EIF, NSP, and SDD methods. In the first set of experiments, these algorithms perform nearly equally well on the different targets and images. In these ROC experiments however, the algorithms respond

differently. The EIF and NSP methods performance is best with Targets 2 and 3. These targets are easy to moderately difficult to detect. For Target 1, the methods perform significantly worse than the ideal case. For Target 4, the NSP method performs nearly the best although this is again significantly less than the ideal. Nevertheless, both algorithms are consistently some of the best methods for background dimension estimation.

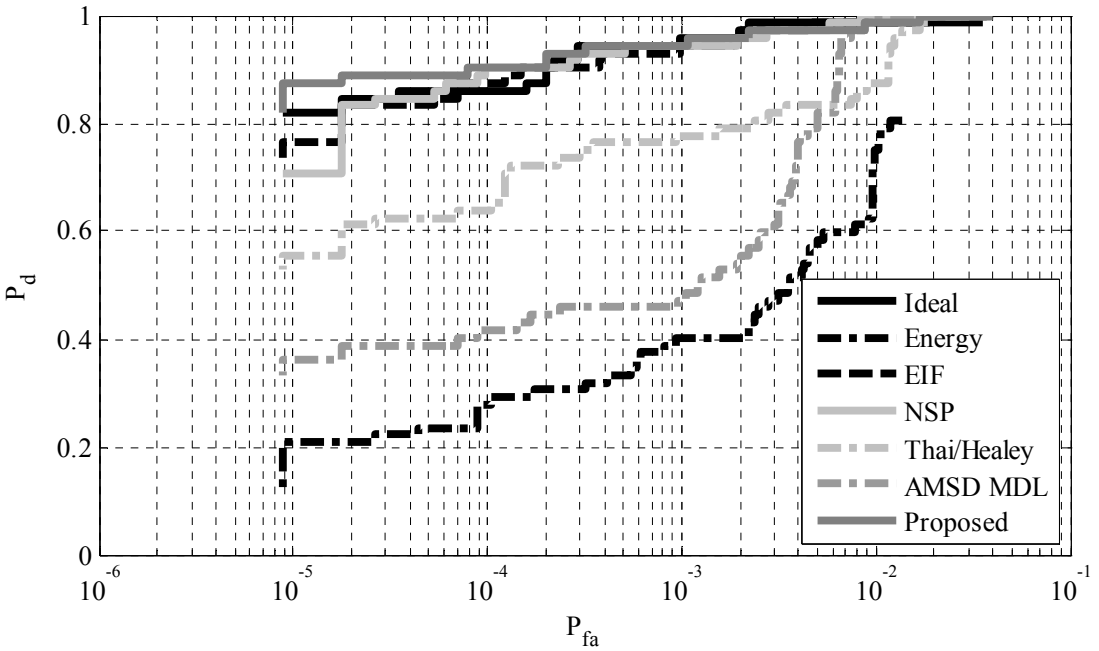


Figure 22: Comparison of Background Dimension Estimates for Target 2

The SDD method demonstrates excellent performance when the target is easy and degrades as the targets become more difficult. This performance is expected given the method is based directly on the performance of the detector using a simulated subpixel target. For Target 1, the SDD method is nearly ideal and substantially better than any other method tested. For Target 2, the method matches the ideal case although the EIF and NSP methods have similar performance. For Target 3, the SDD method degrades slightly as this target is more difficult to detect

due to the spectral variability of the target. On this target, the EIF and NSP methods have a slight advantage. For Target 4 however, the SDD method performs poorly because the detector has difficulty finding such a weak target.

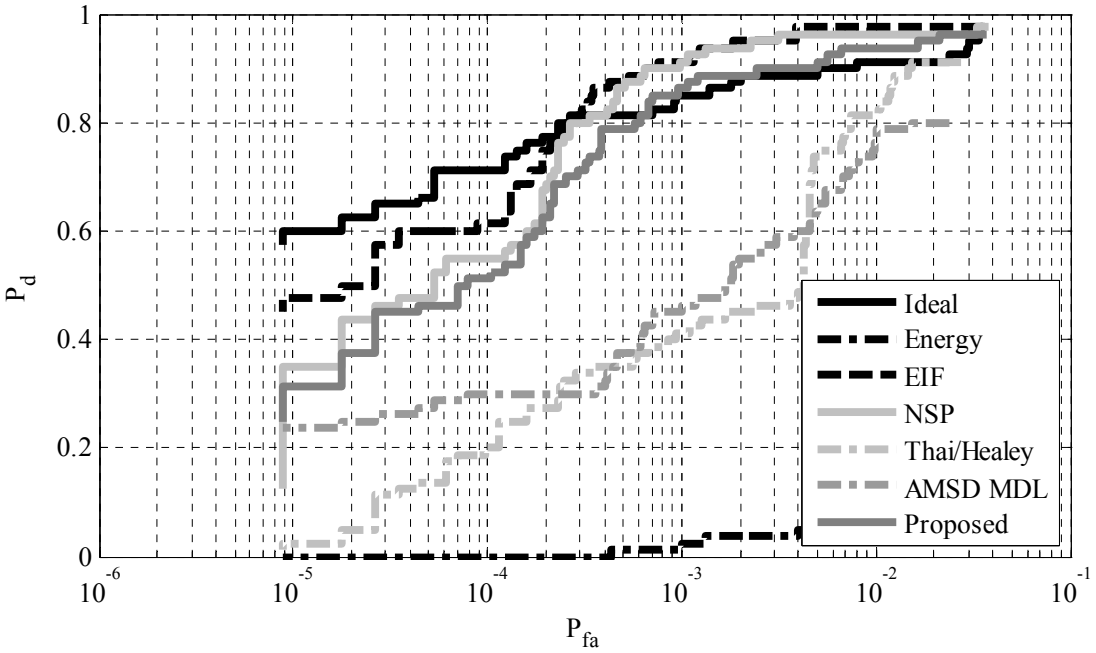


Figure 23: Comparison of Background Dimension Estimates for Target 3

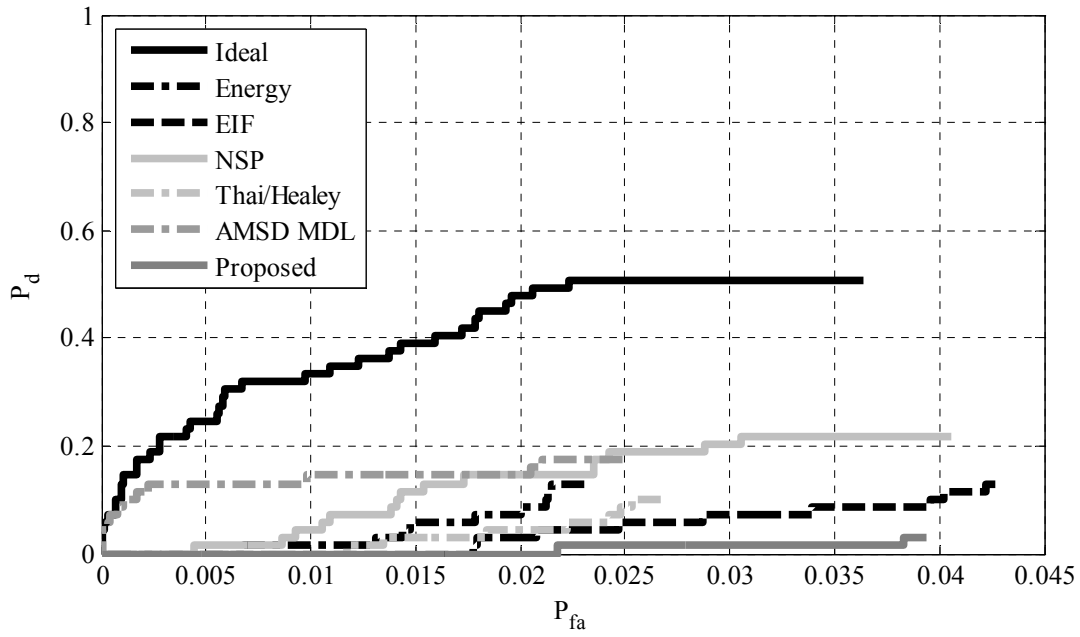


Figure 24: Comparison of Background Dimension Estimates for Target 4

4.4.3. Conclusions

The results of these experiments show that the SDD method has an advantage over the other estimates for all detectable targets. Since the method is based directly on the subpixel detector performance, this result is expected. The EIF and NSP methods are close competitors. These methods show good separation in both the single image and ROC experiments. Since these algorithms were intentionally designed for HSI data, these results are consistent with theory.

The energy, Thai/Healey, and AMSD MDL methods are not good indicators of the background dimension. Energy is the worst indicator although it has been used in numerous papers. The Thai/Healey method does not perform well despite being designed for subpixel processing using the AMSD algorithm. This can most likely be traced to the fact that Thai and Healey used mostly targets that were not subpixel in their paper. For full pixel targets, the method should work well. Unfortunately, the AMSD MDL method did not perform well because it only uses the denominator of the AMSD detector to make its estimate.

4.5. Summary

The estimation of the number of background endmembers for subpixel detection remains a challenging problem. Our work has shown that improvements can be made over the current methods, but these improvements are directly linked to the performance of the detector and the strength of the target signature. In cases where the target signature is well characterized and significantly different from the background, the SDD method we proposed works very well followed closely by the EIF and NSP methods. As the target becomes weaker (or the background becomes more complex), all of the methods degrade.

Further research should be continued to identify better ways to estimate the background dimension. The results clearly show the loss of performance when the background is not correctly identified. Such performance can be significant – especially in the case of weak targets like Target 4. The other direction is to develop detection algorithms that are partially invariant to the number of background endmembers. Such algorithms would show minimal loss in subpixel detection performance due to minor errors in background dimension estimation.

Chapter 5: Physics-Based Hybrid Detectors

A number of different methods have been proposed to address subpixel detection. One of the earliest methods uses array processing techniques to nullify the background signatures as one would nullify an interfering signature when performing beamforming. The Orthogonal Subspace Projection (OSP) [41] and Constrained Energy Minimization (CEM) [20] algorithms are examples of such methods. In order to implement these detectors, the authors assume the noise to be a zero-mean multivariate normal distribution with covariance matrix $\sigma^2\mathbf{I}$. The idea behind this algorithm is that the background can be fully characterized by endmembers and that the remaining noise will meet the aforementioned $\sigma^2\mathbf{I}$ assumption.

Another approach uses the linear mixing model to directly estimate the abundance values and use the estimated target abundances for detection purposes. Two examples of this approach are the Non-Negativity Constrained Least Squares [20] and Fully Constrained Least Squares algorithms [46]. These methods can be considered physics-based methods since they attempt to address all of the phenomenological constraints in the linear mixing model. Others have also incorporated the constraint of a full covariance matrix into these methods which can be considered as the first use of a semi-structured approach [86]. Incorporation of covariance information addresses the fact that most of the spectral bands in HSI data are highly correlated. The estimated covariance is used for designing a whitening transform that decorrelates the bands making the HSI data fit the aforementioned assumption of $\sigma^2\mathbf{I}$. These physics-based methods perform well for both unsupervised estimation of background endmembers and the calculation of the corresponding

abundances; however, they do not provide a statistical hypothesis test – they only provide an estimate of the target abundance.

To develop such a statistical test, a set of hypotheses must be generated to differentiate those pixels containing targets of interest from those pixels that exclusively contain background spectra. The set of hypotheses are

$$\begin{aligned} H_0 &: \mathbf{x} = \mathbf{B}\mathbf{a}_{b,0} + \mathbf{n} \\ H_1 &: \mathbf{x} = \mathbf{S}\mathbf{a}_s + \mathbf{B}\mathbf{a}_{b,1} + \mathbf{n} \end{aligned} \tag{31}$$

where $\mathbf{x}$ is the pixel under test, $\mathbf{B}$ is a $L \times Q$ matrix representing background endmembers, $\mathbf{a}_{b,0}$ and $\mathbf{a}_{b,1}$ are the abundances of the background endmembers under each hypothesis, $\mathbf{S}$ is a $L \times P$ matrix representing target endmembers, $\mathbf{a}_s$ are the abundances of the targets, and $\mathbf{n}$ is a noise model typically assumed to be a zero-mean multivariate normal distribution.

Using this set of hypotheses, a set of detectors has been developed based on structured and unstructured backgrounds. A good example of a structured background detector is the Adaptive Matched Subspace Detector (AMSD) [71]. The AMSD algorithm models the background using the linear mixing model with endmembers and abundances. This statement is a misnomer however since the endmembers in the AMSD algorithm have no physical meaning. Instead the endmembers are the eigenvectors of the image correlation matrix. Thus the abundances are no longer measurements of area. They are simply magnitudes along the eigenvector directions which in general do not satisfy the non-negativity and sum-to-one constraints. So although the linear mixing model is used as the basis for AMSD, all physical considerations are ignored in favor of a purely statistical approach. While AMSD has

shown good performance, research has shown that a purely structured background model does not fully represent the background in real-world HSI data [71].

An example of an unstructured detector is the Adaptive Cosine/Coherent Estimate (ACE) [58]. The ACE algorithm assumes no background signatures opting instead for modeling the background as a multivariate normal distribution. While this removes the need to extract and identify the proper number of background endmembers, it also removes the physical constraints of the linear mixing model. Despite this seemingly simple background model, the ACE detector is one of the more powerful subpixel detectors available for HSI data [70]. Unfortunately, research has shown that an unstructured detector which uses the multivariate normal distribution is not a good model of backgrounds in hyperspectral imagery [103].

Another algorithm that uses the hypotheses in (31) is the Constrained Signal Detector (CSD) [49]. This algorithm was the one of the first to use some of the physical constraints of the linear mixing model within a statistical hypothesis test. The algorithm included the sum-to-one constraint on the abundances, but only required the target abundance to be non-negative arguing that proper estimation of the background abundances was not required for detection purposes. The algorithm was also designed assuming that the noise was zero-mean multivariate normal distribution with covariance $\sigma^2\mathbf{I}$. These assumptions made the algorithm very fast, but still do not account for all of the physical constraints in the linear mixing model or a full covariance for the background noise distribution.

Therefore, we present two new hybrid subpixel detectors based on modeling the background using a physically meaningful linear mixing model within a statistical

hypothesis test. The idea is that the physically-based endmembers and abundances will account for the known physics of the problem while the statistical distribution accounts for unknown quantities due to such phenomena as nonlinear mixing effects and sensor noise. Our hypothesis is that the hybrid detectors which model the background both physically and statistically will provide improved performance over their purely statistical counterparts AMSD and ACE. Section 5.1 describes the FCLS, AMSD, and ACE algorithms that form the basis for our hybrid detectors. Section 5.2 describes the two proposed hybrid detectors. Section 5.3 details the experiments used to test our hypothesis. Section 5.4 presents the results of the experiments showing the hybrid detectors excel in three areas: endmember insensitivity, target/background separation on an image by image basis, and improved ROC performance over multiple images. Section 5.5 summarizes the results and identifies future research directions.

5.1. Current Subpixel Algorithms

This section details the FCLS, AMSD, and ACE algorithms. These algorithms are the foundation on which we derive the hybrid detectors. The FCLS algorithm provides a method to incorporate the sum-to-one and non-negativity constraints on the abundances. The AMSD algorithm provides a detector based on a structured background that uses endmembers to define the background **B**. The ACE algorithm provides a detector based on an unstructured background (i.e., a background modeled by a statistical distribution instead of endmembers).

5.1.1. Fully Constrained Least Squares (FCLS)

The FCLS algorithm directly estimates the abundances in (31). While other algorithms have been developed that handle both the non-negativity and sum-to-one

constraints [4][8][98], these algorithms tend to be computationally intense as the number of endmembers increase. The FCLS algorithm meets both abundance constraints as well, but in an efficient manner that is optimal in terms of least squares error (LSE) [46]. Because of these reasons, we chose to use it in our algorithms. Unfortunately, FCLS does not allow a closed-form mathematical solution due to the non-negativity constraints. Instead, a numerical solution is required.

To calculate the FCLS solution, we begin with the non-negativity constraints. The idea is to minimize the LSE by estimating the non-negative abundance values. Mathematically this is expressed as

$$\min_{a}\left(\mathbf{x}-\mathbf{E}\mathbf{a}\right)^{T}(\mathbf{x}-\mathbf{E}\mathbf{a}),\ a_i \geq 0\ \forall i \tag{32}$$

where **E** is the concatenation of the target **S** and background **B** signatures. Using Lagrange multipliers, a Lagrangian J is defined such that

$$J=\tfrac{1}{2}(\mathbf{x}-\mathbf{E}\mathbf{a})^{T}(\mathbf{x}-\mathbf{E}\mathbf{a})+\boldsymbol{\lambda}^{T}(\mathbf{a}-\mathbf{c}), \tag{33}$$

where $\mathbf{a} = \mathbf{c}$, and each member of the unknown constant $M\times 1$ vector **c** is non-negative to enforce the non-negativity constraint. This construction allows the use of Lagrange multipliers because the non-negativity constraints have been substituted by equality constraints with the unknown vector **c**. To calculate the estimate of **a**, we take the partial derivative of J with respect to **a** to obtain

$$\left.\frac{\partial J}{\partial \mathbf{a}}\right|_{\hat{\mathbf{a}}}=\mathbf{E}^{T}\mathbf{E}\hat{\mathbf{a}}-\mathbf{E}^{T}\mathbf{x}+\boldsymbol{\lambda}=0\,. \tag{34}$$

Equation (34) contains two unknowns: the abundance estimates and the Lagrange multipliers. Solving for these unknown results in

$$\hat{\mathbf{a}}=(\mathbf{E}^{T}\mathbf{E})^{-1}\mathbf{E}^{T}\mathbf{x}-(\mathbf{E}^{T}\mathbf{E})^{-1}\boldsymbol{\lambda} \tag{35}$$

and

$$\boldsymbol{\lambda} = \mathbf{E}^T(\mathbf{x} - \mathbf{E}\hat{\mathbf{a}}). \tag{36}$$

Iterating through (35) and (36) provides the numerical solution for the non-negativity constraints. To begin this iterative method, we set all the Lagrange multipliers to zero and calculate the abundance using (35). Note that this initial calculation is the unconstrained least squares solution for the abundance values. From this solution, we identify those abundance values that are greater than zero and place them in the passive set *P*. The remaining non-positive abundance values are placed in the active set *R*. Equations (35) and (36) are iterated until all Lagrange multipliers in the passive set are zero and all Lagrange multipliers in the active set are either zero or negative. At this point, the Kuhn-Tucker conditions have been met and an optimal solution for the abundance values has been found.

One may note that this solution only accounts for the non-negativity constraints of (1). To handle the sum-to-one constraints, an easy modification of the aforementioned algorithm was developed to retain the optimality guaranteed under the Kuhn-Tucker conditions for numerical optimization on a finite computing machine [42]. In the modification, the endmember matrix and pixel signatures are extended such that

$$\widetilde{\mathbf{E}} = \begin{bmatrix} \delta\mathbf{E} \\ \mathbf{1}^T \end{bmatrix} \tag{37}$$

is the new endmember matrix and

$$\widetilde{\mathbf{x}} = \begin{bmatrix} \delta\mathbf{x} \\ 1 \end{bmatrix} \tag{38}$$

is the new pixel signature where δ is a small number (typically 1×10-5). The δ variable controls how tightly the solution will sum to one so that smaller values provide a better solution, but may need longer convergence time. The new endmember matrix and pixel signature are then used in (35) and (36) to obtain an abundance solution that meets both the non-negativity and sum-to-one constraints.

5.1.2. Adaptive Matched Subspace Detector (AMSD)

While the FCLS algorithm provides an elegant solution to calculating the abundance values in the linear mixing model, the algorithm does not provide a statistical hypothesis test to differentiate between a pixel that contains a target and a pixel that contains only the background. The AMSD algorithm provides such a statistical test using a Generalized Likelihood Ratio Test (GLRT) [71]; however, the non-negativity and sum-to-one constraints on the abundance estimates are in general not satisfied. Thus, the AMSD approach leads to a closed-form solution with CFAR optimality, but has to sacrifice the physical constraints on the abundance estimates.

Since the AMSD algorithm is based on a GLRT, we can use the model in (31) assuming that the noise model is a zero-mean normal distribution with covariance matrix $\sigma^2\mathbf{I}$. Therefore, the AMSD hypotheses are

$$\begin{aligned} H_0 &: \mathbf{x} \sim N(\mathbf{B}\mathbf{a}_{b,0}, \sigma_0^2\mathbf{I}) \\ H_1 &: \mathbf{x} \sim N(\mathbf{S}\mathbf{a}_s + \mathbf{B}\mathbf{a}_{b,1}, \sigma_1^2\mathbf{I}) \end{aligned}. \tag{39}$$

Under these assumptions, we can calculate the remaining unknown parameters using Maximum Likelihood Estimation (MLE) techniques. To do this, we calculate the likelihood equation for the null hypothesis as

$$L(\mathbf{x} \mid H_0) = (2\pi\sigma_0^2)^{-\frac{L}{2}} \exp\left\{-\frac{1}{2\sigma_0^2}(\mathbf{x} - \mathbf{B}\mathbf{a}_{b,0})^T(\mathbf{x} - \mathbf{B}\mathbf{a}_{b,0})\right\}. \tag{40}$$

Taking the derivative of the logarithm of (40) with respect to each of the unknown parameters and setting them equal to zero allows us to arrive at the MLE abundance estimate

$$\hat{\mathbf{a}}_{b,0} = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{x} \tag{41}$$

and the MLE noise variance estimate

$$\hat{\sigma}_0^2 = \frac{1}{L}(\mathbf{x} - \mathbf{B}\hat{\mathbf{a}}_{b,0})^T(\mathbf{x} - \mathbf{B}\hat{\mathbf{a}}_{b,0}). \tag{42}$$

Substituting (41) and (42) back into (40) provides the generalized likelihood equation under the null hypothesis

$$f_0 = \left(\frac{2\pi}{L}\mathbf{x}^T(\mathbf{I} - \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T)\mathbf{x}\right)^{-\frac{L}{2}} \exp\left(-\frac{L}{2}\right). \tag{43}$$

Similarly, the same can be done for the alternative hypothesis to arrive at

$$f_1 = \left(\frac{2\pi}{L}\mathbf{x}^T(\mathbf{I} - \mathbf{E}(\mathbf{E}^T\mathbf{E})^{-1}\mathbf{E}^T)\mathbf{x}\right)^{-\frac{L}{2}} \exp\left(-\frac{L}{2}\right) \tag{44}$$

where **E** is again defined as the concatenation of the target and background signatures.

Having calculated the likelihoods for each hypothesis and using some simple algebra, the GLRT takes the ratio of the two likelihoods to calculate the following detection statistic

$$\frac{f_1}{f_0} \Rightarrow \frac{\mathbf{x}^T(\mathbf{I} - \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T)\mathbf{x}}{\mathbf{x}^T(\mathbf{I} - \mathbf{E}(\mathbf{E}^T\mathbf{E})^{-1}\mathbf{E}^T)\mathbf{x}} = \frac{\mathbf{x}^T\mathbf{P}_\mathbf{B}^\perp\mathbf{x}}{\mathbf{x}^T\mathbf{P}_\mathbf{Z}^\perp\mathbf{x}}. \tag{45}$$

Since **E** and **B** are related, it is difficult to identify the distribution of this detection statistic; so, a new detection statistic is created by subtracting one from (45) to obtain

$$D_{AMSD}(\mathbf{x}) = \frac{\mathbf{x}^T(\mathbf{P}_\mathbf{B}^\perp - \mathbf{P}_\mathbf{Z}^\perp)\mathbf{x}}{\mathbf{x}^T\mathbf{P}_\mathbf{Z}^\perp\mathbf{x}}. \tag{46}$$

Applying this mapping does not change the outcome of the decision statistic, but it does allow the new statistic to be distributed as

$$D_{AMSD}(\mathbf{x}) \sim F_{P,L-P-Q}\left(\frac{\left\|\mathbf{P}_\mathbf{B}^\perp\mathbf{S}\mathbf{a}_s\right\|^2}{\sigma^2}\right)[71]. \tag{47}$$

Under the null hypothesis ($\mathbf{S} = \mathbf{Ø}$ and hence the signal to interference plus noise ratio (SINR) term in the parentheses of (47) is equal to zero), the AMSD statistic is based on the parameters P, L, and Q independent of any estimates. Because of this, the AMSD statistic enjoys the CFAR property and should allow a single threshold to determine the false alarm rate. Of course, the single threshold only holds if the underlying data has a multivariate normal distribution.

5.1.3. Adaptive Cosine/Coherent Detector

The methods described earlier are detectors based on structured backgrounds. The ACE method uses a statistical distribution (namely the multivariate normal distribution) to model the background. Referring to (31), the ACE algorithm sets $\mathbf{B} = \mathbf{Ø}$ thus removing any structured background information. In this algorithm, the background is entirely modeled as a zero-mean Gaussian distribution with scaled covariance $\sigma^2\mathbf{\Gamma}$ giving us the hypotheses

$$\begin{aligned} H_0 &: \mathbf{x} \sim N(0, \sigma_0^2\mathbf{\Gamma}) \\ H_1 &: \mathbf{x} \sim N(\mathbf{S}\mathbf{a}_s, \sigma_1^2\mathbf{\Gamma}). \end{aligned} \tag{48}$$

The scaling term σ^2 is interesting as this term is not typically found empirically. The term is necessary theoretically however to make the ACE detector scale-invariant as will be shown later in this section. Since $\mathbf{B}$ does not exist in this

algorithm, the sum-to-one and non-negativity constraints of (1) cannot be met either as they require a background subspace. Despite these seemingly poor assumptions for hyperspectral data, the ACE detector is one of the more powerful subpixel detectors available [70].

For this derivation, we follow the work by Kelly [53] and Kraut and Scharf [56][57][58]. Besides the information we have in (48), we also assume that we have an independent data set Y such that

$$Y = \left\{ \mathbf{y}_i \middle| \mathbf{y}_i \sim N(0, \mathbf{\Gamma}), i = 1, \ldots, N \right\}. \tag{49}$$

Combining (48) and (49) provides the joint likelihood equation under the null hypothesis

$$\begin{aligned} L(\mathbf{x}, Y \mid H_0) = (2\pi)^{-\frac{1}{2}L(N+1)} \mid \mathbf{\Gamma} \mid^{-\frac{1}{2}(N+1)} (\sigma_0^2)^{-\frac{1}{2}L} \times \\ \exp\left\{ -\frac{1}{2\sigma_0^2} \mathbf{x}^T \mathbf{\Gamma}^{-1} \mathbf{x} - \frac{1}{2} \sum_{i=1}^{N} \mathbf{y}_i^T \mathbf{\Gamma}^{-1} \mathbf{y}_i \right\} \end{aligned} \tag{50}$$

and the joint likelihood equation under the alternate hypothesis

$$\begin{aligned} L(\mathbf{x}, Y \mid H_1) = (2\pi)^{-\frac{1}{2}L(N+1)} \mid \mathbf{\Gamma} \mid^{-\frac{1}{2}(N+1)} (\sigma_1^2)^{-\frac{1}{2}L} \times \\ \exp\left\{ -\frac{1}{2\sigma_1^2} (\mathbf{x} - \mathbf{S}\mathbf{a}_s)^T \mathbf{\Gamma}^{-1} (\mathbf{x} - \mathbf{S}\mathbf{a}_s) \right\} \times \\ \exp\left\{ -\frac{1}{2} \sum_{i=1}^{N} \mathbf{y}_i^T \mathbf{\Gamma}^{-1} \mathbf{y}_i \right\}. \end{aligned} \tag{51}$$

If we assume that N is very large, the covariance estimate from these likelihoods can be simplified to

$$\hat{\mathbf{\Gamma}} = \sum_{i=1}^{N} \mathbf{y}_i \mathbf{y}_i^T \tag{52}$$

which is a standard assumption made in the literature. Note that under this assumption, the covariance under the null hypothesis and alternate hypothesis are equal and greatly simplifying the following mathematics.

Following the derivation of the covariance under each hypothesis using MLE, we obtain the abundance estimate as

$$\hat{\mathbf{a}}_s = (\mathbf{S}^T \boldsymbol{\Gamma}^{-1} \mathbf{S})^{-1} \mathbf{S}^T \boldsymbol{\Gamma}^{-1} \mathbf{x} \tag{53}$$

and the variance estimates under each hypothesis as

$$\hat{\sigma}_0^2 = \tfrac{1}{L} \mathbf{x}^T \boldsymbol{\Gamma}^{-1} \mathbf{x} \tag{54}$$

and

$$\hat{\sigma}_1^2 = \tfrac{1}{L} (\mathbf{x} - \mathbf{S}\mathbf{a}_s)^T \boldsymbol{\Gamma}^{-1} (\mathbf{x} - \mathbf{S}\mathbf{a}_s). \tag{55}$$

The estimates are substituted back into the original likelihood equations in (50) and (51). The updated likelihoods are taken as a ratio to obtain the GLRT as was done in the AMSD derivation. After some algebra and simplification, the ACE detector is

$$D_{ACE}(\mathbf{x}) = \frac{\mathbf{x}^T \hat{\boldsymbol{\Gamma}}^{-1} \mathbf{S} (\mathbf{S}^T \boldsymbol{\Gamma}^{-1} \mathbf{S})^{-1} \mathbf{S}^T \boldsymbol{\Gamma}^{-1} \mathbf{x}}{\mathbf{x}^T \hat{\boldsymbol{\Gamma}}^{-1} \mathbf{x}}. \tag{56}$$

This is a CFAR detector and has the following distribution under the null hypothesis

$$D_{ACE}(\mathbf{x}) \sim Beta(\tfrac{P}{2}, \tfrac{L-P}{2}) \tag{57}$$

where L is the number of spectral bands and P is the number of target signatures [57]. Therefore, the ACE statistic is based only on the parameters P and L independent of any estimates. Because of this, the ACE statistic also enjoys the CFAR property and should allow a single threshold to determine the false alarm rate. Again, the single

threshold only holds if the underlying data is a multivariate normal distribution (or any distribution in the family of elliptically contoured distributions) [57].

5.2. Hybrid Detectors

Using the derivations and ideas in the previous section, we present two hybrid subpixel detectors that incorporate the HSI physical constraints directly into the detector derivation. The first detector uses a structured background and is similar to AMSD. The second detector uses an unstructured background and is similar to ACE.

5.2.1. Hybrid Structured Detector

The hybrid structured detector (HSD) approaches the solution to (31) using a structured background like AMSD, but using physically meaningful endmembers and replacing the abundance estimates with their FCLS counterparts. The HSD hypotheses are

$$\begin{aligned} H_0 &: \mathbf{x} \sim N(\mathbf{B}\mathbf{a}_{b,0}, \sigma_0^2\boldsymbol{\Gamma}) \\ H_1 &: \mathbf{x} \sim N(\mathbf{S}\mathbf{a}_s + \mathbf{B}\mathbf{a}_{b,1}, \sigma_1^2\boldsymbol{\Gamma}) \end{aligned} \tag{58}$$

Since this derivation includes a full covariance matrix, we follow a similar derivation to ACE incorporating the background subspace $\mathbf{B}$ and its abundances $\mathbf{a}_b$ as was done in AMSD. With this new information the likelihood equation under the null hypothesis is

$$\begin{aligned} L(\mathbf{x}, Y \mid H_0) &= (2\pi)^{-\frac{1}{2}L(N+1)} \mid \boldsymbol{\Gamma} \mid^{-\frac{1}{2}(N+1)} (\sigma_0^2)^{-\frac{1}{2}L} \times \\ &\exp\left\{ -\frac{(\mathbf{x} - \mathbf{B}\mathbf{a}_{b,0})^T \boldsymbol{\Gamma}^{-1} (\mathbf{x} - \mathbf{B}\mathbf{a}_{b,0})}{2\sigma_0^2} - \frac{1}{2}\sum_{i=1}^{N} \mathbf{y}_i^T \boldsymbol{\Gamma}^{-1} \mathbf{y}_i \right\} \end{aligned} \tag{59}$$

and the likelihood equation under the alternate hypothesis is

$$L(\mathbf{x}, Y \mid H_1) = (2\pi)^{-\frac{1}{2}L(N+1)} \mid \boldsymbol{\Gamma} \mid^{-\frac{1}{2}(N+1)} (\sigma_1^2)^{-\frac{1}{2}L} \times$$
$$\exp\left\{ -\frac{(\mathbf{x}-\mathbf{Ea})^T \boldsymbol{\Gamma}^{-1}(\mathbf{x}-\mathbf{Ea})}{2\sigma_1^2} - \frac{1}{2}\sum_{i=1}^{N} \mathbf{y}_i^T \boldsymbol{\Gamma}^{-1} \mathbf{y}_i \right\} \tag{60}$$

where $\mathbf{Ea} = \mathbf{S}\mathbf{a}_s + \mathbf{B}\mathbf{a}_{b,1}$.

The covariance estimate is the same as (52) given the assumption that N is large. Under this assumption, we obtain the variance estimates under each hypothesis as

$$\hat{\sigma}_0^2 = \tfrac{1}{L}(\mathbf{x} - \mathbf{B}\mathbf{a}_{b,0})^T \boldsymbol{\Gamma}^{-1}(\mathbf{x} - \mathbf{B}\mathbf{a}_{b,0}) \tag{61}$$

and

$$\hat{\sigma}_1^2 = \tfrac{1}{L}(\mathbf{x} - \mathbf{Ea})^T \boldsymbol{\Gamma}^{-1}(\mathbf{x} - \mathbf{Ea}) . \tag{62}$$

Besides the covariance and variance estimates, the abundance estimates also need to be calculated. At this point instead of using the standard MLEs, we use a variant of the FCLS algorithm to estimate these parameters. Because of the covariance matrix, the variant of the FCLS algorithm attempts to minimize

$$\min_a (\mathbf{x} - \mathbf{Ea})^T \boldsymbol{\Gamma}^{-1}(\mathbf{x} - \mathbf{Ea}), a_i \geq 0 \ \forall i \tag{63}$$

This update leads to a new Lagrangian J such that

$$J = \tfrac{1}{2}(\mathbf{x} - \mathbf{Ea})^T \boldsymbol{\Gamma}^{-1}(\mathbf{x} - \mathbf{Ea}) + \boldsymbol{\lambda}(\mathbf{a} - \mathbf{c}) . \tag{64}$$

Therefore, the new equations that we iterate through to meet the Kuhn-Tucker conditions are

$$\hat{\mathbf{a}} = (\mathbf{E}^T \boldsymbol{\Gamma}^{-1}\mathbf{E})^{-1}\mathbf{E}^T \boldsymbol{\Gamma}^{-1}\mathbf{x} - (\mathbf{E}^T \boldsymbol{\Gamma}^{-1}\mathbf{E})^{-1}\boldsymbol{\lambda} \tag{65}$$

and

$$\boldsymbol{\lambda} = \mathbf{E}^{\mathbf{T}} \boldsymbol{\Gamma}^{-\mathbf{1}} (\mathbf{x} - \mathbf{E}\hat{\mathbf{a}}) . \tag{66}$$

The rest of the algorithm proceeds as in Section 5.1.1 to obtain abundance estimates that incorporate the sum-to-one and non-negativity constraints with a full covariance matrix. While this prevents us from obtaining a closed-form solution for our detector, it enforces all of the known physical constraints.

All of the estimates are substituted back into the original likelihood equations in (59) and (60). The generalized likelihoods are taken as a ratio to obtain the GLRT as was done in the AMSD derivation. After some algebra and simplification, the HSD is

$$D_{HSD}(\mathbf{x}) = \frac{(\mathbf{x} - \mathbf{B}\hat{\mathbf{a}}_b)^T \boldsymbol{\Gamma}^{-1}(\mathbf{x} - \mathbf{B}\hat{\mathbf{a}}_b)}{(\mathbf{x} - \mathbf{E}\hat{\mathbf{a}})^T \boldsymbol{\Gamma}^{-1}(\mathbf{x} - \mathbf{E}\hat{\mathbf{a}})}. \tag{67}$$

The HSD algorithm is similar to our original hybrid detector [16] except for the inclusion of the full covariance matrix.

5.2.2. Hybrid Unstructured Detector

The Hybrid Unstructured Detector (HUD) models the background as a multivariate normal distribution similar to ACE. Since the ACE detector is already white, the HUD algorithm simply replaces the abundance estimates with their whitened FCLS counterparts. To accomplish this, we rewrite (56) such that

$$D_{ACE}(\mathbf{x}) = \frac{\mathbf{x}^T \hat{\boldsymbol{\Gamma}}^{-1} \mathbf{S}\mathbf{a}}{\mathbf{x}^T \hat{\boldsymbol{\Gamma}}^{-1} \mathbf{x}} \tag{68}$$

where the abundance estimate **a** is taken from (53).

To form the HUD algorithm, we simply replace the abundance with its whitened FCLS counterpart. Therefore, the new detector is

$$D_{HUD}(\mathbf{x}) = \frac{\mathbf{x}^T \hat{\boldsymbol{\Gamma}}^{-1} \mathbf{S}\hat{\mathbf{a}}}{\mathbf{x}^T \hat{\boldsymbol{\Gamma}}^{-1} \mathbf{x}} \tag{69}$$

where the abundance estimate **a** is taken from (65) after the Kuhn-Tucker conditions have been satisfied. Note that this solution still requires the extraction of endmembers to define the abundance estimate, but these endmembers are not directly used within the decision statistic. They only serve to provide a better estimate of the target abundance based on the physical constraints of the linear mixing model.

5.3. Experimental Results

Our hypothesis is that the hybrid detectors provide improved performance by taking advantage of the known physics of the linear mixing model within a statistical hypothesis test. To show whether this occurs or not in practice, we have implemented a number of experiments on hyperspectral imagery under real-world conditions. One of the major difficulties in doing such an analysis is being as unbiased as possible. This is a real concern when using real world hyperspectral data as many of the variables are simply out of our control. However, we can develop a series of tests that reduce this bias and provide meaningful results. We argue that these types of tests are more germane to detection performance as real world data collections have to encounter many of the same issues. This section will be devoted to identifying the issues related to data acquisition and the methods we used for each of our detectors. This is not meant to be a full comparison of all the different ways to process hyperspectral data. This comparison is only meant to help understand whether our hypothesis is valid. The following sections identify the experimental design and provide results for three experiments measuring endmember sensitivity, separation performance, and overall ROC performance. Table I summarizes these selections for each of the detectors (AMSD, ACE, HSD, and HUD) used in our experiments.

Table 9: Subpixel Experiment Details

Detector	Background Model	Background Signatures	Target Signatures	Abundance Constraints
AMSD	Structured	Eigenvectors of the Image Correlation Matrix	MODTRAN	No
ACE	Unstructured	Multivariate Normal with Global Covariance	MODTRAN	No
HSD	Structured	Iterative Error Analysis & Global Covariance	MODTRAN	Yes
HUD	Unstructured	Iterative Error Analysis & Global Covariance	MODTRAN	Yes

5.3.1. Experimental Design

These experiments require imagery, background signatures, target signatures, and ground truth information. The imagery used for these experiments comes from the Sensor X data described in Chapter 2. From this sensor, we used Images 1 through 6 because these images contain subpixel targets. The other images are full or multi-pixel targets which provide little challenge for the detectors.

As indicated in Chapter 4, we used two background endmember extraction techniques. The most significant eigenvectors of the global image correlation matrix were used as the "endmembers" for the AMSD algorithm as documented in [71]. Since these do not produce physically meaningful endmembers, we used the IEA algorithm for the hybrid detectors [77]. Additionally, we used the image covariance matrix for the hybrid detectors to whiten the data. In all cases, the endmembers and covariance matrices were estimated from the entire image. We also tried local estimates, but these provided results no better than using global estimates.

To choose the number of endmembers for each detector, we first extracted up to 150 endmembers for AMSD and 60 endmembers for the hybrid detectors. While we could have used our estimation techniques from Chapter 4, we decided to identify

the ideal cases for these results to present the best performance possible for each of the detectors. The concept of best performance turned out to be trickier than we first imagined. For images where we had targets, the best performance was defined as the number of endmembers that maximized the probability of detection while minimizing the number of false alarms. In cases where perfect separation was achieved between targets and false alarms, the best performance was defined using a minimax criterion where the clutter with the highest detection score was minimized. The same minimax criterion was applied to the cases where no targets were present. This method provided the best results independent of detector type both in terms of separation of targets and clutter and setting a fixed threshold for ROC curves.

The target information we received from NVESD were measured in units of reflectance. As discussed in Chapter 3, the images are measured in terms of radiance. There are three approaches to overcome this mismatch: use target signatures directly from the image for the experiments, convert the images to reflectance, or convert the targets to radiance.

Because the images only contained sub-pixel targets, we could not directly use target signatures from the image. If we did, the signatures would be corrupted with background and bias our results. Moreover, using target signatures from the image reduces the pool of targets. Those targets would have to be dropped from the analysis as any target pulled from the imagery would be guaranteed to be detected biasing the results.

Because of these reasons we turned to the atmospheric compensation techniques documented in Chapter 3. For the analysis in this chapter, we relied on the

model-based method MODTRAN to generate the target signatures. We did this to remove variability that may have been introduced using the ARRT method. This was especially true for Target 4 where the low reflectance signature made estimation using in-scene methods difficult.

Ground truth was used to create background and target objects. Following the procedures in Chapter 2, we applied a cluster threshold to each detector output to guarantee 1% of the pixels were above the threshold. This threshold was used knowing that the number of targets in the image was far less than 1% of the pixels in the image. Adjacent pixels above the threshold were assigned to the same cluster. In each cluster, the maximum detection score was assigned as the cluster detection score. These clusters' positions were then compared to ground truth information to label the clusters as either target or false alarms.

Fill factors for the experiment ranged from 10% to 60%. Fill factor describes the percent of the pixel that is occupied by target signature. Fill factors assume that the target lies exactly within the pixel. In numerous cases, subpixel targets can lie across pixel boundaries or be obscured by the competing environment (e.g. tall grass) generating fill factors in the image that are much smaller than expected.

5.3.2. Endmember Sensitivity Analysis

This experiment measures how sensitive the AMSD, HSD, and HUD algorithms are to the number of endmembers. In our experiments, we have ground truth information and hence can determine the "best" number of endmembers as defined in the previous section. In real world applications, this knowledge is not available to us; hence, we have to rely on algorithms to estimate the correct number of endmembers without the associated ground truth. As shown in Chapter 4, these

algorithms can have significant errors. Therefore, detectors that are insensitive to these estimation errors are highly desirable.

For this experiment, we measured the probability of detection and number of false alarms at varying numbers of endmembers from one to 60 across all images, targets, and detectors. We stopped at 60 endmembers because in all cases, the performance for all detectors on all targets degraded well before reaching this number and continued to degrade as will be shown in our results. The only exception to this rule was the number of endmembers used for AMSD. Additional experimentation showed that we needed to extract as many as 150 endmembers to provide good detection results.

We present the results in two ways. First, we provide an example to show how the false alarm density varies with the number of endmembers and type of detector. The results for this experiment are in Figure 25 which shows the performance of the AMSD, HSD, and HUD algorithms on Image 1 and Target 2. We chose this image and target type because it is indicative of the entire set of results we produced. The figure shows the number of false alarms for varying numbers of endmembers on each detector. We did not include the P_d figures simply because all detectors were able to achieve nearly 100% P_d across all numbers of endmembers. Therefore, good performance on this test is achieved if a minimal number of false alarms are detected across multiple numbers of endmembers. This indicates that the detector is partially insensitive to the number of endmembers chosen.

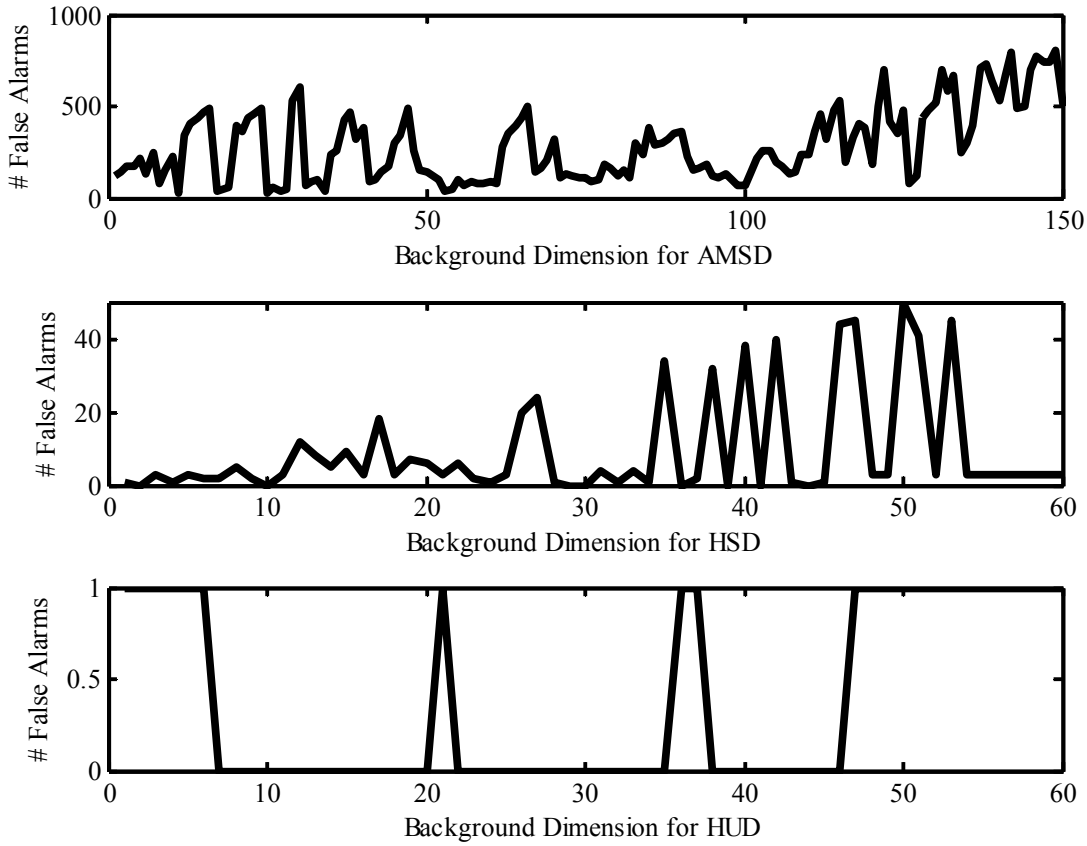


Figure 25: Graphical Comparison of Endmember Sensitivity

Figure 25 shows the hybrid algorithms are more insensitive to the number of endmembers than AMSD. The AMSD results are random and lack the general trend seen in the HSD and HUD results. When using AMSD, even slight changes in endmembers can produce dramatically different results varying from 27 false alarms to nearly 800. The HSD algorithm results show that endmembers numbering less than ten tend to produce better results. Also, the HSD results at higher number of endmembers do not vary as greatly as the AMSD figures. Instead, the worst case number of false alarms is limited to 50. HUD is the best in terms of being insensitive to the number of endmembers. This algorithm provides excellent performance regardless of the number of endmembers. The data shows that the hybrid detectors are

insensitive to the number of endmembers with HUD being nearly independent of them.

Figure 25 only shows the results for one image and one target type. To verify that this occurs for all target types and images, we put together Table 10 that contains the number of times the best performance was achieved across the 60 endmembers for each detector. The more insensitive a detector is to the number of endmembers, the higher the number. Best performance is defined as the instances that achieve 100% P_d with the lowest number of false alarms. Note that this could mean that the lowest number of false alarms is greater than zero. Results are only posted for images where the target is present.

Table 10: Endmember Sensitivity Results

Target	Image	AMSD	HSD	HUD
1	1	54	35	60
	4	39	34	60
2	1	1	8	37
	4	3	42	60
3	2	2	36	59
	3	21	42	60
	5	1	1	2
	6	1	1	59
4	2	0	0	0
	3	0	0	0
	5	0	0	0
	6	0	0	0

The results in Table 10 support the results from the first experiment. Target 1 is the easiest of the targets and this is demonstrated by the high numbers achieved with all the detectors. As the targets become more difficult to identify though, the results start to diverge. AMSD performance drops to single digits as target difficulty increases. HSD maintains good numbers until the hardest images where it too drops

to single digits. HUD fairs the best maintaining nearly perfect performance on all the images except a few. These experiments show the hybrid detectors are partially insensitive to the number of endmembers selected. Since the true number of endmembers is rarely if ever known, detectors with this insensitivity have a significant advantage over those that do not.

The only exception to the rule is Target 4 where none of the detectors are able to achieve 100% Pd. In this case, the performance is poor independent of the number of endmembers. The most likely cause is the target is so weak that target characterization methods are not correctly modeling the signature. This mismatch causes all detectors to perform poorly.

5.3.3. Separability Analysis

Having shown that the hybrid detectors are more insensitive to the number of endmembers selected, the question remains whether they provide improved detection performance over their AMSD and ACE counterparts. This set of experiments answers this question using figures that show the separability between target and background for each image and detector type. The figures were patterned after those found in [69]. These graphs are very useful because they can be used even when few targets are present. This allows us to measure the performance of the detectors on each image and target type.

The figures for each target type are shown in Figure 26 through Figure 29. Each figure contains four sub-figures. Each sub-figure contains black and gray vertical bars. The black bars show the range of detection values for the background. The gray bars show the range of detection values for the targets. Ideally, these bars should not overlap indicating the targets are completely separable from the

background. In cases where overlaps do occur, a number is posted above the black bar. This is the number of false alarms that occur if all of the targets are detected. Within any sub-figure, the ranges of the targets and background can be compared across images to see the consistency of the detector. A good detector will consistently suppress the background into a similar range of values while separating the targets.

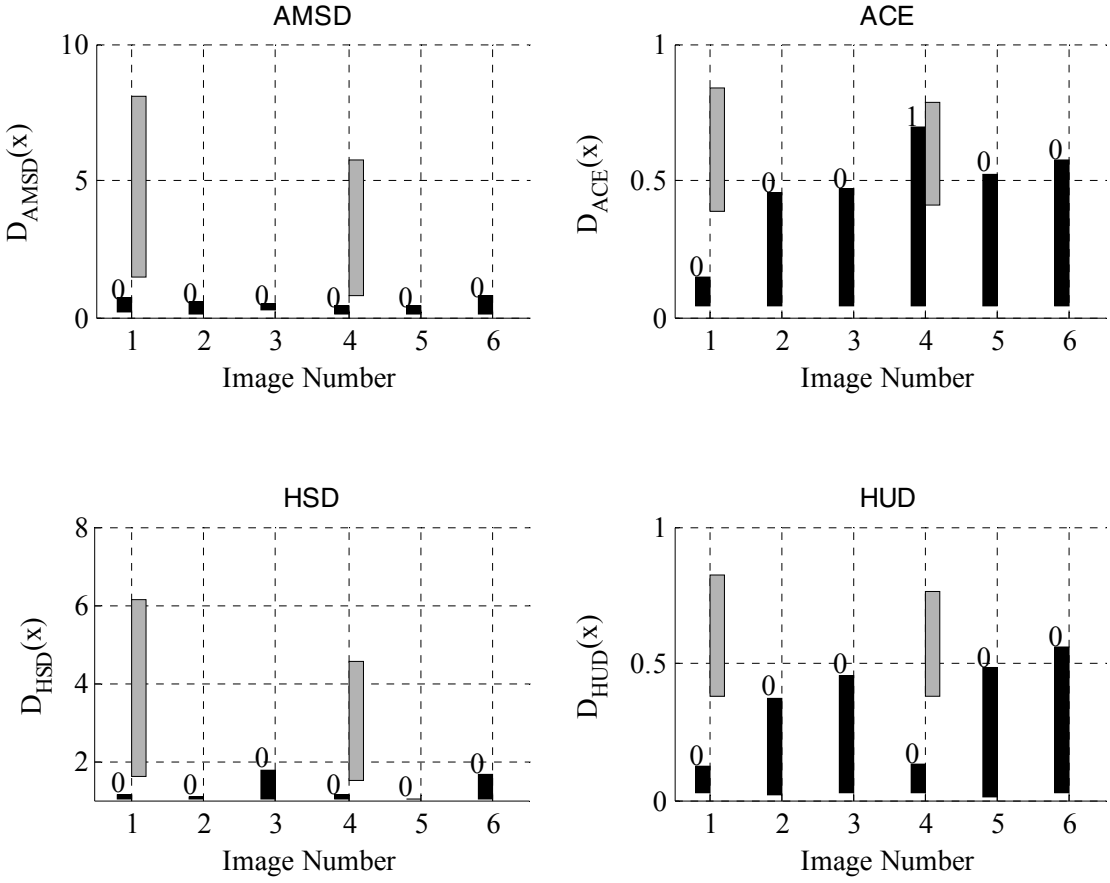


Figure 26: Separability Analysis for Target 1

Figure 26 shows the results for Target 1. This is the easiest target due to its white color that makes it very different from the surrounding background. All the detectors perform well with only ACE picking up one false alarm on Image 4. The structured detectors however perform better than their unstructured counterparts. The ACE and HUD algorithms do separate the target from the background, but have

difficulty suppressing the background in the images where the targets are not present. The structured detectors do not suffer from this problem and suppress the background nearly equally across all images. AMSD has a slight advantage over HSD on Images 3 and 6 where the background values have been compressed a bit farther than with HSD. Nevertheless, all detectors show good performance on this target.

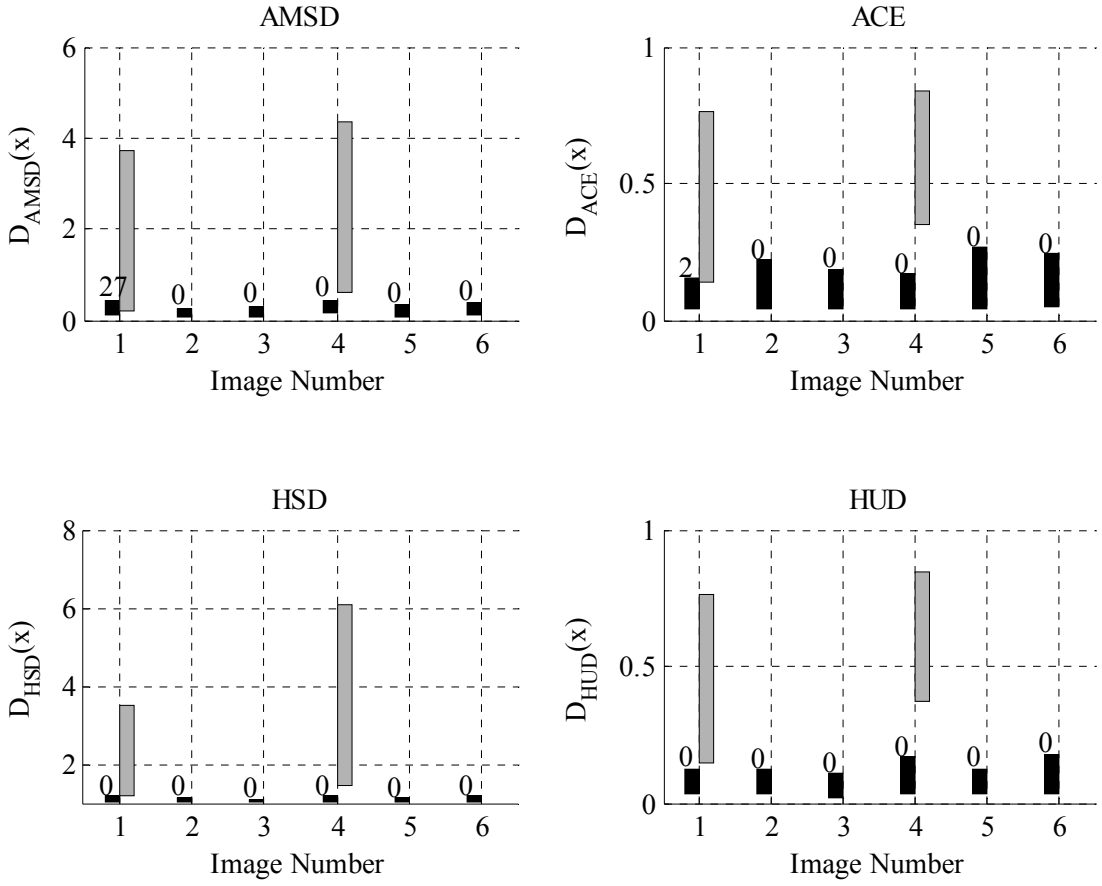


Figure 27: Separability Analysis for Target 2

Figure 27 shows the results for Target 2. This target is painted green and although larger than Target 1 is harder to separate from the green background. It is with this target that the hybrid detectors begin to show a slight performance advantage over the standard detectors. The hybrid detectors maintain zero false alarms across all images as was the case with Target 1. AMSD however picks up 27

false alarms on the first image and ACE picks up 2 false alarms on the same image. The hybrid detectors also do a better job of suppressing the background into similar ranges across the images. AMSD does as well but has the aforementioned 27 false alarms. ACE is the only detector where the background values vary significantly across the images.

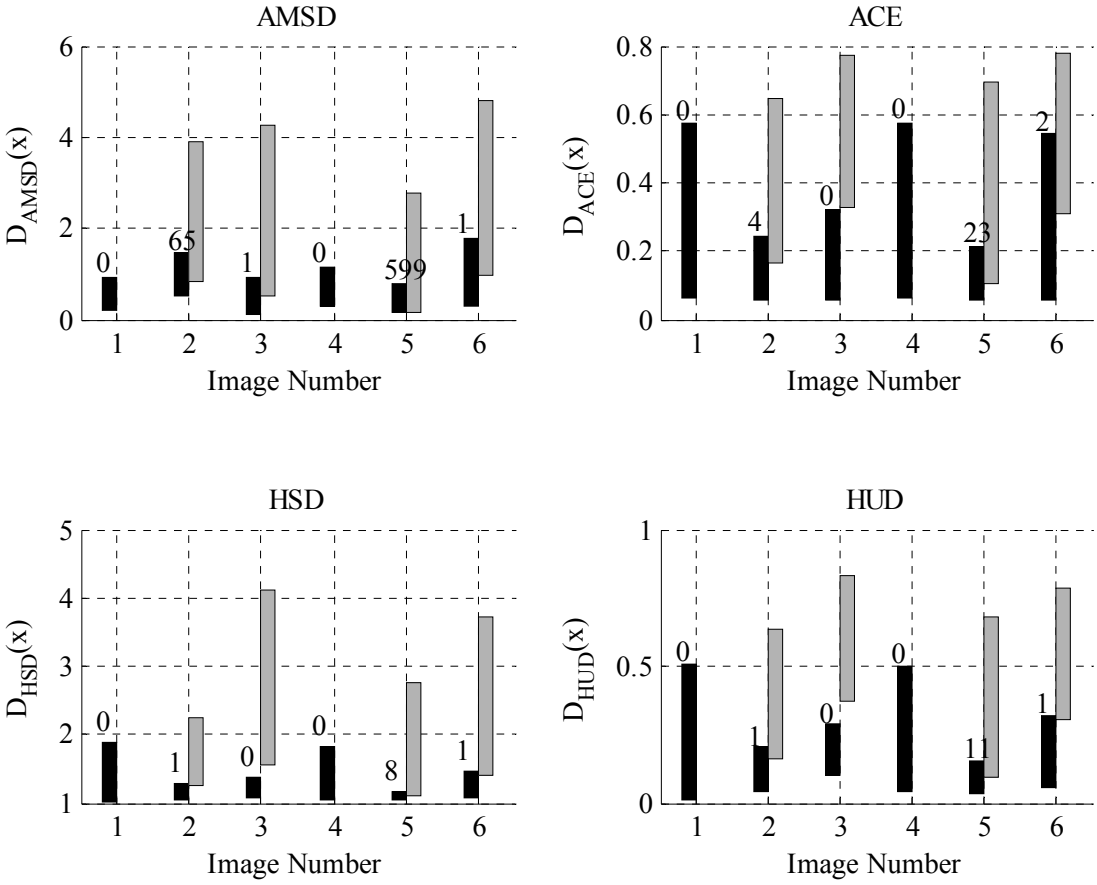


Figure 28: Separability Analysis for Target 3

Figure 28 shows the results for Target 3. This target has multiple reflectance signatures which indicate a significant variability of the spectral signature. Because of this variability, all the detectors have difficulties with this target. The background is no longer being compressed to the same range of values for any detector although the structured detectors do fair better than their unstructured counterparts. The key is the

number of false alarms. AMSD achieves 666 false alarms across all images. ACE drops this number to 29. HUD further reduces the number to 13 while HSD performs the best with only 10 false alarms. These numbers are remarkable in that the hybrid detectors have provided 66 times less false alarms than AMSD and 3 times less false alarms than ACE. When one considers that the hybrid detectors are also the most insensitive to the number of endmembers selected, the performance gains become much more significant.

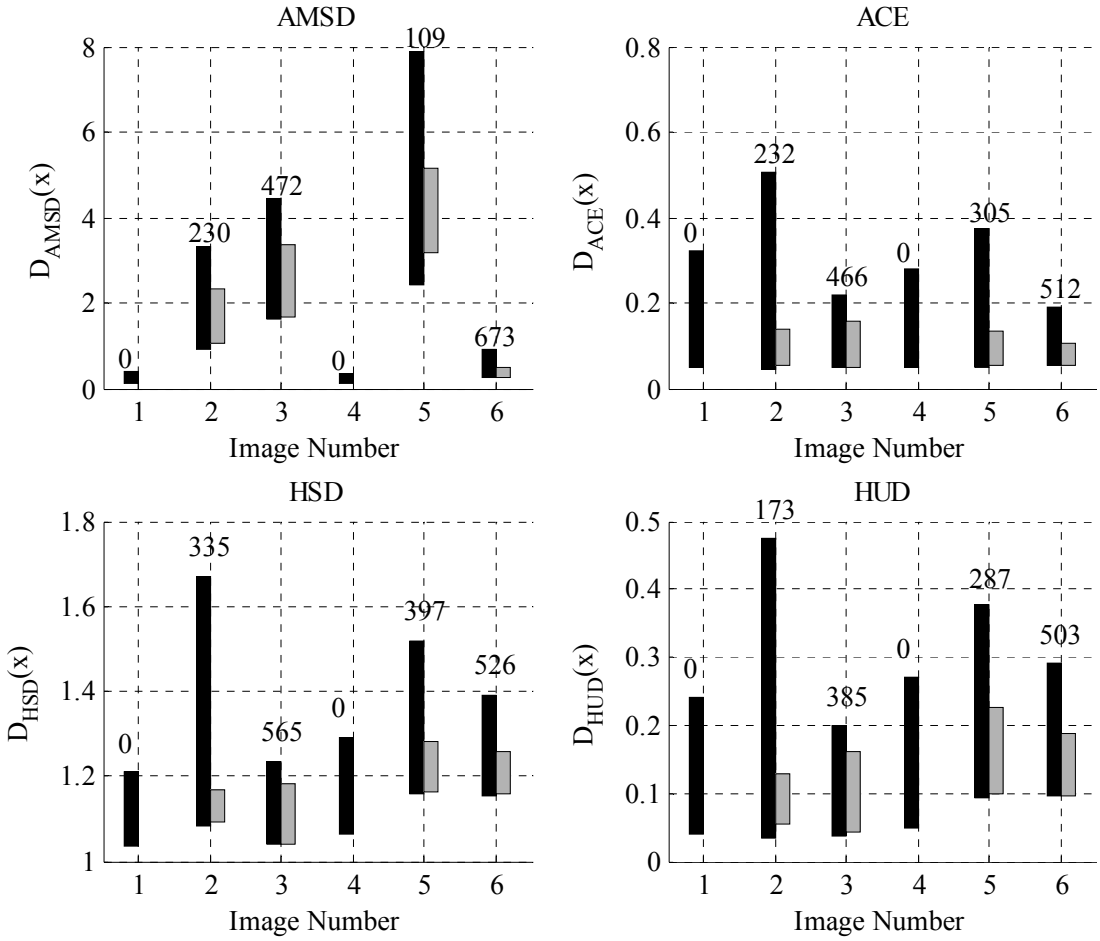


Figure 29: Separability Analysis for Target 4

Figure 29 shows the results for Target 4. As expected, all of the detectors have difficulty with this weak target. This is the only target where the hybrid detectors

show no improvement over the statistical detectors AMSD and ACE. The most likely cause of this result is the incorrect modeling of the target radiance signature as noted in Chapter 3. When the estimated target does not match the target signature in the image, no signature based detector is going to perform well. None of the detectors are able to detect 100% of Target 4 in any of the images. Therefore, this target is not a good example for comparing the different subpixel target detectors, but it does support the need for good target characterization.

5.3.4. Receiver Operating Characteristics

In our separability analysis, we argued that some detectors did a better job consistently pushing the background values into a similar region across all the images. A good way to measure this consistency is to use a ROC curve. The ROC curves we generate are for a single detector and single target across all images. This provides enough target returns to make each ROC statistically significant. Note that a ROC measures the average performance for a fixed threshold across all images; therefore, detectors that consistently separate the targets and background into similar detection values across each image will perform better than those that do not. Theoretically, the CFAR algorithms AMSD and ACE should provide such performance. Our interest is whether the hybrid algorithms will meet or exceed the results of the CFAR algorithms thus giving them CFAR-like properties although this fact cannot be proved theoretically.

Figure 30 shows the ROC curves for Target 1. As expected from our separability analysis, the structured detectors outperform the unstructured detectors. AMSD does have a slight performance improvement over HSD, but the results show

the hybrid detectors are achieving the same CFAR performance as the standard detectors.

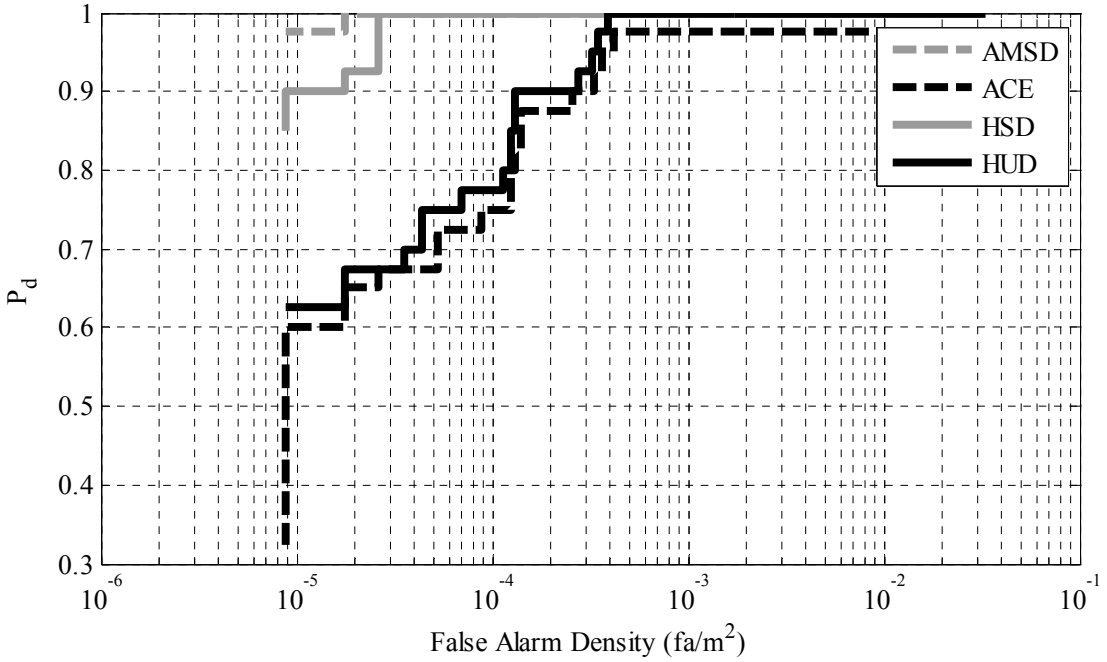


Figure 30: Subpixel Detection ROC Curves for Target 1

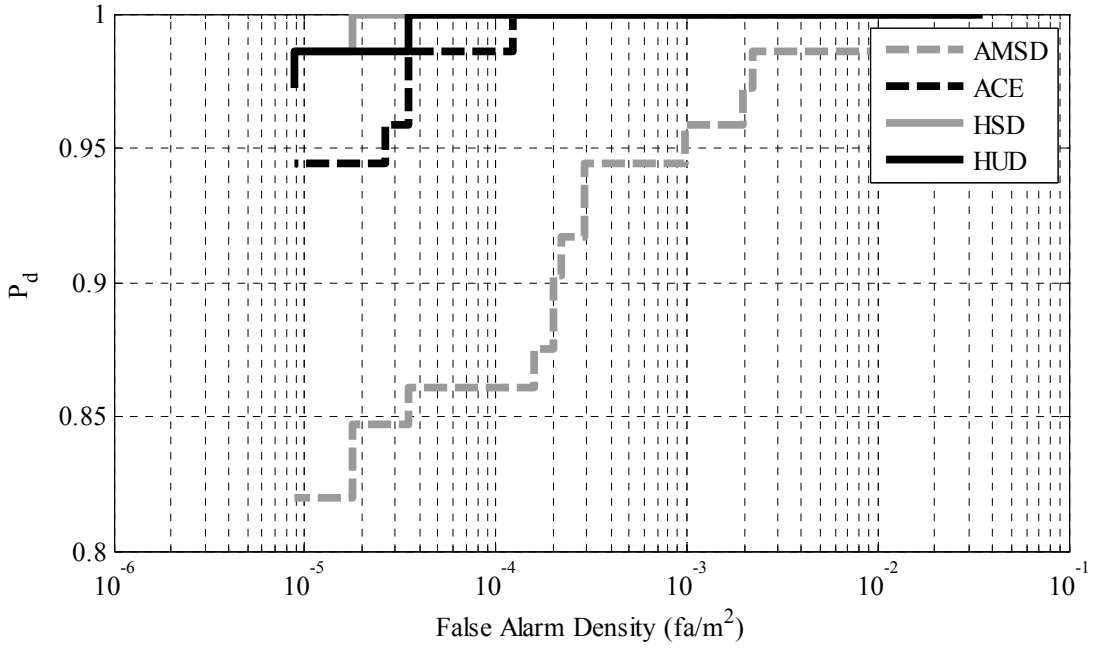


Figure 31: Subpixel Detection ROC Curves for Target 2

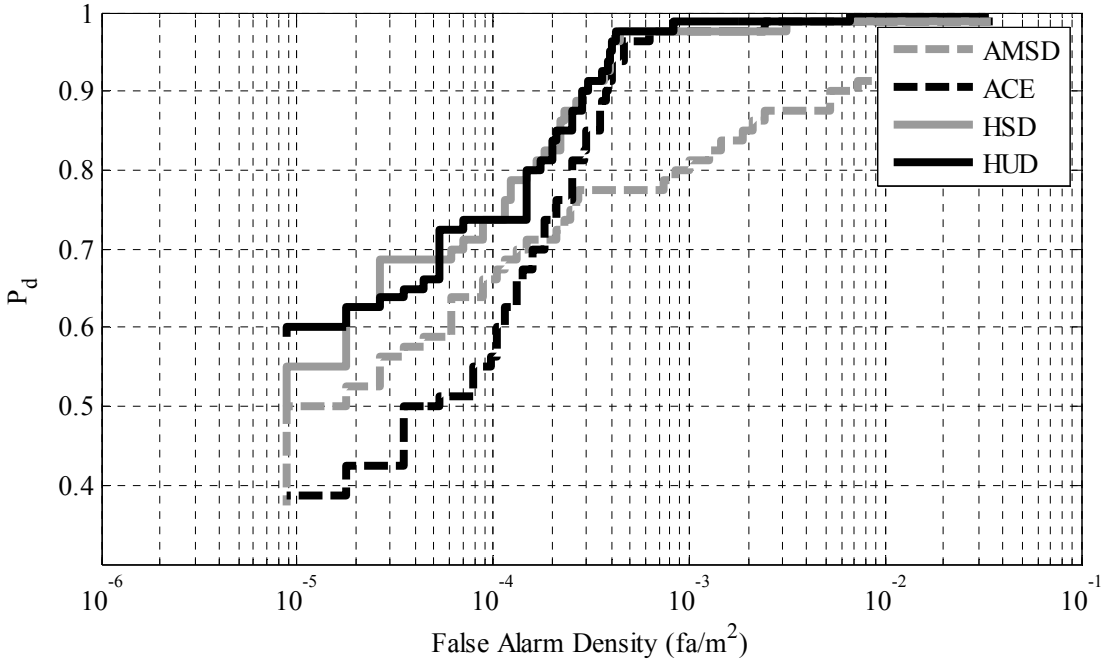


Figure 32: Subpixel Detection ROC Curves for Target 3

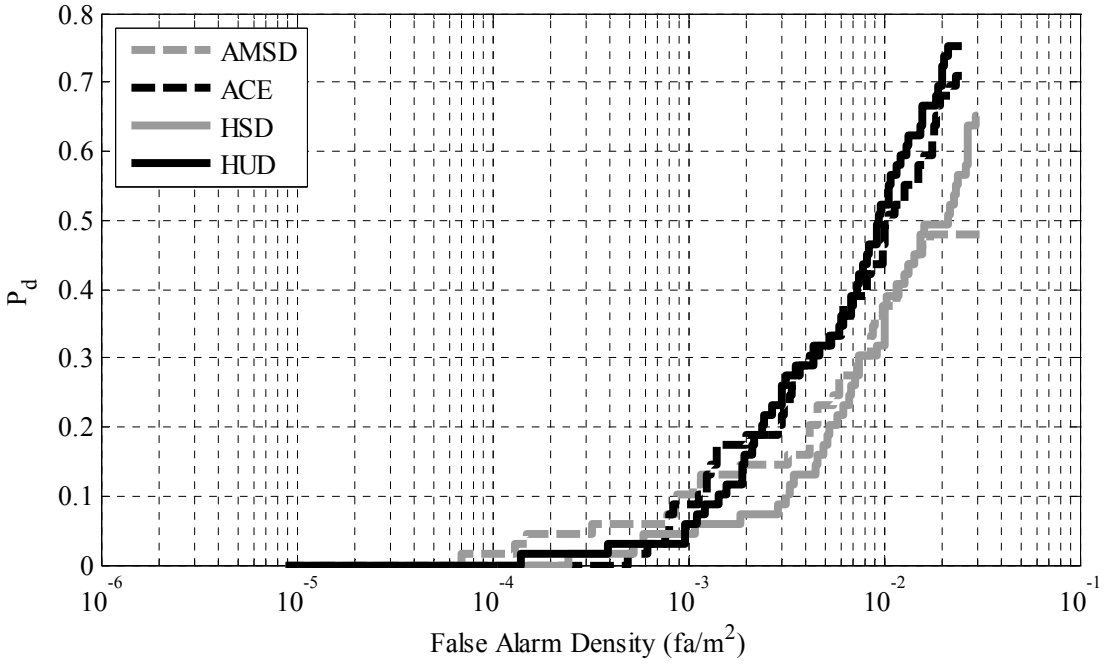


Figure 33: Subpixel Detection ROC Curves for Target 4

Figure 31 shows the ROC curves for Target 2. The hybrid detectors are slightly better than their standard counterparts. While the figure seems to show a great

improvement in performance, the range of Pd values is measured from 0.8 to 1.0. This gives the impression of much better performance. Nevertheless, the hybrid algorithms are again performing as well if not better than their CFAR counterparts.

Figure 32 shows the ROC curves for Target 3. In the separability analysis, the hybrid algorithms showed great performance improvements over AMSD and ACE. What was noted in that section was that none of the detectors were able to suppress the background into a consistent range of values. The ROC curves show this fact. The hybrid algorithms are performing better than their CFAR counterparts, but the performance improvement is not as significant as in the separability analysis. The conclusion that can be drawn from this result is that the background and target are similar making the background harder to suppress. Nevertheless, the hybrid detectors are modeling the background better than AMSD and ACE which provides the gains in performance.

Figure 33 shows the ROC curves for Target 4. As expected, none of the detectors perform well. This is the only target for which the acceptable performance criteria of 50% P_d at 10^{-3} false alarms/m^2 is not met. As mentioned before, the reason is due to incorrect modeling of the target radiance signature.

5.3.5. Conclusions

Our set of experiments demonstrates the usefulness of the hybrid detectors. These detectors have a three-fold gain over their standard counterparts. First, they are tolerant of slight errors in the number of endmembers. Second, they show greater separability between targets and background – especially as the target becomes more difficult to detect. Third, they maintain a slightly more consistent threshold across the images than the known CFAR detectors AMSD and ACE. This result argues the

hybrid detectors' ability to better model the background and therefore detect subpixel targets.

What has not been mentioned so far is the efficiency of the hybrid algorithms. The algorithms require very little extra processing time when compared to either AMSD or ACE. ACE was perhaps the fastest of the detectors since we estimated the covariance matrix from the entire image. Results were also generated for ACE using local neighborhoods, but the performance showed little to no improvement over using the entire image. AMSD was nearly as quick as ACE except for the extraction of endmembers using an eigenvalue decomposition of the image correlation matrix. The hybrid detectors took the longest, but only because of the IEA endmember extraction algorithm. Once the endmembers were extracted, the performance was no different than that achieved with AMSD. The reason for this is the efficient FCLS algorithm which only took ten minutes to process an image when using 60 endmembers. With endmembers less than 20, the FCLS algorithm took less than a minute. Since most of the hybrid detectors prefer endmembers numbering less than 20, the processing times were similar to AMSD.

One final note is on the difference between the HSD and the HUD algorithms. Both of these algorithms performed well, but the HSD algorithm has a slight performance advantage. On all targets it was able to achieve false alarm densities smaller than HUD. HSD was also more consistent in suppressing the background into a similar range of detection values. The tradeoff is that the HSD algorithm is more sensitive to the number of endmembers. For example, the HSD algorithm requires an estimate of the number of endmembers that is close to the ideal. HUD on the other

hand can simply set the number of endmembers to some fixed number and achieve the same results for nearly all images. Therefore, the HUD algorithm is not as dependent on the number of endmembers, but has slightly lower performance than HSD because of this fact.

5.4. Summary

In this chapter we argue that better characterization of the background through physics-based knowledge can improve subpixel detection performance. To this end, we develop two hybrid detectors which use physically meaningful endmembers and abundances within a statistical hypothesis test. We compare these detectors to their purely statistical counterparts AMSD and ACE.

Our results show that the improved background models of the hybrid detectors provide improved performance in three different ways. First, the hybrid detectors are less sensitive to the number of endmembers used. Thus, endmember estimation algorithms can allow some error without significantly degrading subpixel detection performance. Second, the hybrid algorithms provide better separation between the targets and background per individual image. This is especially the case with weaker targets like Target 3 where AMSD and ACE have false alarm densities well over 30 compared to 10 for the hybrid detectors. Finally, the hybrid detectors provide a more consistent separation of target and background that leads to improved ROC performance.

While this research shows the importance of modeling the background on subpixel target detection algorithms, further research is required. On Target 3, the hybrid detectors did outperform their statistical counterparts, but Figure 28 shows that the background detection scores can still vary significantly from image to image. One

way to counteract this phenomenon is to better characterize the background using more appropriate density functions or non-parametric techniques in conjunction with physics-based knowledge. Another means to counteract this phenomenon is to use adaptive threshold techniques. Either way, our research suggests much more can be done to model and understand the complex background inherent in hyperspectral imagery to improve subpixel target detection performance.

Chapter 6: Adaptive Detection Thresholds via Extreme Value Theory

Subpixel detectors present a significant challenge in determining the detector threshold for a desired probability of false alarm. For example, the most common threshold estimation method is a theoretical calculation for used for CFAR detectors. CFAR detectors are designed such that the distribution of the detector given the background is independent of any estimates needed to derive the detector [70]; therefore, the conditional background distribution is independent of the data. This independence of the clutter distribution from the data allows a theoretical calculation of a fixed false alarm density α_0. CFAR detectors achieve this goal by making an assumption about the underlying distribution of the data. Typically this assumption is that the underlying distribution is a normal distribution (or at least any zero-mean elliptically contoured distribution [57]), which makes the mathematics tractable enough to determine the detector's statistical distribution. Additionally, CFAR detectors typically assume independent and identically distributed (iid) samples. For instance, a standard detector for HSI data is the Adaptive Cosine Estimate (ACE) detector which assumes the underlying distribution is multivariate normal [58]. ACE is a CFAR detector whose threshold can be calculated theoretically for a desired false alarm density. In practice though, HSI data has been shown to be rarely multivariate normal [103] and hence any theoretically calculated threshold for the ACE detector is most likely inaccurate.

In recent publications, the use of elliptically contoured distributions has been explored to model the outputs of detectors [69]. This method is similar to the theoretical threshold calculations for CFAR detectors except the method models the

output of the detector as an elliptically contoured distribution. The detector data is then used to estimate parameters which in turn provide a distribution from which a detection threshold can be theoretically calculated. The usefulness of this method is currently being investigated, but its applications are limited to CFAR detectors. This prevents us from using these techniques for our hybrid detectors where the output distribution is difficult at best to determine due to the non-negativity constraints. Therefore, we must rely on methods that directly use the output detection statistics.

A standard non-parametric approach for determining the desired detector threshold is to use order statistics. The detector output is sorted in descending order to create an ordered list. The number of detection values N is multiplied by the desired α_0 and rounded to the nearest integer. This integer is used to identify the position in the ordered list that will be used as the detection threshold. The strength of this approach is that any detector output can be used – not just those that are CFAR. Even if the detection threshold varies significantly from image to image, the use of this method adjusts the threshold automatically to track such deviations. Unfortunately, the method is very sensitive to outliers when low false alarm densities are required. For example, a typical detection image will contain both targets and clutter. The order statistic algorithm will count the targets as clutter and this will skew the detection threshold. We can think of this as a Monte Carlo (MC) method where instead of estimating the probability of false alarm density from the detector samples, we use the samples to estimate the threshold for a desired false alarm density. In subsequent discussions, we will call this the MC method.

Another method of determining the detector threshold is based on importance sampling (IS). Importance sampling is a forced Monte Carlo method that is used to simulate rare events [101]. IS has been mostly used to test system responses to rare events in an efficient manner. There are a number of papers that prove its ability to provide unbiased estimates of rare event probabilities with low variance [91][99][102]. These rare events simulate the distribution tails of the system and hence are closely related to the design and measurement of detectors.

Srinivasan showed that IS could be used to determine a detector threshold for a desired fixed false alarm probability α_0 [101]. This method is called inverse importance sampling. Initially, these thresholds were determined for standard background distributions that a detector may encounter such as the normal, Rayleigh, or Weibull distributions. Bucklew extended this research to handle situations where the underlying probability density function was unknown [17]. Unfortunately, these methods are designed for sums of random variables. In [101], Srinivasan shows that blind importance sampling when applied to data from a single random variable provides no gains over MC methods. Since the detector output is from a single random variable, blind IS methods are not ideal.

Therefore, we turn to the use of Extreme Value Theory (EVT). EVT concerns problems where the probability of a rare event must be estimated even if such a rare event has never occurred [39]. This type of research has wide applicability in such fields as climatology [100], detection theory [74], anomaly detection [89], and financial analysis [25]. It is in the last field where most of the theory has been applied to estimate stock market anomalies, insurance rates for catastrophic events, and

management of risk. These applications are very similar to our problem of estimating a threshold for rare events even if they have not occurred. This makes EVT a variance reduction technique similar to IS, but far more applicable to wider class of problems [38].

In target detection, the presence of targets can significantly impact the performance of threshold estimates. A variety of methods have been developed to remove outliers (e.g., isolation of target returns from the background) [47]. These methods vary widely from simple sample statistics to advanced classification techniques based on Support Vector Domain Descriptions [108]. Interestingly, EVT theory can also be used to identify outliers in a data sample [89]. Thus, EVT can both estimate detection thresholds for a given false alarm density and simultaneously be used to remove the influence of outliers on the sample.

Therefore, we present a novel adaptive threshold technique based on extreme value theory. The new technique is able to set thresholds for desired probabilities of false alarm densities similar to the MC technique. Unlike the MC technique, we develop an outlier rejection capability using the Generalized Pareto Distribution (GPD) that can identify samples that do not belong to the same distribution as the background. These outlier samples can be removed such that desired false alarm densities in the presence of target returns can be calculated with some confidence. The rest of the chapter is structured as follows. Section 6.1 presents an overview of Extreme Value Theory. Section 6.2 describes our adaptive threshold algorithm based on GPD estimates. Experimental results are given in Section 6.3. A summary concludes the chapter in Section 6.4

6.1. Extreme Value Theory

6.1.1. The Fisher-Tippett Theorem

Assume there is a set $X = \{x_1, x_2, \ldots, x_m\}$ of m i.i.d. samples drawn from the same unknown and continuous cdf $F(x)$. Denote the maximum of the set X as

$$x_{(m)} = \max(X). \tag{70}$$

with cdf

$$H(x) = \left[F(x)\right]^m. \tag{71}$$

Fisher and Tippett [28] show that if $H(x)$ is stable in the limit as $m \to \infty$, then an affine transformation exists such that

$$x_{(m)} \overset{d}{=} \sigma_m x + \mu_m \tag{72}$$

for a given scale parameter σ_m and location parameter μ_m. Equation (72) states that the maximum of the set X converges in distribution to the affine transform. Using the affine transformation given, Fisher and Tippett show that

$$H(x_{(m)} \le x) \overset{d}{=} H(\sigma_m^{-1}(x - \mu_m)) \tag{73}$$

the normalized form is the only form for the limit distribution of $x_{(m)}$ given any $F(x)$.

Now assume that $H(x)$ is a non-degenerate limit distribution for normalized maxima of the form $\sigma_m^{-1}(x - \mu_m)$, then $H(x)$ is only one of three forms. This theorem is the famous Fisher-Tippett theorem [28] and is the foundation for extreme value theory. Denoting $y_m = \sigma_m^{-1}(x - \mu_m)$, the "reduced variate", the three forms are

$$H_1(y_m) = \exp(-\exp(-y_m))$$
$$H_2(y_m) = \begin{cases} 0 & \text{if } y_m \le 0 \\ \exp(-y_m^{-\alpha}) & \text{if } y_m > 0 \end{cases} \tag{74}$$
$$H_3(y_m) = \begin{cases} \exp(-(-y_m)^{\alpha}) & \text{if } y_m \le 0 \\ 1 & \text{if } y_m > 0 \end{cases}$$

for $\alpha > 0$ which are the Gumbel, Fréchet, and Weibull distributions respectively. What this theorem states is that as $m \to \infty$, the maximal distribution $H(x)$ is in the domain of attraction of one of the three limit forms in (74) for any $F(x)$. Therefore, much like the central limit theorem for sums of random variables, the Fisher-Tippett theorem provides a known limiting distribution for the maxima from any set of i.i.d. samples.

6.1.2. EVT for the Exponential Class

Most research has focused on the type I or Gumbel distribution. This limiting distribution occurs for all samples that are drawn from a distribution in the exponential class [35][39] which contains such well-known distributions as the normal, lognormal, and K distributions. A number of researchers have developed theory to identify whether data samples belong in the exponential class such as Gumbel [39], Gnedenko [35], and von Mises [112]. From this theory, Weinstein [114] introduced the generalized extreme value theory (GEVT) such that

$$\lim_{m\to\infty} H\left(\left(a_m^{v} + c_m y\right)^{1/v}\right) = \exp(-\exp(y)) \tag{75}$$

where $a_m > 0$, $v > 0$, and

$$x_m = \left(a_m^{v} + c_m y\right)^{1/v}. \tag{76}$$

When considering tail estimates based on data from the exponential class, the Gnedenko criterion states that (75) holds if and only if

$$\lim_{n\to\infty} n\left\{1 - F\left(\left(a_n^v + c_n y\right)^{1/v}\right)\right\} = -\exp(y), \quad \forall y\,. \tag{77}$$

Using (75) through (77), we can estimate the tail of the unknown exponential class *F(x)* by

$$Q(x) = 1 - F(x) = \frac{1}{n}\exp\left(\frac{-(x^v - a_n^v)}{c_n}\right). \tag{78}$$

Having defined the unknown tail probability, we need to estimate the four parameters: a_n, c_n, v, and n. Guida, Iovino, and Longo present a way to find these parameters using numerical optimization of the maximum likelihood estimates [38]. These estimates are

$$\hat{c}_n = \frac{1}{L}\sum_{i=1}^{L} x_{n(i)}^{\hat{v}} - \frac{\sum_{i=1}^{L} x_{n(i)}^{\hat{v}} \exp\left(-x_{n(i)}^{\hat{v}}/\hat{c}_n\right)}{\sum_{i=1}^{L} \exp\left(-x_{n(i)}^{\hat{v}}/\hat{c}_n\right)}, \tag{79}$$

$$\hat{a}_n = -\hat{c}_n \ln\left(\frac{1}{L}\sum_{i=1}^{L} \exp\left(-x_{n(i)}^{\hat{v}}/\hat{c}_n\right)\right), \tag{80}$$

and

$$\hat{v} = L\left\{\sum_{i=1}^{L}\left\{\ln x_{n(i)} + \left(\frac{x_{n(i)}^{\hat{v}} \ln x_{n(i)}}{\hat{c}_n}\right)\left(\exp\left(\frac{x_{n(i)}^{\hat{v}} - \hat{a}_n}{\hat{c}_n}\right) - 1\right)\right\}\right\}^{-1} \tag{81}$$

where $x_{n(i)}$ is the maximum value from the i^{th} set of n samples. These can be iteratively solved using numerical techniques such as the Kimball procedure [39].

The only other parameter to be estimated is n. Unfortunately, this parameter cannot be estimated using MLEs. Instead, Guida, Iovion, and Longo perform a number of trials to see the effect of this parameter on the final solution [38]. Their results show that n should be on the order of tens of samples to maximize the number

of L sets. If n becomes to large, L decreases leading to poor estimates of the tail distribution.

6.1.3. Generalized Pareto Distribution

Pickands [79] noted that classical EVT (Fisher-Tippett theorem) has a number of difficulties when applied in practice. First, most research has focused on only one of the three limiting distributions – namely, the distribution for data from the exponential class as noted in the previous section. Unfortunately, if the data does not come from the exponential class, a practitioner must use his/her intuition and subjective reasoning to choose the correct parametric model. Second, classical EVT requires partitioning the data into n set of m samples. As noted in [38], there is no direct way to identify the best partitioning a-priori. To this end, Pickands [79] and Balkema and de Haan [5] introduce a new way to estimate the tail of a distribution based on modeling the distribution of samples above some high threshold.

Following the work of Pickands [79], assume that we have n i.i.d. samples from a continuous and unknown distribution $F(x)$. Pickands assumes for some c, $-\infty < c < \infty$, there exists

$$\lim_{n\to x_\infty} \inf_{0<\alpha<\infty} \sup_{0\le x<\infty} \left| \frac{1-F(u+x)}{1-F(u)} - \exp\left(-\int_0^{x/a} \left[(1+ct)_+ \right]^{-1} dt \right) \right| = 0 \qquad (82)$$

where x_∞ = greatest lower bound $\{x: F(x) = 1\}$ = lowest upper bound $\{x: F(x) < 1\}$, and $y_+ = max(0,y)$. For any u and x, the $[1 - F(u+x)]/[1 - F(u)]$ is the conditional probability that an observation is greater than $x+u$ where u is some high threshold. Therefore,

$$P(X \mid X \geq u) = 1 - G(x) = \exp\left(- \int_0^{x/a} \left[(1 + ct)_+ \right]^{-1} dt \right). \quad (83)$$

Von Mises [112] showed for EVT that the extremal distribution functions have the form

$$\Lambda(x) = \exp\left(- \int_0^{y} \left[(1 + ct)_+ \right]^{-1} dt \right), \quad y = \sigma_m^{-1}(x - \mu_m). \quad (84)$$

Therefore, *P(X X ≥ u)* is in the domain of attraction of the classical EVT distributions without having to partition the data into *n* sets of *m* samples.

If *F(x)* is continuous, then *G(x)* is a generalized Pareto distribution (GPD) of the form

$$G(x) = \begin{cases} 1 - \left(1 + c\frac{x}{a} \right)^{-\frac{1}{c}} & if\ c \neq 0 \\ 1 - e^{-\frac{x}{a}} & if\ c = 0 \end{cases} \quad (85)$$

for all *x* such that $0 < x < \infty$. Depending on the shape factor *c*, the GPD embeds a number of other distributions. When $c = 0$, the GPD is an exponential distribution. When $c > 0$, the GPD is the ordinary Pareto distribution. When $c < 0$, the GPD is the Pareto II distribution. Pickands also shows that the estimated GPD is consistent and converges in probability to the true tail distribution such that

$$\lim_{n \to \infty} P\left\{ \sup_{0 \leq x < \infty} \left| \frac{1 - F(\hat{u} + x)}{1 - F(\hat{u})} - [1 - \hat{G}(x)] \right| > \varepsilon \right\} = 0, \quad \forall \varepsilon > 0. \quad (86)$$

Therefore, the GPD is a consistent estimate of the tail distribution based on samples above some high threshold *u* for an unknown *F(x)*. The importance of this research is

that it removes the subjective selection of one of the extremal distributions in (74) and removes the need to partition the data set into *n* set of *m* samples.

6.2. EVT Adaptive Threshold Algorithm

Having described the main theorems for extreme value theory, we now proceed to describe how this theory can be used to estimate detection thresholds. Detection thresholds are typically set by fixing the threshold at a desired probability of false alarm (α_0). In CFAR detectors, this threshold can be calculated directly assuming the data fits the statistical distribution of the detector. In subpixel detection, the HSI data rarely fits the standard CFAR assumption of normal statistics. MC methods shown in (88) can be used to estimate the threshold from the data, but they are inaccurate for very small α_0 and are sensitive to outliers.

We can use the theory based on GPD to calculate the threshold for a tail distribution. Following the derivations in [33], we can redefine the unknown cdf as

$$F(x) = (1 - \Pr(X \le t))F_t(x - t) + \Pr(X \le t). \tag{87}$$

where t is a sufficiently high threshold. The probability that the set of data is less than t is easy to find using MC methods. The estimate is

$$\Pr(X \le t) = \frac{N - n}{N} \tag{88}$$

where N is the total number of samples and n is the number of samples above t. Thus, the threshold needs to be high enough such that the remaining samples are in the tail of the distribution, but not so high that very few samples exist above the threshold. A good rule of thumb is to use either a threshold that captures 90% or 95% of the data. Note that this metric is a simple MC method and will provide unbiased, consistent estimates as the number of samples increase.

The remaining term in (87) is the cdf of the tail of the distribution $F_t(x-t)$. For this estimate, we use the GPD given in (85). To use the GPD, we must estimate the parameters a and c from the data. To perform this estimation, we calculate the log likelihood function from (85). To begin, we first calculate the probability density function (pdf) as the derivative with respect to x of (85) to obtain

$$g(x) = \begin{cases} a^{-1}\left(1 + c\frac{x}{a}\right)^{-1/c} & if\ c \neq 0 \\ a^{-1}\exp\left(-\frac{x}{a}\right) & if\ c = 0 \end{cases} \tag{89}$$

If we assume i.i.d. samples from the distribution, the likelihood equation is

$$g(X) = \prod_{i=1}^{n} g(x_i) \tag{90}$$

Taking the natural logarithm of (83), we obtain the log likelihood function

$$\log g(X) = \begin{cases} -n\log a - \frac{1+c}{c}\sum_{i=1}^{n}\log\left(1 + c\frac{x_i}{a}\right) & if\ c \neq 0 \\ -n\log a - \frac{1}{a}\sum_{i=1}^{n} x_i & if\ c = 0 \end{cases} \tag{91}$$

Unfortunately, the log likelihood equation is nonlinear and solving for each of the parameters results in coupled nonlinear equations. Therefore instead of trying to directly estimate the parameters using MLEs, we turn to the Nelder-Mead Simplex Method which is an implementation of unconstrained nonlinear optimization [62]. This method finds the minimum of a function; thus, instead of maximizing *log g(X),* we minimize *–log g(X).* Using this technique, we obtain estimates of a and c.

Having calculated all the parameters, we can rewrite (87) for the tail samples such that

$$\hat{F}(x) = 1 - \frac{n}{N}\left(1 + \hat{c}\frac{x-t}{\hat{a}}\right)^{-1/\hat{c}}. \tag{92}$$

Conversely, we can rewrite (92) to find the threshold for a given cdf value to obtain

$$t_\alpha = t + \frac{\hat{a}}{\hat{c}}\left(\left(\frac{N\alpha_0}{n}\right)^{-\hat{c}} - 1\right) \tag{93}$$

where t_α is the threshold for a desired $\alpha_0 = 1\text{-}F(x)$ beyond threshold t.

This is a very useful result for our application. After setting a clustering threshold t, we can estimate a detection threshold t_α from the data samples for a desired α_0 value. The problem here as with the MC method is the GPD method assumes that all the data samples come from the same underlying distribution. In the case where targets are present, this assumption is invalid and suffers from the same problems as MC techniques.

The GPD method, however, is based on the knowledge that the tails of a distribution will converge in probability to the generalized Pareto distribution [79]. This only occurs though if the data samples come from the same distribution. When the data contains samples from multiple distributions, the tail will not converge to a GPD. We can use this knowledge to identify when target samples are present in the data and remove them before estimating a threshold for a desired α_0.

To identify the presence of samples from two different distributions, we use the confidence bounds of the GPD. The idea is based on the fact that if the data comes from a single distribution, it should fall within the confidence bounds. Therefore, if we set 90% confidence bounds, 90% of the samples should fall between the bounds. If a higher percentage of samples fall outside these bounds, we hypothesize that the samples must come from multiple distributions.

To generate the confidence bounds, we rely on either numerical optimization or Monte Carlo simulation. Both provide reliable estimates of the GPD bounds, but we found the Monte Carlo simulations to be much quicker. To create these Monte Carlo estimates of the confidence bounds, we generate hundreds of random samples for each data sample of the GPD using the estimates found from (84). This provides a range of estimated *F(x)* values at each data sample. The estimated samples are ordered. The confidence bound for the particular data sample is then calculated by taking the two estimated samples such that 90% of the remaining samples fall between them. This is done at every data sample to calculate the confidence bounds.

To help describe how we use the confidence bounds, we construct two simple examples. For the first example we generate 10,000 samples from a standard normal distribution. For the second example, we generate 9,900 samples from a standard normal distribution and 100 samples from a normal distribution with a mean value of 6 as "target" detections. We fit a GPD to the top 10% of the data for both examples. From these points, we estimate the tail cdf according to (83). We compare the results to the cdf calculated using MC techniques in (88) (also called the Kaplan-Meier empirical cdf [22]).

Figure 34 shows the estimated GPD with associated 90% confidence bounds compared to the empirical cdf for the first example. The solid gray red curves represent the 90% confidence bounds. The black points are the empirical cdf and the dashed gray line is the best fit using the GPD. The empirical cdf fits well between the confidence bounds having only 4 samples fall outside the bounds. This represents 0.4% of the samples which is much less than the 10% limits enforced by the bounds.

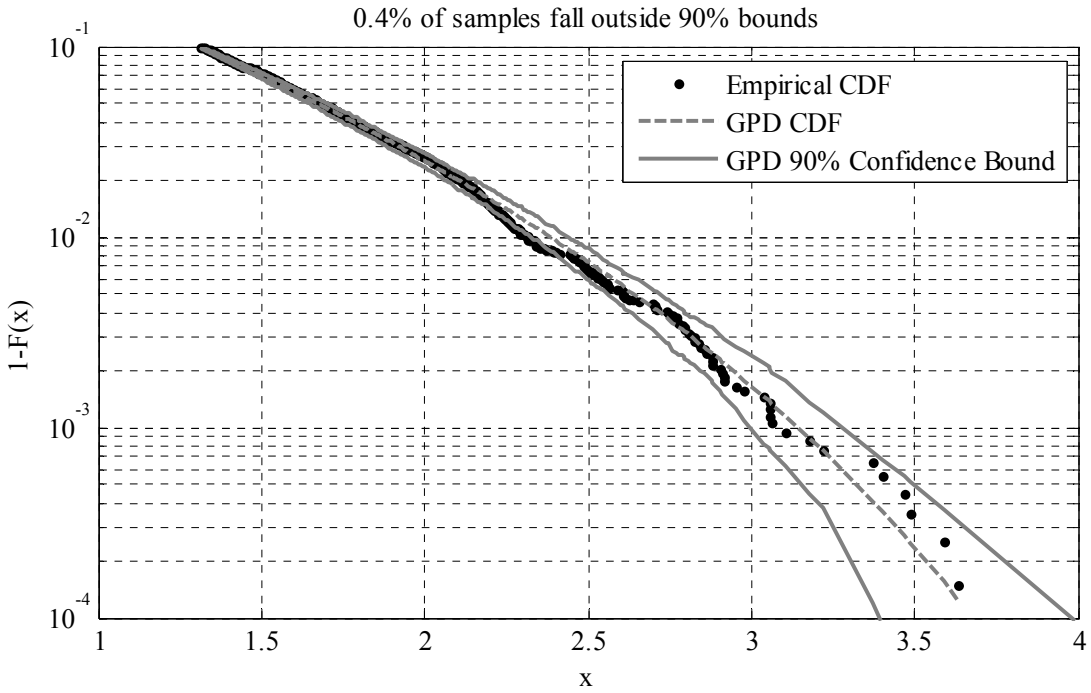


Figure 34: Comparison of the GPD to the Empirical CDF for Example 1

Figure 35 shows the estimated GPD with associated 90% confidence bounds compared to the empirical CDF for the second example. The empirical cdf falls well outside the bounds with over 30% of its samples beyond the 90% confidence limits. This example is therefore considered as having come from multiple distributions. This can be seen clearly in the empirical cdf. The 100 samples from the normal distribution with mean value 6 cause a hump in the cdf centered at 6. These are our fictional "target" detections. The challenge now is to identify these samples and remove them.

Upon further examination of Figure 35, the empirical cdf curve does follow a GPD distribution until it begins flattening out near values of 3. At this point, it intersects the lower bound. Therefore, we can use the lower bound as a threshold for outlier rejection. Any samples in the empirical cdf beyond the lower bound are removed from the data sample. Because the GPD method is a variance reduction

method, it is acceptable to remove some of the non-target samples from the data. This allows us some flexibility in choosing which samples will be used to estimate the new generalized Pareto distribution.

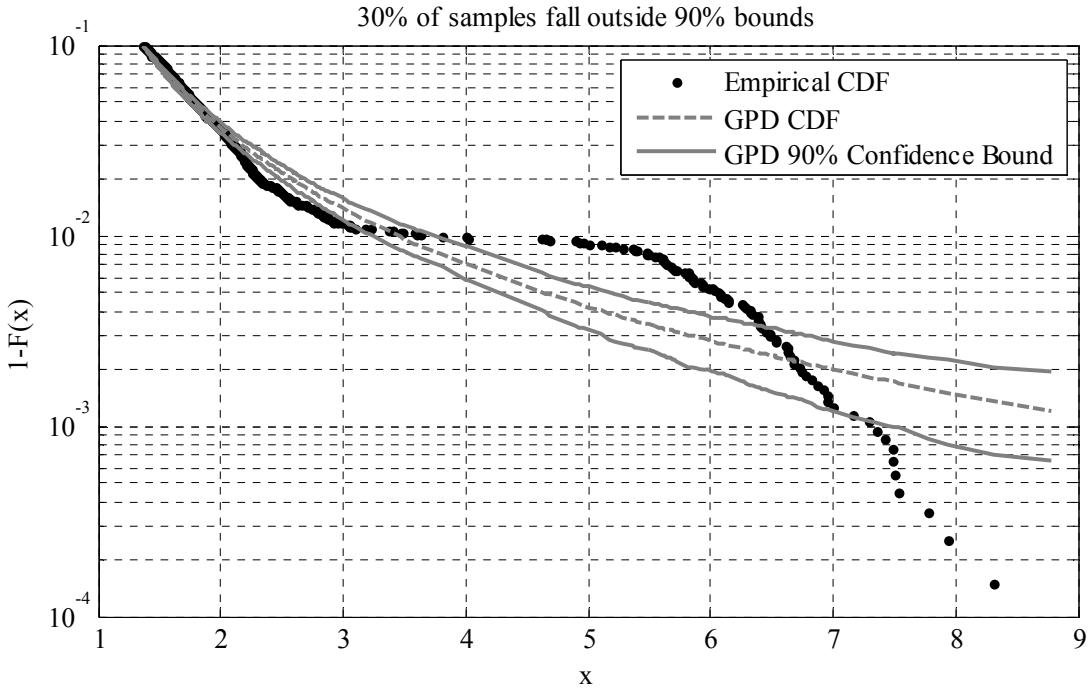


Figure 35: Comparison of the GPD to the Empirical CDF for Example 2

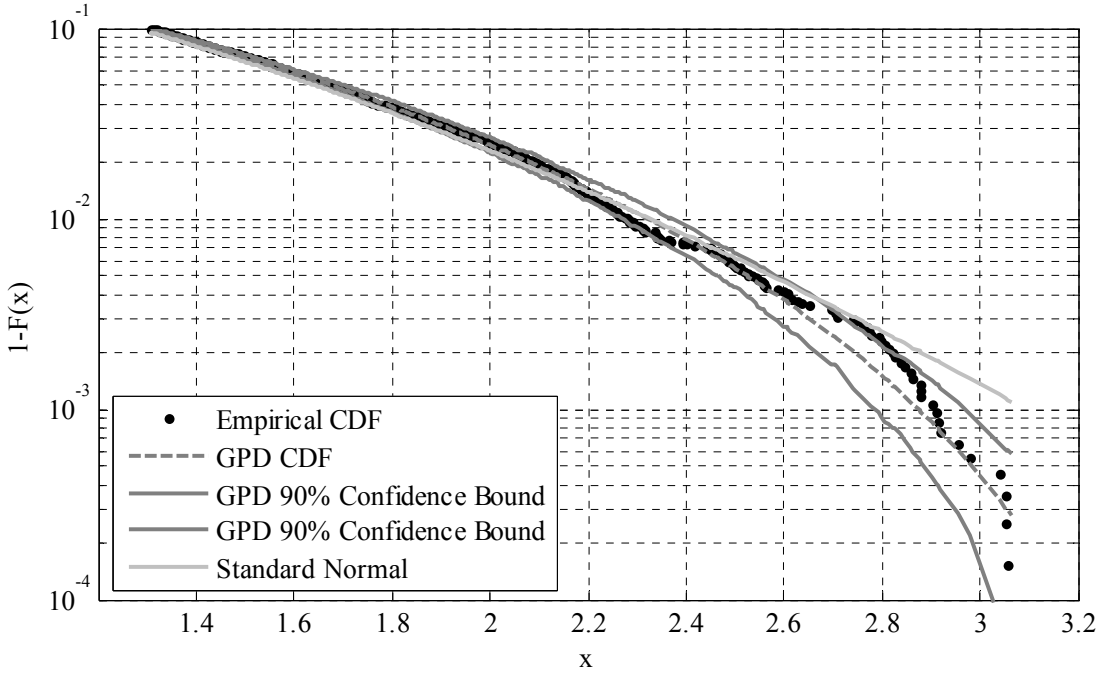


Figure 36: Comparison of Corrected Samples

Using the lower bound to identify the samples to keep, we recalculate the GPD and display the results in Figure 36. Along with the edited empirical cdf and GPD estimates, we include the true cdf of a standard normal distribution. The edited samples now approximate the true normal cdf well –especially at lower samples. The results only diverge at the highest samples and even then, they differ only by 0.0005. This shows that the algorithm can identify samples with "targets", prune the "target" samples, and then recompute a new tail distribution that is close to the original "background" samples. All of this can be done without any knowledge of the underlying background distribution or knowledge of the target samples. A block diagram of the proposed algorithm is given in Figure 37.

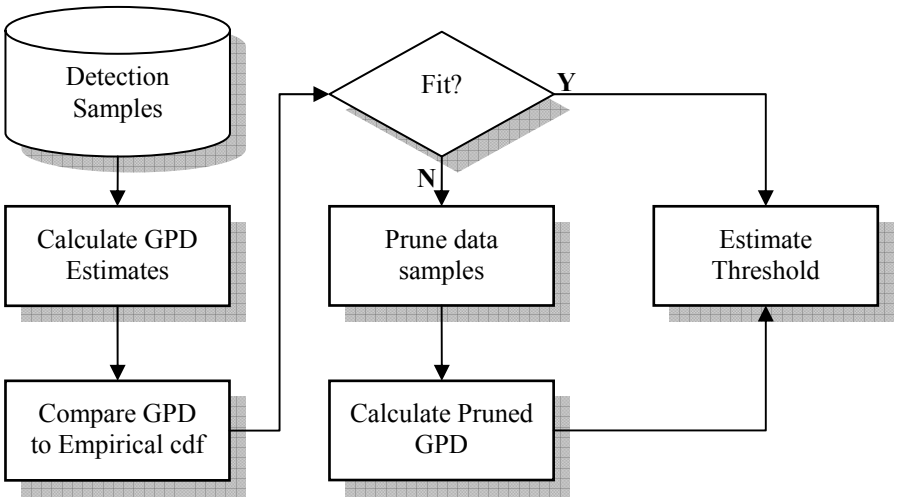


Figure 37: Block Diagram of the EVT Adaptive Threshold Algorithm

6.3. Experimental Results

Our hypothesis is we can detect and eliminate the influence of target samples to adaptively threshold detection results. Not only can we eliminate the influence of the target samples, but by using the generalized Pareto distribution, we can accurately estimate a threshold for a desired false alarm density. To show whether this occurs or

not in practice, we have implemented a number of experiments on both known distributions and on subpixel detector results from real-world hyperspectral imagery. The following sections describe the experimental design philosophy and provide results for two experiments measuring the accuracy of the GPD against known distributions and the ability of the EVT Adaptive Threshold algorithm to determine the thresholds for desired false alarm densities on subpixel detection results.

6.3.1. Experiments with Known Distributions

The first set of experiments shows the ability of the GPD to accurately estimate thresholds on known distributions. We use three distributions for this experiment: the normal distribution, the chi-squared distribution with 169 degrees of freedom, and a beta distribution with parameters 0.5 and 84. The normal distribution was used as a statistical benchmark. The chi-squared distribution was used because it represents the detection output of the well-known RX anomaly detector [84]. Finally, the beta distribution represents the statistical output of the ACE detector introduced in Chapter 5.

Another reason for using these distributions is because they all represent different ranges and limits. The normal distribution is valid for the entire real line. The chi-squared distribution is only valid for non-negative values of the real-line. The most limiting distribution is the beta distribution whose range is restricted between 0 and 1. All of these distributions test the ability of the GPD estimate to adapt to different statistical properties. Again, the GPD knows nothing about the true underlying distribution – only that the various tails of the distributions should converge in probability to a generalized Pareto distribution.

For each of the distributions listed above, a set of experiments was conducted to measure the accuracy and precision of the MC and GPD methods. The experiments were developed to estimate thresholds for false alarm densities at 10^{-2}, 10^{-3}, and 10^{-4} given 1000 samples from the distribution in question. Note that these experiments should task each of the methods by attempting to find thresholds as low as 10^{-4} with only 1000 samples – a threshold beyond the MC method's abilities. At each of the thresholds, 1000 runs were performed to achieve reasonable measurements of the mean and variance. The results of these experiments are given in Table 11. The table includes estimates for the MC method, the GPD method with clustering threshold of 10%, and the theoretical ideal for each false alarm probability α_0. For the MC and GPD methods, the table includes the mean with the variance in parentheses for each α_0.

Table 11: Comparison of MC and GPD on Known Distributions

Distribution	α_0	Ideal	MC	GPD
N(0,1)	10^{-2}	2.326	2.348 (0.016)	2.331 (0.009)
	10^{-3}	3.090	3.233 (0.125)	3.038 (0.053)
	10^{-4}	3.719	3.239 (0.122)	3.517 (0.205)
χ^2_{169}	10^{-2}	187.5	187.8 (5.967)	187.6 (3.556)
	10^{-3}	203.4	206.9 (56.83)	202.3 (24.57)
	10^{-4}	217.0	206.9 (56.48)	213.6 (109.4)
Beta(0.5,84)	10^{-2}	0.0386	0.0393 ($1.1\cdot10^{-5}$)	0.0384 ($0.6\cdot10^{-5}$)
	10^{-3}	0.0622	0.0675 ($1.6\cdot10^{-4}$)	0.0612 ($0.7\cdot10^{-4}$)
	10^{-4}	0.0859	0.0685 ($1.7\cdot10^{-4}$)	0.0875 ($5.1\cdot10^{-4}$)

The results from these experiments demonstrate the theoretical gains of using the GPD method. For all distributions, the GPD method obtains a better estimate of the threshold with nearly half the variance of the MC method. This is expected given the variance reduction benefits of using the generalized Pareto distribution. The GPD is also able to provide an estimate for $\alpha_0 = 10^{-4}$. While the estimate does have some bias, it shows the ability of the GPD to take advantage of its variance reduction property to estimate thresholds beyond that of MC methods.

6.3.2. Experiments on Subpixel Target Detectors

The simulated results are good for comparing the GPD method with its MC counterpart, but these experiments do not take into account situations that occur in real HSI data. In these cases, the data may not be necessarily homogeneous and can contain numerous outliers. This is especially true when targets are present in the imagery. To measure the usefulness of the GPD-based EVT adaptive threshold method on such data, we applied it and a number of other well-known techniques to the ACE and HSD detector results from Chapter 5 on Target 2. The ACE results were chosen because ACE has a known output distribution (assuming normal statistics). We chose HSD because the detector's output statistics cannot be easily quantified. Target 2 was chosen because it is not the strongest or weakest target signature and provides a good challenge for the algorithms.

6.3.2.1. ACE Threshold Results

For the experiments with the ACE detector, we tested four different algorithms. The parameters for this experiment were set such that the desired false alarm density varied from 10^{-3} to 10^{-5}, P is 1, and L is 169. The first algorithm is based on a theoretical calculation using (57). The second algorithm is a parametric

algorithm based on (57); however, instead of using the theoretical parameters, the parameters are estimated directly from the data. The third algorithm is the MC algorithm. The last algorithm is the proposed EVT method. For the EVT method, we use the clustering threshold of 1% to select the samples for estimation of the GPD parameters. On Images 2, 3, 5, and 6, no targets are present; therefore, the MC method should be ideal. On Images 1 and 4, however, where numerous targets are present in the data, we expect the EVT method to perform best. The results for the ACE detector are in Table 12 through Table 14.

Table 12: Comparison of Threshold Estimates for ACE Results

α_0	Image	Theoretical	Parametric	MC	EVT	Ideal
10^{-3}	1	0.0626	0.0664	0.1136	0.0736	0.0759
	2	0.0626	0.0610	0.0681	0.0695	0.0681
	3	0.0626	0.0668	0.0740	0.0751	0.0740
	4	0.0626	0.0656	0.0970	0.0711	0.0750
	5	0.0626	0.0600	0.0690	0.0707	0.0690
	6	0.0626	0.0682	0.0804	0.0823	0.0804
10^{-4}	1	0.0864	0.0922	0.6428	0.1171	0.1146
	2	0.0864	0.0843	0.1111	0.1063	0.1111
	3	0.0864	0.0923	0.1161	0.1146	0.1161
	4	0.0864	0.0910	0.6951	0.1126	0.1203
	5	0.0864	0.0830	0.1449	0.1123	0.1449
	6	0.0864	0.0944	0.1334	0.1305	0.1334
10^{-5}	1	0.1100	0.1177	0.7644	0.1737	0.1533
	2	0.1100	0.1075	0.2201	0.1515	0.2201
	3	0.1100	0.1175	0.1876	0.1630	0.1876
	4	0.1100	0.1162	0.8396	0.1710	0.1684
	5	0.1100	0.1057	0.2637	0.1669	0.2637
	6	0.1100	0.1201	0.2435	0.1935	0.2435

In each table, there are seven columns. The first column identifies the desired false alarm rate we want to achieve. The second column identifies the image that is being processed. The next four columns give the results for the theoretical, parametric, MC, and EVT methods. The last column presents the ideal results for the

desired false alarm rate. This ideal setting was found using the ground truth information to identify target clusters as described in Chapter 2. These target samples were then removed and the rest of the pixels were ordered by detection score. The MC method was then applied to this reduced set to identify the "ideal" threshold.

Table 13: Comparison of P_d Estimates for ACE Results

α_0	Image	Theoretical	Parametric	MC	EVT	Ideal
10^{-3}	1	1.00	1.00	1.00	1.00	1.00
	2	0.00	0.00	0.00	0.00	0.00
	3	0.00	0.00	0.00	0.00	0.00
	4	1.00	1.00	1.00	1.00	1.00
	5	0.00	0.00	0.00	0.00	0.00
	6	0.00	0.00	0.00	0.00	0.00
10^{-4}	1	1.00	1.00	0.24	1.00	1.00
	2	0.00	0.00	0.00	0.00	0.00
	3	0.00	0.00	0.00	0.00	0.00
	4	1.00	1.00	0.33	1.00	1.00
	5	0.00	0.00	0.00	0.00	0.00
	6	0.00	0.00	0.00	0.00	0.00
10^{-5}	1	1.00	1.00	0.02	0.98	1.00
	2	0.00	0.00	0.00	0.00	0.00
	3	0.00	0.00	0.00	0.00	0.00
	4	1.00	1.00	0.03	1.00	1.00
	5	0.00	0.00	0.00	0.00	0.00
	6	0.00	0.00	0.00	0.00	0.00

The results show the usefulness of the EVT method even when the detector distribution can be assumed. The theoretical calculation using the beta distribution underestimates the thresholds consistently. This leads to false alarm rates that are significantly higher than the desired rates. In the most extreme case of 10^{-5}, the false alarm rate is nearly an order of magnitude greater than the desired rate. While the ACE detector is a CFAR detector, the high false alarm rates occur because the underlying HSI data is rarely normally distributed [103]. This assumption of

normality leads to a mismatch between theory and real HSI data causing the higher false alarms and incorrect thresholds.

The parametric method performs slightly better than the theoretical case. Instead of using the predicted parameters for the beta distribution, the parameters are estimated using the maximum likelihood technique. These estimates do improve the results, but the underlying assumption that the data comes from a normal distribution (thus leading to the beta distribution of the ACE detector) does not match the true distribution of the HSI data. Therefore even with estimated parameters, the parametric method does not perform well.

Table 14: Comparison of False Alarms for ACE Results

α_0	Image	Theoretical	Parametric	MC	EVT	Ideal
10^{-3}	1	256	202	11	120	102
	2	155	180	102	94	102
	3	226	169	102	95	102
	4	216	182	29	125	102
	5	147	181	102	91	102
	6	327	220	102	96	102
10^{-4}	1	50	42	0	9	10
	2	26	29	10	11	10
	3	45	32	10	10	10
	4	55	38	0	13	10
	5	42	45	10	20	10
	6	73	43	10	12	10
10^{-5}	1	13	8	0	0	1
	2	10	10	1	2	1
	3	11	9	1	2	1
	4	16	11	0	0	1
	5	22	23	1	5	1
	6	22	14	1	4	1

The MC estimates are more interesting. As expected, the MC estimates are ideal when no targets are present. If only a few targets are present, the MC estimates will continue to provide good thresholds for larger desired false alarm rates. In these

experiments however, the targets span tens of pixels. While this may not be significant at 10^{-2}, it does affect the P_d and desired false alarm rates at 10^{-3} and below. Because the MC method has no mechanism to identify possible target samples, it degrades as the desired false alarm density becomes small. This has the unfortunate effect of removing target detections first before removing clutter (assuming the detector has done an adequate job of separating the targets from the background). The final result is threshold estimates much higher than the ideal which penalize the P_d.

The EVT method performs well in these experiments. The method was able to isolate the influence of the target signatures in Images 1 and 4 before calculating the threshold. The result is a threshold that is near ideal for false alarm rates of 10^{-3} and 10^{-4}. At these false alarm rates, the method provides P_d and false alarm numbers that are unmatched by any other algorithm when targets are present. At the 10^{-5} false alarm rate, the EVT method begins to diverge from the ideal cases; however, the EVT method still provides thresholds that exceed the ability of the MC method. This is an intriguing result as the EVT method is using less than 10,000 samples to estimate a 10^{-5} desired false alarm rate with good accuracy. When targets are not present, the MC method provides the best results as expected; however, the EVT method provides results that are close to ideal. When considering the EVT method's ability to estimate thresholds close to ideal in the presence or absence of targets, the slight errors in threshold level are acceptable to maintain good performance in all conditions.

6.3.2.2. HSD Threshold Results

For the experiments with the HSD detector, we tested only two algorithms because HSD's use of non-negativity constraints precludes the derivation of a theoretical distribution for the detector. The parameters for this experiment were set

such that the desired false alarm density varied from 10^{-3} to 10^{-5} as in the ACE experiment. The two algorithms tested are the MC and EVT methods. For the EVT method, we use the clustering threshold of 1% to select the samples for estimation of the GPD parameters. On Images 2, 3, 5, and 6, no targets are present; therefore, the MC method should be ideal. On Images 1 and 4, however, where numerous targets are present in the data, we expect the EVT method to perform best. The results for the HSD detector are in Table 15 through Table 17.

Table 15: Comparison of Threshold Estimates for HSD Results

α_0	Image	MC	EVT	Ideal
10^{-3}	1	1.0912	1.0540	1.0529
	2	1.0750	1.0738	1.0750
	3	1.0266	1.0207	1.0266
	4	1.1199	1.0884	1.0934
	5	1.0669	1.0668	1.0669
	6	1.0706	1.0709	1.0706
10^{-4}	1	2.4647	1.1011	1.0912
	2	1.1061	1.1142	1.1061
	3	1.0455	1.0416	1.0455
	4	3.1925	1.1395	1.1491
	5	1.0898	1.0973	1.0898
	6	1.1064	1.1131	1.1064
10^{-5}	1	3.7026	1.1759	1.1124
	2	1.1439	1.1632	1.1439
	3	1.0862	1.0773	1.0862
	4	6.5592	1.2100	1.2148
	5	1.1148	1.1312	1.1148
	6	1.1614	1.1687	1.1614

In each table, there are five columns. The first column identifies the desired false alarm rate we want to achieve. The second column identifies the image that is being processed. The next two columns give the results for the MC method and EVT method. The last column provides the ideal results for the desired false alarm rate. This ideal setting was found using the ground truth information to identify target

clusters as described in Chapter 2. These target samples were then removed and the rest of the pixels were ordered by detection score. The MC method was then applied to this reduced set to identify the "ideal" threshold.

Table 16: Comparison of P_d Estimates for HSD Results

α_0	Image	MC	EVT	Ideal
10^{-3}	1	1.00	1.00	1.00
	2	0.00	0.00	0.00
	3	0.00	0.00	0.00
	4	1.00	1.00	1.00
	5	0.00	0.00	0.00
	6	0.00	0.00	0.00
10^{-4}	1	0.24	1.00	1.00
	2	0.00	0.00	0.00
	3	0.00	0.00	0.00
	4	0.33	1.00	1.00
	5	0.00	0.00	0.00
	6	0.00	0.00	0.00
10^{-5}	1	0.02	0.93	1.00
	2	0.00	0.00	0.00
	3	0.00	0.00	0.00
	4	0.03	1.00	1.00
	5	0.00	0.00	0.00
	6	0.00	0.00	0.00

The results for this experiment support the results found using the ACE detector. In this case, however, the detector statistics are entirely unknown and have to be estimated from the data. As expected, the MC method is ideal when no targets are present in the imagery. Once target detections are present, the MC method performs poorly setting the threshold based on target detection scores. This effect, of course, removes targets while giving improper false alarm rates.

The EVT method is able to isolate the target detections and provide good detection thresholds across all images. In images with targets, the EVT method is able to remove the influence of the target samples and calculate thresholds that are near

ideal. The corresponding P_d and false alarm statistics show good performance across all desired false alarm rates. When targets are not present, the EVT method achieves thresholds close to ideal. Again, the GPD method gives good performance across all images regardless of the detection of targets.

Table 17: Comparison of False Alarm Rates for HSD Results

α_0	Image	MC	EVT	Ideal
10^{-3}	1	10	95	102
	2	102	108	102
	3	102	201	102
	4	27	134	102
	5	102	104	102
	6	102	99	102
10^{-4}	1	0	4	10
	2	10	5	10
	3	10	14	10
	4	0	13	10
	5	10	5	10
	6	10	8	10
10^{-5}	1	0	0	1
	2	1	0	1
	3	1	1	1
	4	0	2	1
	5	1	0	1
	6	1	0	1

6.3.3. Conclusions

The EVT adaptive threshold method was developed to work well across all types of detectors and in the presence of targets. The experimental results demonstrate this ability across two different detectors and at multiple desired false alarm rates – even at rates lower than the number of samples present. Strikingly, the method also excels above the theoretical and parametric methods which are based on the known distribution of the detector (unless the data distribution matches the assumed detector distribution).

The other benefit of the EVT method is the speed of calculation. The method takes less than a second to estimate a threshold given a 256x400 pixel image. The method is scalable to any size image and performs as quickly as any of the other methods. This makes the EVT method accessible to a wide range of target applications beyond subpixel detection.

6.4. Summary

We present a new way to adaptively estimate detector thresholds via extreme value theory. The method can be used on any detector type – not just those that are CFAR algorithms. In most real-world cases, the EVT adaptive threshold algorithm can outperform CFAR algorithms due to the inherent mismatch between the model assumptions and the real data. Additionally, the EVT method can work in the presence of target detections while still estimating an accurate threshold for a desired false alarm rate. This ability makes it useful to any number of detection applications – not just physics-based subpixel target detection in HSI data.

Chapter 7: Summary

In this dissertation, we have introduced a number of new algorithms for detection of subpixel targets in hyperspectral imagery. Our approach has been to incorporate the known physics of the problem while taking advantage of statistics to account for the unknown variables. Till this point, we have introduced each algorithm separately to isolate their performance. In this chapter, we introduce how these algorithms work together. From this analysis, we identify new areas of research for subpixel detection. We conclude this chapter by summarizing the new algorithms introduced in this dissertation.

7.1. Cumulative Performance Results

In Chapter 1, we presented a block diagram for subpixel target detection in Figure 2. Using that block diagram, we identified the various areas of subpixel detection where we developed new algorithms. These algorithms were independently updated to identify their performance without the influence of the other algorithms. Unfortunately, this never allowed us to bring all the algorithms together to measure their cumulative performance. This section presents an experiment designed to test the cumulative performance of the proposed algorithms.

Figure 38 presents the proposed subpixel detection system. For target characterization, we use the ARRT algorithm introduced in Chapter 3. For background characterization, we use the IEA algorithm and the SDD algorithms described in Chapter 4. The subpixel detector is the HSD algorithm introduced in Chapter 5. Finally, the EVT Adaptive Threshold Algorithm applies a detection threshold based on a desired false alarm density to the HSD detection scores.

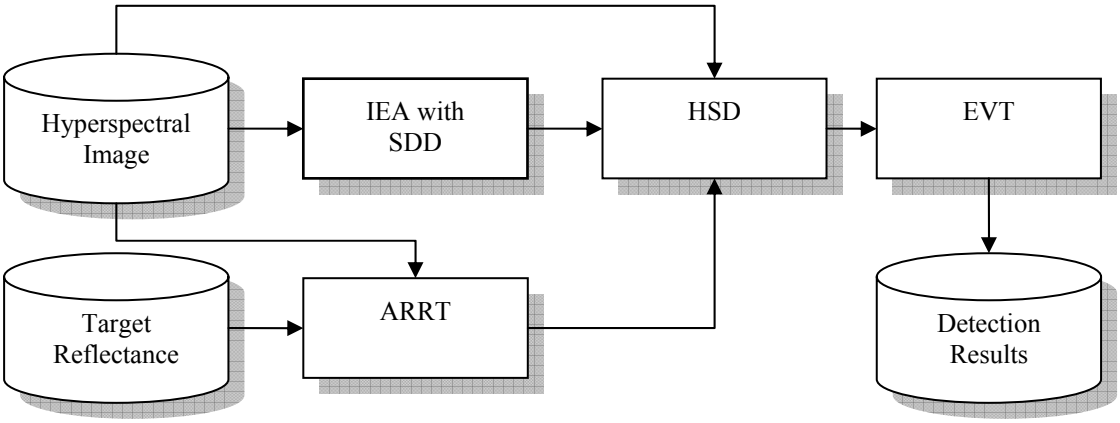


Figure 38: Proposed Subpixel Detection Block Diagram

To show how all of our proposed algorithms work together, we designed an experiment on Target 2. We chose Target 2 because it is not the easiest or hardest target to detect providing a moderate challenge for subpixel detection. We used Images 1 through 6 from Sensor X because these images contain true subpixel targets. The images were left uncalibrated for this experiment to test the ability of the ARRT algorithm to adjust to such conditions. For the target and background reflectance signatures, we used Target 2 and vegetation signatures measured in the field using hand-held spectrometers. No other information was needed to run the system.

The results of the experiment are shown in Figure 39. For reference, we included the best case results for the HSD algorithm operating on Target 2 (as shown in Chapter 5). This best case result assumes the imagery has been vicariously calibrated and target signatures are generated using the MODTRAN algorithm. Additionally, the number of endmembers has been chosen to maximize performance based on ground truth information. This curve represents what a subpixel detector could achieve if all other variables were known.

The dashed gray line is the performance of the HSD algorithm using the EIF background dimension estimate. The EIF method provides consistently good results as shown in Chapter 4. We included this performance curve to show the need for good background dimension estimates even with HSD – a detector partially invariant to the number of background endmembers used.

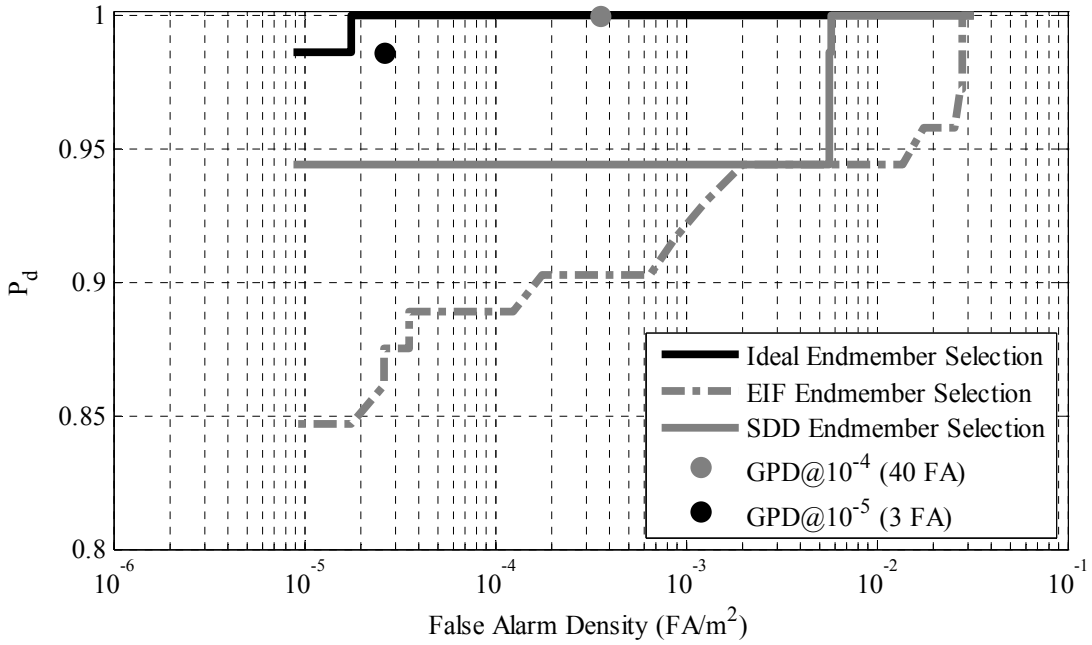


Figure 39: Subpixel Detection System ROC Curves

The solid gray line represents the results of our combined subpixel detector system in Figure 38. This curve shows the system achieves nearly ideal performance. Only two targets are missed at false alarm densities less than 10^{-5}. Even though HSD is partially insensitive to the number of background endmembers chosen, the SDD algorithm is able to produce better results than the EIF algorithm.

Perhaps the most impressive results are the two points calculated by using the EVT Adaptive Threshold Algorithm. The EVT algorithm was applied to the results of the HSD detector (gray line). As noted in Chapter 5, the HSD algorithm sometimes

does not suppress the background into similar ranges of values. The EVT algorithm automatically adapts the threshold for each image taking into account the different background ranges. The result of applying the EVT algorithm provides performance that almost perfectly matches the ideal case. Even though the EVT algorithm is not able to fix the false alarm density exactly, it provides estimates that are very close to the ideal.

The final result is that the proposed combined subpixel detection system is able to obtain performance that is nearly identical to the case where all parameters are known. When one considers the proposed system only uses a target reflectance signature, a reference reflectance signature, and the hyperspectral image without any knowledge of ground truth, the combined performance result is striking. Moreover, the proposed subpixel detection system is able to process each image in less than five minutes making it applicable for near-real time applications.

7.2. Future Work

While this work demonstrates good results for subpixel detection, there are many more interesting topics that spring from the research within this dissertation. Perhaps the most immediate need is improved characterization of target signatures as demonstrated by the subpixel detection results on Target 4. The ARRT and MODTRAN methods both have difficulty handling low reflectance targets. They both produced signatures for Target 4 that underestimated the actual target signature in the SWIR bands. Work should focus on providing better estimates of the upwelled radiance signature using shadow zones as indicated by [80]. These shadow zones can be automatically identified using [1]. Methods can also focus on improved estimates

of the aerosol content of the imagery to help characterize scattering loses at different altitudes.

Estimation of the background dimension remains an active area of research. As shown in Chapter 4, this topic has been only partially treated in the literature. New methods that incorporate target, background, and detector characteristics need to be developed to help improve this area. While our research has produced an improved method to estimate the background dimension, much more could be done.

Another interesting area of research is using the contextual information gained by using physically meaningful endmembers and abundances. For example, when looking for a white automobile, you can remove detections that are not on roads or parking lots. This information can be used to build site models that lead to improved spectral object level change detection (SOLCD) studies [44].

An interesting branch of subpixel detection was proposed by Kwon and Nasrabadi using kernel-based methods [60][61]. The reason for using kernel methods is to project the data into a space that can account for nonlinearities in the data not covered by first and second order moments. They show promising results although their work uses the energy algorithm to estimate the number of background endmembers for the AMSD algorithm [60]. Thus, we cannot identify how well the kernel methods improve detection performance because AMSD performance has been degraded unintentionally.

Nevertheless, the kernel methods open up the possibility of physics-based kernel methods. Just as we created the hybrid detectors by incorporating the known physics of the linear mixing model, we can take the same approach with their kernel

counterparts. For example, research has proposed a new method to extract endmembers based on Support Vector Data Description [6]. This method extracts endmembers in the kernel space that identify the endmembers as the vertices of the enclosing hypersphere. From this work, we developed a Kernel FCLS method to accurately estimate the abundances of those endmembers in the kernel space allowing for the possibility of greater separation between similar spectral signatures [11]. The next step is to modify the Kernel AMSD and Kernel ACE detectors to use the new physics-based kernel parameters. This work will produce a Kernel Hybrid Structured Detector and Kernel Hybrid Unstructured Detector. These algorithms will then be assessed relative to their hybrid counterparts presented in [12]. Other interesting work in kernel methods is the development of algorithms to estimate the kernel parameters – a challenging subject in all kernel methods [92].

While this dissertation focused on the reflective region of the electromagnetic spectrum, hyperspectral sensors have been developed for the Mid-Wave Infrared (MWIR) from 3.0 to 7.0 microns and the Long Wave Infrared (LWIR) from 7.0 to 15.0 microns regions as well. At these wavelengths, emissivity dominates the spectral signature. Emissivity is "the ratio of the emission from [a] material to that of a blackbody at the same temperature" [93]. Therefore, emissivity is a measure of the energy an object emits instead of reflects. Initial work has already been finished applying the hybrid detectors to LWIR sensors [13]. However, target characterization is much more difficult in MWIR and LWIR because temperature has to be accounted for as well as the emissivity [93]. These topics should be pursued however because LWIR sensors provide the opportunity to work in either day or night conditions.

7.3. Contributions

In this dissertation, we present a physics-based approach to subpixel detection in hyperspectral imagery. This physics-based approach required the development of new techniques at all levels of subpixel detection from target characterization to threshold estimation. In this section, we summarize the contributions of this thesis:

- We have developed a new target characterization method based on principles of radiative transfer theory and detection theory. Results show this method matches the results by model-based methods, but requires no ancillary data such as weather information, source-target-receiver information, or calibrated sensor responses.
- We have developed a new method to estimate the number of endmembers for subpixel detection applications. We show that the proposed SDD method performs well when compared to the state-of-the-art methods.
- More importantly, we show that for the first time how poor estimates of background dimension lead to significantly reduced subpixel detection performance.
- We created two new physics-based subpixel detectors. The HSD and HUD detectors are the combination of physics-based knowledge to produce physically meaningful parameter estimates and detection theory to account for unknown quantities in the data. Results show these detectors have three advantages: insensitivity to the number of endmembers, improved performance on an image to image basis, and consistent performance across images better than that of known CFAR detectors.

- We developed an adaptive threshold technique based on extreme value theory. This technique is applicable to a wide variety of detectors – not just those that are CFAR. Additionally, the method is able to suppress the influence of target detections to make accurate estimates of the detection threshold without any knowledge of the underlying distribution of the data.

Bibliography

[1] S.M Adler-Golden, Mw. Matthew, G.P. Anderson, G.W. Felde, and J.A. Gardner, "Algorithm for de-shadowing spectral imagery," *Proceedings of the SPIE, Imaging Spectrometry VIII*, vol. 4816, pp. 203-210, November 2002.

[2] H. Akaike, "A new look at the statistical identification model," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716-723, December 1974.

[3] G.P. Anderson, B. Pukall, C.L. Allred, L.S. Jeong, M. Hoke, J.H. Chetwynd, S.M. Adler-Golden, A. Berk, L.S. Bernstein, S.C. Richtsmeier, P.K. Acharya, and M.W. Matthew, "FLAASH and MODTRAN4: state-of-the-art atmospheric correction for hyperspectral data," *IEEE Proceedings of the1999 Aerospace Conference*, vol. 4, pp 177-181, March 1999.

[4] E. A. Ashton and A. Schaum, "Algorithms for the detection of sub-pixel targets in multispectral imagery," *Photogramm. Eng. Remote Sensing*, pp. 723–731, July 1998.

[5] A.A. Balkema and L. de Haan, "Residual life time at great age," *The Annals of Probability*, vol. 2, no. 5, pp. 792-804, October 1974.

[6] A. Banerjee, P. Burlina, and J. Broadwater, "A machine learning approach for finding hyperspectral endmembers," accepted to *IEEE International Geoscience and Remote Sensing Symposium 2007*, Barcelona, Spain, July 2007.

[7] C.A. Bateson, G.P. Asner, and C.A. Wessman, "Endmember bundles: A new approach to incorporating endmember variability into spectral mixture analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, pp. 1083-1094, March 2000.

[8] J. Boardman, "Inversion of high spectral resolution data," *Proc. SPIE*, vol. 1298, pp. 222–233, 1990.

[9] J.W. Boardman, F.A. Kruse, and R.O. Green, "Mapping target signatures via partial unmixing of AVIRIS data," *Summaries of the VI JPL Airborne Earth Science Workshop*, Pasadena, CA 1995.

[10] J. Bowles, D. Gillis, and P. Palmadesso, "New improvements in the ORASIS algorithm," *Proceedings of the IEEE Aerospace Conference*, Big Sky, MT, vol. 3, pp. 293-298, March 2000.

[11] J. Broadwater, A. Banerjee, P. Burlina, and R. Chellappa, "Kernel Fully Constrained Least Squares Abundance Estimates," accepted to *IEEE*

International Geoscience and Remote Sensing Symposium 2007, Barcelona, Spain, July 2007.

[12] J. Broadwater and R. Chellappa, "Hybrid Detectors for Subpixel Targets," accepted to *IEEE Transactions on Pattern Analysis and Machine Intelligence*, December 2006.

[13] J. Broadwater and R. Chellappa, "Physics-based detectors applied to long-wave infrared hyperspectral data," *Proceedings of the 2006 Army Science Conference*, November 2006.

[14] J. Broadwater and R. Chellappa, "An adaptive threshold method for hyperspectral target detection," *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. V-1201 – V-1204, May 2006.

[15] J. Broadwater, R. Meth, and R. Chellappa, "Average relative radiance transform for subpixel detection," *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium 2005*, vol. 5, Seoul, South Korea, pp. 3565-3568, July 2005.

[16] J. Broadwater, R. Meth, and R. Chellappa, "A Hybrid Algorithm for Subpixel Detection in Hyperspectral Imagery," *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium 2004*, vol. 3, Anchorage, AK, pp. 1601-1604, September 2004.

[17] J.A. Bucklew, "The blind simulation problem and regenerative processes," *IEEE Transactions on Information Theory*, vol. 44, no. 7, November 1998, pp. 2877-2891.

[18] G. Casella and R.L. Berger, *Statistical Inference*, Second Ed. Duxbury Press, 2001.

[19] C-I Chang and Q. Du, "Estimation of Number of Spectrally Distinct Signal Sources in Hyperspectral Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 3, pp. 608-619, March 2004.

[20] C-I Chang and D.C. Heinz, "Constrained Subpixel Target Detection for Remotely Sensed Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 3, pp. 1144-1159, May 2000.

[21] R.N. Clark, G.A. Swayze, K.E. Livo, R.F. Kokaly, T.V.V. King, J.B. Dalton, J.S. Vance, B.W. Rockwell, T. Hoefen, and R.R. McDougal, "Surface Reflectance Calibration of Terrestrial Imaging Spectroscopy Data: A Tutorial Using AVIRIS," *Proceedings of the 10th Airborne Earth Science Workshop*, JPL Publication 02-1, 2002.

[22] D.R. Cox and D. Oakes, *Analysis of Survival Data*, Chapman & Hall, London, 1984.

[23] R. Davis and S. Resnick, "Tail estimates motivated by extreme value theory," *The Annals of Statistics*, vol. 12, no. 4, pp. 1467-1487, December 1984.

[24] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification*, Second Edition, John Wiley & Sons, Inc., 2001.

[25] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modeling Extremal Events for Insurance and Finance*, Springer, Berlin, 1997.

[26] W.H. Farrand, R.B. Singer, and E. Merenyi, "Retrieval of apparent surface reflectance from AVIRIS data – a comparison of empirical line, radiative-transfer, and spectral mixture methods," *Remote Sensing of Environment*, Vol. 47, pp. 311-321, 1994.

[27] M.D. Farrell and R.M Mersereau, "On the Impact of Covariance Contamination for Adaptive Detection in Hyperspectral Imaging," *IEEE Signal Processing Letters*, vol. 12, no. 9, pp. 649-652, September 2005.

[28] R.A. Fisher and L.H.C. Tippett, "Limiting forms of the frequency distribution of the largest or smallest member of a sample," *Proceedings of the Cambridge Philosophical Society*, vol. 24, pp. 180-190, 1928.

[29] G. Foody and D. Cox, "Sub-pixel land cover composition estimation using a linear mixture model and fuzzy membership model and fuzzy membership function," *International Journal of Remote Sensing*, vol. 15, pp. 619-631, 1994.

[30] M. Frechet, "Sur la loi de probabilité de l'écart maximum," *Ann. Soc. Pol. Math. (Krakow)*, vol. 6, pp. 93-116, 1927.

[31] J.R. Freemantle, R. Pu, and J.R. Miller, "Calibration of imaging spectrometer data to reflectance using pseudo-invariant features," *Proceedings of the 15th Candian Symposium on Remote Sensing*, pp. 452-455, 1992.

[32] B.C. Gao and A.F.H. Goetz, "Column atmospheric water vapor and vegetation liquid water retrievals from airborne imaging spectrometer data," *Journal of Geophysical Res. – Atm.,* Vol. 95, pp. 3549-3564, 1990.

[33] R. Gençay, F. Selçuk, and A. Ulugülyağci, "EVIM: A software package for extreme value analysis in MATLAB," *Studies in Nonlinear Dynamics and Econometrics*, vol. 5, no. 3, pp. 213-244, 2001.

[34] M. Gianinetto and G. Lechi, "The Development of Superspectral Approaches for the Improvement of Land Cover Classification," *IEEE Transactions on*

Geoscience and Remote Sensing, vol. 42, no. 11, pp. 2670-2679, November 2004.

[35] B. Gnedenko, "Sur la distribution limite du terme maximum d'une série aléatoire," *Annals of Mathematics*, vol. 44, pp. 423-453, 1943.

[36] A.A. Green, M. Berman, P. Switzer, and M.D. Craig, "A Transformation for Ordering Multispectral Data in Terms of Image Quality with Implications for Noise Removal," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 26, no. 1, January 1988.

[37] J.M. Grossman, J. Bowles, D. Haas, J.A. Antoniades, M.R. Grunes, P. Palmadesso, D. Gillis, K.Y. Tsang, M. Baumback, M. Daniel, J. Fisher, and I. Triandaf, "Hyperspectral analysis and target detection system for the adaptive spectral reconnaissance program (ASRP)," *SPIE, Algorithms for Multispectral and Hyperspectral Imagery IV*, Orlando, FL, vol. 3372, pp. 2-13, 13-14 April 1998

[38] M. Guida, D. Iovino, and M. Longo, "Comparative performance analysis of some extrapolative estimators of probability tails," *IEEE Journal on Selected Areas in Communications*, vol. 6, no. 1, pp. 76-84, January 1988.

[39] E. Gumbel, *Statistics of Extremes*, Columbia University Press, New York, 1958.

[40] B. Hapke, *Introduction to the Theory of Reflectance and Emittance Spectroscopy*, Cambridge University Press: Cambridge, UK, 1993.

[41] J.C. Harsanyi and C-I Chang, "Hyperspectral Image Classification and Dimensionality Reduction: An Orthogonal Subspace Projection Approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 4, pp. 779-785, July 1994.

[42] K.H. Haskell and R.J. Hansen, "An algorithm for linear least squares problems with equality and non-negativity constraints generalized," *Math Prog.*, vol. 21, pp. 98-118, 1981.

[43] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer-Verlag, New York, NY, 2001.

[44] G.G. Hazel, "Object-level change detection in spectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 3, pp. 553-561, March 2001.

[45] G. Healey and D. Slater, "Models and Methods for Automated Material identification in Hyperspectral Imagery Acquired Under Unkonwn

Illumination and Atmospheric Conditions," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 6, pp 2706-2717, November 1999.

[46] D.C. Heinz and C-I Chang, "Fully Constrained Least Squares Linear Spectral Mixture Analysis Method for Material Quantification in Hyperspectral Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 3, pp. 529-545, March 2001.

[47] V.J. Hodge and J. Austin, "A survey of outlier detection methodologies," *Artificial Intelligence Review*, vol. 22, pp. 85-126, 2004.

[48] A.R .Huete, "A soil-adjusted vegetation index (SAVI)," *Remote Sensing of Environment*, vol. 25, no. 3, pp. 295-309, August 1988.

[49] S. Johnson, "The Constrained Signal Detector," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 6, pp. 1326-1337, June 2002.

[50] R. Kashyap, "A Bayesian Comparison of Different Classes of Dynamic Models Using Empirical Data," *IEEE Transactions on Automatic Control*, vol. AC-22, no. 5, pp. 715-727, October 1977.

[51] R. Kashyap, "Inconsistency of the AIC Rule for estimating the order of autoregressive models," *IEEE Transactions on Automatic Control*, vol. AC-25, no. 5, pp. 996-998, October 1980.

[52] Y.J. Kaufman, "Solution of the equation of radiative transfer for remote sensing over nonuniform surface reflectivity," *Journal of Geophysical Research*, vol. 87, no. C6, pp. 4137-4147, 1982.

[53] E.J. Kelly, "An Adaptive Detection Algorithm," *IEEE Transactions in Aerospace and Electronic Systems*, vol. 22, pp. 115-127, March 1986.

[54] N. Keshava and J.F. Mustard, "Spectral Unmixing," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 44-57, January 2002.

[55] G.R. Kirchhoff and R. Bunsen, "Chemical Analysis by Observation of Spectra," *Annalen der Physik und der Chemie*, vol. 110, pp. 161-189, 1860.

[56] S. Kraut and L.L. Scharf, "The CFAR adaptive sub-space detector is a scale-invariant GLRT," *IEEE Transactions on Signal Processing*, vol. 47, pp. 2538-2541, Sept. 1999.

[57] S. Kraut, L.L. Scharf, and L.T. McWhorter, "Adaptive subspace detectors," *IEEE Transactions in Signal Processing*, vol. 49, no. 1, pp. 1-16, January 2001.

[58] S. Kraut, L.L. Scharf, and R.W. Butler, "The Adaptive Coherence Estimator: A Uniformly Most-Powerful-Invariant Adaptive Detection Statistic," *IEEE*

Transactions on Signal Processing, vol. 53, no. 2, pp. 427-438, February 2005.

[59] F.A. Kruse, "Use of Airborne Imaging Spectrometer Data to Map Minerals Associated with Hydrothermally Altered Rocks in the Northern Grapevine Mountains, Nevada, and California," *Remote Sensing of the Environment*, vol. 24, no. 1, pp. 31-51, February 1988.

[60] H. Kwon and N.M. Nasrabadi, "Kernel matched subspace detectors for hyperspectral target detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 2, pp. 178-194, February 2006.

[61] H. Kwon and N.M. Nasrabadi, "Kernel adaptive subspace detector for hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 2, pp. 271-275, April 2006.

[62] J.C. Lagarias, J.A. Reeds, M.H. Wright, and P.E. Wright, "Convergence properties of the Nelder-Mead simplex method in low dimensions," *SIAM Journal of Optimization*, vol. 9, no. 1, pp. 112-147, 1998.

[63] D. Landgrebe, "The evolution of landsat data analysis," *Photogrammetric Engineering and Remote Sensing*, vol. LXIII, no. 7, pp 859-867, July 1997.

[64] D. Landgrebe, "Multispectral Land Sensing: Where From, Where to?," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 414-421, March 2005.

[65] K. Lee, "A Subpixel Scale Target Detection Algorithm for Hyperspectral Imagery," Ph.D. Dissertation, Rochester Institute of Technology, Center for Imaging Science, 2002.

[66] M. Lewis, V. Jooste, and A.A. de Gasparis, "Discrimination of Arid Vegetation with Airborne Multispectral Scanner Hyperspectral Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 7, pp. 1471-1479, July 2001.

[67] Lord Rayleigh (J.W. Strutt), "On the Light from the Sky, Its Polarization and Colour," *Philosophical Magazine*, vol. 41., pp. 107-120, 274-279, 1871.

[68] E.R. Malinowski, "Theory of error in factor analysis," *Analytical Chemistry*, vol. 49, no. 4, pp. 606-612, 1977.

[69] D. Manolakis, "Realistic matched filter performance prediction for hyperspectral target detection," *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, 2004*, vol. 2, pp. 953-955, September 2004.

[70] D. Manolakis and G. Shaw, "Detection Algorithms for Hyperspectral Imaging Applications," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 29-43, January 2002.

[71] D. Manolakis, C. Siracusa, and G. Shaw, "Hyperspectral Subpixel Target Detection Using the Linear Mixing Model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 7, pp. 1392-1409, July 2001.

[72] R. Marion, R. Michel, and C. Faye, "Measuring Trace Gases in Plumes from Hyperspectral Remotely Sensed Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 4, pp. 854-864, April 2004.

[73] S.S. McArdle, J.R. Miller, and J.R. Freemantle, "Airborne acquisition under clouds: preliminary comparisons with clear-sky scene radiance and reflectance imagery," *Proceedings of the 15th Canadian Symposium on Remote Sensing*, pp. 446-449, 1992.

[74] L.B. Milstein, D.L. Schilling, and J.K. Wolf, "Robust detection using extreme-value theory," *IEEE Transactions on Information Theory*, vol. IT-15, no.3, pp. 370-375, May 1969.

[75] P. K. Murphy and M. A. Kolodner, "Classification of Hyperspectral Data for Landuse Applications, Final Report", JHU/APL Report, SRM-01-063, February 8, 2002.

[76] J.F. Mustard and C.M. Pieters, "Photometric phase functions of common geologic minerals and applications to quantitative analysis of mineral mixture reflectance spectra," *J. Geophys. Res.*, vol. 94, pp. 13619-13634, 1989.

[77] R.A. Neville, K. Staenz, T. Szeredi, J. Lefebvre, and P. Hauff, "Automatic endmember extraction from hyperspectral data for mineral exploration," in *4th Int. Airborne Remote Sensing Conf. Exhibition/21st Canadian Symposium on Remote Sensing*, Ottawa, Ontario, Canada, 21-24, pp. 891-896, June 1999.

[78] J. Oberg, "High-tech sensor in the shuttle search," *MSNBC News*, http://www.msnbc.msn.com/id/3077570/, Houston, March 30, 2003

[79] J. Pickands, "Statistical inference using extreme order statistics," *The Annals of Statistics*, vol. 3, no. 1, pp. 119-131, 1975.

[80] K.R. Piech and J.E. Walker, "Interpretation of Soils," *Photogrammetric Engineering and Remote Sensing*, vol. 40, pp. 87-94, 1974.

[81] A. Plaza, P. Martínez, R. Pérez, and J. Plaza, "Spatial/spectral endmember extraction by multidimensional morphological operations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, pp. 2025-2041, September 2002.

[82] A. Plaza, P. Martínez, R. Pérez, and J. Plaza, "A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 3, pp. 650-663, March 2004.

[83] R. Price, C.D. Anger, and S. Mah, "Preliminary evaluation of *casi* preprocessing techniques," *Proceedings of the 17th Canadian Symposium on Remote Sensing*, pp. 694-697, 1995.

[84] I.S. Reed and X. Yu, "Adaptive Multiple-Band CFAR Detection of an Optical Pattern with Unknown Spectral Distribution," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 10, pp. 1760-1770, October 1990

[85] H. Ren and C-I Chang, "Automatic Spectral Target Recognition in Hyperspectral Imagery," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1232-1249, October 2003.

[86] H. Ren, Q. Du, and J. Jensen, "Constrained Weighted Least Squares Approaches for Target Detection and Classification in Hyperspectral Imagery," *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium 2002*, Toronto, Canada, pp. 3426-3428, June 2002.

[87] J. Rissanen, "A universal prior for integers and estimation by minimum description length," *Annals of Statistics*, vol. 11, pp. 416-431, 1983.

[88] D.A. Roberts, M.O. Smith, and J.B. Adams, "Green Vegetation, Nonphotosynthetic Vegetation, and Soils in AVIRIS Data," *Remote Sensing of the Environment*, vol. 44, pp. 3039-3054, 1993.

[89] S.J. Roberts, "Novelty detection using extreme value statistics," *IEE Proceedings of Vision, Image, and Signal Processing*, vol. 146, no. 3, pp. 124-129, June 1999.

[90] J.W. Rouse, R.H. Haas, J.A. Schell, and D.W. Deering, "Monitoring vegetation systems in the great plains with ERTS," *Third ERTS Symposium*, NASA SP-351, pp. 309-317, 1973.

[91] J.S. Sadowsky and J.A. Bucklew, "On large deviations theory and asymptotically efficient Monte Carlo simulation," *IEEE Transactions on Information Theory*, vol. 36, issue 3, May 1990, pp. 579-588.

[92] B. Scholkopf and A.J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, 2002.

[93] J.R. Schott, *Remote Sensing: The Image Chain Approach*, Oxford, U.K.: Oxford Univ. Press, 1997.

[94] J.R. Schott, K. Lee, R. Raqueno, and G. Hoffman, "Use of Physics Based Models in Hyperspectral Image Exploitation," *IEEE Proc. of the 31st Applied Imagery Pattern Recognition Workshop (AIPR '02)*, pp. 36-42, October 2002.

[95] R.A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*, San Diego, CA: Academic, 1997.

[96] G. Schwartz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 5, no. 2, pp. 461-464, 1978.

[97] S. Seager, E.L. Turner, J. Schafer, and E.B. Ford, "Vegetation's Red Edge: A Possible Spectroscopic Biosignature of Extraterrestrial Plans," *Astrobiology*, vol. 5, no. 3, pp. 372-390, 2005.

[98] J. J. Settle and N. A. Drake, "Linear mixing and estimation of ground cover proportions," *Int. J. Remote Sensing*, vol. 14, no. 6, pp. 1159–1177, 1993.

[99] D. Siegmund, "Importance sampling in the Monte Carlo study of sequential tests," *Annals of Statistics*, vol. 4, pp. 673-684, 1976.

[100] R. Smith, "Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone," *Statistical Science*, vol. 4, pp. 367-393, 1989.

[101] R. Srinivasan, *Importance Sampling: Applications in Communications and Detection*, Springer-Verlag, Berlin, Germany, 2002.

[102] R. Srinivasan, "Simulation of CFAR detection algorithms for arbitrary clutter distributions," *IEE Proceedings, Radar, Sonar, and Navigation*, vol. 147, no. 6, pp. 31-40, December 2000.

[103] D.W. Stein, S.G. Beaven, L.E. Hoff, E.M. Winter, A.P. Shaum, and A.D. Stocker, "Anomaly Detection from Hyperspectral Imagery," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 58-69, January 2002.

[104] A.D. Stocker and P. Schaum, "Applications of stochastic mixing models to hyperspectral detection problems," *Proceedings of SPIE, Algorithms for Multispectral and Hyperspectral Imagery III*, vol. 3071, Orlando, FL, pp. 47-60, April 1997.

[105] A. Stratton, *Electromagnetic Theory*, McGraw-Hill, New York, NY, 1941.

[106] S. Subramanian and N. Gat, "Subpixel Object Detection Using Hyperspectral Imaging for Search and Rescue Operations," *Proc. SPIE for Automatic Target Recognition VIII*, vol. 3371, pp. 216-225, November 1998.

[107] G. Swayze, R.N. Clark, F. Kruse, S. Sutley, and A. Gallagher, "Ground-truthing AVIRIS Mineral Mapping at Cuprite, Nevada," *Summaries of the Third Annual JPL Airborne Geoscience Workshop*, vol. 1, pp. 47-49, 1992.

[108] D.M.J. Tax and R.P.W. Duin, "Support vector domain description," *Pattern Recognition Letters*, vol. 20, pp. 1191-1199, 1999.

[109] B. Thai and G. Healey, "Invariant Subpixel Material Detection in Hyperspectral Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 3, pp. 599-608, March 2002.

[110] S. Tompkins, J.F. Mustard, C.M. Pieters, and D.W. Forsyth, "Optimization of endmembers for spectral mixture analysis," *Remote Sensing of Environment*, vol. 59, pp. 472-489, 1997.

[111] G. Vane, R.O. Green, T.G. Chrien, H.T. Enmark, E.G. Hansen, and W.M. Porter, "The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)," *Remote Sensing of the Environment*, vol. 44, pp. 127-143, 1993.

[112] R. von Mises, "La distribution de la plus grande de n valeurs," *Rev. Math. Union Interbalkan*, vol. 1, pp. 141-160, 1936.

[113] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 2, pp. 387-392, April 1985.

[114] S.B. Weinstein, "Theory and application of some classical and generalized asymptotic distributions of extreme values," *IEEE Transactions on Information Theory*, vol. IT-19, no. 2, pp. 148-154, March 1973.

[115] M.E. Winter, "Fast autonomous spectral endmember determination in hyperspectral data," *Proc. 13th Int. Conf. Applied Geologic Remote Sensing*, Vol. II, Vancouver, BC, Canada, pp. 337-344, 1999.

[116] M.E. Winter, "N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," *Proceedings of SPIE*, vol. 3753, pp. 266-275, 1999.

HYPERSPECTRAL ENDMEMBER DETECTION AND BAND SELECTION USING BAYESIAN METHODS

By

ALINA ZARE

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2008

© 2008 Alina Zare

To Mom and Dad

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Paul Gader, for all of his guidance, support and the numerous opportunities he provided me throughout my studies and research. I would also like to thank my committee members, Dr. Jeffery Ho, Dr. Gerhard Ritter, Dr. Clint Slatton, and Dr. Joseph Wilson, for all of their help and valuable suggestions.

Thank you to Miranda Schatten Silvious of NVESD, Dr. Russell Harmon of ARL/ARO, Dr. William Clark of ARO, and Dr. Michael Cathcart of GTRI for their support throughout this research.

Additionally, thank you to my many former and current labmates. I am particularly grateful to Jeremy Bolton and Xuping Zhang for the countless number of times they provided insight and discussion during my studies.

Thank you to my parents, Roobik and Emik Zare, my sister, Anahita Zare, my boyfriend, Michael Black, and all of my family for their never-ending love, support and understanding. Finally, many thanks to my godfather, Artoosh Avanessian, whose influence first led me to computer science.

TABLE OF CONTENTS

page

ACKNOWLEDGMENTS 4

LIST OF TABLES 7

LIST OF FIGURES 8

LIST OF SYMBOLS AND ABBREVIATIONS 10

ABSTRACT 12

CHAPTER

1 INTRODUCTION 14

1.1 Hyperspectral Image Data and Analysis 14
1.1.1 Endmember Detection 15
1.1.2 Hyperspectral Band Selection 17
1.2 Statement of Problem 17
1.3 Overview of Research 18

2 LITERATURE REVIEW 20

2.1 Existing Endmember Detection Algorithms 20
2.1.1 Pixel Purity 20
2.1.2 Convex Hull 23
2.1.3 Nonnegative Matrix Factorization 25
2.1.4 Morphological Associative Memories 30
2.1.5 Evolutionary Search 35
2.1.6 Independent Components Analysis 36
2.1.7 Estimating the Number of Endmembers 38
2.2 Existing Hyperspectral Band Selection Algorithms 44
2.3 Summary of Literature Review 47

3 TECHNICAL APPROACH 56

3.1 Review of Sparsity Promotion Techniques 57
3.2 Review of the Iterated Constrained Endmembers Detection Algorithm . . . 58
3.3 New Endmember Detection Algorithm Using Sparsity Promoting Priors . . 60
3.4 New Band Selection Algorithm Using Sparsity Promoting Priors 62
3.5 New Endmember Distribution Detection Algorithm 65
3.6 Review of Markov Chain Monte Carlo Sampling Algorithms 69
3.7 Review of the Dirichlet Distribution and the Dirichlet Process 71
3.7.1 Dirichlet Process Mixture Model 73
3.7.2 Gibbs Sampling for the Dirichlet Process Mixture Model 74

3.8 New Piece-Wise Convex Endmember Detection Algorithm using the Dirichlet Process . . . 76

4 RESULTS . . . 82

4.1 Sparsity Promoting Iterated Constrained Endmember (SPICE) Algorithm Results . . . 82
4.1.1 The SPICE Two-Dimensional Example Results . . . 82
4.1.2 The SPICE AVIRIS Cuprite Data Results . . . 83
4.1.3 The SPICE AVIRIS Indian Pines Results . . . 85
4.1.4 The SPICE AHI Vegetation Detection Results . . . 86
4.2 Band Selecting SPICE (B-SPICE) Algorithm Results . . . 91
4.2.1 The B-SPICE AVIRIS Cuprite Data Results . . . 91
4.2.2 The B-SPICE AVIRIS Indian Pines Results . . . 92
4.2.3 The B-SPICE AVIRIS Indian Pines Results using Sampled Parameter Values . . . 94
4.3 Endmember Distribution (ED) Detection Results . . . 95
4.3.1 Results on Two-Dimensional Data using ED . . . 95
4.3.2 Results on AVIRIS Cuprite data using ED . . . 96
4.4 Piece-wise Convex Endmember (PCE) Detection Results . . . 96
4.4.1 Detection Results on Two-Dimensional Data using PCE . . . 96
4.4.2 Detection Results on the AVIRIS Indian Pines Data using PCE . . . 97

5 CONCLUSION . . . 130

REFERENCES . . . 132

BIOGRAPHICAL SKETCH . . . 140

LIST OF TABLES

Table page

4-1 Number of endmembers found by SPICE and ICE on AVIRIS Cuprite data . . . 102

4-2 Mine distributions in overlap regions of AHI and Lynx imagery 102

4-3 False alarm rate reduction using blackbody mask in AHI image 1 102

4-4 False alarm rate reduction using blackbody mask in AHI image 2 102

4-5 False alarm rate reduction using blackbody mask in AHI image 3 103

4-6 Mean and standard deviation of the number of endmembers and bands retained using SPICE and B-SPICE on the simulated AVIRIS Cuprite data set 103

4-7 Statistics of the averaged squared error per abundance value using SPICE and B-SPICE 103

4-8 Indian Pines data set classification results using SPICE and B-SPICE 104

4-9 The AVIRIS Indian Pines data B-SPICE results using sampled parameter values 104

4-10 Parameter values used to generate ED results on two-dimensional data sets . . . 104

4-11 Parameter values used to generate ED results on hyperspectral data sets 104

4-12 Parameter values used to generate PCE results 105

LIST OF FIGURES

Figure page

2-1 Three-dimensional data points and endmember results using convex cone analysis 49

2-2 Subset of AVIRIS Cuprite data scene 49

2-3 The CCA endmember results on AVIRIS Cuprite data 50

2-4 Morphological associative memories endmember detection results using the min memory on two-dimensional data 50

2-5 Morphological associative memories endmember detection results using both memories on two-dimensional data 51

2-6 Three dimensional data set generated from two endmembers with Gaussian noise. 52

2-7 Normalized AVIRIS Cuprite spectra 53

2-8 Data set generated from AVIRIS Cuprite endmembers with a small amount of Gaussian noise 54

2-9 Data set generated from AVIRIS Cuprite endmembers with a large amount of Gaussian noise 55

3-1 Endmember distribution algorithm's abundance prior 80

3-2 Endmember distribution algorithm's abundance prior as a function of $\mathbf{c}$ 80

3-3 Data points generated from two endmember distributions 81

4-1 Two-dimensional example data set 105

4-2 Comparison of ICE and SPICE algorithm results on two-dimensional data . . . 105

4-3 The SPICE results on two-dimensional data 106

4-4 Endmembers found using SPICE on AVIRIS Cuprite data 106

4-5 Comparison of SPICE endmember and USGS Alunite spectrum 107

4-6 Selected endmembers from AVIRIS Cuprite data 107

4-7 Test pixels from AVIRIS Cuprite data 108

4-8 Endmember SPICE results on AVIRIS Cuprite scene 108

4-9 The AVIRIS Indian Pines data set and ground truth 109

4-10 Abundance maps on AVIRIS Indian Pines data generated by SPICE 110

4-11 Abundance maps on labeled AVIRIS Indian Pines data generated by SPICE . . 111

4-12 Histograms of SPICE endmember results on labeled AVIRIS Indian Pines data . 112
4-13 Subset of AHI hyperspectral image 113
4-14 Vegetation masks for AHI hyperspectral images generated using SPICE 114
4-15 Selected AVIRIS Cuprite endmembers 115
4-16 The BSPICE AVIRIS Cuprite simulated data endmember results 115
4-17 Histograms of the number of endmembers and bands found using B-SPICE . . . 116
4-18 Results on two-dimensional triangle data found using ED 116
4-19 Data points generated from three endmember distributions 117
4-20 Results on two dimensional data using ED 117
4-21 Results on two dimensional data generated from endmember distributions using SPICE 118
4-22 Results on simulated AVIRIS Cuprite data found using ED 118
4-23 Results on subset of AVIRIS Cuprite data found using ED 119
4-24 Results on subset of AVIRIS Cuprite data found using ED 120
4-25 Two-dimensional data generated from three sets of endmembers 120
4-26 Two-dimensional data results found using PCE 121
4-27 Abundance maps found using PCE on labeled PCA-reduced AVIRIS Indian Pines data 122
4-28 Histograms of PCE endmember results on labeled PCA-reduced AVIRIS Indian Pines data 123
4-29 Abundance maps found using SPICE on labeled PCA-reduced AVIRIS Indian Pines data 124
4-30 Histograms of SPICE endmember results on labeled PCA-reduced AVIRIS Indian Pines data 125
4-31 Abundance maps found using PCE on labeled AVIRIS Indian Pines data with hierarchical dimensionality reduction 126
4-32 Histograms of PCE endmember results on labeled AVIRIS Indian Pines data with hierarchical dimensionality reduction 127
4-33 Abundance maps found using PCE on labeled AVIRIS Indian Pines data 128
4-34 Histograms of PCE endmember results on labeled AVIRIS Indian Pines data . . 129

LIST OF SYMBOLS AND ABBREVIATIONS

AMEE	Automated morphological endmember extraction algorithm
AVIRIS	Airborne visible/infrared imaging spectrometer
BSPICE	Band selecting sparsity promoting iterated constrained endmember detection algorithm
CCA	Convex cone analysis
D	Dimensionality of the input hyperspectral data set
DECA	Dependent components analysis
DPEM	Dirichlet process endmember detection algorithm
DWT	Discrete wavelet transform
$\mathbf{E}$	Matrix of endmembers where the i^{th} column contains the spectral signature for the i^{th} endmember, $\mathbf{e}_i$
ED	Endmember distribution detection algorithm
EMD	Euclidean minimum distance
FCM	Fuzzy c-means clustering algorithm
HFC	Harsanyi-Farrand-Chang method
ICA	Independent components analysis
ICE	Iterated constrained endmembers algorithm
iid	Independently and identically distributed
M	True or estimated number of endmembers for a given hyperspectral scene
$\mathbf{m}$	Vector containing the mean of each band of a given hyperspectral image
MCMC	Markov chain monte carlo
MEI	Morphological eccentricity index
MNF	Maximum noise fraction transform
MVC-NMF	Minimum volume constrained non-negative matrix factorization algorithm
MVT	Minimum volume transform
NATGD	Noise-adjusted transformed Gerschogorin disk
NMF	Non-negative matrix factorization

NSP	Noise subspace projection method
NWHFC	Noise whitened Harsanyi-Farrand-Chang method
P	Matrix of proportion (abundance) values where each element, p_{ik}, is the abundances of endmember k in pixel i
PCA	Principal components transform
PCE	Piece-wise convex endmember detection algorithm
PNACP	Partitioned noise-adjusted principal components algorithm
PPI	Pixel purity index
RVM	Relevance vector machine
SAM	Spectral angle mapper
SIE	Single individual evolutionary strategy
SPICE	Sparsity promoting iterated constrained endmembers algorithm
SSEE	Spatial-spectral endmember extraction algorithm
SVD	Singular value decomposition
TGD	Transformed Gerschogorin disk
VD	Virtual dimensionality
X	Input hyperspectral data set

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

HYPERSPECTRAL ENDMEMBER DETECTION AND BAND SELECTION USING BAYESIAN METHODS

By

Alina Zare

December 2008

Chair: Paul Gader
Major: Computer Engineering

Four methods of endmember detection and spectral unmixing are described. The methods determine endmembers and perform spectral unmixing while simultaneously determining the number of endmembers, representing endmembers as distributions, partitioning the input data set into several convex regions, or performing hyperspectral band selection. Few endmember detection algorithms estimate the number of endmembers in addition to determining their spectral shape. Also, methods which treat endmembers as distributions or treat hyperspectral images as piece-wise convex data sets have not been previously developed.

A hyperspectral image is a three-dimensional data cube containing radiance values collected over an area (or scene) in a range of wavelengths. Endmember detection and spectral unmixing attempt to decompose a hyperspectral image into the pure - separate and individual - spectral signatures of the materials in a scene, and the proportions of each material at every pixel location. Each spectral pixel in the image can then be approximated by a convex combination of proportions and endmember spectra.

The first method, the Sparsity Promoting Iterated Constrained Endmembers (SPICE) algorithm, incorporates sparsity-promoting priors to estimate the number of endmembers. The algorithm is initialized with a large number of endmembers. The sparsity promotion process drives all proportions of some endmembers to zero. These endmembers can be removed by SPICE with no effect on the error incurred by representing the image

with endmembers. The second method, the Endmember Distribution detection (ED) algorithm, models each endmember as a distribution rather than a single spectrum incorporating an endmember's inherent spectral variation or the variation due to differing environmental conditions. The third method, the Piece-wise Convex Endmember (PCE) detection algorithm, partitions the input hyperspectral data set into convex regions while simultaneously estimating endmember distributions for each partition and proportion values for each pixel in the image. The number of convex regions are determined autonomously using the Dirichlet process. The fourth method is known as the Band Selecting Sparsity Promoting Iterated Constrained Endmember (B-SPICE) algorithm and is an extension of SPICE that performs hyperspectral band selection in addition to all of SPICE's endmember detection and spectral unmixing features. This method applies sparsity promoting priors to discard those hyperspectral bands which do not aid in distinguishing between endmembers in a data set. All of the presented algorithms are effective at handling highly-mixed hyperspectral images where all of the pixels in the scene contain mixtures of multiple endmembers. These methods are capable of extracting endmember spectra from a scene that does not contain pure pixels composed of only a single endmember's material. Furthermore, the methods conform to the Convex Geometry Model for hyperspectral imagery. This model requires that the proportions associated with an image pixel be non-negative and sum to one.

Results indicate that SPICE and B-SPICE consistently produce the correct number of endmembers and the correct spectral shape for each endmember. The B-SPICE algorithm is shown to significantly decrease the number of hyperspectral bands while maintaining competitive classification accuracy for a data set. The ED algorithm results indicate that the algorithm produces accurate endmembers and can incorporate spectral variation into the endmember representation. The PCE algorithm results on hyperspectral data indicate that PCE produces endmember distributions which represent the true ground truth classes of the input data set.

CHAPTER 1
INTRODUCTION

1.1 Hyperspectral Image Data and Analysis

Hyperspectral imaging sensors capture both the spatial and spectral information of a scene. A hyperspectral sensor collects radiance data in hundreds of contiguous wavelengths. As a sensor collects data over a region, a three-dimensional data cube is generated. The data cube can be interpreted as a stack of two-dimensional images captured over a range of wavelengths. Each element of the three-dimensional data cube corresponds to the radiance measured in a particular wavelength at one ground location (Keshava and Mustard, 2002; Manolakis, Marden, and Shaw, 2003).

Radiance measured by a hyperspectral sensor is a combination of radiation that is reflected and/or emitted by materials on the ground (Manolakis et al., 2003). In passive systems, the reflected portion of the signal is the amount of radiation reflected from sunlight shining on ground materials (Keshava and Mustard, 2002). The atmosphere between the sensor and materials on the ground affects the radiance measurements. Water vapor and oxygen in the atmosphere cause the largest effect. In certain wavelengths, those known as *absorption bands*, water vapor and oxygen absorb a large portion of the signal, causing poor signal-to-noise ratios (Manolakis et al., 2003). In addition to absorption characteristics, the different wavelengths across which radiance can be measured have varying properties. For example, in the 0.4 to 2.5 μm range, sunlight or another active illumination source is needed since reflected radiance dominates this portion of the spectrum. In contrast, the thermal infrared region from 8 to 14 μm is dominated by emitted radiance and can, therefore, be measured during the night without an active illumination source (Manolakis et al., 2003).

The main appeal for hyperspectral imaging is the concept that different materials reflect and emit varying amounts of radiance across the electromagnetic spectrum. In other words, different materials generally have unique spectral signatures. It is for this

reason that hyperspectral sensors can be used to identify and distinguish between different materials in a scene (Manolakis et al., 2003).

Two important characteristics of a hyperspectral sensor are its spectral and spatial resolution. Spectral resolution of a sensor corresponds to the range of wavelengths over which radiance values are measured and combined to become a single band in a hyperspectral image. Spatial resolution corresponds to the size of the physical area on the ground from which radiance measurements are taken for a single image pixel. As the area corresponding to a pixel increases, the spatial resolution of the image decreases (Keshava and Mustard, 2002; Manolakis et al., 2003). For airborne systems, spatial resolution is generally constant across an image. However, for many forward-looking ground-based systems, the spatial resolution may vary within an image. The varying spatial resolution is a result of the angle from which a hyperspectral sensor images a region. Pixels closer to the sensor have higher spatial resolution than those farther away.

Spatial resolution is one of the causes of *mixed pixels* in a hyperspectral data set (Keshava and Mustard, 2002; Manolakis et al., 2003). A mixed pixel is a pixel which combines the radiance values of multiple materials. A *pure pixel* corresponds to a single material's radiance values. Mixed pixels can occur from low spatial resolution since, as a pixel's corresponding area on the ground increases, neighboring materials are likely to be combined into the image pixel. Mixed pixels also occur when the different materials are mixed on the ground. Beach sand is a common example for this type of mixed pixel since grains of different materials are intermingled (Keshava and Mustard, 2002).

1.1.1 Endmember Detection

Pure spectral signatures, or the *constituent spectra*, in an imaged scene are referred to as *endmembers* (Keshava and Mustard, 2002). Due to the presence of mixed pixels in a hyperspectral image, *spectral unmixing* is often performed to decompose mixed pixels into their respective endmembers and *abundances*. Abundances are the proportions of

the endmembers in each pixel in a hyperspectral image. Spectral unmixing relies on the definition of a mixing model.

Complex mixing models for hyperspectral imagery can be defined. These complex models can take into account the atmospheric effects and the orientation, size, and shape of objects in a scene. They can also consider the incident angles of sunlight and the sensor on ground materials. Despite the large number of variables that can be included into a mixing model, the most popular model is the *convex geometry model* (also known as the *linear mixing model*) (Keshava and Mustard, 2002; Nascimento and Bioucas-Dias, 2005a).

The convex geometry model assumes that every pixel is a convex combination of endmembers in the scene. This model can be written as shown in Equation 1–1 (Keshava and Mustard, 2002; Manolakis et al., 2003; Nascimento and Bioucas-Dias, 2005a),

$$\mathbf{x}_i = \sum_{k=1}^{M} p_{ik}\mathbf{e}_k + \epsilon_i \quad i = 1, \ldots, N \tag{1–1}$$

where N is the number of pixels in the image, M is the number of endmembers, ϵ_i is an error term, p_{ik} is the proportion of endmember k in pixel i, and $\mathbf{e}_k$ is the k^{th} endmember. The proportions of this model satisfy the constraints in Equation 1–2,

$$p_{ik} \geq 0 \quad \forall k = 1, \ldots, M; \quad \sum_{k=1}^{M} p_{ik} = 1. \tag{1–2}$$

The convex geometry model has been found to effectively describe regions where the various pure materials are separated into regions dominated by a single endmember. Generally, mixed pixels in these types of regions are caused by a sensor's inadequate spatial resolution. In cases where materials are mixed on the ground, nonlinear mixing models have been found to be more effective (Keshava and Mustard, 2002).

The endmember detection problem is the task of determining the pure spectral signatures in a given hyperspectral scene. Endmember detection algorithms often assume the convex geometry model and perform spectral unmixing to return the endmembers and abundances in an image (Keshava and Mustard, 2002).

1.1.2 Hyperspectral Band Selection

Since, for each pixel, hyperspectral sensors measure radiance values at a very large number of wavelengths, this imagery contains an immense amount of data (Manolakis et al., 2003). Although the resolution provided allows for the extraction of material spectra, the volume of data poses many challenging problems such as data storage, computational efficiency, and the *curse of dimensionality* (Chang, Du, Sun, and Althouse, 1999; Huang and He, 2005). One method to overcome these challenges is the use of data reduction techniques (Huang and He, 2005). Hyperspectral band selection is one method of data reduction that also retains the physical meaning of the data set (Guo, Gunn, Damper, and Nelson, 2006). Hyperspectral band selection selects a set of bands from the input hyperspectral data set which retain the information needed for subsequent hyperspectral image spectroscopy.

1.2 Statement of Problem

Most endmember detection algorithms require the knowledge of the number of endmembers for a given hyperspectral scene. Also, many existing algorithms rely on the *pixel purity assumption* which assumes that pure pixels for each endmember exist in the data set. Some existing endmember detection algorithms do not unmix the data set and do not provide abundance values that conform to the non-negativity and sum-to-one constraints in Equation 1–2. Existing endmember detection algorithms represent endmembers as single spectral points, which does not incorporate the spectral variability that occurs due to differing environmental conditions. Furthermore, existing endmember detection algorithms generally assume that the hyperspectral data points lie in a single convex region with one set of endmembers. However, it may be the case that multiple sets of endmembers, defining several overlapping convex regions, can better describe the hyperspectral image.

Existing hyperspectral band selection algorithms often require an input of the number of hyperspectral bands to be retained. Furthermore, many hyperspectral data

reduction techniques perform a projection or merging of the bands which removes each band's physical meaning. Many hyperspectral band selection algorithms are also tied to a classification problem which requires labeled training data to determine the bands which distinguish between the identified classes.

This study examines methods to tackle endmember detection and hyperspectral band selection. Algorithms that autonomously estimate the number of endmembers and hyperspectral bands while simultaneously estimating endmember spectral shapes, and which yield abundances which conform to the convex geometry model's constraints, retain physically meaningful bands, and avoid reliance on the pixel purity assumption are investigated. Methods which determine endmember distributions and autonomously learn the number of convex regions needed to describe an input hyperspectral scene are presented.

1.3 Overview of Research

The conducted research involves the development and analysis of three novel endmember detection algorithms. These methods either determine the number of endmembers required for a scene, learn endmember distributions, or determine the number of convex regions needed to describe a hyperspectral image while simultaneously estimating the endmembers' spectral distributions. Furthermore, all of the presented methods also simultaneously determine appropriate abundance values for every pixel and do not rely on the pixel purity assumption. Additionally, a novel hyperspectral band selection algorithm is developed that: determines the needed number of hyperspectral bands to distinguish between endmembers for a given scene, performs unsupervised band selection, and retains the physical meaning of the hyperspectral bands.

The endmember detection algorithms determine the spectral shape of each endmember and the proportion of each endmember in each pixel. These algorithms are based on the convex geometry model in Equation 1–1 and thus constrain the proportion values to be non-negative and sum to one. The general approach involves integrating

state-of-the-art machine learning approaches based on Bayesian methods into the framework of hyperspectral image spectroscopy. The Sparsity Promoting Iterated Constrained Endmembers (SPICE) algorithm determines the number of endmembers by beginning with a large number of initial endmembers and removing endmembers as they become superfluous. The number of endmembers are determined by applying a sparsity promoting prior to the proportions for each endmember. The Endmember Distributions (ED) algorithm estimates the distribution of each endmember for an input data set rather than single spectra. The Piece-wise Convex Endmember (PCE) detection algorithm uses the Dirichlet process to determine the number of convex regions needed to describe an input hyperspectral image while simultaneously performing spectral unmixing and determining endmember distributions for each convex region. The sparsity promoting, endmember distributions, and Dirichlet process techniques utilize Bayesian machine learning approaches to estimate the number of endmembers, learn endmember distributions, or partition the data set into convex regions while estimating proportion values and values for the endmembers themselves.

In addition to the endmember detection and spectral unmixing algorithms, a simultaneous band selection and endmember detection algorithm is developed. This method, the Band Selecting Sparsity Promoting Iterated Constrained Endmember (B-SPICE) algorithm, determines the number of required bands for a data set needed to distinguish between the endmembers in a scene. This is in contrast to previous approaches which incorporate separate metrics into the objective function to find bands that aid in discriminating between labeled regions in a scene. In addition to performing band selection and determining the number of bands needed, this algorithm performs endmember determination and spectral unmixing. The unnecessary bands and endmembers in this method are removed using sparsity promoting priors.

CHAPTER 2
LITERATURE REVIEW

This chapter provides a review of existing hyperspectral endmember detection algorithms followed by a summary of existing hyperspectral band selection algorithms.

2.1 Existing Endmember Detection Algorithms

Many endmember detection algorithms are described in the literature. The majority of these algorithms rely on the convex geometry model described in Equation 1–1 (Keshava and Mustard, 2002). Most existing algorithms require advance knowledge of the number of endmembers in a given scene. However, this value is often unknown for a given data set. Several methods make the *pixel purity assumption* and assume that pure pixels exist in the input data set for every endmember in the scene. This assumption causes algorithms to be inaccurate for highly-mixed data sets where pure pixels for each material cannot be found in the imagery. Additionally, some methods do not encompass all of the data points and, therefore, either prevent spectral unmixing with abundance values that conform to the constraints in Equation 1–2 or have large reconstruction errors using the estimated endmember and abundance matrices. The existing methods generally represent each endmember as a single spectrum which does not account for the spectral variation that may occur due to varying environmental conditions. The majority of these methods also assume that the hyperspectral data points lie in a single convex region and can be described by a single set of endmembers which encompass the data set. In this chapter, a summary of many of these existing endmember detection algorithms is provided.

2.1.1 Pixel Purity

Many endmember detection algorithms rely on the assumption that the spectral signature for each endmember can be found without performing spectral unmixing on the data set. This assumes that there exists at least one pixel for each endmember which consists of only that endmember's material. Furthermore, the hyperspectral imaging

device must be operating at a spatial resolution that does not combine endmember spectra with the spectra of neighboring materials. Algorithms relying on the pixel purity assumption include the NFindr algorithm (Winter, 1999) and the Pixel Purity Index algorithm (Boardman, Kruse, and Green, 1995) both of which are described in detail below.

Additionally, the Automated Morphological Endmember Extraction (AMEE) algorithm defines multispectral dilation and erosion operators used to compute the *morphological eccentricity index* (MEI) (Plaza, Martinez, Perez, and Plazas, 2002). The MEI is used to identify spectrally pure pixels in the image which are returned as endmembers (Plaza et al., 2002). The Spatial-Spectral Endmember Extraction (SSEE) algorithm projects the image onto eigenvectors computed from the Singular Value Decomposition (SVD) of subsets in the input data set (Rogge, Rivard, Zhang, Sanchez, Harris, and Feng, 2007). SSEE identifies candidate endmembers as those that fall on the extreme ends of the projection and returns either the pixels or the mean of pixels that are spatially close and spectrally similar. The method based on Morphological Associative Memories described by Grana, Sussner, and Ritter (2003) also depends on the pixel purity assumption for endmember extraction as described in Section 2.1.4. Vertex Component Analysis adds endmembers sequentially by selecting pixels which project farthest in a direction orthonormal to the space spanned by the current endmember set (Nascimento and Bioucas-Dias, 2005b). Thus, Vertex Component Analysis also relies on the pixel purity assumption (Nascimento and Bioucas-Dias, 2005b).

NFindr. The NFindr algorithm is a well-known, established method of endmember detection that searches for endmembers within an input hyperspectral data set (Winter, 1999). NFindr seeks the set of input pixels that encompass the largest volume (Winter, 1999).

The algorithm begins by randomly selecting a set of pixels from the image to be the initial endmember set. Then, each endmember is replaced, in succession, by all other

pixels in the image. After each replacement, the volume of the space defined by the current set of potential endmembers is computed. When a replacement increases the volume, the replacement is maintained. The algorithm cycles through image pixels and endmembers until no further replacements are made (Winter, 1999).

The volume enclosed by each set of potential endmembers is computed using Equation 2–1,

$$V\left(\mathbf{E}^*\right) = \frac{1}{(M-1)!} abs\left(|\mathbf{E}^*|\right) \tag{2–1}$$

where

$$\mathbf{E}^* = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \mathbf{e}_1 & \mathbf{e}_2 & \dots & \mathbf{e}_M \end{pmatrix} \tag{2–2}$$

M is the number of endmembers and $\mathbf{e}_i$ is a column vector containing an endmember. If $(M-1)$ is not the dimensionality of the data, then a dimensionality reduction method, such as Principal Components Analysis or Maximum Noise Fraction, must be employed (Green, Berman, Switzer, and Craig, 1988; Lee, Woodyatt, and Berman, 1990). The data dimensionality must be one less than the desired number of endmembers since the determinant of a non-square matrix is not defined (Winter, 1999).

This algorithm works by maximizing the volume by the endmembers inscribed within the hyperspectral data cloud. Since the endmembers are found within the data cloud, the endmembers may not enclose all the data points. In addition to assuming pure pixels can be found in the image, this algorithm requires knowledge of the number of endmembers in advance (Winter, 1999).

Pixel purity index. The Pixel Purity Index, PPI, is a commonly used algorithm for determining the purest pixels in an input image (Boardman et al., 1995). The PPI algorithm ranks image pixels based on their *pixel purity indices*. Then, the M pixels with the highest pixel purity values are returned as potential endmembers. The number of endmembers, M, is not determined by this algorithm. PPI is often used for generating candidate endmembers which are then used as inputs to other endmember extraction algorithms (Berman, Kiiveri, Lagerstrom, Ernst, Donne, and Huntington, 2004) or loaded

into a visualization tool for users to hand select endmembers from the candidates (Rogge et al., 2007).

The PPI algorithm assigns each pixel a pixel purity value by repeatedly projecting all of the pixels onto randomly directed vectors. The algorithm is initialized by assigning all pixels a pixel purity value of zero. The pixel purity values are updated following each random projection by adding one to the values of the pixels that fall near either end of every projection. Since PPI values are generated using random vectors, the results are dependent on the number of random projections and the threshold for determining if a pixel's projection is considered near an end-point (Boardman et al., 1995).

2.1.2 Convex Hull

The convex geometry model defines endmembers to be the vertices of a simplex that surround the pixels in an image. The Minimum Volume Transform and Convex Cone Analysis methods are based on this model and search for points that lie at the corners of a simplex surrounding the data.

The Minimum Volume Transform (MVT) (Craig, 1994) finds the smallest simplex that circumscribes the hyperspectral data points. This is in contrast to the NFindr method which obtains the largest simplex inscribed within the input data set (Winter, 1999). MVT searches for hyperplanes that minimize their enclosed volume while encompassing all of the data. The algorithm then iteratively varies the hyperplanes using linear programming methods to provide a progressively tighter fit around the data. After minimizing the volume enclosed by the hyperplanes while encompassing all of the input data, the intersections of the planes are returned as endmembers. Although MVT does not require pure pixels to be in the data set, it does require the number of endmembers in advance. The method performs the Maximum Noise Fraction transform (Green et al., 1988) to reduce the data dimensionality to $(M - 1)$ where M is the number of endmembers.

Convex cone analysis. The convex cone analysis (CCA) method (Ifarraguerri and Chang, 1999) of endmember extraction also searches for the boundaries of a convex, non-negative region that enclose the input data points. This method relies on the fact that radiance values are non-negative and, therefore, can restrict the endmembers to be non-negative points. The method requires an input of the number of desired endmembers, M. Given M, the eigenvectors of the sample correlation matrix that correspond to the M largest eigenvalues are computed (Ifarraguerri and Chang, 1999).

$$\mathbf{X}_N^T \mathbf{X}_N = \mathbf{C} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T \tag{2–3}$$

where $\mathbf{X}_N$ is the normalized input data matrix, $\mathbf{C}$ is the sample correlation matrix, $\mathbf{\Lambda}$ is a diagonal matrix of the sample correlation matrix's sorted eigenvalues, and $\mathbf{U}$ is the matrix of sorted eigenvectors. The data is normalized such that every data point has a constant L2-norm. This normalization retains the spectral shape of each data point while constraining each point to the same hyper-sphere. Let $\{\mathbf{u}_i\}_{i=1}^{M}$ be the set of the M most significant eigenvectors. Given the eigenvectors, the endmembers in this method are defined to be linear combinations of the eigenvectors that satisfy Equation 2–4,

$$\mathbf{e} = \mathbf{u}_1 + a_1\mathbf{u}_2 + \cdots + a_{M-1}\mathbf{u}_M \geq \mathbf{0} \tag{2–4}$$

where $\mathbf{0}$ is the zero vector (Ifarraguerri and Chang, 1999). The first eigenvector of the sample correlation matrix, $\mathbf{u}_1$, will point towards the data set. Equation 2–4 can be interpreted as perturbing the first eigenvector by a linear combination of the other orthogonal eigenvectors while constraining the endmembers to be non-negative.

Since each eigenvector is of dimension D, solving for the $(M-1)$ a_j coefficients in Equation 2–4 is an over-determined problem. Because of this, the CCA method iterates through each subset of bands of size $(M-1)$ and solves a set of $(M-1)$ linear equations for the a_j coefficients,

$$\mathbf{u}_1(\gamma^i) + a_1\mathbf{u}_2(\gamma^i) + \cdots + a_{M-1}\mathbf{u}_M(\gamma^i) = 0 \tag{2–5}$$

where γ^i represents the indices of the i^{th} set of $(M-1)$ hyperspectral bands. After solving for the a_j coefficients, they are plugged back into Equation 2–4. The vectors computed by plugging back into Equation 2–4 will contain $(M-1)$ zero values. The remaining band values are checked to ensure that they are non-negative. If a potential endmember is found to be non-negative, then it is kept as an endmember, otherwise, that vector is discarded (Ifarraguerri and Chang, 1999).

The CCA method searches through $\binom{D}{M-1}$ potential endmembers which can be prohibitive for data sets with a large number of hyperspectral bands. Furthermore, since $\binom{D}{M-1}$ may be greater than M, more endmembers than specified may be found. Ifarraguerri and Chang (1999) list potential methods for removing the additional endmembers such as removing endmembers that are collinear with other endmember spectra.

This algorithm does not provide endmembers which tightly surround the data points. This is an artifact from the $(M-1)$ zeros in each endmember spectra as can be seen in Figure 2-1. Figure 2-1 shows the three-dimensional data set and the three endmembers found using CCA. Since each endmember has two zeros in their spectra, the endmember spectra lie along the x-, y- and z-axis rather than tightly surrounding the data set. This is further illustrated in Figures 2-2 and 2-3. The normalized data points in Figure 2-2 are the first twenty-five bands (approximately 1978 to 2228nm) from a subset of pixels in the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Cuprite "Scene 4" data set (AVIRIS). Figure 2-3 shows the nine endmembers determined using CCA with M set to three. The normalized data set values range from 0.138 to 0.244 while the range of endmember values is 0 to 0.904.

2.1.3 Nonnegative Matrix Factorization

Several methods for endmember extraction have been developed based on Non-Negative Matrix Factorization (NMF). Non-Negative Matrix Factorization searches for two non-negative matrices, $\mathbf{E} \in \mathbf{R}^{D\times M}$ and $\mathbf{P} \in \mathbf{R}^{M\times N}$ that approximate an input

non-negative matrix $\mathbf{X} \in \mathbf{R}^{D \times N}$ (Lee and Seung, 1999; Miao and Qi, 2007),

$$\mathbf{X} \approx \mathbf{EP}. \tag{2–6}$$

The non-negative assumptions in this method are appropriate for endmember detection since hyperspectral radiance data is nonnegative (Miao and Qi, 2007; Pauca, Piper, and Plemmons, 2005). One NMF algorithm proposed by Lee and Seung (2000) minimizes the objective function in Equation 2–7 (Pauca et al., 2005),

$$f(\mathbf{E}, \mathbf{P}) = \frac{1}{2}\|\mathbf{X} - \mathbf{EP}\|_F^2 = \sum_{i=1}^{D}\sum_{j=1}^{N}(\mathbf{X}_{ij} - (\mathbf{EP})_{ij})^2. \tag{2–7}$$

The NMF update developed by Lee and Seung (2000) uses the *multiplicative update rules* in Equations 2–8 and 2–9,

$$p_{ij}^{k+1} = p_{ij}^{k}\frac{(\mathbf{E}^T\mathbf{X})_{ij}^{k}}{(\mathbf{E}^T\mathbf{EP})_{ij}^{k}} \tag{2–8}$$

$$e_{ij}^{k+1} = e_{ij}^{k}\frac{(\mathbf{X}^T\mathbf{P})_{ij}^{k}}{(\mathbf{EPP}^T)_{ij}^{k}} \tag{2–9}$$

where k indicates the iteration. The elements of the $\mathbf{P}$ and $\mathbf{E}$ matrices are updated simultaneously by iterating between the elements of the two matrices (Pauca et al., 2005). Lee and Seung (2000) prove that the distance in Equation 2–7 does not increase when using the updates in Equations 2–8 and 2–9. As shown by Lee and Seung (2000), the multiplicative update rules are equivalent to standard gradient descent updates when the step size parameter is set to $\frac{p_{ij}^{k}}{(\mathbf{E}^T\mathbf{EP})_{ij}^{k}}$ when updating proportion values and $\frac{e_{ij}^{k}}{(\mathbf{EPP}^T)_{ij}^{k}}$ when updating endmember values.

The basic NMF algorithm has been modified to include constraints and initialization methods for better performance in endmember detection. Three endmember extraction algorithms based on NMF are described.

Minimum volume constrained nonnegative matrix factorization. The Minimum Volume Constrained Nonnegative Matrix Factorization (MVC-NMF) algorithm

for endmember detection (Miao and Qi, 2007) attempts to minimize the objective function in Equation 2–10 solving for endmembers and the abundance values for each pixel (Miao and Qi, 2007).

$$\min f(\mathbf{E}, \mathbf{P}) = \frac{1}{2} \sum_{i=1}^{D} \sum_{j=1}^{N} (\mathbf{X}_i j - \mathbf{E}\mathbf{P}_{ij})^2 + \lambda \frac{1}{2(M-1)!} \left| \begin{pmatrix} \mathbf{1}_M^T \\ \tilde{\boldsymbol{E}} \end{pmatrix} \right|^2 \tag{2–10}$$

where

$$\tilde{\boldsymbol{E}} = \mathbf{U}^T \left(\mathbf{E} - \mu \mathbf{1}_M^T \right) \tag{2–11}$$

The $\mathbf{U}$ matrix consists of the $(M-1)$ most significant principal components of the input data, $\mathbf{X}$, and $\mathbf{1}_M^T$ is an M length vector of ones. The size of the endmember matrix is reduced using Equation 2–11 in order to be able to compute the determinant in Equation 2–10 (Miao and Qi, 2007).

The first term of the objective function is a squared error term. By minimizing the first term, the error between the input data set and the estimated pixels computed from the abundance values and endmembers are minimized. The second term of the objective function is the volume of the space defined by the endmembers. By minimizing the second term, the endmembers provide a tight fit around the data. These two terms can be seen as an "internal force" and an "external force" (Miao and Qi, 2007). The first term can be interpreted as an outward force that prefers endmembers which completely encompass the data and the second term is an inward force that wants to minimize the volume enclosed by the endmembers (Miao and Qi, 2007).

In MVC-NMF, the objective function in Equation 2–10 is minimized using gradient descent with clipping. The values for the endmembers and their proportions are updated in an alternating fashion. In other words, in each iteration of the algorithm, either the endmembers or the proportions are updated while the other is held constant (Miao and Qi, 2007).

In order to enforce the non-negativity constraint in Equation 1–2, after solving for either the endmember or the proportions, any negative values are set to zero.

$$\mathbf{E}^{k+1} = \max\left(\mathbf{0}, \mathbf{E}^k - \alpha^k \nabla_{\mathbf{E}} f\left(\mathbf{E}^k, \mathbf{P}^k\right)\right) \tag{2–12}$$

$$\mathbf{P}^{k+1} = \max\left(\mathbf{0}, \mathbf{P}^k - \beta^k \nabla_{\mathbf{P}} f\left(\mathbf{E}^k, \mathbf{P}^k\right)\right) \tag{2–13}$$

where α and β are the gradient descent learning rates (Miao and Qi, 2007).

To promote the sum-to-one constraint of the proportions, when updating proportion values, the endmember and data matrices are augmented by a row of constant positive values. The larger the constant, the more emphasis is placed on the sum-to-one constraint (Miao and Qi, 2007).

The algorithm seeks endmembers that minimize the squared reconstruction error. The algorithm also allows for some resilience to noise and selects endmembers that provide a tight fit around the data. Still, the MVC-NMF algorithm does have some drawbacks. The algorithm requires knowledge of the number of endmembers in advance and does not strictly enforce the sum-to-one constraint.

Constrained non-negative matrix factorization. Pauca, Piper, and Plemmons (2005) also develop a method of endmember extraction based on the NMF algorithm. Their constrained NMF algorithm incorporates smoothness constraints into the NMF objection function described in Equation 2–7. The resulting objective function is shown in Equation 2–14,

$$\min_{\mathbf{E},\mathbf{P}} \{\|\mathbf{X} - \mathbf{EP}\|_F^2 + a\|\mathbf{E}\|_F^2 + \beta\|\mathbf{P}\|_F^2\} \tag{2–14}$$

where a and b are regularization parameters balancing the error and smoothness terms (Pauca et al., 2005). The smoothness terms encourage sparsity within the matrices and are equivalent to applying a Gaussian prior on the endmembers and abundances. This objective is minimized using gradient descent. Following Lee and Seung (2000), the step size parameters are set to $\frac{p_{ij}^k}{(\mathbf{E}^T\mathbf{EP})_{ij}^k}$ when updating proportion values and $\frac{e_{ij}^k}{(\mathbf{EPP}^T)_{ij}^k}$ when updating endmember values (Pauca et al., 2005).

Pauca, Piper, and Plemmons (2005), after computing endmembers using the Constrained NMF method, retain endmembers based on their similarity to laboratory measured spectra. Endmembers are compared to laboratory spectra using the *symmetric Kullback-Leibler Divergence*,

$$K_s(\mathbf{e}_1^N, \mathbf{e}_2^N) = K(\mathbf{e}_1^N || \mathbf{e}_2^N) + K(\mathbf{e}_2^N || \mathbf{e}_1^N) \tag{2–15}$$

where

$$K(\mathbf{x}||\mathbf{y}) = \sum_{i=1}^{D} \mathbf{x}(i) \log\left(\frac{\mathbf{x}(i)}{\mathbf{y}(i)}\right) \tag{2–16}$$

and $\mathbf{e}^N$ are endmembers normalized to sum to one. The symmetric Kullback-Leibler divergence is computed between each endmember and every library spectra. An endmember is associated to the library spectrum for which the endmember has the smallest divergence value given that the divergence is below a threshold, τ. If all of the symmetric Kullback-Leibler divergences computed for an endmember are greater than τ, the endmember is pruned (Pauca et al., 2005). Like the previous NMF-based algorithms, this method required an estimate of the number of endmembers prior to running the NMF algorithm. However, this number may change based on the final pruning step which requires access to a spectral library containing signatures that can be found in the image.

Fuzzy c-means initialized non-negative matrix factorization. Liou and Yang (2005) also developed an endmember extraction method based on NMF. Their method relies on the basic NMF multiplicative update rule (Lee and Seung, 1999) but provides a method of initializing the two non-negative matrices. The **E** and **P** matrices are initialized using the Fuzzy C-Means (FCM) clustering method. The FCM clustering algorithm clusters the data into M clusters with each input point having varying degrees of membership in each cluster. The objective function for FCM is defined in Equation 2–17,

$$J = \sum_{k=1}^{M}\sum_{i=1}^{N} u_{ki}^m d_{ik} + \sum_{i=1}^{N} \lambda_i \left(\sum_{k=1}^{M} u_{ki} - 1\right) \tag{2–17}$$

where M is the number of clusters (in this case, the number of endmembers), N is the number of data points, m is the fuzzifier, u_{ki} is the membership of the i^{th} points to the k^{th} cluster, d_{ik} is the squared distance between the i^{th} data point and the k^{th} cluster center and λ_i is the Lagrange multiplier (Liou and Yang, 2005; Theodoridis and Koutroumbas, 2003). FCM minimizes the objective function by updating membership values and cluster centers.

Since FCM provides both cluster centers and membership values for each data point, the matrices for the NMF algorithm are initialized using the cluster centers as the initial endmembers and the membership values as the initial abundance values. Since NMF is dependent on initialization, well chosen initial matrices can improve performance (Liou and Yang, 2005).

This method also requires advance knowledge of the number of endmembers to perform both the FCM and the NMF algorithms. Liou and Yang (2005) utilize the Partitioned Noise-Adjusted Principal Component Analysis method (Tu, Huang, and Chen, 2001) to try to estimate the number of endmembers prior to applying the endmember detection algorithm. This method of estimating the number of endmembers is described in Section 2.1.7.

2.1.4 Morphological Associative Memories

Many methods for endmember detection have been based on the use of Morphological Associative Memories (Ritter and Gader, 2006). There are two types of morphological memories that can be computed, the *min memory* and the *max memory*. Given a set of input vectors, $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ and associated desired outputs, $Y = \{\mathbf{y}_1, \ldots, \mathbf{y}_N\}$. The min and max morphological associative memories, $\mathbf{W}_{XY}$ and $\mathbf{M}_{XY}$, are computed using

Equations 2–18 and 2–19,

$$\mathbf{W}_{XY} = \bigwedge_{i=1}^{N} [\mathbf{y}_i + (\mathbf{x}_i)^*] \tag{2–18}$$

$$\mathbf{M}_{XY} = \bigvee_{i=1}^{N} [\mathbf{y}_i + (\mathbf{x}_i)^*] \tag{2–19}$$

where $\mathbf{x}^*$ is the lattice conjugate transpose of $\mathbf{x}$ which is defined to be $\mathbf{x}^* = (-\mathbf{x})^T$. Auto-associative morphological memories are the morphological associative memories which associate a set X to itself. The min and max auto-associative memories are related to each other using the conjugate transpose operator, $\mathbf{W}_{XX} = \mathbf{M}_{XX}^*$.

Patterns are recalled using associative morphological associative memories through either the *max product* or the *min product*, $\mathbf{y} = \mathbf{W}_{XY} \boxtimes \mathbf{x}$ and $\mathbf{y} = \mathbf{M}_{XY} \boxtimes \mathbf{x}$. The max product between an $m \times p$ matrix $\mathbf{A}$ and an $p \times n$ matrix $\mathbf{B}$ is defined in Equation 2–20,

$$c_{ij} = \bigvee_{k=1}^{N} a_{ik} + b_{kj}. \tag{2–20}$$

The min product is similarly defined using a minimum operator.

Morphological associative memories have been used for endmember detection because they can be used to find affinely independent points in D-dimensional space or points that are *morphologically independent* (Grana, Sussner, and Ritter, 2003; Myers, 2005). The convex hull of endmembers that follow the convex geometry model in Equation 1–1 defines a volume in D-dimensional space which surround the hyperspectral data points in an image. The convex hull of $D + 1$ affinely independent points defines a simplex in D-dimensional space (Myers, 2005; Ritter and Urcid, 2008). Therefore, the motivation to find affinely independent points using morphological associative memories is that the simplex defined by these points bounds a volume is D-dimensional space which can be used to try to surround the hyperspectral image points (Myers, 2005). Morphological associative memories can also be used to determine whether points are morphologically independent (Grana et al., 2003). Some endmember detection algorithms

using morphological associative memories search for morphologically independent endmembers which surround the points of the hyperspectral image (Grana et al., 2003).

Morphological associative memory method 1. Grana, Sussner, and Ritter (2003) developed a method using Morphological Associative Memories to extract endmembers. The image pixels are all shifted by their mean, $X' = \{\mathbf{x}'_i | \mathbf{x}'_i = \mathbf{x}_i - \mu\}$. Then, the algorithm begins by randomly selecting a single input pixel to be the initial endmember. Using this initial endmember's binary representation, min and max auto-associative memories are created. The binary representation of pixel $\mathbf{x}$ is defined to be $sgn(\mathbf{x})$ (Grana and Gallego, 2003; Grana et al., 2003).

After selecting the initial pixel, all other image pixels are sequentially considered to be endmembers. When being considered, a pixel is shifted and tested for *morphological independence* against all of the current endmembers' binary representations (Grana et al., 2003).

$$\mathbf{x}_i^+ = \mathbf{M}_{BB} \boxtimes (\mathbf{x}'_i - a\sigma) \tag{2–21}$$

$$\mathbf{x}_i^- = \mathbf{W}_{BB} \boxtimes (\mathbf{x}'_i + a\sigma) \tag{2–22}$$

where a is a constant value, $\mathbf{B}$ matrix of the current endmembers' binary representations, and σ is the vector of variances of each band of the input image. A pixel is determined to be morphologically independent if $\mathbf{x}_i^+ \notin B$ and $\mathbf{x}_i^- \notin B$. If a shifted pixel is found to be morphologically independent, it is added to the set of endmembers and new auto-associative memories are computed. If it is not morphologically independent, then the pixel is compared against existing endmembers to see if it is more extreme than the current endmember. If the pixel is more extreme, then it is replaces that endmember; otherwise, the pixel is discarded (Grana et al., 2003). Using D-length vectors, there are 2^D possible binary vectors. This algorithm will return one endmember for each set of shifted input data points with the same binary representations. Therefore, up to 2^D endmembers may be returned.

Although this algorithm does not require prior knowledge of the number of endmembers, the algorithm assumes that pure pixels exist in the input hyperspectral image and it does not compute abundances values. Like NFindr, this algorithm relies on the pixel purity assumption. This method differs from NFindr since the volume encompassed by the endmembers is not computed while performing endmember detection.

Morphological associative memory method 2. Grana, Hernandez, and d'Anjou (2005) developed an algorithm that combines an evolutionary search and endmember detection using Morphological Associative Memories. This algorithm uses an evolutionary search to find a set of morphologically independent endmembers that minimize the fitness function in Equation 2–25 (Grana et al., 2005). The algorithm proceeds by evolving a set of binary vectors using a mutation operator and *roulette wheel selection* based on the fitness function (Grana et al., 2005). Every mutation is tested for morphological independence. If the mutated set of binary vectors is not morphologically independent, it is rejected. Given a set of binary vectors, the corresponding endmembers are the extreme pixels in the direction identified by the binary vectors (Grana et al., 2005).

This algorithm provides both endmember spectra and abundance values. However, this algorithm requires prior knowledge of the desired number of endmembers and does not strictly enforce the non-negativity and sum-to-one constraints on the abundance values.

Morphological associative memory method 3. Ritter and Urcid (2008) developed a method of extracting endmembers that uses the columns of the min auto-associative memories in Equation 2–18. This method is similar to the one presented by Myers (2005) which also returns the *strong lattice independent* columns of the min or max auto-associative memories. The auto-associative memories in this method are created using the points of the input hyperspectral image.

After computing the auto-associative memories, any duplicate columns of the min memory are removed ensuring that the remaining column are linearly independent as

proven by Ritter and Urcid (2008). Linearly independent sets are created since linear independence implies affine independence. The members of the linearly independent sets are then shifted by the elements of the *bright point* (the component-wise maximum of all input points)

$$\mathbf{w}^j = \mathbf{W}_{XX}^j + \mathbf{u}^j \tag{2–23}$$

where $\mathbf{W}_{XX}^j$ is the j^{th} column of the $\mathbf{W}_{XX}$ memory and $\mathbf{u}^j = \bigvee_{i=1}^{N} \mathbf{x}_i^j$ is the maximum value in the j^{th} spectral band over all input data points (Ritter and Urcid, 2008). The elements along the diagonal of the memory are equal to zero. Therefore, after shifting, these values are set to the maximum value of the data points in the corresponding hyperspectral band. This provides physical meaning between the endmembers and the input data set (Ritter and Urcid, 2008)

After shifting, the unique linearly independent vectors, $\mathbf{W} = \{\mathbf{w}^1, \ldots, \mathbf{w}^2\}$, are returned as endmembers. Additionally, the *shade point* (the component-wise minimum of all input points) is returned as an endmember (Myers, 2005; Ritter and Urcid, 2008). Using the min-memory and the shade point provides up to $D + 1$ endmembers.

This method is very efficient; it requires only a single pass through all of the input pixels (Myers, 2005; Ritter and Urcid, 2008). However, this algorithm does not compute abundance values and it does not guarantee that all pixels will be encompassed by the selected endmembers. Figure 2-4 displays the endmembers determined using this method on two-dimensional data. The data set was generated from four endmembers ([10,30], [13, 24], [15, 31], [22, 25]). Also, Gaussian random noise was added to each coordinate of the data set. The min memory, prior to shifting, found endmembers (0,1.38) and (-21.42, 0). These were shifted by (the 1^{st} and 2^{nd} coordinates of the *bright point*, the component-wise max of the data) 22.64 and 32.97, respectively, to obtain (22.64, 24.02) and (11.55, 32.97). As shown in the figure, all of the pixels are not encompassed by the endmembers.

Using the bright point, the shade point and the unique columns of both the min- and max-memories as endmembers guarantees that all input data points will be encompassed

(Ritter and Urcid, 2008). This is shown in Figure 2-5. However, this does not provide a tight fit around the data points and it will return up to $2D + 2$ endmembers. With high dimensional data sets, the method would return a very large number of endmembers. Methods to reduce the number of endmembers when using either both memories or only the min-memory are discussed by Ritter and Urcid (2008). For example, every other column of the memory may be discarded since contiguous columns are often highly correlated. Another method presented is to compute linear correlation coefficients between each of the endmembers and retain a subset of endmembers whose correlation coefficients fall below a set threshold (Ritter and Urcid, 2008).

2.1.5 Evolutionary Search

In addition to the Morphological Associative Memory method which incorporates evolutionary search strategies to perform hyperspectral endmember detection, the single individual evolutionary strategy (SIE) for endmember detection also uses an evolutionary algorithm to determine an endmember set for a given hyperspectral image (Grana, Hernandez, and Gallego, 2004). The SIE algorithm begins by sampling M endmembers from a Gaussian distribution centered at the mean of the hyperspectral data set

$$\mathbf{e}_i \sim \mathcal{N}\left(\mathbf{m}, diag\left(\sigma\right)\right) \tag{2–24}$$

where $\mathbf{m}$ is vector containing the mean of each band of the input image and $diag\left(\sigma\right)$ is a diagonal covariance matrix with the elements equal to the variance of each band of the input image. The initial *mutation variances* are set to the variance of each band in the input data set. The mutation variances are used to create the distribution from which mutations for the evolutionary step of the algorithm are generated.

Given the initial set of endmembers, the *global population fitness* can be computed.

$$F(\mathbf{E}, \mathbf{X}) = \sum_{i=1}^{N} \|\mathbf{x}_i - \mathbf{E}^T\mathbf{p}_i\|^2 + \sum_{i=1}^{N} \left(1 - |\mathbf{p}_i|\right)^2 + \sum_{i=1}^{N} \sum_{\{k:p_{ik}<0\}} |p_{ik}| \tag{2–25}$$

where $\mathbf{p}_i = \left(\mathbf{E}^T\mathbf{E}\right)^{-1}\mathbf{E}^T\mathbf{x}_i$ is the vector of unconstrained abundance values for the pixel $\mathbf{x}_i$. In addition to the global fitness function, individual fitness functions, $F(\mathbf{e}_i, \mathbf{X})$, are computed for each endmember using only that endmember's abundance fractions.

The algorithm proceeds by iteratively selecting endmembers for mutation. An endmember is chosen for mutation based on its individual fitness function using *roulette wheel* selection (Grana et al., 2004; Whitley, 2001). Once chosen, an endmember is mutated by adding a random Gaussian perturbation.

$$\mathbf{e}_i^{j+1} = \mathbf{e}_i^j + \zeta_i \tag{2–26}$$

where

$$\begin{aligned} \zeta_i &\sim \mathcal{N}(\mathbf{0}, \sigma_i^{j+1}) && \text{(2–27)} \\ \sigma_i^{j+1} &= \sigma_i^j \exp(\tau \cdot \xi) && \text{(2–28)} \\ \xi &\sim \mathcal{N}(0, 1) && \text{(2–29)} \end{aligned}$$

τ is a step size constant. After mutating the endmember, the global fitness function in Equation 2–25 is recomputed. If the fitness function improves, then the mutated endmember replaces the original endmember. An endmember can be mutated up to λ times.

This algorithm searches for endmembers that minimize the squared error between the pixels and their estimation and minimize the amount abundance values are negative or do not prescribe to the sum-to-one constraint. Like many of other of existing endmember detection algorithms, this algorithm requires the input of the number of desired endmembers.

2.1.6 Independent Components Analysis

Several methods have been developed based on Independent Component Analysis (ICA) (Chiang, Chang, and Ginsberg, 2000; Tu, 2000; Tu, Huang, and Chen, 2001; Wang and Chang, 2006). Independent Component Analysis performs unsupervised separation of statistically independent sources in a data set (Nascimento and Bioucas-Dias, 2005a). The

data points, $\{\mathbf{x}_i\}_{i=1}^{N}$, are assumed to be linear mixtures of these independent components (Hyvarinen and Oja, 2000).

$$\mathbf{x}_i = \mathbf{E}\mathbf{p}_i \tag{2–30}$$

One method of determining components is by minimizing the mutual information between sources and ensuring sources effectively describe the data using $\mathbf{E}\mathbf{p}$ (Hyvarinen and Oja, 2000). Equivalently, ICA can be performed by searching for the components which match the data and are "non-Gaussian" (Hyvarinen and Oja, 2000). In ICA, the number of signal sources found is the same as the dimensionality of the data. Therefore, either D signal sources are found where D is the data dimensionality, or dimensionality reduction is used to find $M \leq D$ signals (Hyvarinen and Oja, 2000; Tu et al., 2001).

Chiang et al. (2000) directly apply ICA to the problem of determining the endmembers for a hyperspectral image where the abundance values are assumed to be statistically independent "random signal sources". Tu (2000) also applied the ICA algorithm for endmember extraction. Prior to running ICA, Tu estimates the number of endmembers and whitens the data to reduce the data dimensionality using the Noise-Adjusted Transformed Gerschgorin disk (NATGD) (Tu, 2000). Similarly, the Spectral Data Explorer algorithm (SDE) uses the Partitioned Noise-Adjusted Principal Components Algorithm (PNAPCA) to whiten the data and determine the number of endmember after which ICA is performed (Tu et al., 2001). NATGD and PNAPCA methods for estimating the number of endmembers are described in Section 2.1.7. Wang and Chang (2006) apply ICA for endmember extraction where, prior to running ICA, the number of endmembers, M, is estimated using the Virtual Dimensionality (described in Section 2.1.7) of the data set. Given the estimated number of endmembers, the independent components determined using the ICA algorithm are prioritized using the 3^{rd} and 4^{th} order statistics of the component (Wang and Chang, 2006). For each of the M highest priority components, image pixels with the largest absolute value of the abundance are returned as endmembers (Wang and Chang, 2006).

Nascimento and Bioucas-Dias (2005a) argue that the use of Independent Components Analysis for endmember detection is not an accurate method since the sum-to-one constraint on the abundance values cause the sources in the corresponding ICA problem to be dependent. This dependency violates the basic ICA assumption of statistically independent sources (Nascimento and Bioucas-Dias, 2005a). Nascimento and Bioucas-Dias (2005a) provide results to argue that the some endmembers are incorrectly unmixed using ICA methods. As an alternative, the Dependent Component Analysis method (DECA) was developed which assumes the abundance values are drawn from a Dirichlet distribution (Nascimento and Bioucas-Dias, 2007a,b). The Dirichlet enforces the non-negativity and sum-to-one constraints on the abundance values. DECA determines abundances and endmember values using the Expectation-Maximization (EM) method (Nascimento and Bioucas-Dias, 2007a,b). However, like ICA-based methods, DECA also requires the number of endmembers to be known in advance.

2.1.7 Estimating the Number of Endmembers

The concept of *virtual dimensionality* is used by some methods, prior to endmember extraction, to estimate the number of endmembers for a given scene (Chang and Du, 2004; Wang and Chang, 2006). Also, Tu (2000) relies on the Transformed Gerschgorin Disk (TGD) and the Noise-Adjusted TGD method of estimating the number of endmembers. The Partitioned Noise-Adjusted Principal Components Analysis (PNAPCA) method of computing the number of endmembers is based on partitioning and transforming the noise-adjusted covariance matrix (Liou and Yang, 2005; Tu et al., 2001).

Virtual dimensionality. Virtual Dimensionality (VD) is defined as the "minimum number of spectrally distinct signal sources" in a hyperspectral data set (Chang and Du, 2004). VD is computed using the eigenvalues of the covariance and correlation matrices of the input data set. Let $\left\{\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_d\right\}$ be the eigenvalues from the sample correlation matrix and let $\{\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d\}$ be the eigenvalues from the sample

covariance matrix. The VD is computed as the number of corresponding correlation and covariance eigenvalues that differ (Chang and Du, 2004).

$$\hat{\lambda}_r \geq \lambda_r, \qquad for \quad r = 1, \ldots, VD \tag{2–31}$$
$$\hat{\lambda}_r = \lambda_r, \qquad for \quad r = VD + 1 \ldots, d \tag{2–32}$$

The eigenvalues computed from the sample covariance matrix equal the variances of the transformed data (Theodoridis and Koutroumbas, 2003) and the eigenvalues of the sample correlation matrix are related to the variance of the data from the origin. Assuming that the noise has zero-mean and unit-variance and signals in the data have non-zero values, the eigenvalues of the sample correlation matrix corresponding to signals in the data will have a larger value than the corresponding eigenvalues from the sample covariance matrix. The eigenvalues of the sample correlation and covariance matrix corresponding to noise will be equal (Chang and Du, 2004),

$$\hat{\lambda}_r \geq \lambda_r \geq \sigma^2_{n_l} \quad for \quad r = 1, \ldots, VD \tag{2–33}$$
$$\hat{\lambda}_r = \lambda_r = \sigma^2_{n_l} \quad for \quad r = VD + 1 \ldots, d \tag{2–34}$$

where $\sigma^2_{n_l}$ is the noise variance.

An example of this concept is shown using the data in Figure 2-6. The three-dimensional data set was generated using two endmembers, [2, 5, 0] and [3, 6, 1]. Zero-mean Gaussian noise with a variance of 0.03 was added to each coordinate of the data. The eigenvalues of the covariance and correlation matrices were computed. The eigenvalues from the covariance matrix were found to be 0.1814, 0.0010, and 0.0008. The eigenvalues for the correlation matrix were found to be 36.4991, 0.0647, and 0.0008. Therefore, the virtual dimensionality correctly determines the number of endmembers to be two, since the third eigenvalue from both the covariance and correlation matrices are equal.

In order to determine that the eigenvalues differ, their differences are thresholded. Chang and Du (Chang and Du, 2004) describe three thresholding methods for determining the virtual dimensionality. These methods include the Harsanyi-Farrand-Chang (HFC)

method, the Noise Whitened HFC (NWHFC) method, and the Noise Subspace Projection (NSP) method.

The HFC method thresholds the differences of the covariance and correlation eigenvalues based on a Neyman-Pearson detector. The probability of detection is maximized while the probability of false alarm is held to a constant value, a.

$$P_F = \int_{\tau_l}^{\infty} \mathcal{N}(0, \sigma_{z_l}^2) dz = a \tag{2–35}$$

$$P_D = \int_{\tau_l}^{\infty} \mathcal{N}(\mu_l, \sigma_{z_l}^2) dz \tag{2–36}$$

Solving for τ_l gives the threshold for the eigenvalue differences. This thresholding method requires an estimate of the variance of the difference between the eigenvalues at each band, σ_{z_l}. The HFC method uses $\sigma_{z_l}^2 \approx \frac{2\hat{\lambda}_l^2}{N} + \frac{2\lambda_l^2}{N}$ as the estimate of this variance.

The HFC method assumes that the data has white noise with zero mean. The Noise-Whitened HFC method (NWHFC) attempts to improve VD estimates by performing noise-whitening to the correlation and covariance matrices prior to computing the differences between their eigenvalues. The NWHFC method requires an estimate of the noise covariance. After estimating the noise covariance and whitening the correlation and covariance matrices, the thresholds are computed using the same method as HFC (Chang and Du, 2004).

The Noise Subspace Projection method is similar to the NWHFC. An estimate of the noise covariance for the data is used to whiten the covariance matrix. However, instead of computing differences. This method recognizes that the eigenvalues corresponding to noise should be equal to one. Therefore, the computed threshold is applied directly to the eigenvalues to determine if their difference from one is significant (Chang and Du, 2004).

This method is sensitive to the variance and covariance estimates used in determining the thresholds for each eigenvalue. Therefore, the VD estimate of the number of endmembers is sensitive to noise in the input data set. The VD was run on two data sets generated from the three AVIRIS Cuprite spectra shown in Figure 2-7. The a small

amount of Gaussian noise added to the first data set, shown in Figure 2-8. The second data set, shown in Figure 2-9, has more added noise. The three thresholding methods were applied to both sets of data. On the first data set, the NSP method correctly determined the number of endmembers by estimating 3 signals. The HFC and NWHFC methods incorrectly estimated the number of endmembers with 2 and 5, respectively. On the second data set with larger amounts of noise, none of the thresholding methods correctly estimated three endmembers. The HFC method estimated 2, the NWHFC method estimated 7, and the NSP method estimated 2 endmembers.

Maximum noise fraction. The Partitioned Noise Adjusted Principal Components Analysis (PNAPCA) method (Tu et al., 1999, 2001) of estimating the number of endmembers is based on the Maximum Noise Fraction (MNF) (also known as the Noise Adjusted Principal Components Analysis method) (Green et al., 1988; Lee et al., 1990).

The MNF transform uses an estimate of the noise covariance matrix to transform the data into components which are sorted based on their signal-to-noise ratio. Using the convex geometry model in Equation 1–1 and assuming the noise and signal components of the data are uncorrelated, the covariance matrix of the data can be written as (Lee et al., 1990; Tu et al., 1999)

$$\begin{aligned}\mathbf{\Sigma}_X &= E\left[\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{x}_i'^T\right] = E\left[\sum_{i=1}^{N}(\mathbf{E}'\mathbf{a}_i^T)(\mathbf{E}'\mathbf{a}_i^T)^T\right] + E\left[\sum_{i=1}^{N}\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i^T\right] \\ &= \mathbf{\Sigma}_S + \mathbf{\Sigma}_N\end{aligned} \tag{2–37}$$

where $X' = \{\mathbf{x}_i' | \mathbf{x}_i' = \mathbf{x}_i - \mu\}$. The MNF determines a transformation which maximizes the signal-to-noise ratio (Lee et al., 1990; Tu et al., 1999),

$$\begin{aligned}arg\max_{\mathbf{W}} SNR &= arg\max_{\mathbf{W}} \frac{\mathbf{W}^T\mathbf{\Sigma}_S\mathbf{W}}{\mathbf{W}^T\mathbf{\Sigma}_N\mathbf{W}} \\ &= arg\max_{\mathbf{W}} \frac{\mathbf{W}^T\mathbf{\Sigma}_X\mathbf{W}}{\mathbf{W}^T\mathbf{\Sigma}_N\mathbf{W}} - 1.\end{aligned} \tag{2–38}$$

The signal-to-noise ratio is maximized by assigning $\mathbf{W} = \mathbf{\Phi}_N\mathbf{\Lambda}_N^{-\frac{1}{2}}\mathbf{\Phi}_A$ where $\mathbf{\Phi}_N$ is the eigenvector matrix of the noise covariance matrix, $\mathbf{\Lambda}_N$ is the diagonal eigenvalue matrix of

the noise covariance matrix, and $\mathbf{\Phi}_A$ is the eigenvector matrix of noise-adjusted covariance matrix, $\mathbf{\Sigma}_A = \left(\mathbf{\Phi}_N \mathbf{\Lambda}_N^{-\frac{1}{2}}\right)^T \mathbf{\Phi}_X \left(\mathbf{\Phi}_N \mathbf{\Lambda}_N^{-\frac{1}{2}}\right)$ (Lee et al., 1990; Tu et al., 1999). Using the matrix $\mathbf{W}$, the MNF transform simultaneously diagonalizes the data covariance matrix and whitens the noise covariance matrix.

The MNF transformation requires an estimate of the noise covariance matrix, $\mathbf{\Sigma}_N$ (Tu et al., 1999). As described by Tu et al. (1999, 2001), PNAPCA method partitions the noise adjusted covariance matrix found by MNF and diagonalizes the two partitions. Tu et al. (1999, 2001) claim that by examining the eigenvalues of the two partitions simultaneously, the effects of incorrectly estimating the noise covariance matrix are lessened.

Transformed Gerschgorin disk and the noise adjusted transformed Gerschgorin disk. Wu et al. (1995) and Tu (2000) developed methods of estimating the number of signals in a data set based on Gerschgorin's disk theorem (Horn and Johnson, 1985). Gerschgorin's disk theorem provides a method of estimating the locations of eigenvalues of a matrix. The theorem states that the eigenvalues of a matrix, $\mathbf{A}$, are located within the union of the disks defined by

$$G_i = \{z : |z - a_{ii}| \leq r_i\} \tag{2–39}$$

where

$$r_i = \sum_{j=1, j \neq i}^{D} |a_{ij}| \tag{2–40}$$

and $a_{ii}, i = 1, \ldots, D$, the centers of the Gerschgorin disks, are the elements along the diagonal of matrix $\mathbf{A}$ (Horn and Johnson, 1985). The theorem also states that if the union of k of the D disks form a connected region and if the connected region is disjoint from all of the remaining disks, then k eigenvalues are located within the region defined by the union of the k disks (Horn and Johnson, 1985).

The transformed Gerschgorin Disk method developed by Wu et al. (1995) defines a transformation on the covariance matrix of an input data set so that the Gerschgorin disks

associated with noise in the data set have small radii and are located far from the signal disks. In other words, the goal of this transformation is to cause the disks containing eigenvalues corresponding to signals in the data to have large radii. Tu (2000) applies the transformed Gerschgorin disk method to the noise-adjusted covariance matrix.

The transformation matrix used in the Transformed Gerschgorin Disk and the Noise-Adjusted Transformed Gerschgorin Disk methods is determined by diagonalizing the $D-1 \times D-1$ leading sub-matrix of the input covariance matrix,

$$\mathbf{\Sigma}_X = \begin{bmatrix} \mathbf{C}_1 & \mathbf{c} \\ \mathbf{c}^T & \mathbf{c}_{DD} \end{bmatrix} = \begin{bmatrix} \mathbf{U}_1^T \mathbf{V}_1 \mathbf{U}_1 & \mathbf{c} \\ \mathbf{c}^T & \mathbf{c}_{DD} \end{bmatrix} \tag{2–41}$$

where $\mathbf{C}_1$ is the $D-1 \times D-1$ leading sub-matrix of the input covariance matrix, $\mathbf{c}^T = \left[c_{1D}, \ldots, c_{(D-1)D}\right]$, $\mathbf{U}_1$ is the matrix of eigenvectors of $\mathbf{C}_1$ and $\mathbf{V}_1$ is the diagonal matrix of eigenvalues. Given the matrix of eigenvectors of the leading sub-matrix, $\mathbf{U}_1$, the following transformation, $\mathbf{U}$, is defined and applied to the input covariance matrix,

$$\mathbf{U} = \begin{bmatrix} \mathbf{U}_1 & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix} \tag{2–42}$$

$$\mathbf{U}^T \mathbf{\Sigma}_X \mathbf{U} = \begin{bmatrix} \mathbf{U}^T \mathbf{C}_1 \mathbf{U} & \mathbf{U}^T \mathbf{c} \\ \mathbf{c}^T \mathbf{U} & \mathbf{c}_{DD} \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & 0 & . & . & . & 0 & \rho_1 \\ 0 & \lambda_2 & 0 & . & . & . & 0 & \rho_2 \\ 0 & 0 & \lambda_3 & . & . & . & 0 & \rho_3 \\ . & & . & & & & . & . \\ . & & & . & & & . & . \\ . & & & & . & & . & . \\ 0 & . & . & . & . & . & \lambda_{D-1} & \rho_{D-1} \\ \rho_1 & \rho_2 & \rho_3 & . & . & . & \rho_{D-1} & c_{DD} \end{bmatrix}. \tag{2–43}$$

Using the Gerschgorin disk theorem in Equation 2–39, the Gerschgorin disks of the transformed input covariance matrix in Equation 2–43 have radii equal to $|\rho_1|, |\rho_2|, \ldots, |\rho_{D-1}|$ and centers at $\lambda_1, \lambda_2, \ldots, \lambda_{D-1}$. Assuming that the noise in the

data is uncorrelated and has zero mean, the radii associated with noise will be equal to zero. Therefore, the Transformed Gerschgorin Disk method returns, as the number of endmembers, the number of large radii.

The Noise Adjusted Transformed Gerschgorin Disk method applies this method to the noise-adjusted covariance matrix. This requires an estimate of the noise covariance matrix which is used to whiten the noise in the input data set. Like the VD and NAPCA methods of estimating the number of endmembers, the Transformed Gerschgorin Disk methods are sensitive to noise.

2.2 Existing Hyperspectral Band Selection Algorithms

In addition to endmember extraction, a hyperspectral band selection method that determines the required number of bands, performs unsupervised band selection, and retains bands that help to distinguish between endmembers in a scene is proposed. The presented method performs these tasks while simultaneously determining endmembers and the number of endmembers needed.

Many data reduction techniques such as Principal Components Analysis (PCA) and Maximum Noise Fraction transform (MNF) (Green et al., 1988; Lee et al., 1990; Theodoridis and Koutroumbas, 2003) have been used to project the data into a lower dimensional space and thus reduce the dimensionality of the data. Although these methods are effective at data reduction, they do not retain physically meaningful bands that correspond to wavelengths in the original data set. Harsanyi and Chang (1994) provide an orthogonal subspace projections approach that also transforms the data. Bruce et al. (2002) conducts dimensionality reduction by extracting features that distinguish between labeled classes in a training set using the Discrete Wavelet Transform (DWT). This method extracts features that incorporate both frequency information and detailed localized features of the input hyperspectral signal. This feature set is further reduced using Fisher's Linear Discriminant Analysis. Since this method using the DWT and Fisher's Linear Discriminant, extracted features do not correspond to wavelengths in the

original data set. DeBacker et al. (2005) and Kumar et al. (2001) both present methods that merge many adjacent hyperspectral bands. DeBacker et al. (2005) merges bands into groups which optimize the Bhattacharya distance between labeled classes in a training set. The supervised band selection method presented by Riedmann and Milton (2003) merges neighboring bands to improve accuracy in a classification task. The hyperspectral dimensionality reduction method based on Localized Discriminant Bases (Venkataraman et al., 2005) also merges adjacent bands for feature extraction. Martinez-Uso et al. (2007) present a band merging algorithm using information measures and hierarchical clustering. A divergence measure between every pair of bands is computed and used to perform hierarchical clustering of the bands. A band representative is then computed for each cluster. Lin and Bruce (2004) use a Projection Pursuits methods to reduce the dimensionality of a hyperspectral data set by determining a projection matrix that aids in distinguishing between classes in the data set. Instead of merging bands or transforming the data, this method maintains only those bands that are useful for the hyperspectral image analysis task. The advantage of physically meaningful bands is to identify useful wavelengths for a particular classification task. Identifying important wavelengths can also be used in the design of hyperspectral sensors. By reducing the number of wavelengths that need to be collected, data collection will be performed faster and with less required storage space.

Additionally, most of the previously mentioned band selection algorithms (DeBacker et al., 2005; Green et al., 1988; Harsanyi and Chang, 1994; Lee et al., 1990; Martinez-Uso et al., 2007) require the knowledge of the desired number of bands. Serpico and colleagues' search method for band selection, Du and colleagues' method of band prioritization based on the Independent Component Analysis' weight matrix, Han and colleagues' eigenvalue weighted band prioritization method, and Guo and colleagues' (Du et al., 2003; Guo et al., 2006; Han et al., 2004; Serpico and Bruzzone, 2001) mutual information based band selection method require the desired number of bands. Petrie et al. (1998) outlines

four band selection strategies based on maximizing spatial autocorrelation, maximizing a distance measure between targets, or merging neighboring bands. All of the methods described by Petrie et al. (1998) require the desired number of bands. The band selection method based on the NFindr algorithm (Wang et al., 2006) retains the bands which maximize the volume between the endmembers found using the NFindr algorithm (Winter, 1999). This method attempts to find bands which aid in spectral unmixing, however, the number of bands to retain must be known in advance. Often, the number of required bands is not known.

Keshava presented a method based on the Spectral Angle Mapper (SAM) distance or the Euclidean Minimum Distance (EMD) measures (Keshava, 2001, 2004). The algorithm incrementally adds bands that increase the SAM or EMD measure between two labeled classes until some stopping criterion is reached. Although this method does not require the number of bands in advance, the method is limited to distinguishing between two labeled classes in the data set. Similarly, the Sparse Linear Filters algorithm (Theiler and Glocer, 2006) develops sparse linear filters to distinguish between two labeled classes in the data. The filters use a sparse set of the hyperspectral bands by utilizing an L1-penalty term to select the bands and the number of bands (Tibshirani, 1996). Although the number of bands is estimated, the method requires two labeled classes (Theiler and Glocer, 2006). Chang et al. (1999) ranks all bands based on loading factors constructed using maximum-variance PCA, MNF, orthogonal subspace projection and minimum misclassification canonical analysis methods. Following ranking, Chang et al. (1999) remove correlated bands using a divergence measure. Chang and Wang (2006) employ a method based on constrained energy minimization. The method selects bands that have the minimal correlation with each other and uses the concept of virtual dimensionality to determine the number of chosen spectral bands.

2.3 Summary of Literature Review

The majority of endmember detection algorithms require the number of endmembers needed for a hyperspectral image prior to running the detection algorithm. Methods of estimating the number the of endmembers in a data set, such as Virtual Dimensionality, are sensitive to noise in the data. Furthermore, many endmember detection algorithms assume that pure pixels for each endmember can be found in the data set. In highly mixed data sets, this assumption does not hold causing the algorithms to return mixed pixels as endmembers. Also, many existing endmember detection algorithms do not estimate abundance values that conform to the non-negativity and sum-to-one constraints in Equation 1–2. These algorithm also do account for spectral variability in their endmember representations. Furthermore, the existing algorithms do not consider cases in which multiple convex regions and sets of endmembers may more accurately describe the data.

Hyperspectral band selection algorithms often require the number of needed hyperspectral bands prior to running the band selection algorithm. Data reduction techniques which perform projections are often used to reduce the dimensionality of a hyperspectral image. However, the projection methods lose the physical meaning associated with the hyperspectral bands. Hyperspectral band selection algorithms are also often tied to classification problems. In these cases, labeled training data is needed to determine the bands which distinguish between the classes.

Band selection and endmember detection methods presented autonomously determine the number of endmember and hyperspectral bands needed for an image. Methods are presented which account for an endmember's spectral variability and can autonomously determine the number of convex regions needed to describe an input data set. Furthermore, all presented methods provide abundance values which conform to the constraints in Equation 1–2. The new algorithms are also capable of determining endmembers for highly mixed data sets since the pixel purity assumption is not employed. The presented hyperspectral band selection algorithm also retains physically meaningful

bands and does not require labeled training data. The method, instead, determines the hyperspectral bands which distinguish between the endmembers in a data set.

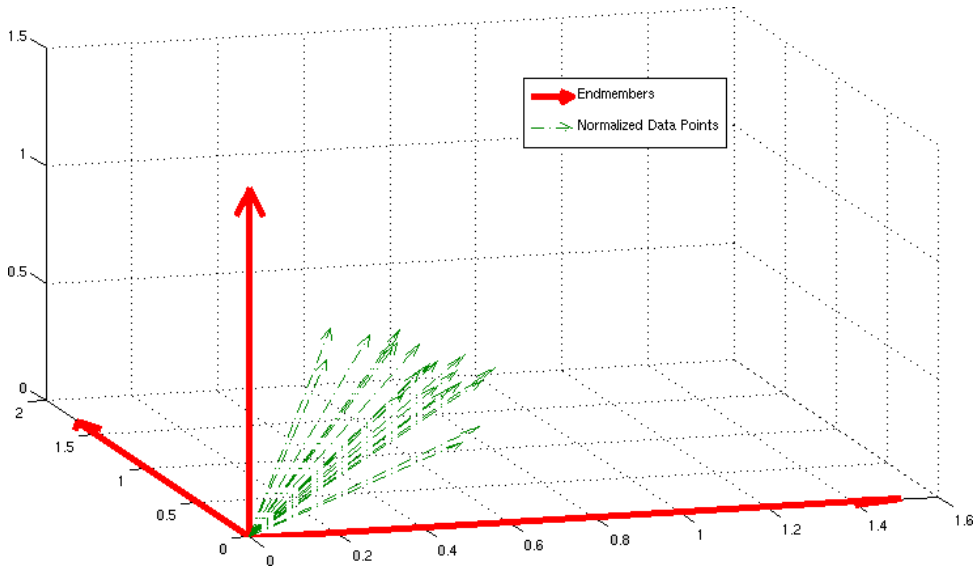


Figure 2-1. Three-dimensional data points and endmember results using convex cone analysis.

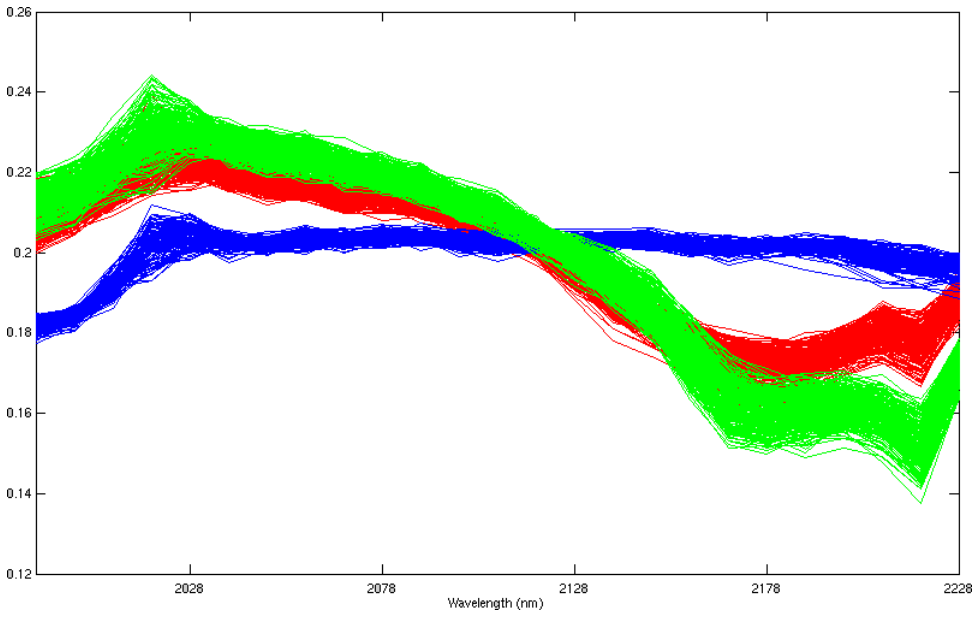


Figure 2-2. First 25 bands (1978 to 2228 nm) of a subset of normalized pixels from the AVIRIS cuprite "scene 4" data set

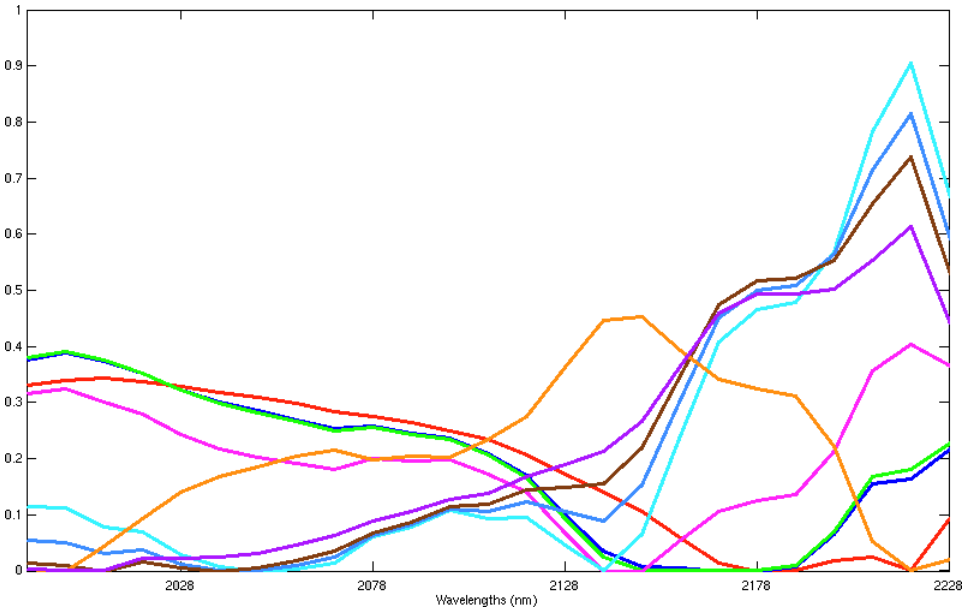


Figure 2-3. CCA endmember results on a subset of AVIRIS Cuprite data

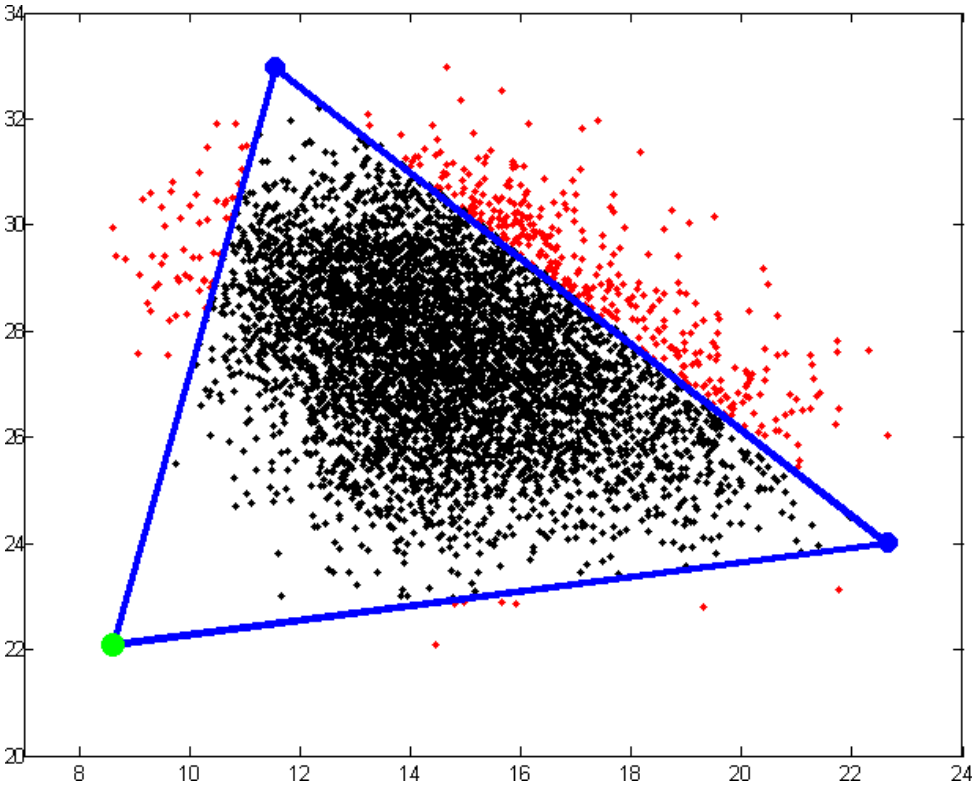


Figure 2-4. Morphological associative memories endmember results using the min memory on two-dimensional data. Endmembers found from the columns of the min memory are shown in blue. The shade point is green. Data points within the area defined by the endmembers are in black. Data points outside of the area defined by the endmembers are in red.

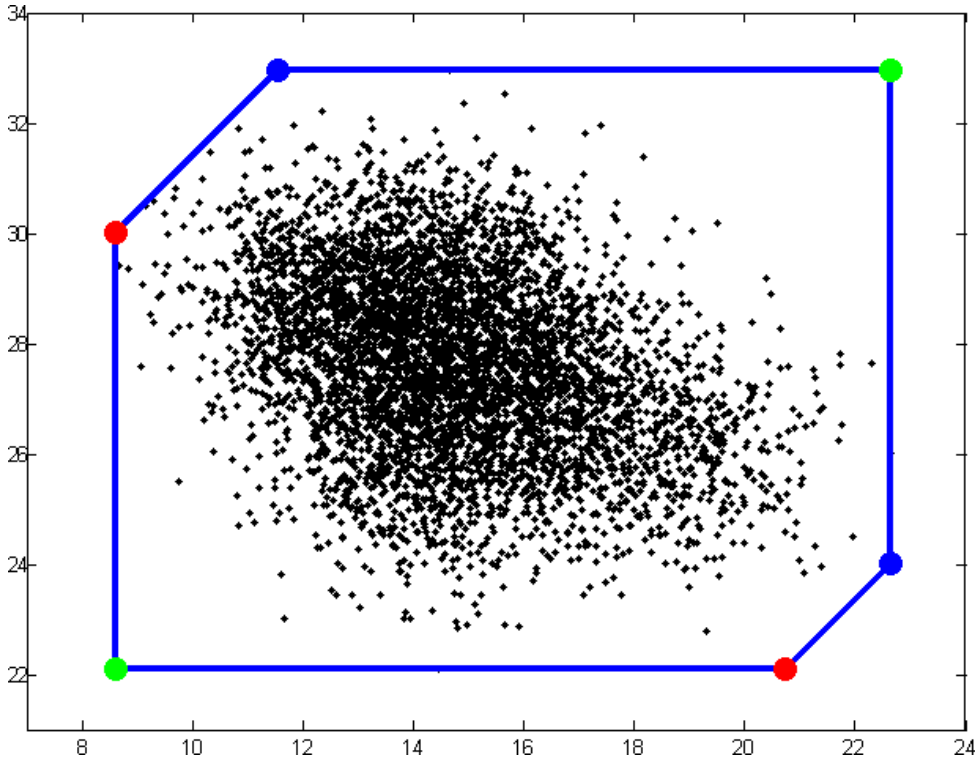


Figure 2-5. Morphological associative memories endmember results using both memories on two-dimensional data. Endmembers found from the columns of the max memory are shown in red. Endmembers from the columns of the min memory are shown in blue. The bright point and shade point are green.

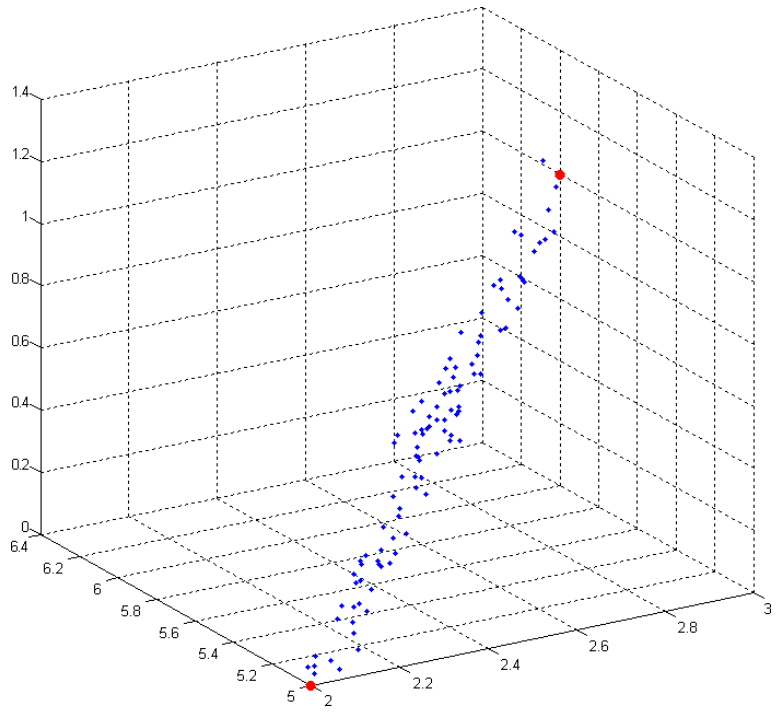


Figure 2-6. Three dimensional data set generated from two endmembers with Gaussian noise.

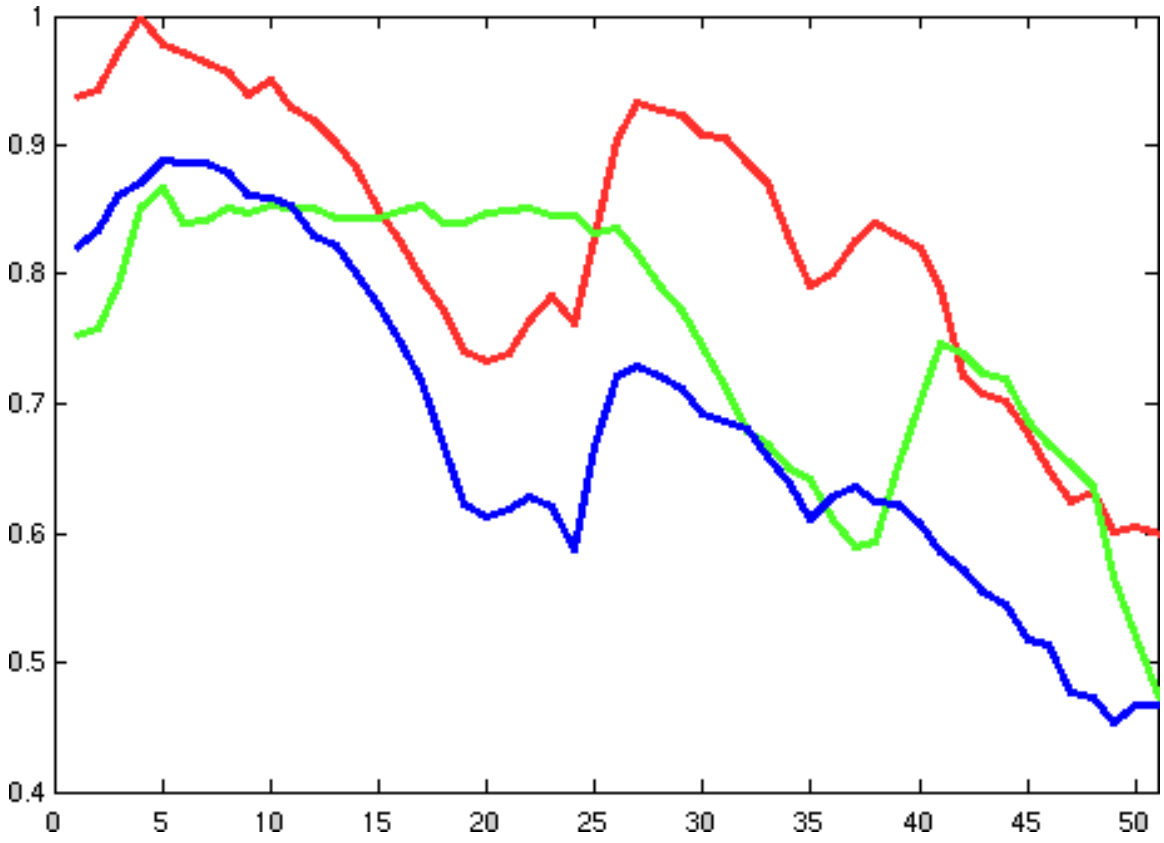


Figure 2-7. Normalized AVIRIS Cuprite spectra

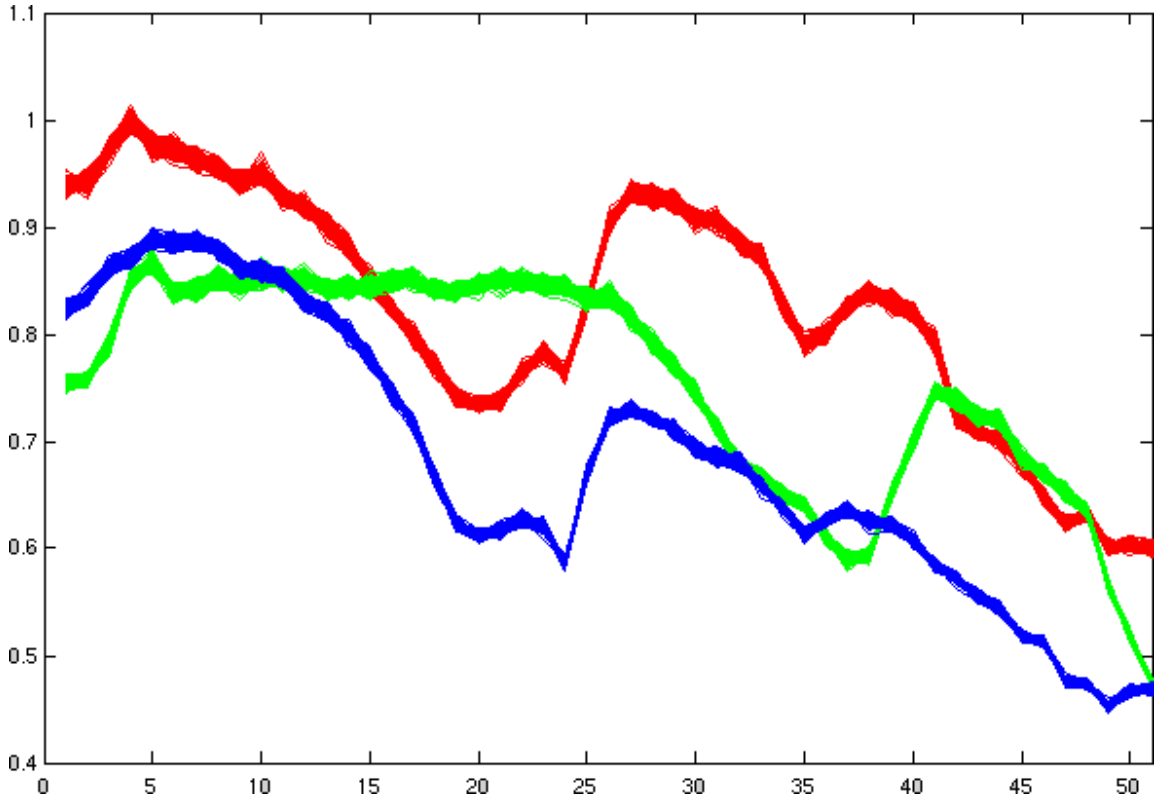


Figure 2-8. Data set generated from AVIRIS Cuprite endmembers with a small amount of Gaussian noise

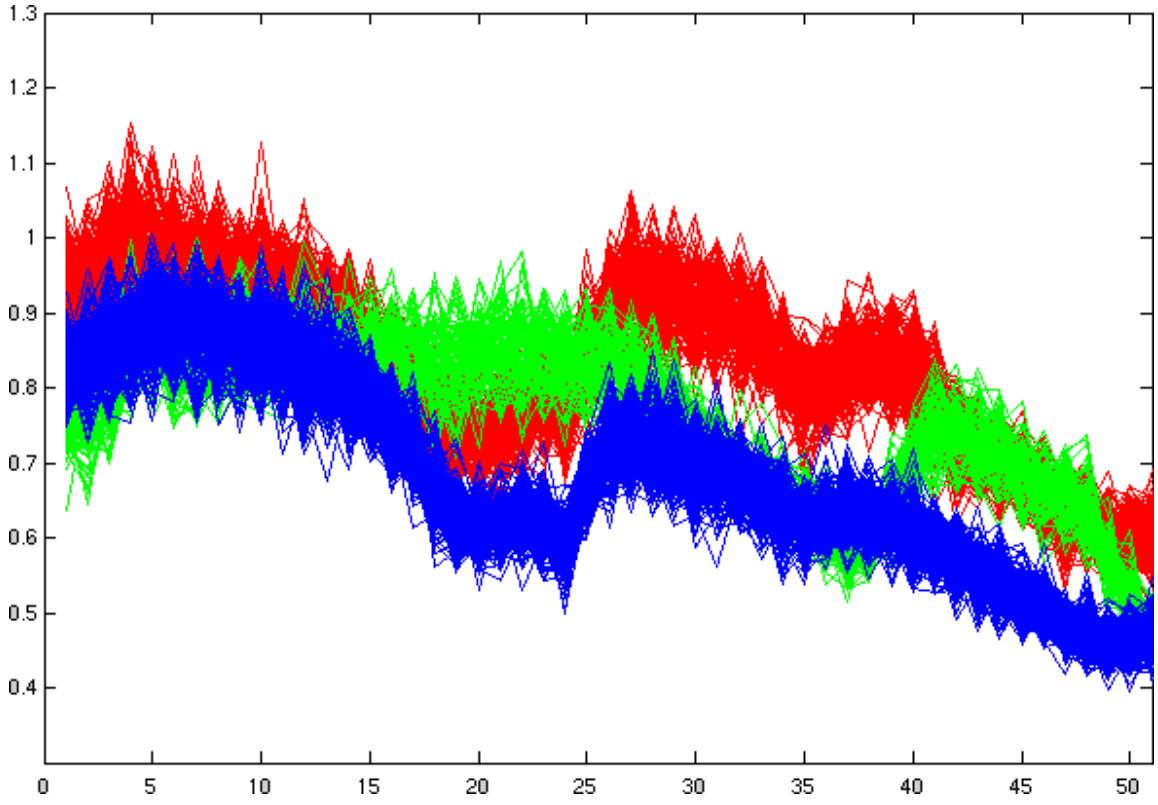


Figure 2-9. Data set generated from AVIRIS Cuprite endmembers with a large amount of Gaussian noise

CHAPTER 3
TECHNICAL APPROACH

The focus of the presented research is on a set of algorithms that utilize Bayesian methods to perform endmember detection and spectral unmixing simultaneously. To this end, four specific approaches are considered. These algorithms determine the number of endmembers, learn endmember distributions, determine the number of convex regions needed to describe the input hyperspectral data set, or determine the number of needed hyperspectral bands. Sparsity Promoting Iterated Constrained Endmembers (SPICE) is an endmember detection and spectral unmixing algorithm which uses sparsity promoting priors to remove unneeded endmembers. Band Selecting Sparsity Promoting Iterated Constrained Endmembers (B-SPICE) is an extension of the SPICE algorithm which incorporates band selection. The Endmember Distributions (ED) detection algorithm determines full endmember distributions for each endmember rather than single endmember spectra thus incorporating the spectral variation which occurs due to varying environmental conditions. The Piece-wise Convex Endmember (PCE) detection algorithm uses the Dirichlet process to autonomously determine the number of convex regions needed to describe a data set while simultaneously learning endmember distributions and abundances for each convex region. SPICE, B-SPICE, ED and PCE determine the spectral shape of the endmembers and compute abundance values which conform to the non-negative and sum-to-one constraints. All methods do not rely on the pixel purity assumption and are, therefore, capable of handling highly mixed data sets. Furthermore, the endmembers determined by these methods are capable of enveloping all of the hyperspectral data points while providing a tight fit around the data. The B-SPICE algorithm, in addition to determining the number of endmembers and their spectral shape, also performs band selection and determines the number of hyperspectral bands required. B-SPICE retains the physical meaning of each hyperspectral band and does not rely on a projection to perform data reduction. Also, B-SPICE does not require labeled data to

determine the useful hyperspectral bands but, instead, retains the bands which help to distinguish between the endmembers in the data set.

3.1 Review of Sparsity Promotion Techniques

The goal of sparsity promotion is to minimize the number of parameters. This is generally done by encouraging parameter values to be driven to zero and, thus, minimizing the number of non-zero parameter values. A common method to promote small parameter values is to add a weight decay term to the objective function (Williams, 1995). Weight decay terms has been previously applied in neural network applications to promote regularization (Haykin, 1999; Williams, 1995).

Consider a least squares objective with a weight matrix, $\mathbf{P}$. A weight decay term applied to the $\mathbf{P}$ parameters attempts to prevent them from becoming large.

$$LSWD = \ln \exp \left\{ -\frac{1}{2} \sum_{i=1}^{N} \left(\mathbf{X}_i - \sum_{k=1}^{M} p_{ik} \mathbf{E}_k \right)^2 - \gamma \sum_{i=1}^{N} \sum_{k=1}^{M} p_{ik}^2 \right\} \tag{3–1}$$

$$= \ln \exp \left\{ -\frac{1}{2} \sum_{i=1}^{N} \left(\mathbf{X}_i - \sum_{k=1}^{M} p_{ik} \mathbf{E}_k \right)^2 \right\} \exp \left\{ -\gamma \sum_{i=1}^{N} \sum_{k=1}^{M} p_{ik}^2 \right\} \tag{3–2}$$

where $\gamma \geq 0$. Equation 3–2 can be interpreted in a probabilistic manner where the second exponential can be seen as a zero-mean Gaussian (Williams, 1995). Therefore, Equation 3–2 can be viewed as the log of the product in Equation 3–3 (Williams, 1995),

$$p\left(\mathbf{P}|\mathbf{X}\right) \propto p\left(\mathbf{X}|\mathbf{P}\right) p\left(\mathbf{P}\right) \tag{3–3}$$

where $p\left(\mathbf{X}|\mathbf{P}\right)$ is the probability of the data given the parameters and $p\left(\mathbf{P}\right)$ is the prior on the parameters.

Unfortunately, the Gaussian prior is not effective at sparsity promotion. The Gaussian does not prefer to set parameter values to zero which would promote sparsity. Instead, the Gaussian prefers several small valued non-zero parameters (Williams, 1995). Instead of using a Gaussian distribution for the parameters' prior, a zero-mean Laplacian distribution can be used which is more effective at sparsity promotion (Figueiredo, 2003; Williams,

1995),

$$LSSP = -\frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{X}_i - \sum_{k=1}^{M} p_{ik}\mathbf{E}_k\right)^2 - \sum_{k=1}^{M}\gamma_k \sum_{i=1}^{N}|p_{ik}|\,. \tag{3–4}$$

The application of a Laplacian distribution is related the to *least absolute shrinkage and selection operator* (LASSO) (Tibshirani, 1996). As described by Tibshirani (1996), the LASSO applies the constraint that the sum of the absolute values of the weights, $\sum_{k=1}^{M}\sum_{i=1}^{N}|p_{ik}|$ must be less than a threshold. This is equivalent to applying a Laplacian prior (Tibshirani, 1996).

Often during optimization, the derivative of an objective function must be computed. In the case where the Laplacian prior has been used to penalize large parameter values, the derivative is not defined at zero due to the absolute value function (Williams, 1995). In these cases, the Laplacian prior on the parameters can be defined in a hierarchical fashion where the parameters are distributed according to a Gaussian distribution whose variance has an exponential hyper-prior (Figueiredo, 2003). By integrating over all possible values for the variance given the hyper-prior, the hierarchical expression is equivalent the to the Laplacian distribution (Figueiredo, 2003).

Sparsity promotion has been applied to a number of applications including neural networks (Williams, 1995), classification and regression using expectation-maximization (EM) (Figueiredo, 2003), feature selection and classification (Krishnapuram et al., 2004), classification and regression using the Choquet integral (Mendez-Vazquez and Gader, 2007) and others.

3.2 Review of the Iterated Constrained Endmembers Detection Algorithm

The presented methods for endmember detection and band selection using sparsity promoting priors are based on the Iterated Constrained Endmembers (ICE) algorithm. The ICE algorithm (Berman et al., 2004) performs a minimization of a residual sum of squares (RSS) term based on the convex geometry model in Equation 1–1. The error between the pixel spectra and the pixel estimate found by the ICE algorithm using the endmembers and their proportions is minimized when the residual sum of squares (RSS)

term is minimized (Berman et al., 2004),

$$RSS = \sum_{i=1}^{N} \left(\mathbf{X}_i - \sum_{k=1}^{M} p_{ik} \mathbf{E}_k \right)^T \left(\mathbf{X}_i - \sum_{k=1}^{M} p_{ik} \mathbf{E}_k \right). \tag{3–5}$$

As described by Berman et al. (2004), the minimizer for the RSS term is not unique. Therefore, the ICE algorithm adds a sum of squared distances (SSD) term to the objective function.

$$SSD = \sum_{k=1}^{M-1} \sum_{l=k+1}^{M} (\mathbf{E}_k - \mathbf{E}_l)^T (\mathbf{E}_k - \mathbf{E}_l) \tag{3–6}$$

This term is related to the volume bounded by the endmembers. Therefore, by adding this term to the objective function, the algorithm finds endmembers that provide a tight fit around the data. Berman et al. (2004) show that the SSD is equivalent to Equation 3–7,

$$SSD = M(M-1)V \tag{3–7}$$

where V is the sum of variances (over the bands) of the simplex vertices. ICE uses V in the objective function instead of $M(M-1)V$ in an effort to make this term independent of the number of endmembers, M (Berman et al., 2004).

The objective function used in the ICE algorithm is shown in Equation 3–8,

$$RSS_{reg} = (1-\mu)\frac{RSS}{N} + \mu V \tag{3–8}$$

where μ is a regularization parameter that balances the RSS and SSD terms of the objective function.

The ICE algorithm minimizes this objective function iteratively. First, given endmember estimates, the proportions for each pixel are estimated. For the first iteration of the algorithm, endmember estimates may be set to randomly chosen pixels from the image. Estimating the proportions requires a least squares minimization of each term in Equation 3–5. Since each of these terms is quadratic and subjected to the linear constraints in Equation 1–2, the minimization is done using quadratic programming. After solving for the proportions, the endmembers are updated using the current proportion

estimates,

$$\mathbf{e}_j = \left\{ \mathbf{P}^T\mathbf{P} + \lambda \left(\mathbf{I}_M - \frac{\mathbf{1}\mathbf{1}^T}{M} \right) \right\}^{-1} \mathbf{P}^T\mathbf{x}_j \tag{3–9}$$

where $\mathbf{P}$ is the $N \times M$ proportion matrix, $\mathbf{e}_j$ is the vector of endmember values in the j^{th} band, $\mathbf{x}_j$ is the vector of all the pixel values in the j^{th} band, $\mathbf{I}_M$ is the $M \times M$ identity matrix, $\mathbf{1}$ is the M-vector of ones and $\lambda = N\mu/\{(M-1)(1-\mu)\}$. This iterative procedure is continued until the value of RSS_{reg} is smaller than a tolerance value. Although the ICE algorithm is an effective algorithm for finding endmembers when the number of endmembers is known, there is no automated mechanism in ICE to determine the correct number of endmembers.

3.3 New Endmember Detection Algorithm Using Sparsity Promoting Priors

The proposed Sparsity Promoting Iterated Constrained Endmember (SPICE) algorithm is based on the Iterated Constrained Endmembers algorithm by Berman et al. (2004) described in Section 3.2. The SPICE algorithm, which is a sparsity promoting extension of ICE, is developed in this section.

The RSS term of the ICE objective function is a least squares term whose minimization is equivalent to the maximization of Equation 3–10 (Williams, 1995)

$$-\frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{X}_i - \sum_{k=1}^{M} p_{ik}\mathbf{E}_k\right)^2 = \ln \exp\left\{ -\frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{X}_i - \sum_{k=1}^{M} p_{ik}\mathbf{E}_k\right)^2 \right\}. \tag{3–10}$$

When examining the exponential in Equation 3–10, it can be seen that this is proportional to the Gaussian density with mean $\sum_{k=1}^{M} p_{ik}\mathbf{E}_k$ and variance 1,

$$\begin{aligned} \mathcal{N}\left(\sum_{k=1}^{M} p_{ik}\mathbf{E}_k, 1\right) &= \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{X}_i - \sum_{k=1}^{M} p_{ik}\mathbf{E}_k\right)^2 \right\} \\ &\propto \exp\left\{ -\frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{X}_i - \sum_{k=1}^{M} p_{ik}\mathbf{E}_k\right)^2 \right\}. \end{aligned} \tag{3–11}$$

Given Equation 3–4, the sparsity promoting term to be added to the ICE objective function should be of the form shown in Equation 3–12,

$$SPT = \sum_{k=1}^{M} \gamma_k \sum_{i=1}^{N} |p_{ik}| = \sum_{k=1}^{M} \gamma_k \sum_{i=1}^{N} p_{ik} \tag{3–12}$$

where the last equality follows due to the constraints in Equation 1–2. For this work, γ_k is set as shown in Equation 3–13,

$$\gamma_k = \frac{\Gamma}{\sum_{i=1}^{N} p_{ik}}. \tag{3–13}$$

Γ is a constant associated with the degree that the proportion values are driven to zero. The advantage of this expression for γ_k is that as the proportion values change during the minimization of the objective function, the weight associated with each endmember adjusts accordingly. If the sum of a particular endmember's proportion values becomes small, then the weight, γ_k, for that endmember becomes larger. This weight change accelerates the minimization of those proportion values. Furthermore, since the objective function is minimized in an iterative fashion, the change in the γ_k values does not disrupt the minimization.

Incorporating this sparsity promoting term into ICE's objective function yields the objective function for SPICE (Zare and Gader, 2007a),

$$RSS^*_{reg} = (1-\mu)\frac{RSS}{N} + \mu V + SPT \tag{3–14}$$

$$= \frac{1-\mu}{N} \sum_{i=1}^{N} \left(\mathbf{X}_i - \sum_{k=1}^{M} p_{ik}\mathbf{E}_k \right)^T \left(\mathbf{X}_i - \sum_{k=1}^{M} p_{ik}\mathbf{E}_k \right) + \mu V + \sum_{k=1}^{M} \gamma_k \sum_{i=1}^{N} p_{ik} \tag{3–15}$$

$$= \frac{1-\mu}{N} \sum_{i=1}^{N} \left[\left(\mathbf{X}_i - \sum_{k=1}^{M} p_{ik}\mathbf{E}_k \right)^T \left(\mathbf{X}_i - \sum_{k=1}^{M} p_{ik}\mathbf{E}_k \right) + \frac{N}{1-\mu} \sum_{k=1}^{M} \gamma_k p_{ik} \right] + \mu V. \tag{3–16}$$

To minimize this new objective function, the iterative procedure in ICE can still be used. The endmembers are still found by solving Equation 3–9 since the SPT term does not depend on the endmembers. When solving for the proportion values given the

endmember estimates, each of the N terms of the sum in Equation 3–17 need to be minimized given the constraints in Equation 1–2 using quadratic programming.

$$RSS^*_{reg,term1} = \frac{1-\mu}{N}\sum_{i=1}^{N}\left[\left(\mathbf{X}_i - \sum_{k=1}^{M} p_{ik}\mathbf{E}_k\right)^T\left(\mathbf{X}_i - \sum_{k=1}^{M} p_{ik}\mathbf{E}_k\right) + \sum_{k=1}^{M}\gamma^*_k p_{ik}\right] \tag{3–17}$$

where

$$\gamma^*_k = \frac{\Gamma^*}{\sum_{i=1}^{N} p_{ik}}, \quad \Gamma^* = \frac{N\Gamma}{1-\mu} \tag{3–18}$$

During the iterative minimization process, endmembers can be removed as their proportion values drop below a threshold. After every iteration of the minimization process, the maximum proportion values for every endmember can be calculated,

$$MAXP_k = max_i \left\{p_{ik}\right\}. \tag{3–19}$$

If the maximum proportion for an endmember drops below a threshold, then the endmember can be removed from the endmember set.

3.4 New Band Selection Algorithm Using Sparsity Promoting Priors

The proposed Band Selecting Sparsity Promoting Iterated Constrained Endmembers (B-SPICE) performs band selection using sparsity promoting priors applied to band weights. This method is developed by extending the SPICE algorithm to perform simultaneous band selection. In order to perform simultaneous band selection, band weights and a band sparsity promoting term are added to the SPICE objective function in Equation 3–14 (Zare and Gader, 2008). Incorporating the band weights and the band sparsity promoting term yields Equation 3–20,

$$J = \eta\frac{RSS_B}{N} + \beta SSD_B + SPT + BST \tag{3–20}$$

where

$$RSS_B = \sum_{i=1}^{N}\left(\mathbf{W}\mathbf{x}_i - \sum_{k=1}^{M} p_{ik}\mathbf{W}\mathbf{e}_k\right)^T\left(\mathbf{W}\mathbf{x}_i - \sum_{k=1}^{M} p_{ik}\mathbf{W}\mathbf{e}_k\right) \tag{3–21}$$

$$SSD_B = \sum_{k=1}^{M-1} \sum_{l=k+1}^{M} \left(\mathbf{We}_k - \mathbf{We}_l\right)^T \left(\mathbf{We}_k - \mathbf{We}_l\right) \tag{3–22}$$

$$\mathbf{W} = diag\left(w_1, \ldots, w_D\right) \tag{3–23}$$

w_i is the weight for the i^{th} band, D is the number of bands, η and β are the constant coefficient parameters for the RSS and SSD terms, and the BST term is the band sparsity promoting term defined in Equation 3–24.

The band sparsity promoting term (BST) is defined as a weighted sum of band weights with one term for each band,

$$BST = \sum_{j=1}^{D} \lambda_j \left|w_j\right| = \sum_{j=1}^{D} \lambda_j w_j \tag{3–24}$$

where

$$\lambda_j = \frac{\Lambda\left(\left(\sum_{k=1}^{M} \frac{1}{N} \sum_{i=1}^{N} p_{ik}\left(x_{ij} - e_{kj}\right)^2\right) + 1\right)}{\left(\sum_{k=1}^{M}\left(e_{kj} - \mu_0\right)^2\right) + 1} \tag{3–25}$$

Λ is a tunable parameter controlling the degree of sparsity among the band weights, μ_0 is the global data mean, x_{ij} is the j^{th} band of the i^{th} pixel, and e_{kj} is the j^{th} band of the k^{th} endmember.

The band weights are subject to the constraints in Equation 3–26,

$$w_j \geq 0, \quad j = 1 \ldots d, \quad \sum_{j=1}^{d} w_j = d \tag{3–26}$$

where d is the number of bands. The non-negativity constraint in Equation 3–26 allows for the second equality in Equation 3–24.

The λ_j values are related to the method of ranking bands according to the Minimum Misclassification Canonical Analysis (MMCA) used by Chang et al. (1999). Chang:1999 rank bands according to the MMCA value which is derived from Fisher's discriminant function. Although the proposed method uses a weight that is related to the Fisher's discriminant value, this algorithm differs from the method used by Chang et al. (1999) by

performing simultaneous endmember detection and using sparsity promoting priors rather than a divergence threshold to determine the number of useful spectral bands.

Note that if a λ_j value is small, then the associated band weights can be large and still yield a small value in the objective function. Conversely, if a λ_j value is large, then the associated weight must be small. Hence, a large λ_j value for a particular band should lead to small weights for that band. The λ_j values are defined to depend on the ratio of the with-in class to between-class scatter. Each endmember has one class that consists of those points with high abundances with respect to the corresponding endmember. So, bands with small ratios separate the data and endmembers well and are therefore encouraged to have large weights. In contrast, bands with large ratios do not separate the data and endmembers well and are encouraged to be removed.

In order to minimize the new objective function in Equation 3–20, the iterative procedure used in SPICE can still be applied. The minimization process iterates between solving for the proportions, endmembers and band weights. The endmembers can be solved for directly as was done in Equation 3–9. When solving for the proportion values given endmember and band weight estimates, N quadratic programming steps, one for each data point, can be employed to minimize Equation 3–20 with respect to the constraints in Equation 1–2.

Similarly, when solving for the band weights given the proportion and endmember estimates, Equation 3–20 can be minimized using a single quadratic programming step given the constraints in Equation 3–26. After updating band weights, bands are removed from data points and endmembers when the corresponding band weight drops below a prescribed threshold.

Since the band weights and endmember values depend on each other, an optimization schedule needs to be employed. An estimate of the endmembers is needed before determining which bands are useful in distinguishing between the endmembers. Therefore, an update schedule allows the endmembers and proportions to settle before determining

band weights. The optimization schedule consists of a starting iteration for band selection, the frequency of band weight updates and a stopping criterion for band weight updates. This iterative schedule is summarized in the following pseudo-code.

BSPICE(X)

$iteration \leftarrow 1$
$n \leftarrow$ iteration frequency of band updates
$StartBandUpdate \leftarrow$ iteration to begin band weight updates
while $(ObjValue - PreviousObjValue)^2 > ChangeThreshold$ **do**
  Update Proportion Values
  Update Endmember Values
  $BandUpdateFlag \leftarrow 0$
  if $(iteration > StartBandUpdate)$ and $(mod_n(iteration) = 0)$ and $(BandUpdateFlag = 0)$ **then**
    Update Band Weights
    Remove Bands
    if $norm(PrevBandWeight - CurrentBandWeight) < BandChangeThreshold$ **then**
      $BandUpdateFlag \leftarrow 1$
    end if
  end if
  Update Objective Function Value, $ObjValue$
  $iteration \leftarrow iteration + 1$
end while

3.5 New Endmember Distribution Detection Algorithm

The new Endmember Distribution (ED) detection algorithm has the unique property of representing endmembers as random vectors, thereby calculating endmembers distributions rather than single spectra. Endmember distributions are found by assuming a model for each endmember and iteratively updating endmember distributions and proportion vectors for each pixel. ED was developed for use within the Piece-wise Convex Endmember detection algorithm in Section 3.8. However, since ED incorporates spectral variability when performing spectral unmixing and endmember determination, applications for ED may extend beyond use within the PCE algorithm.

Assuming the convex geometry model in Equation 1–1, each input hyperspectral pixel is a linear combination of the endmembers. In the following, all endmember distributions

are assumed to be Gaussian distributions with mean spectra, $\mathbf{e}_k$ and known covariance matrices, $\mathbf{V}_k$. It follows that each pixel is a multivariate Gaussian random variable whose distribution is defined by the linear combination of the endmembers' Gaussian distributions,

$$P(\mathbf{x}_j|\mathbf{E},\mathbf{p}_j) \propto \exp\left\{-\frac{1}{2}\left(\mathbf{x}_j - \sum_{k=1}^{M} p_{jk}\mathbf{e}_k\right)^T \left(\sum_{k=1}^{M} p_{jk}^2\mathbf{V}_k\right)^{-1} \left(\mathbf{x}_j - \sum_{k=1}^{M} p_{jk}\mathbf{e}_k\right)\right\} \quad (3\text{–}27)$$

where $\mathbf{e}_k$ and $\mathbf{V}_k$ are the mean spectrum and covariance matrix for the k^{th} endmember distribution, M is the number of endmember distributions being determined, and p_{jk} is the j^{th} data point's proportion value for the k^{th} endmember (Wackerly et al., 1996). The joint likelihood for all the hyperspectral pixels is assumed to be the product of the individual likelihoods,

$$P(\mathbf{X}|\mathbf{E},\mathbf{P}) \propto \prod_{j=1}^{N} \exp\left\{-\frac{1}{2}\left(\mathbf{x}_j - \sum_{k=1}^{M} p_{jk}\mathbf{e}_k\right)^T \left(\sum_{k=1}^{M} p_{jk}^2\mathbf{V}_k\right)^{-1} \left(\mathbf{x}_j - \sum_{k=1}^{M} p_{jk}\mathbf{e}_k\right)\right\}. \quad (3\text{–}28)$$

Each hyperspectral data point has a unique abundance vector. Although all the data points share the same set of endmember distributions, their unique abundance vectors result in each data point having a unique Gaussian distribution. In Equation 3–28, the maximum likelihood value of the data point $\mathbf{x}_j$ is $\mathbf{p}_j\mathbf{E}$.

In order to provide a tight fit around the input hyperspectral data set, the prior on the endmembers is defined using the sum of squared distances between the means of the endmember distributions. This is similar to the prior on the endmembers used by SPICE algorithm.

$$P(\mathbf{E}) = \frac{1}{(2\pi)^{\frac{D}{2}}\,|\mathbf{S}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{4}\sum_{k=1}^{M}\sum_{l=1}^{M}(\mathbf{e}_k - \mathbf{e}_l)^T\mathbf{S}^{-1}(\mathbf{e}_k - \mathbf{e}_l)\right\} \quad (3\text{–}29)$$

Initially, the Dirichlet distribution was considered for the prior on the abundance values. However, since the Dirichlet distribution is not a conjugate prior to $P(\mathbf{X}|\mathbf{E},\mathbf{P})$, a simple update formula cannot be used. Instead, constrained non-linear optimization is required when updating abundance values. As abundances approach zero (which

is very desirable and common), the log of the Dirichlet distribution is very steep and approaches $-\infty$ causing instability when using non-linear optimization techniques. Therefore, Equation 3–30 was developed for the prior on the abundance vectors.

$$P(\mathbf{p}_j) = \frac{1}{Z}\left(\sum_{k=1}^{M} b_k + 1 - \sum_{k=1}^{M} b_k (p_{jk} - c_k)^2\right) \tag{3–30}$$

where Z is a normalization constant given by

$$Z = \frac{\sqrt{M}\left(\sum_{k=1}^{M} b_k + 1\right)}{(M-1)!} - \sqrt{M}\sum_{k=1}^{M} \frac{b_k}{(M-1)!}\left(\left(c_k - \frac{1}{M}\right)^2 + \frac{M-1}{(M+1)M^2}\right).$$

The $\mathbf{p}$ and $\mathbf{c}$ vectors are constrained to be non-negative and sum-to-one,

$$p_{jk} \geq 0 \quad \forall k = 1, \ldots, M; \quad \sum_{k=1}^{M} p_{jk} = 1,$$

$$c_k \geq 0 \quad \forall k = 1, \ldots, M; \quad \sum_{k=1}^{M} c_k = 1.$$

The vector $\mathbf{c}$ is the maximum likelihood value for $\mathbf{p}$. The b_k terms control the steepness of the prior. This abundance prior prefers abundance vectors which are binary; that is, vectors with a single abundance with value 1 and the rest with value 0. This is a result of the normalization constant, Z.

The numerator of the abundance prior is maximized when $\mathbf{c}$ is equal to $\mathbf{p}$. The normalization constant in the denominator is minimized when $\mathbf{c}$ is binary. Thus, when both the $\mathbf{p}$ and $\mathbf{c}$ vectors are binary, the abundance prior is maximized. This property introduces sparsity within abundance vectors which, when combined with the flexibility achieved by representing endmembers by distributions, represents a major advance in automated determination of meaningful endmembers and abundances.

If several endmembers adequately describe a data point, the abundance prior will place all weight on one endmember rather than spreading the abundance across endmembers encouraging the method to use the minimum number of endmembers needed. Furthermore, many different points can be assigned abundance values of one with respect

to a given endmember because of the variance of the endmember distribution. Examples of this prior for abundance vectors of length two are shown in Figure 3-1. Also, plots showing the abundance prior as a function of $\mathbf{c}$ are shown in Figure 3-2.

The algorithm proceeds by iteratively maximizing $P(\mathbf{X}|\mathbf{E},\mathbf{P})P(\mathbf{E})P(\mathbf{P})$ where $P(\mathbf{P})$ is the joint likelihood of all the abundance vectors. Given initial estimates of the endmember distributions and $\mathbf{c}$ from the abundance prior, abundance vectors are updated by maximizing the log of the product of Equations 3–28 and 3–30 with respect to $\mathbf{P}$. This is a constrained non-linear optimization problem. In the current Matlab implementation, this is maximized using Matlab's fmincon function in the optimization toolbox. Following an update of the abundance vectors, the product of Equations 3–28 and 3–29 are maximized with respect to means of the endmember distributions, $\mathbf{e}_k$ for $k = 1, \ldots, M$. This maximization is performed directly by taking the derivative of the log of the product and setting it equal to zero.

$$\begin{aligned}\mathbf{e}_k^T = &\left(\sum_{j=1}^{N}\left(\mathbf{x}_j - \sum_{l\neq k} a_{jl}\mathbf{e}_l\right)^T\left(\sum_{l\neq k} a_{jl}^2\mathbf{V}_l\right)^{-1} a_{jk} + \sum_{l\neq k}\mathbf{e}_l^T\mathbf{S}^{-1}\right) \\ &\left(\sum_{j=1}^{N} a_{jk}^2\left(\sum_{l\neq k} a_{jl}^2\mathbf{V}_l\right)^{-1} + (M-1)\mathbf{S}^{-1}\right)^{-1}.\end{aligned} \tag{3–31}$$

The third step of the iteration updates the $\mathbf{c}$ vector in the abundance prior given the abundance vectors for all the data points. The third step is also a non-linear optimization problem solved using Matlab's fmincon function.

Although the ED algorithm was developed for use within the PCE algorithm, applications of the ED algorithm may extend beyond this. This may occur since, using endmember distributions, the spectral variation which occurs due varying environmental conditions or inherent variability can be measured in controlled environments and then incorporated and utilized during endmember detection or spectral unmixing. For example, if endmember means and covariances are estimated from a spectral library, these can be

held constant during the ED algorithm while spectral unmixing is performed. Additional endmember distributions are learned if necessary.

The use of the endmember distribution model can represent a wide variety of data. For example, the data points in Figure 3-3 were generated using two endmember distributions. The standard model using convex combinations of single endmember spectra would require three endmembers to represent the data while maintaining a small reconstruction error.

3.6 Review of Markov Chain Monte Carlo Sampling Algorithms

The presented Piece-Wise Convex Endmember (PCE) detection algorithm uses the Dirichlet Process to sample the number convex regions needed to describe a data set. Before developing the new algorithm, a review of MCMC sampling methods is provided.

Markov Chain Monte Carlo (MCMC) sampling methods provide a means of generating samples from complicated *target distributions* without needing to enumerate every possible outcome and its probability (Chib and Greenberg, 1995; MacKay, 2003). Samples are produced in a sequence where each new sample is generated based on the previous one using a *transition kernel*. The transition kernel defines the conditional probability of moving to a particular sample (or any subset of samples) given the current sample value (Chib and Greenberg, 1995). Since each sample in the sequence is produced based on the previous one, consecutive samples generated using MCMC methods are not independent (MacKay, 2003).

One MCMC sampling method is the Metropolis-Hastings algorithm which uses a normalized *candidate-generating density* to provide potential samples (Chib and Greenberg, 1995). These candidate samples are then evaluated using an *acceptance ratio* which defines the probability of retaining or rejecting the candidate sample (Chib and Greenberg, 1995). The Metropolis-Hastings algorithm can be used when it is difficult to generate samples directly from the target distribution but samples can be easily evaluated in the distribution (MacKay, 2003).

The Metropolis-Hastings sampling method is initialized with an arbitrary starting point, s_0. Then, a candidate, c_1, is generated from the candidate-generating distribution, $q(s_0, \cdot)$. Given s_0 and c_1, the acceptance ratio is computed according to Equation 3–32 (Chib and Greenberg, 1995),

$$a(s_{i-1}, c_i) = \begin{cases} \min\left[\frac{f(c_i)q(c_i, s_{i-1})}{f(s_{i-1})q(s_{i-1}, c_i)}, 1\right] & \text{if } f(s_{i-1})q(s_{i-1}, c_i) > 0 \\ 1 & \text{otherwise} \end{cases} . \tag{3–32}$$

where q is the candidate-generating distribution which can rely on a previous sample and f is the target density from which the samples are desired (Chib and Greenberg, 1995).

The candidate sample is accepted with probability $a(s_{i-1}, c_i)$. If the sample is rejected, then $s_1 = s_0$ otherwise $s_1 = c_1$. Samples are generated in this sequential manner for a large number of iterations. In Metropolis-Hastings and all MCMC methods, samples generated during an initial period of running the algorithm are discarded. These samples generated during the *burn-in period* are discarded since convergence to the desired target distribution has not yet been reached and a bias based on the arbitrary starting point is present (Casella and George, 1992; Chib and Greenberg, 1995).

The number of samples that need to be discarded is difficult to determine. One technique to determine the length of the burn-in period is described by Chib and Greenberg (1995). This technique uses several Metropolis-Hastings generated sample sequences with varying initialization points. As samples are collected in each sequence, variances across samples are compared between the chains. The technique of using several chains is also described by Casella and George (1992) to generate independent samples from an MCMC method. Sequences with distinct starting points are generated for large number of iterations. The final samples from each chain are then used as independent and identically distributed (iid) samples from the target distribution (Casella and George, 1992). Another technique uses a single sequence and returns every kth position in the

sequence. When k is "large enough," each retained sample can be regarded as an iid samples from the target density (Casella and George, 1992).

The popular Gibbs sampler is a significant special case of the Metropolis-Hastings algorithm (Casella and George, 1992; Chib and Greenberg, 1995). The Gibbs sampler produces samples from a multi-variate distribution by iteratively sampling from the conditional distribution of each variable given all the others (Casella and George, 1992). In this case, the candidate-generating distribution is the conditional distribution for the variable being sampled (Chib and Greenberg, 1995; MacKay, 2003). It can be shown that using the conditional distributions for producing candidate samples causes the acceptance ratio for every transition to be 1 (Chib and Greenberg, 1995).

Consider the multi-variate joint density of the random variables R, S and T, $f(r, s, t)$. The Gibbs sequence, $r_0, s_0, t_0, r_1, s_1, t_1, \ldots$ is generated by iterating between the conditionals in Equation 3–33 given initial values for r_0 and s_0 (Casella and George, 1992),

$$\begin{aligned} t_j &\sim f(t|R = r_j, S = s_j) \\ r_{j+1} &\sim f(r|S = s_j, T = t_j) \\ s_{j+1} &\sim f(s|R = r_{j+1}, T = t_j). \end{aligned} \tag{3–33}$$

Furthermore, the acceptance ratio for this example can be computed and shown to be equal to 1,

$$\frac{f(r_j, s_j, t_j)q((r_j, s_j, t_j), (r_j, s_j, t_{j-1}))}{f(r_j, s_j, t_{j-1})q((r_j, s_j, t_{j-1}), (r_j, s_j, t_j))} = \frac{f(t_j|r_j, s_j)f(r_j, s_j)}{f(t_{j-1}|r_j, s_j)f(r_j, s_j)} \frac{f(t_{j-1}|r_j, s_j)}{f(t_j|r_j, s_j)} = 1. \tag{3–34}$$

3.7 Review of the Dirichlet Distribution and the Dirichlet Process

Markov Chain Monte Carlo (MCMC) sampling techniques have been applied with the Dirichlet Process to clustering problems and determining the required number of clusters (Neal, 1991; Teh et al., 2006; Xing et al., 2007). A review of the Dirichlet Process Mixture Model and its application to a data set for clustering is described in this section. First, the definitions for the Dirichlet distribution, the Dirichlet Process, and the Dirichlet Process

Mixture Model are provided. Then, a method of sampling from a Dirichlet Process Mixture Model using Gibbs sampling is developed.

Definition 1. *The Dirichlet distribution with a base distribution,* $\mathbf{m} = \{m_1, m_2, \ldots, m_n\}$*, and a concentration parameter,* α*, on* $\boldsymbol{\pi} = \{\pi_1, \pi_2, \ldots, \pi_n\}$ *is defined to be (Devroye, 1986)*

$$D(\boldsymbol{\pi}; \alpha\mathbf{m}) = \frac{\Gamma(\alpha)}{\prod_{i=1}^{n} \Gamma(\alpha m_i)} \prod_{i=1}^{n} \pi_i^{\alpha m_i - 1} \tag{3–35}$$

where $\sum_{i=1}^{n} \pi_i = 1$ *and* $\sum_{i=1}^{n} m_i = 1$.

The mean and covariance of the Dirichlet distribution are given by Equations 3–36, 3–37 and 3–38,

$$E[\pi_i] = \frac{\alpha m_i}{\sum_{j=1}^{n} \alpha m_i} = m_i \tag{3–36}$$

$$V[\pi_i] = \frac{m_i(1 - m_i)}{1 + \alpha} \tag{3–37}$$

$$C[\pi_i, \pi_j] = \frac{-m_i m_j}{1 + \alpha}. \tag{3–38}$$

By examining Equation 3–36, it can be seen that as α, the concentration parameter, is varied, the mean of the Dirichlet distribution does not change. In contrast, as α is increased, the covariance decreases.

Given the definition of a Dirichlet Distribution, the Dirichlet Process can be defined.

Definition 2. *Given a set* π *with a* σ*-algebra,* $\mathcal{B}$*, let* α *be a positive constant and let* αG_0 *be a finite, non-null, non-negative, finitely additive measure on* $(\pi, \mathcal{B})$*. Then, the random probability measure* G *on* $(\pi, \mathcal{B})$ *is a Dirichlet Process on* $(\pi, \mathcal{B})$ *with parameters* G_0 *and* α *if for every measurable partition of the set,* $(B_1, B_2, \ldots, B_r)$*, the joint distribution of* $(G(B_1), G(B_2), \ldots, G(B_r))$ *is a Dirichlet distribution with parameters* α *and* $(G_0(B_1), G_0(B_2), \ldots, G_0(B_r))$*. Therefore,* $(G(B_1), G(B_2), \ldots, G(B_r)) \sim \mathscr{D}(\alpha G_0)$ *(Antoniak, 1974; Ferguson, 1973; Teh et al., 2004).*

3.7.1 Dirichlet Process Mixture Model

Mixture models are often used to describe data which is distributed according to some set of "underlying mechanisms" where each data point is assumed to be independently generated by only one of these underlying distributions (Neal, 1991). Finite mixture models can be expressed using Equation 3–39,

$$p(\mathbf{x}_i|\boldsymbol{\pi},\boldsymbol{\theta}) = \sum_{k=1}^{M} \pi_k p(\mathbf{x}_i|\boldsymbol{\theta}_k) \tag{3–39}$$

where $\boldsymbol{\pi} = \{\pi_1, \pi_2, \ldots, \pi_M\}$ is the set of mixing proportions for component distributions such that $\sum_{k=1}^{M} \pi_k = 1$ and $\pi_k \geq 0$ and $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots, \theta_M\}$ where $\boldsymbol{\theta}_k$ is a vector of parameters for the k^{th} component distribution for $k = 1, \ldots M$.

The Dirichlet Process Mixture Model extends the basic mixture model by applying a Dirichlet Process prior to the mixing proportions. This extension allows for a countably infinite number of mixture components (Jain and Neal, 2000). Consider N data points, $\{x_1, \ldots, x_N\}$ each of which are assumed to have been independently generated by some distribution $f_i(\cdot, \boldsymbol{\phi}_i)$ where $\boldsymbol{\phi}_i$ is the vector of parameters that defines the process generating observation $\mathbf{x}_i$. Under the Dirichlet Process Mixture Model, $\boldsymbol{\phi}_i$ is generated by some unknown distribution G (West et al., 1994). Then, G is distributed according to the Dirichlet process, $\mathscr{D}(\alpha G_0)$ where G_0 is the base distribution and α is the concentration parameter (Jain and Neal, 2000). Therefore, the complete model can be written as (Jain and Neal, 2000; Neal, 1998)

$$\begin{aligned} \mathbf{x}_i &\sim f(\cdot|\boldsymbol{\phi}_i) \\ \boldsymbol{\phi}_i &\sim G \\ G &\sim \mathscr{D}(\alpha G_0). \end{aligned} \tag{3–40}$$

Under this model, the values $\boldsymbol{\phi}_i$, $i = 1, \ldots, N$, are generated from G are members of a set of $M \leq N$ distinct values denoted as $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M\}$ corresponding to the parameters for each mixture components. More precisely, $\mathbf{X}$ can be partitioned into M subsets, $\mathbf{X} = \mathbf{X}_1 \cup \mathbf{X}_2 \cup \ldots \cup \mathbf{X}_M$ with the property that $\mathbf{x}_i \in \mathbf{X}_j$ if and only if $\boldsymbol{\phi}_i = \boldsymbol{\theta}_j$.

In other words, several data points can be generated from the same mixture component (West et al., 1994).

To simplify the model, G can be integrated out to express the prior of each $\boldsymbol{\phi}_i$ in terms of the base distribution, G_0, and all other parameter sets (Jain and Neal, 2000; Neal, 1998; Rasmussen, 2000; West et al., 1994),

$$\boldsymbol{\phi}_i|\boldsymbol{\phi}_{-i} \sim \frac{1}{\alpha + N - 1} \sum_{j=1, i\neq j}^{N} \delta(\boldsymbol{\phi}_j) + \frac{\alpha}{\alpha + N - 1} G_0 \tag{3–41}$$

where $\boldsymbol{\phi}_{-i}$ is the set of component distributions for all data points other than i, N is the number of data points, $\delta(\boldsymbol{\phi}_i)$ is the distribution over parameters with all weight concentrated at parameter set $\boldsymbol{\phi}_i$, and G_0 is the prior distribution for the component parameters (Neal, 1998; Ranganathan, 2006).

3.7.2 Gibbs Sampling for the Dirichlet Process Mixture Model

As shown by Neal (1998), the likelihood of a data point given component parameters can be combined with the probability of a class label given all other labels in Equation 3–41. Then, the Gibbs sampler can be used to sample indicator variable values and component parameter values. The conditional probabilities for an indicator variable are defined in Equation 3–42,

$$\begin{aligned} P(c_i = c_j \text{ for some } j \neq i|\mathbf{c}_{-i}, \mathbf{x}_i, \theta) &= C\frac{n_{-i,j}}{\alpha + N - 1} f(\mathbf{x}_i|\boldsymbol{\theta}_{c_j}) \\ P(c_i \neq c_j \forall j \neq i|\mathbf{c}_{-i}, \mathbf{x}_i) &= C\frac{\alpha}{\alpha + N - 1} \int f(\mathbf{x}_i|\boldsymbol{\theta}) G_0(\boldsymbol{\theta}) d\boldsymbol{\theta} \end{aligned} \tag{3–42}$$

where C is a normalizing constant computed by Equation 3–43,

$$\frac{1}{\sum_j \left(\frac{n_{-i,j}}{\alpha+N-1} f(\mathbf{x}_i|\boldsymbol{\theta}_{c_j})\right) + \frac{\alpha}{\alpha+N-1} \int f(\mathbf{x}_i|\boldsymbol{\theta}) G_0(\boldsymbol{\theta})}. \tag{3–43}$$

The Markov Chain for the Gibbs sampler using the conditionals in Equation 3–42 consists of all the indicator variables c and all component distribution parameters $\boldsymbol{\theta}$ (Neal, 1998). If G_0 is a conjugate prior to the likelihood distributions (component distributions) $f(\cdot|\boldsymbol{\theta})$, then the integral in Equation 3–42 can be analytically computed (Neal, 1998). Assuming

that the posterior can be integrated, when G_0 is a conjugate prior to the likelihood, integrating over the component parameters in both conditionals in Equation 3–42 requires that only the indicator variables of the observations need to be sampled. In this case, the conditional distributions are expressed as in Equation 3–44 (Jain and Neal, 2000),

$$\begin{aligned} P(c_i = c_j \text{ for some } j \neq i | \mathbf{c}_{-i}, \mathbf{x}_i) &= C \frac{n_{-i,j}}{\alpha + N - 1} \int f(\mathbf{x}_i|\boldsymbol{\theta}) H_{-i,c_j}(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ P(c_i \neq c_j \forall j \neq i | \mathbf{c}_{-i}, \mathbf{x}_i) &= C \frac{\alpha}{\alpha + N - 1} \int f(\mathbf{x}_i|\boldsymbol{\theta}) G_0(\boldsymbol{\theta}) d\boldsymbol{\theta} \end{aligned} \tag{3–44}$$

where C is a normalizing constant and H_{-i,c_j} is the posterior distribution of the component parameters given prior G_0 and current indicator values $\mathbf{c}_{-i}$ (Jain and Neal, 2000). These integrals remove the need to include component parameters in the Markov Chain which significantly reduces the search space for the Gibbs sampler (Jain and Neal, 2000).

For cases in which G_0 is not a conjugate prior to the likelihood functions, techniques have been developed to either estimate the integral values or use sampling techniques to avoid the need to compute the integral values. Some of these techniques are discussed by Neal (1998). One method uses the Metropolis-Hastings algorithm where the candidate distribution for a parameter set, $\boldsymbol{\theta}_i$, is $\frac{1}{N-1+\alpha} \sum_{j \neq i} \delta(\boldsymbol{\theta}_j) + \frac{\alpha}{N-1+\alpha} G_0$ and the acceptance probability is $a(\boldsymbol{\theta}_i^*, \boldsymbol{theta}_i) = \min \left[1, \frac{f(\mathbf{x}_i, \boldsymbol{\theta}_i^*)}{f(\mathbf{x}_i, \boldsymbol{\theta}_i)}\right]$. Another method avoids the need to evaluate the integral by introducing temporary auxiliary variables into a Gibbs sampling scheme. In this method, the temporary auxiliary variables are parameter sets drawn independently from G_0 (Neal, 1998).

Specific cases of this Gibbs algorithm for the Dirichlet Process Mixture Model are derived in the literature. Rasmussen (2000) derives the algorithm where the component distributions and priors are all Gaussian. Neal (1991) derives the method for categorical data using a Bernoulli distributions for the component distributions and a Beta distributions for the priors.

As described by Rasmussen (2000) and West et al. (1994), this model can also be extended by adding hyper-priors for the α parameter and the parameters of the prior distribution G_0.

3.8 New Piece-Wise Convex Endmember Detection Algorithm using the Dirichlet Process

In this section, a novel method for endmember detection using the Dirichlet process is presented. Existing endmember detection algorithms generally assume that all pixels in a hyperspectral image are convex combinations of a single set of endmembers. However, some hyperspectral images may be better represented using several sets of endmembers. The new algorithm partitions the input hyperspectral data set into convex regions each with a its own set of endmember distributions. Using the Dirichlet process, the Piece-wise Convex Endmember (PCE) detection algorithm learns the number of convex regions needed to represent an input hyperspectral image while simultaneously learning endmember distributions and proportion values for each partition.

This method differs from the Dirichlet process mixture model since each convex region is represented with a set of endmember distributions for which each data point has a unique abundance vector. Thus, as previously shown in Equation 3–27, each data point is a random variable with a unique distribution. Each data point having a unique distribution contrasts with the DPMM approach where data points from each cluster are assumed to be identically distributed.

The PCE algorithm performs Gibbs sampling with Dirichlet process priors to sample the partition to which each data point belongs. The probability of sampling a partition is computed using the likelihood of a data point belonging to a convex combination of the

associated endmember distributions,

$$
\begin{aligned}
P(r_i = r_j \quad j \neq i | \mathbf{r}_{-i}, \mathbf{x}_i) &= C \frac{n_{-i,j}}{\alpha + N - 1} \int f(\mathbf{x}_i | \mathbf{p}_i^{r_j}, \mathbf{V}^{r_j}, \mathbf{E}^{r_j}) H_{-i,r_j}(\mathbf{E}^{r_j}, \mathbf{V}^{r_j}, \mathbf{P}^{r_j}) d\mathbf{p}_i^{r_j} \mathbf{E}^{r_j} \\
&= \mathscr{N}\left(T(T+S)^{-1}\mathbf{p}\mathbf{E}^{r_j} + S(T+S)^{-1}\mathbf{c}\mathbf{E}^{r_j}, S + T\left(T+S\right)^{-1} S\right) \\
P(r_i \neq r_j \quad \forall j \neq i | \mathbf{r}_{-i}, \mathbf{x}_i) &= C \frac{\alpha}{\alpha + N - 1} \int f(\mathbf{x}_i | \mathbf{E}^*) G_0(\mathbf{E}^*) d\mathbf{E}^* \\
&= \mathscr{N}\left(\mathbf{V}_0(\mathbf{V}_0 + \mathbf{V})^{-1}\mathbf{x}_i + \mathbf{V}(\mathbf{V}_0 + \mathbf{V})^{-1}\mu_0, \left(\mathbf{V}_0^{-1} + \mathbf{V}^{-1}\right)^{-1} + \mathbf{V}\right)
\end{aligned}
\tag{3–45}
$$

where r_i is the indicator variable for the current data point, $\mathbf{x}_i$, C is a normalization constant, $n_{-i,j}$ is the number of data points excluding $\mathbf{x}_i$ in partition r_j, N is the total number of data points, and α is the innovation parameter for the Dirichlet process. The matrices, T and S, correspond to $\sum_k c_k^2 V_k$ and $\sum_k p_{ik}^2 V_k$, respectively. The matrices $\mathbf{V}$ and $\mathbf{V}^{r_j}$ are all the covariance matrices associated with new and existing endmember distributions. In the current implementation of this algorithm, all covariance matrices for endmember distributions are set to the same constant matrix value.

The prior distribution, G_0, is Gaussian where the mean, μ_0, is set to the mean of the input data set and the covariance, $\mathbf{V}_0$, is constant,

$$
G_0 = \mathscr{N}\left(\mu_0 = \frac{1}{N}\sum_{j=1}^{N} \mathbf{x}_j, \mathbf{V}_0\right) \tag{3–46}
$$

The prior distribution combined with α, the innovation parameter in the Dirichlet process prior, dictates the probability of generating a new partition. The covariance matrix, $\mathbf{V}_0$, is set to a large value to approximate a broad uniform prior over the data set.

Assuming that each endmember distribution is Gaussian with a known covariance matrix, the likelihood for an existing partition, $f(\mathbf{x}_i | \mathbf{p}_i^{r_j}, \mathbf{V}^{r_j}, \mathbf{E}^{r_j})$, is determined by Equation 3–27. The vector, $\mathbf{p}_i^{r_j}$, contains the proportion values for the current data point in partition r_j. These proportion values are determined by maximizing the product of Equations 3–30 and 3–27 given the endmembers of the partition, $\mathbf{E}^{r_j}$. The likelihood

value, f, measures the ability of a set of endmembers to represent a data point by computing the distance between the data point and **pE**.

The distribution, $H_{-i,r}(\mathbf{E}^r, \mathbf{V}^r, \mathbf{P}^r)$ is the prior distribution updated based on the data points assigned to the r^{th} partition,

$$H_{-i,r}(\mathbf{E}^r, \mathbf{P}^r) = \mathcal{N}\left(\mathbf{cE}^r, \sum_{k=1}^{M} c_k^2 \mathbf{V}_k^r\right) \tag{3–47}$$

where **c** is the center from abundance prior determined by maximizing Equation 3–30 given $\mathbf{P}^r$ and $\mathbf{V}^r$. By incorporating this updated prior, the likelihood depends not only on the distance to **pE** but also to **cE**. When the covariance matrices for endmember distributions are equal, the updated prior depends on the distance to a point on the line segment connecting **pE** and **cE**, namely, $w_1\mathbf{pE} + w_2\mathbf{cE}$ where $w_1 = \frac{\sum_{k=1}^{M} c_k^2}{\sum_{k=1}^{M}(c_k^2+p_{jk}^2)}$ and $w_2 = \frac{\sum_{k=1}^{M} p_{jk}^2}{\sum_{k=1}^{M}(c_k^2+p_{jk}^2)}$.

As stated above and shown in line 12 of the following pseudo-code, in each iteration of the algorithm a partition is sampled for the current data point. In this step, a partition is sampled by computing the likelihood of a data point belonging to each existing partition and the likelihood of a data point generating a new partition using Equation 3–45. The unit interval is then divided into regions whose lengths are equal to each partition's normalized likelihood value. A random value from the unit interval is then generated. The corresponding partition whose region includes the generated random value is the partition which is sampled for the current data point.

After a partition is sampled, the parameters of the sampled partition are updated. This is done by updating the prior on the abundances, Equation 3–30, with respect to **c** for the given partition. After one or more iterations of the partition sampling scheme using the Dirichlet process, the endmember distributions and all proportion values are updated using a designated number of iterations of the ED algorithm.

Several items in the following PCE pseudo-code differ from the standard DPMM method. As stated in lines 10 and 13 of the pseudo-code, in PCE, a partition's parameters

are updated when a data point is removed or added to the partition by updating the partition's **c** vector in the abundance prior. In contrast, for the standard Gaussian DPMM method, the mean of the Gaussian cluster would be updated instead. Lines 16 to 18 of the pseudo-code also differ from the standard DPMM method. After a set number of Gibbs sampling iterations in PCE, each partition's endmembers and proportion matrices are updated. In the standard DPMM, all values associated with each cluster are updated in each Gibbs iteration. PCE essentially is performs a series of several Gibbs sampling runs each with a new set of endmembers.

PCE(X)
1: Initialize Partitions
2: **for** $r \leftarrow 1$ to $R_{initial}$ partitions **do**
3: Initialize $\mathbf{E}^r$ and $\mathbf{P}^r$ using ED
4: **end for**
5: **for** $k \leftarrow 1$ to number of total iterations **do**
6: **for** $i \leftarrow 1$ to number of Gibbs sampling iterations **do**
7: Randomly reorder data points in X
8: **for** $j \leftarrow 1$ to number of data points **do**
9: Remove $\mathbf{x}_j$ from its current partition
10: Update the partition's **c**
11: Compute Dirichlet process partition probabilities for $\mathbf{x}_j$ using Equation 3–45.
12: Sample a partition for $\mathbf{x}_j$ based on the Dirichlet process partition probabilities
13: Update new partition's **c**
14: **end for**
15: **end for**
16: **for** $r \leftarrow 1$ to R_k partitions **do**
17: Update $\mathbf{E}^r$ and $\mathbf{P}^r$ using ED
18: **end for**
19: **end for**
20: $R_{final} = R_k$

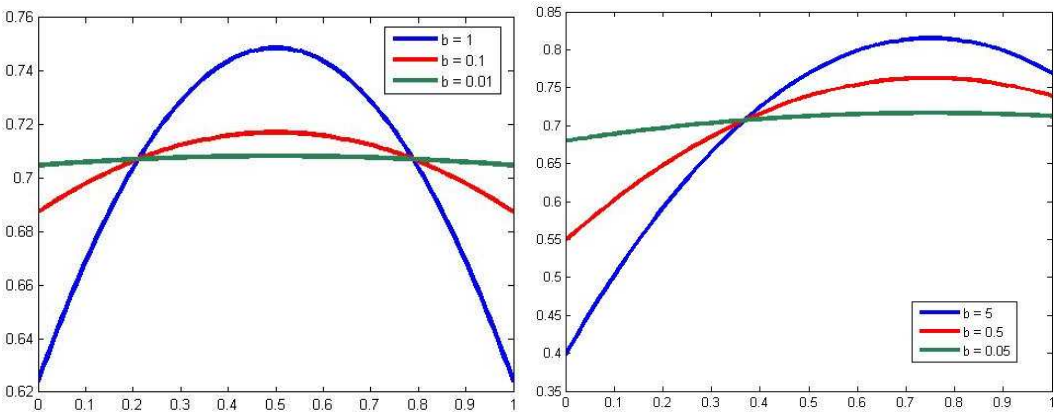


Figure 3-1. Plots of ED's abundance prior for $M = 2$ and various $\mathbf{c}$ and $\mathbf{b}$ values. The x-axis is the 1^{st} abundance value. The y-axis is the prior probability value for the abundance vector. A) $\mathbf{c} = [.5, .5]$ B) $\mathbf{c} = [.75, .25]$.

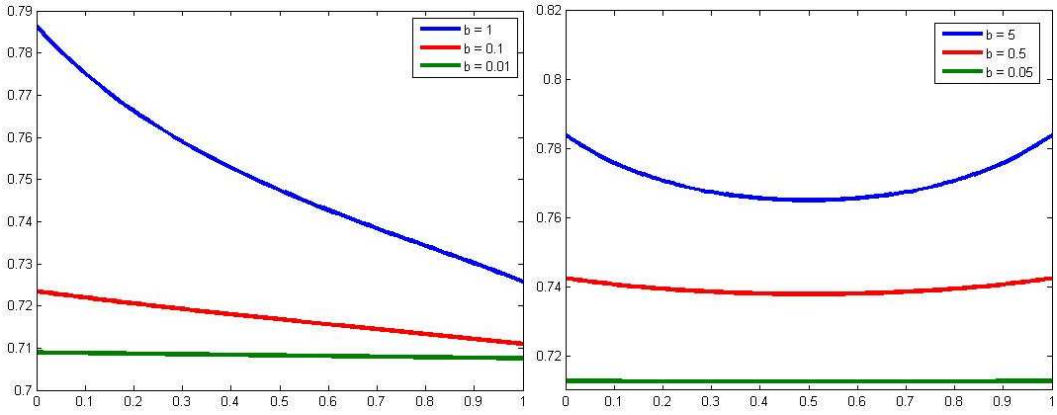


Figure 3-2. Plots of ED's abundance prior for $M = 2$ and various $\mathbf{p}$ and $\mathbf{b}$ values. The x-axis is the 1^{st} $\mathbf{c}$ value. The y-axis is the prior probability value for $\mathbf{c}$. A) $\mathbf{p} = [.45, .55]$ B) $\mathbf{p} = [.5, .5]$.

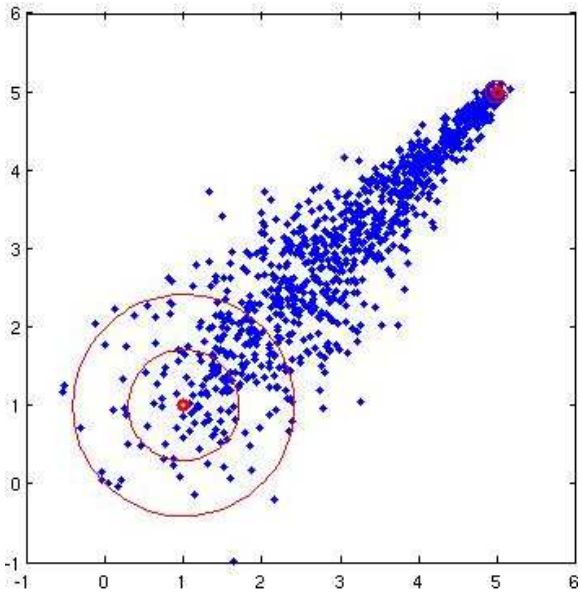


Figure 3-3. Data points generated from linear combinations of 2 endmember distributions. The endmember distribution centered at (5,5) has a diagonal covariance whose elements are all equal to 0.005. The endmember distribution centered at (1,1) has a diagonal covariance whose elements are all equal to 0.5. Data points are shown in blue. Mean spectra and standard deviation curves for the endmember distributions are shown in red.

CHAPTER 4
RESULTS

The presented algorithms were applied to a variety of data sets including two-dimensional data, simulated data, and hyperspectral imagery collected using AVIRIS, the Airborne Visible/Infrared Image Spectrometer, and AHI, the Airborne Hyperspectral Imager. Both the AVIRIS Cuprite and Indian Pines data sets were used.

4.1 Sparsity Promoting Iterated Constrained Endmember (SPICE) Algorithm Results

The SPICE algorithm has been applied to a variety of simulated and real hyperspectral data sets. SPICE results are shown and discussed in Sections 4.1.1 to 4.1.4.

4.1.1 The SPICE Two-Dimensional Example Results

A two-dimensional example was initially used for testing the SPICE algorithm. Figure 4-1 shows the data set and the endmembers from which the data were generated. These data points were generated in the same fashion as the two-dimensional example shown by Berman et al. (2004). The endmembers that were used to generate the 100 data points were $(-10\sqrt{2}, 0), (10\sqrt{2}, 0)$ and $(0, 20)$. The maximum proportions of the bottom two endmembers were 0.80 and 0.60, respectively. Zero-mean independent Gaussian random noise with a variance of 1 was added to each coordinate of the data points.

When running ICE, Berman et al. (2004) assign the number of endmembers for this example to three. In SPICE, the number of endmembers does not need to be known in advance. Therefore, the algorithm can be initialized with a large number of endmembers. The results of three experiments comparing the ICE and SPICE algorithms are shown in Figure 4-2. The parameters for each algorithm, other than the sparsity-promoting term, were set to be the same during the experiments. The initial number of endmembers for all three runs was 20, and μ was set to 0.001. The γ parameter for the SPICE algorithm was set to 10, 20, and 5 for the three runs, respectively. ICE and SPICE were initialized to the same endmembers for each experiment. These initial endmembers were chosen randomly from the data set.

An endmember was pruned from either algorithm when the endmember's maximum proportion over the data points dropped below 0.0005. In these three experiments, proportion values were averaged over the iterations in which an endmember was pruned.

$$MINMAXP_k = \min_k \left\{ \max_i \{p_{ik}\} \right\} \tag{4–1}$$

These were found to be 3.3×10^{-4}, 2.4×10^{-4}, and 2.4×10^{-4} for ICE, respectively. In comparison, 4.1×10^{-6}, 8.3×10^{-17}, and 7.8×10^{-17} are these mean values for SPICE in the three experiments, respectively. As shown, these values are significantly lower in SPICE compared to the pruning threshold than the values in ICE. SPICE consistently drives proportion values for unnecessary endmembers well below a 0.0005 pruning threshold. Despite this high pruning threshold, ICE did not find the correct number of endmembers with pruning without the use of a sparsity-promoting term.

As shown by the results, SPICE consistently determined that three endmembers was an appropriate number of endmembers to represent the data set. ICE ended the algorithm with six endmembers. In Figure 4-2D, two of the endmembers that are found by ICE were (-3.62, 7.94) and (-3.68, 7.94); they appear as one endmember in the figure.

SPICE was also applied to the two-dimensional data used to the test the Morphological Associative Memory method in Section 2.1.4. The SPICE results on this data set are shown in Figure 4-3. For the results shown, Γ was set to 20 and μ was set to 0.01.

4.1.2 The SPICE AVIRIS Cuprite Data Results

SPICE was also tested on real hyperspectral imagery. The AVIRIS hyperspectral image data from Cuprite, NV was used. This data contained 51 contiguous spectral bands (in the range of 1978 - 2477 nm) from "Scene 4" of the AVIRIS Cuprite data set (AVIRIS). This data set was chosen to be able to compare the results with the NFINDR results presented by Winter (1999).

Following Berman et al. (2004), SPICE was run on a subset of pixels from the image to reduce computational time. These *candidate points* were selected using the

pixel purity index (PPI) method (Berman et al., 2004; Boardman et al., 1995). In these experiments, the candidate points were chosen from 10,000 random projections. Points within a distance of two from the boundary of the projection received increased purity indices. The 1011 pixels with the highest PPI were used as the candidate points. 1000 pixels were used by Berman et al. (2004) during the experiments on the real image sets. A PPI threshold that produced as close to 1000 pixels as possible (many pixels had the same PPI) was chosen. Also, fast implementations for the algorithm can be created as was done by Berman et al. (2004) to avoid the need to select a subset of the pixels.

The spectral profiles of the nine endmembers that were found by SPICE to represent this image are shown in Figure 4-4. The three endmembers in the Figures 4-4C, 4-4G, and 4-4I compare well to three endmembers that are found and identified as kaolinite, alunite, and calcite by Winter (1999), respectively. Figure 4-5 shows a comparison of 4-4I to the U.S. Geological Survey (USGS) spectral library data on alunite (Clark et al., 2004).

Although it is clear that SPICE was able to find some of the same endmembers that are identified by Winter (1999), it is not clear if the correct number of endmembers was found. The difficulty of using real image data is that the correct number of endmembers in the scene is unknown. To overcome this problem, a subset of the Cuprite data was used for further testing of the method.

Three endmembers, shown in Figure 4-6, were selected from the hyperspectral image by hand. The squared Euclidean distance was calculated from every pixel in the image to these three endmembers. The pixels within 500,000 squared Euclidean distance from these three hand-selected endmembers were collected and used as a test set for SPICE. The test set was normalized and is shown in Figure 4-7.

Table 4-1 shows the number of endmembers that are found using SPICE for a range of Γ's and initial number of endmembers. As shown, SPICE consistently finds three endmembers for this data set. The results in Table 4-1 and in Figure 4-2 show that the SPICE algorithm is fairly stable with respect to Γ. SPICE is also very stable with respect

to the initial number of endmembers. Therefore, the initial number of endmembers should be set to a large value.

Figure 4-8 shows the endmembers that are found using SPICE in these experiments. These endmembers are clearly very similar to the three hand-selected endmembers used for this experiment.+

4.1.3 The SPICE AVIRIS Indian Pines Results

SPICE was also run on the June 1992 AVIRIS data set collected over the Indian Pines Test site in an agricultural area of northern Indiana. The image has 145×145 pixels with 220 spectral bands and contains approximately two-thirds agricultural land and one-third forest and other elements. The soybean and corn crops in the image are in early growth stages and, thus, have only about a 5% crop cover (Grana and Gallego, 2003; Serpico and Bruzzone, 2001). The remaining field area is soil covered with residue from the previous crop. The no till, min till, and clean till labels indicate the amount of previous crop residue remaining. No till corresponds to a large amount of residue, min till has a moderate amount, and clean till has a minimal amount of residue (Serpico and Bruzzone, 2001). Figure 4-9 shows band 10 (approximately 0.49 μm) and the ground truth of the data set. Only 49% of the pixels in the image have ground truth information (Serpico and Bruzzone, 2001).

SPICE was run on a subset of the image pixels. 1100 pixels were randomly selected from the image. Before running SPICE, these pixels were normalized. The initial number of endmembers, μ, and Γ were set to 60, 0.1, and 1, respectively. Ten endmembers were found for this data set using SPICE. The resulting abundance maps are shown in Figure 4-10. SPICE pruned unnecessary endmembers and provided interpretable results that compare well to previously published results on this data set (Grana and Gallego, 2003; Grana et al., 2003; Miao et al., 2006).

In Figure 4-10, the images were found to roughly correspond to the following: (A) and (I) are woods and tree canopies; (B), (C), and (J) are a mixture of soybean and corn

crops; (D) and (E) are grass and background; (F) is hay windrows; (G) is steel towers, roads, and other man-made objects; and (H) is grass/pasture and wheat.

Since 49% of the pixels in the scene are unlabeled, SPICE was also tested on only the labeled pixels of the AVIRIS Indian Pines scene. A total of 1037 normalized pixels (every 10^{th} labeled pixel) was selected from the image and used to determine the endmembers. The initial number of endmembers, μ, and Γ, were set to 20, .01 and .1, respectively. Six endmembers were found for the labeled pixels of the Indian Pines scene. The abundance maps are shown in Figure 4-11. The endmembers roughly correspond to the following classes: (A) grass/pasture and woods, (B) hay-windrowed, alfalfa and grass/pasture-mowed, (C) and (E) correspond to corn and soybean, (D) stone-steel towers, and (F) grass/trees, wheat, woods.

Normalized histograms showing the distribution of abundances values among endmembers in each ground truth class are shown in Figure 4-12. These histograms were computed by summing all the abundance values associated with an endmember in each ground truth class. Each histogram was normalized by dividing by the number of points in the corresponding ground truth class,

$$h_{lk} = \frac{\sum_{i:\mathbf{x}_i \in G_k} a_{il}}{N_k} \tag{4–2}$$

where G_k is the set of pixels in ground truth class k, N_k is the number of points in ground truth class k, a_{il} is the i^{th} data points' abundance value for the l^{th} endmember, and h_{lk} is the k^{th} histogram's value corresponding to the l^{th} endmember.

4.1.4 The SPICE AHI Vegetation Detection Results

The SPICE algorithm was run on data collected by AHI, the Airborne Hyperspectral Imager (Lucey et al., 1998; Zare and Gader, 2007b). AHI collects 256 spectral bands of data from the long wave infrared region in the range of 7.88 to 11.49 microns (Lucey et al., 1998). The AHI data set collected is trimmed and binned down to 70 bands over the same wavelengths (Lucey et al., 1998). The data set used for these results was collected from

an arid testing site containing both surface and buried landmines. Fiducial markers are also contained in the imagery for alignment and ground truthing purposes. SPICE was applied to this data to extract vegetation endmembers and create a vegetation mask for the reduction of false alarms during landmine detection (Zare and Gader, 2007b).

Three AHI images were used and scoring was performed to determine the reduction in false alarm rate. Scoring for the results in this paper was carried out over regions of interest in the imagery. The regions of interest for this study were defined as the areas where collected Lynx Synthetic Aperture Radar and AHI imagery intersect (LYNXSAR). Four mine types were distributed in the intersecting regions. Two of the mine types were plastic cased (PC) and two were metal cased (MC). The distribution of mines types in the intersecting regions of the AHI and Lynx images are displayed in Table 4-2 (Zare and Gader, 2007b).

Vegetation mapping in the long wave infrared. Vegetation detection in the LWIR is based on the emissive properties of vegetation. Vegetation behaves similar to a blackbody in the LWIR exhibiting a high mean emissivity and a low standard deviation of emissivity across spectral bands (French et al., 2000). Additionally, skewness of emissivity across spectral bands has been seen to be helpful in distinguishing vegetation (Zare et al., 2008). To exploit this information, the SPICE algorithm can be run on the emissivity spectra calculated from LWIR hyperspectral data. For this study, the emissivity spectrum of each pixel in the image is calculated using the Emissivity Normalization Method (Kealy and Hook, 2000).

After applying the SPICE algorithm to the emissivity spectra, the endmembers determined by the algorithm are examined. The endmember with the highest mean and the lowest standard deviation is determined to be the blackbody endmember,

$$\mathbf{E}_B = \begin{cases} \mathbf{E}_k & \text{if } \arg\max_{\mathbf{E}_k}(\mu_k) = \arg\min_{\mathbf{E}_k}(\sigma_k) \\ \emptyset & \text{otherwise} \end{cases} \tag{4–3}$$

where μ_k is the mean and σ_k is the standard deviation across the spectral bands of the k^{th} endmember, $\mathbf{E}_k$, found by SPICE.

Since the proportion maps generated by the SPICE algorithm represent the amount of a particular endmember in a pixel, the proportion map associated with the blackbody endmember is used as the blackbody map, ν, for the image,

$$\nu_j = \begin{cases} 0 & \text{if } \mathbf{E}_B = \emptyset \\ p_{jB} & \text{otherwise} \end{cases} \tag{4–4}$$

where j corresponds to the j^{th} pixel in the image. A mask $\mathbf{V}$, is defined using the blackbody map by inverting the values and enhancing the map using a local 3 minimum filter,

$$V_j = localmin(1 - \nu_j) \tag{4–5}$$

where j corresponds to the j^{th} pixel in the image. Following the local 3×3 minimum filter, a partial threshold was applied to the mask,

$$V_j^{thresh} = \begin{cases} V_j & \text{if } V_j \leq t \\ 1 & \text{otherwise} \end{cases} \tag{4–6}$$

where t is the threshold determined using Otsu's thresholding method (Otsu, 1979). The partial threshold is applied so that the only values modified by the mask are those that are associated with pixels that behave like a blackbody (Zare and Gader, 2007b).

Following Berman et al. (2004), SPICE was run on a subset of pixels from the image. The subset was selected using the Pixel Purity Index (PPI) algorithm (Boardman et al., 1995). The subset was chosen using 30,000 random projections. Points within a distance of three from the boundary of the projection received increased pixel purity values. A threshold was selected to allow as close to 1000 pixels as possible (many pixels have the same PPI). The number of points selected was 1095, 767 and 1103 for AHI Images 1, 2,

and 3, respectively. In order to compute proportion maps for the entire image, the entire image was unmixed using the endmembers found on the image subsets.

The results were compared to those generated using the vegetation mapping method described by Zare et al. (2008). Since Zare et al. (2008) used only the statistics of emissivity (mean, standard deviation and skewness across spectral bands) instead of the full emissivity curve, the results displayed are those generated by running SPICE on only the statistics of emissivity instead of the full emissivity spectra. This method was used to be able to compare performance of the clustering method by Zare et al. (2008) and the SPICE method directly without adding confusion over whether the difference in performance resulted from the methods or the input data. Furthermore, a partial threshold as defined in Equation 4–6 was also applied to the mask generated by the clustering method by Zare et al. (2008).

In contrast to SPICE, which finds the desired number of endmembers for a data set, the method by Zare et al. (2008) requires the number of clusters to be supplied to the algorithm. The method was run on this data with the number of clusters ranging from three to six. The results displayed are the best results obtained over this range of number of cluster values. Figure 4-14 shows the blackbody mask generated using SPICE and the clustering method for four and five clusters. When comparing the two masks generated by the clustering method, it can be seen that when five clusters is chosen instead of four, many of the vegetation pixels are being split between multiple clusters and, thus, are farther from the selected vegetation cluster center. Since SPICE automatically selects the desired number of endmembers, this difficulty is eliminated. When examining the SPICE mask and the clustering mask generated with 4 clusters, it can be seen that the SPICE mask provides more solid vegetation regions thus providing a better mapping of the vegetation pixels than the clustering method.

Points of Interest, POIs, in the overlap regions of the imagery were found using the RX detector algorithm (Yu et al., 1993). The RX detector applied was an implementation

of the well-known anomaly detection algorithm by Winter (2004). The RX algorithm was applied to detect buried mines in the LWIR hyperspectral imagery. The blackbody mask is incorporated by multiplying the RX confidence of every POI with their corresponding blackbody mask value. This differs from the detection algorithms used by Zare et al. (2008) where the blackbody mask is applied to the output of a Choquet fusion system incorporating several detection algorithms. In these results, only the comparative performance of the two blackbody masks are being examined.

The results in each of the three overlap regions are shown in Tables 4-3, 4-4 and 4-5. The *probability of detection*, PD, is defined as the number of mines with a confidence above the threshold divided by the total number of mines. The *false alarm rate*, FAR, is defined as the number of non-mines above the threshold divided by the number of square meters in the overlap region. Although RX was applied to detect buried mines, the results are shown over all mine types in the overlap regions. If detected, fiducial markers in the scene are considered false alarms.

The first line in each table displays the false alarm rates without any blackbody mask being used on the RX values. The second line displays the FARs after applying the blackbody mask generated using the clustering method. The third line shows the reduction in the FAR after using the blackbody mask from the clustering method. The fourth line displays the FAR after applying the blackbody mask generated using SPICE. Finally, the fifth line shows the reduction in FAR after using the blackbody mask from SPICE when compared to the results without using a blackbody mask (Zare and Gader, 2007b).

The blackbody mask generated using the SPICE algorithm can provide false alarm reduction during landmine detection. In comparison to the clustering method by Zare et al. (2008), the SPICE method provides improved vegetation detection and eliminates the need to set the number of clusters or endmembers needed.

4.2 Band Selecting SPICE (B-SPICE) Algorithm Results

Results on hyperspectral imagery using the B-SPICE algorithm are shown with comparisons to other band selection methods in Sections 4.2.1 to 4.2.3.

4.2.1 The B-SPICE AVIRIS Cuprite Data Results

The B-SPICE algorithm was applied to a simulated data set generated using four normalized endmembers selected from the AVIRIS Cuprite data set (AVIRIS). The chosen endmembers are shown in Figure 4-15. The data set was generated from the endmembers following the convex geometry model in Equation 1–1.

A simulated data set was used to verify that the method can recover the endmembers, perform effective band selection, and produce accurate abundance values for each pixel. These can be tested using simulated data since the true endmembers and abundances are known. B-SPICE and SPICE were run on this data set for a range of Λ values. All parameters, other than those involved with band selection, were held constant for each run of the algorithm. The η parameter was set to 2000, β was 0.3, Γ was 0.3 for the first 150 iterations and then set to 0, the initial number of endmembers was set to 20, and the endmember pruning threshold was 1×10^{-8}. The initial endmembers were selected randomly from the data set. When running B-SPICE, band selection was not started until the 100th iteration, after which, the band weights were updated every fifth iteration. The band pruning threshold was set to 1×10^{-5}, and the band weight change threshold was set to 1×10^{-5}. Λ was set to 0 (for SPICE), 0.25, 0.5, 0.75, and 1 (for B-SPICE). B-SPICE and SPICE were run on the data 50 times for each parameter set. An example of the endmembers found using each Λ value is shown in Figure 4-16.

Table 4-6 shows the mean and standard deviation of the number of endmembers and the number of bands retained for each parameter set over the 50 runs of the algorithm. As can be seen, both SPICE and B-SPICE are able to consistently determine the correct number of endmembers.

To evaluate the effectiveness of the band selection performed, the average squared error per abundance value was calculated. The mean, median and standard deviation of the error values are shown in Table 4-7. The median average squared error per abundance value was computed by taking the median over 50 runs of the algorithm of the average squared error between each pixels true and computed abundance values. As shown, the median average squared error per abundance value is fairly stable across the Λ values, indicating that B-SPICE can be as effective at determining the true abundance values as the SPICE algorithm. Therefore, by using B-SPICE, the number of bands can be reduced while maintaining the ability to determine abundances. However, when examining the standard deviation of the average squared error per abundance value, it is seen that SPICE is more consistent than B-SPICE. There is an order of magnitude difference between the standard deviations of SPICE and B-SPICE.

4.2.2 The B-SPICE AVIRIS Indian Pines Results

The B-SPICE algorithm was also run on the June 1992 AVIRIS Indian Pines data set described in Section 4.1.3. SPICE and B-SPICE were run twice for five different Λ values. All parameters, other than the Λ parameter, were held constant for each run of the algorithm. To reduce run time, SPICE and B-SPICE were run on 1000 pixels randomly chosen from the data set. After determining the endmembers and selected bands using the subset, unmixing was performed on the entire data set to find abundance values for every pixel. The η parameters was set to 5000, β to 0.3, and Γ to 0.2 for the first 100 iterations and then to 0. The initial number of endmembers was set to 20 and the endmember pruning threshold was 1×10^{-8}. Initial endmembers were selected randomly from the data set. When running B-SPICE, band selection was not started until the 100th iteration, after which, the band weights were updated every fifth iteration. The band pruning threshold was set to 1×10^{-5}, and the band weight change threshold was set to 1×10^{-5}. Λ was set to 0 (for SPICE), 0.5, 1, 5, and 10 (for B-SPICE). The number of endmembers and the number of bands found are shown in Table 4-8.

In order to compare these results to those presented by Guo et al. (2006), supervised classification was performed. The features used for supervised classification were the abundance values computed for each pixel in the 16 classes of the data set. The unlabeled pixels were not included in these experiments. Since the abundance values were the features used for classification, the dimensionality of the feature vectors is equal to the number of endmembers found for the data set.

Two-fold cross-validation was performed on the data set using a 1-versus-1 Relevance Vector Machine (RVM) classification method (Tipping, 2001). The training and testing sets were defined by randomly splitting each of the 16 classes in half. An RVM was trained for each pair of classes. Since there are 16 classes, 120 RVMs were trained for each test set. Test pixels were classified by counting the number of RVMs that assigned the pixel to each class.

$$L_p = [v_1^p, \ldots, v_{16}^p] \tag{4–7}$$

where v_i^p is the number of times the pixel p was assigned to class i by the trained RVMs. After every pixel was run through the entire set of trained RVMs, spatial smoothing was performed to assign a label to each pixel. Spatial smoothing was done by summing over the neighborhood of pixel p and assigning the class with the largest number of votes

$$C_p = \underset{i \in \{1,\ldots,16\}}{\arg\max} \left\{ \sum_{j \in N_p} v_i^j \right\} \tag{4–8}$$

where C_p is the label for pixel p and N_p is a set of pixels in the eight-connected neighborhood of pixel p. The overall classification accuracies for each run of the B-SPICE algorithm are shown in Table 4-8. Since the classification accuracies depend on the random splitting of the data into training and testing sets, classification was performed three times for each run of the B-SPICE algorithm.

Wang et al. (2006) provide supervised classification results with band selection on the Indian Pines data set. The results shown by Wang et al. (2006) show very good classification accuracies ranging from 90% to 94.5% with less than 50 bands; however,

their classification method was not described. Band selection results on the Indian Pines data set are also shown by Archibald and Fann (2007) and Huang and He (2005), but the results are provided on only a subset of the labeled classes. Table 4-8 also shows results from Martinez-Uso et al. (2006); only the results with less than 50 bands were provided.

4.2.3 The B-SPICE AVIRIS Indian Pines Results using Sampled Parameter Values

In order to reduce the need to set parameters by hand, parameters can be sampled from prior distributions. This was implemented by sampling η, β, Γ, and Λ from gamma distributions with means of 6000, 0.3, 0.2, and 1.240 parameter value sets that were sampled, respectively. B-SPICE was run on the Indian Pines data set using each of the 240 sets of sampled parameters. The results of 240 runs can be combined to determine the number of endmembers, the number of bands, and the bands to retain for the data set. Figure 4-17 shows the histograms of the number of endmembers, the number of bands, and the number of times each band was retained. Modes of the histograms in Figure 4-17 are 7 and 114, respectively. The most frequently retained bands over the 240 runs were 1-57, 61-76, 81-100, and 118-138. By using these modes and the most frequently retained bands, ICE can be run to find endmembers and abundance values. In other words, the number of endmembers and the bands to retain were determined using the histograms found by running B-SPICE over sampled parameter values. These values were then used to set the number of endmembers and the bands to use for the ICE algorithm.

The classification accuracies using the sampled parameter values were determined using the same classification method done in the previous Indian Pines experiment in Section 4.2.2. Table 4-9 shows two runs of the ICE algorithm and with three runs of 1-versus-1 classification.

4.3 Endmember Distribution (ED) Detection Results

Endmember Distribution detection results are shown on two-dimensional and hyperspectral imagery. Comparisons between results found using SPICE and ED are discussed in Section 4.3.1

4.3.1 Results on Two-Dimensional Data using ED

The ED algorithm was initially tested on the two-dimensional data shown in Figure 4-1. The results found on this data set using the ED algorithm with the parameters values listed in Table 4-10 are shown in Figure 4-18. After running the algorithm, the final **c** vector found for the abundance prior was [.47 .27 .26] where the values correspond to the endmember distributions centered at (-11.9, 1.9), (-0.1, 18.5) and (7.5, 6.9), respectively. As can be seen, ED performed as expected. The endmember distributions surround the data points and compare well to the endmember results found by SPICE in Figure 4-2.

ED was also run on the two-dimensional data shown in Figure 4-19. This data was generated by sampling endmembers from three endmember distributions and computing the data points as convex combinations of the sampled endmembers using randomly generated abundance values. The ED results on this data are shown in Figure 4-20. Parameters used to generate these results are shown in Table 4-10. Again, ED generated the expected results. The endmember distributions that were found are very similar to those used to generate the data.

For comparison, SPICE was also run on the two-dimensional data in Figure 4-19. The SPICE was run on the data set with $\mu = 0.01$, $\gamma = 1$ and 10 initial endmembers. The resulting endmembers are shown in Figure 4-21A. As shown in the figure, SPICE needed four endmembers to represent the data set.

SPICE results found using $\mu = 0.001$, $\gamma = 1$ and 10 initial endmembers are shown in Figure 4-21B. By decreasing μ, less emphasis is placed on the sum-of-squared distances term in the SPICE objective function which may result in SPICE requiring a smaller

number of endmembers. However, in this case, SPICE still required four endmembers to represent the data set.

The SPICE μ parameter was reduced rather than adjusting γ because, by adjusting μ, the residual error incurred by representing data points using endmembers is kept low. A smaller number of endmembers can be found by increasing γ, however, the residual error will increase.

4.3.2 Results on AVIRIS Cuprite data using ED

To examine ED's capabilities on hyperspectral imagery, ED was run on the simulated AVIRIS Cuprite data generated from the endmembers shown in Figure 4-15. The data set was generated based on the convex geometry model in Equation 1–1. The results using ED on this data set are shown in Figure 4-22.

Parameter values used to generate these results are shown in 4-11. As can be seen in Figure 4-22, the means of the endmember distributions match the true endmembers well.

ED was also run on the subset of AVIRIS Cuprite data shown in Figure 4-7. This data set is a compilation of the pixels spectrally similar to three endmembers selected from the AVIRIS Cuprite data. The endmembers are shown in Figure 4-6. Results on this data set found using the ED algorithm are shown in Figure 4-23. As can be seen in the figure, the means of the endmember distributions closely match the true endmembers and the data set. These results superimposed on the input data set are shown in Figure 4-24.

4.4 Piece-wise Convex Endmember (PCE) Detection Results

The PCE algorithm was tested on two-dimensional data and the AVIRIS Indian Pines hyperspectral data set. Results are presented and compared to SPICE algorithm results.

4.4.1 Detection Results on Two-Dimensional Data using PCE

The PCE algorithm was initially tested on two-dimensional data. The data set used is shown in Figure 4-25. This data set was generated from three sets of endmembers, each set forming a triangle of data points.

Results on this data, after running PCE for 100 iterations, are shown in Figure 4-26. Prior to running the algorithm, partitions were initialized using the Kernel Global Fuzzy C-Means (KG-FCM) algorithm and the Dirichlet Process Mixture Model algorithm resulting in 8 partitions (Heo and Gader, 2008). Endmembers and abundances for each partition were then initialized by running the ED algorithm on each cluster. The parameters used to generate these results are shown in Table 4-12. As can be seen, PCE partitioned the data set into the correct number of convex regions. Furthermore, PCE was able to identify an appropriate set of endmembers for each convex region.

4.4.2 Detection Results on the AVIRIS Indian Pines Data using PCE

The PCE algorithm was further tested on the labeled pixels of the AVIRIS Indian Pines hyperspectral data set described in Section 4.1.3. Prior to running the PCE algorithm, the data dimensionality was reduced from 220 bands to 6 dimensions using principal components analysis. A total of 1037 pixels (every 10^{th} labeled pixel) were selected from the data set and used in the PCE algorithm. Partitions on this data were initialized using the KG-FCM algorithm and the DPMM algorithm resulting in 3 partitions. After initial partitions were found, endmembers for each partition were initialized using the ED algorithm. Each partition was restricted to 3 endmembers. The parameters used to generate results shown are listed in Table 4-12.

In order to compute abundance maps for the entire image, every data point was unmixed using each partitions' set of endmembers and the likelihood under each partition was computed. Every data point was then assigned to partition with the largest likelihood value. Also, all endmembers whose maximum proportion value was less than 0.05 were removed. Following these steps, 13 clusters were found with a total of 14 endmembers. Figure 4-27 shows the abundances maps associated with each endmember.

For comparison with the SPICE results in Figure 4-12, normalized histograms showing the distribution of abundance values across each endmember were computed using Equation 4–2. The histograms found are shown in Figure 4-28. When comparing

the SPICE and PCE histograms, the PCE results for each ground truth class are more concentrated than the SPICE results. This fact can be measured by computing Shannon's entropy for the normalized histogram associated with each ground truth class (Bishop, 2006). A smaller entropy value indicates that a fewer number of endmembers are being used to describe each ground truth class and that the endmembers are better representatives of the ground truth classes. The sum of the Shannon entropies for the SPICE histograms comes to 19.0. In contrast, the sum of the Shannon entropies for the PCE histograms is significantly lower at 9.4. This indicates that PCE produces endmembers which better represent the ground truth classes.

The histograms and abundance maps associated with several of the ground truth classes verify that PCE is producing endmembers which provide a better representation of the data than the endmembers produced by SPICE. Some of these ground truth classes are wheat, stone-steel towers, hay-windrowed.

Consider the wheat ground truth class in the SPICE and PCE results. The SPICE abundance map associated with the most amount of wheat is shown in Figure 4-11F and the corresponding histogram is found in Figure 4-12M. By examining the abundance map, it can be seen that many pixels other than wheat have non-zero abundance values associated with wheat's SPICE endmember. In contrast, very few pixels outside of the wheat ground truth class share wheat's endmember. This is shown in the PCE abundance map in Figure 4-27J. Furthermore, by examining the SPICE histogram for wheat, only about 60% of the wheat pixels' abundance values are associated with that endmember whereas 100% of wheat's abundance values are placed with the associated endmember found using PCE.

For the stone-steel towers ground truth class, more than 70% of the pixels assigned to a single endmember using PCE and that endmember is not associated with any other ground truth classes. The SPICE endmember most associated with the stone-steel towers ground truth class is also used by every other ground truth class.

The hay-windrowed (Figure 4-28H), grass/pasture-mowed (Figure 4-28G) and alfalfa (Figure 4-28A) PCE histograms show that they are associated with the same endmember. This can also be seen in the abundance map in Figure 4-27I. The corresponding SPICE histograms for hay-windrowed, grass/pasture-mowed and alfalfa in Figures 4-12H, 4-12G, and 4-12A show that the three ground-truth classes have similar histogram shapes and share the same endmembers. However, the abundances found by SPICE are spread among three endmembers whereas PCE placed their full weight with one endmember.

Soybean and corn ground truth classes constitute a large majority of the Indian Pines scene. In the SPICE results, abundance values associated with the soybean and corn classes are spread over all of the six endmember found. In contrast, the PCE endmember results places nearly all soybean and corn abundances with the 2^{nd}, 6^{th}, and 10^{th} endmembers.

Another indication that PCE is producing representative endmembers is found with the Building/Grass/Trees/Drive ground truth class. This class is composed of a variety of material types. Interestingly, this is clearly shown in the class' PCE histogram (Figure 4-28O). The abundance values for the class are spread across many endmembers.

In order to verify that the difference in the results between PCE and SPICE are not due to different data dimensionality and a different number of endmembers, the ICE algorithm was run on the same AVIRIS PCA-reduced Indian Pines data set discussed in this section. The ICE algorithm was employed rather than SPICE since the number of endmembers can be set to the same number found by PCE. The ICE algorithm was restricted to 14 endmembers and the μ parameter was set to 0.01. The resulting abundance maps are shown in Figure 4-29 and the corresponding histograms are in Figure 4-30.

The sum of the ICE histogram entropies from these results is 29.2. In comparison, PCE's value was 9.4. Therefore, although ICE was restricted to the same number of endmembers found using PCE, ICE did not produce endmembers that represent the

ground truth classes as well as PCE. The comparison between the SPICE histograms in Figures 4-30I, 4-30J, 4-30K, and 4-30L and the PCE histograms in Figures 4-28I, 4-28J, 4-28K, and 4-28L are indicative of this. These histograms correspond to the oats and soybean classes. In the SPICE histograms, the abundance values are spread across all of the endmembers. In contrast, the PCE histograms for these ground truth classes concentrate the abundance values to a few endmembers. The PCE results in this section strongly indicate that the algorithm produces endmembers which correspond very well to the true ground truth classes.

The PCE results on AVIRIS Indian Pines data with hierarchical dimension reduction. The PCE algorithm was run again on the AVIRIS Indian Pines data set. However, rather than reducing dimensionality using PCA, hierarchical dimensionality reduction was used (Martinez-Uso et al., 2006). The data dimensionality was reduced from 220 to 3 dimensions. The hierarchical dimensionality reduction computed the pair-wise KL-divergences between the bands' normalized histograms of intensity values. The KL-divergences were then used to hierarchically group similar bands. The average value across each group of bands was used to form the reduced dimensionality data set.

Partitions were initialized using the KG-FCM algorithm and the DPMM algorithm resulting in 3 clusters. Initial endmembers were found for each partition using the ED algorithm. A total of 1037 pixels (every 10^{th} labeled pixel) were selected from the data set and used in the PCE algorithm. Parameter values used to generate the results shown on this data set are listed in Table 4-12.

In order to compute abundance maps for the entire image, after finding endmembers on the subset of pixels using PCE, every data point was unmixed using each clusters' set of endmembers and the likelihood under each cluster is computed. Every data point was assigned to cluster with the largest likelihood value. Furthermore, partitions with less than 5 assigned pixels were pruned. Following these steps, 2 clusters were found with a total of 6 endmembers. Figure 4-31 shows the abundances maps associated with each endmember.

Abundances maps from this experiment are shown in Figure 4-31 and histograms are shown in Figure 4-32. The first partition found using PCE on the three-dimensionality data corresponded to the majority of the corn and soybean ground truth classes. Hay and alfalfa were also associated with the first partition. The second partition included the majority of the grass, trees and woods. The sum of the Shannon entropy values over the histograms from the PCE results came to 16.3 compared to SPICE's 19.0 value. Again, PCE provided more compact histograms and SPICE indicating that the endmembers are better representatives of the true ground truth classes.

The PCE results on full spectra AVIRIS Indian Pines data. The PCE algorithm was run again on the AVIRIS Indian Pines data set. In this run, the data dimensionality was not reduced; the full 220 bands were used. Partitions were initialized using the KG-FCM algorithm and the DPMM algorithm to 3 clusters. Initial endmembers were found for each partition using the ED algorithm. A total of 1037 pixels (every 10^{th} labeled pixel) were selected from the data set and used in the PCE algorithm. Parameter values used to generate the results shown on this data set are listed in Table 4-12.

After finding endmembers on the subset of pixels using PCE, every data point in the image was unmixed using each clusters' set of endmembers and, for every data point, the likelihood under each cluster was computed. Each data point was then assigned to the partition with the maximum likelihood value. Partitions with less than 3 points were removed. Following these steps, two partitions were found with a total of six endmembers. Figure 4-33 shows the abundances maps associated with each endmember. Figure 4-34 contains the normalized histograms for this set of results.

The first partition roughly corresponds to the various tended fields in the imagery whereas the second partition has many of the abundances associated with trees, grass and woods. The sum of the entropies of the histograms from this results came to 15.4. This is value smaller than the SPICE results of 19.0 indicating that the endmembers are better representatives of the ground truth classes than the endmembers found by SPICE.

Table 4-1. Number of endmembers found by SPICE and ICE on test pixels from AVIRIS Cuprite data over a range of Γ values and initial number of endmembers. Each experiment had the same initialization for ICE and SPICE. The μ parameter was 0.1 for all experiments. The pruning threshold was set to 1×10^{-9}.

Experiment	Initial number of endmembers	Gamma constant for SPICE	Number of endmembers found, SPICE	Number of endmembers found, ICE
1	5	1.0	3	5
2	10	0.5	3	9
3	10	0.5	3	8
4	10	10.0	3	9
5	10	10.0	3	8
6	15	1.0	3	12
7	30	1.0	3	12
8	40	1.0	3	13
9	50	1.0	3	11

Table 4-2. Mine distributions in overlap regions of AHI and Lynx imagery

Mine type	Depth	AHI image 1 quantity	AHI image 2 quantity	AHI image 3 quantity
PC1	10 cm	44	17	17
MC1	10 cm	57	48	26
MC1	Flush	34	34	20
MC1	Surface	16	16	16
MC2	Surface	14	14	0
PC2	Surface	5	0	0
Total		**170**	**129**	**79**

Table 4-3. False alarm rate reduction using blackbody mask in AHI image 1

	Probability of detection				
	20%	**30%**	**40%**	**50%**	**60%**
RX without BB mask	2.3×10^{-3}	3.3×10^{-3}	5.8×10^{-3}	6.7×10^{-3}	9.0×10^{-3}
Clustering BB mask	2.1×10^{-3}	3.0×10^{-3}	5.3×10^{-3}	6.1×10^{-3}	8.3×10^{-3}
FAR reduction	**8.7**%	**9.1**%	**8.6**%	**9.0**%	**7.78**%
SPICE BB mask	1.0×10^{-3}	1.2×10^{-3}	2.3×10^{-3}	2.8×10^{-3}	4.2×10^{-3}
FAR reduction	**56.5**%	**63.6**%	**60.3**%	**58.2**%	**53.3**%

Table 4-4. False alarm rate reduction using blackbody mask in AHI image 2

	Probability of detection				
	20%	**30%**	**40%**	**50%**	**60%**
RX without BB mask	1.7×10^{-3}	2.6×10^{-3}	3.8×10^{-3}	6.1×10^{-3}	8.5×10^{-3}
Clustering BB mask	1.4×10^{-3}	2.2×10^{-3}	2.8×10^{-3}	4.2×10^{-3}	6.1×10^{-3}
FAR reduction	**17.6**%	**15.4**%	**26.3**%	**31.1**%	**28.2**%
SPICE BB mask	1.2×10^{-3}	1.9×10^{-3}	2.5×10^{-3}	3.8×10^{-3}	5.8×10^{-3}
FAR reduction	**29.4**%	**26.9**%	**34.2**%	**37.7**%	**31.8**%

Table 4-5. False alarm rate reduction using blackbody mask in AHI image 3

	Probability of detection				
	20%	**30%**	**40%**	**50%**	**60%**
RX without BB mask	3.7×10^{-3}	5.2×10^{-3}	9.2×10^{-3}	1.2×10^{-3}	1.6×10^{-3}
Clustering BB mask	3.7×10^{-3}	5.2×10^{-3}	9.2×10^{-3}	1.2×10^{-3}	1.6×10^{-3}
FAR reduction	**0**%	**0**%	**0**%	**0**%	**0**%
SPICE BB mask	3.3×10^{-3}	4.4×10^{-3}	6.8×10^{-3}	1.0×10^{-2}	1.4×10^{-2}
FAR reduction	**10.8**%	**15.4**%	**26.1**%	**16.7**%	**12.5**%

Table 4-6. Mean number and standard deviation of endmembers and bands found over 50 runs of SPICE or B-SPICE on the simulated data set. The true number of endmembers for this data set is 4.

Λ	Mean number of endmembers	Standard deviation of number of endmembers	Mean number of bands retained	Standard deviation of number of bands retained
0 (SPICE)	4	0	51.0	0.0
0.25	4	0	34.6	1.1
0.50	4	0	25.0	1.4
0.75	4	0	20.9	1.2
1.00	4	0	16.2	3.7

Table 4-7. Statistics of the averaged squared error per abundance value between calculated and true abundance values

Λ	Median average squared error per abundance	Mean average squared error per abundance	Standard deviation of average squared error per abundance
0 (SPICE)	0.005	0.005	0.0005
0.25	0.004	0.008	0.0066
0.50	0.004	0.007	0.0057
0.75	0.004	0.007	0.0050
1.00	0.006	0.010	0.0069

Table 4-8. Indian Pines Data Set Results and Comparison. Comparison Values Estimated from Graphs in (Guo et al., 2006) and (Martinez-Uso et al., 2006)

Exp.	Λ	Num. of endmembers	Num. of bands	Classification accuracy in percentage			Comparison results in percentage	
				Run 1	Run 2	Run 3	Ref. Guo, et al.	Ref. M.-Uso, et al.
1	0.0	8	220	93.6	93.9	93.9	90	-
2	0.0	7	220	93.1	93.1	92.9	90	-
3	0.5	7	124	93.3	93.7	93.7	90	-
4	0.5	7	122	93.0	92.9	93.2	90	-
5	1.0	7	89	93.4	93.3	93.6	90	-
6	1.0	7	103	93.3	93.3	93.5	90	-
7	5.0	7	34	86.4	86.4	86.3	88	80
8	5.0	8	34	86.5	86.0	86.4	88	80
9	10.0	7	19	83.4	83.9	82.5	82	81
10	10.0	8	18	77.8	80.0	78.3	82	82

Table 4-9. Indian Pines Data Set results using sampled parameter values and comparison with (Guo et al., 2006)

Experiment number	Number of endmembers	Number of bands kept	Classification accuracy in percentage			Comparison results in percentage
			Run 1	Run 2	Run 3	Ref. Guo, et al.
1	7	114	92.1	92.1	92.2	90
2	7	114	92.6	92.4	92.5	90

Table 4-10. Parameter values used to generate ED results on two-dimensional data sets. All covariance matrices are diagonal with elements equal to the values shown in the table.

Data Set	Variance of data	Likelihood variance	SSD variance	b_k
Triangle Data (Fig. 4-1)	55.9	0.1	0.5	0.01
Data from Dists. (Fig. 4-19)	5.2	0.5	1.0	0.01

Table 4-11. Parameter values used to generate ED results on hyperspectral data sets. All covariance matrices are diagonal with elements equal to the values shown in the table.

Data Set	Dimensionality of data	Variance of data	Likelihood variance	SSD variance	b_k
Cuprite Data	51	0.01	0.200	0.20	0.001
Indian Pines Data	220	0.02	0.001	0.01	0.010

Table 4-12. Parameter values used to generate PCE results. All covariance matrices are diagonal with elements equal to the values shown in the table.

Data Set	Data dimen.	Variance of data	Likelihood variance	α	ED likelihood variance	ED SSD variance	b_k
2D Data	2	2.16	0.005	2	0.010	1.000	0.001
PCA IP	6	0.05	0.005	1	0.005	0.001	0.001
Hierarchical IP	3	0.03	0.001	1	0.010	0.010	0.010
Full Spectra IP	220	0.03	0.001	1	0.010	0.010	0.010

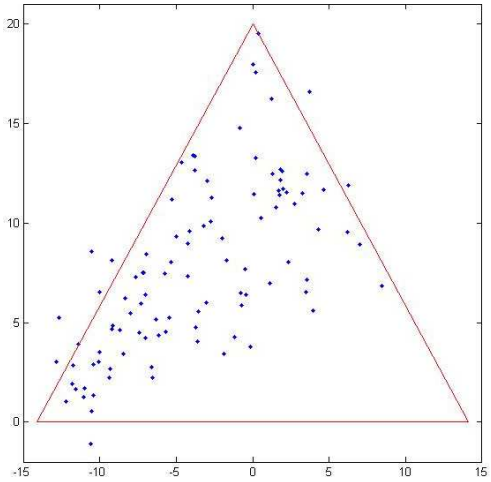


Figure 4-1. Two-dimensional SPICE example data set. 100 data points generated from the corners of the simplex shown.

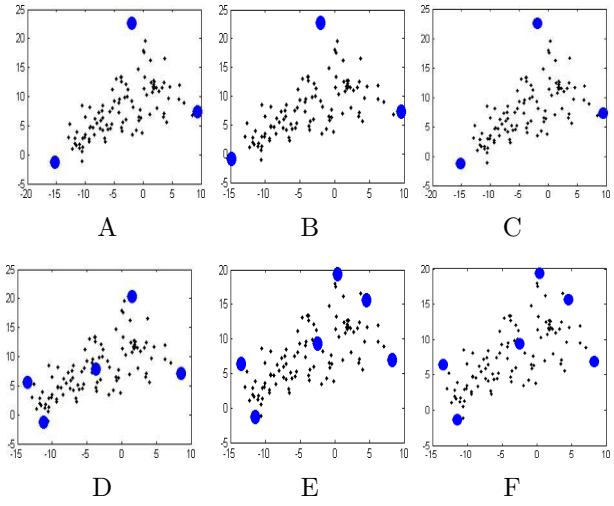


Figure 4-2. Comparison of SPICE (top) and ICE with pruning (bottom). In these three experiments, $\mu = 0.001$ and the pruning threshold was set to 0.0005. Initial number of endmembers was 20.

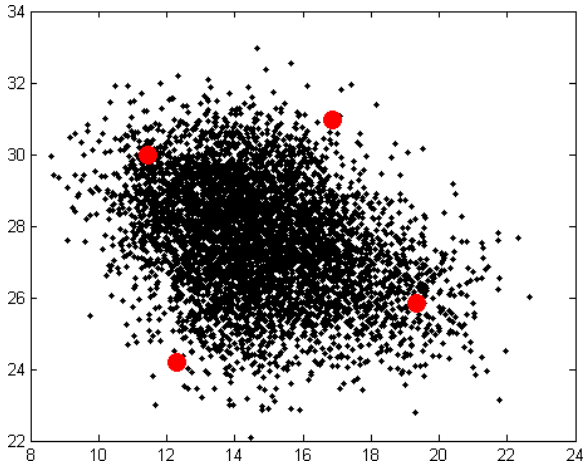


Figure 4-3. The SPICE results on two-dimensional data

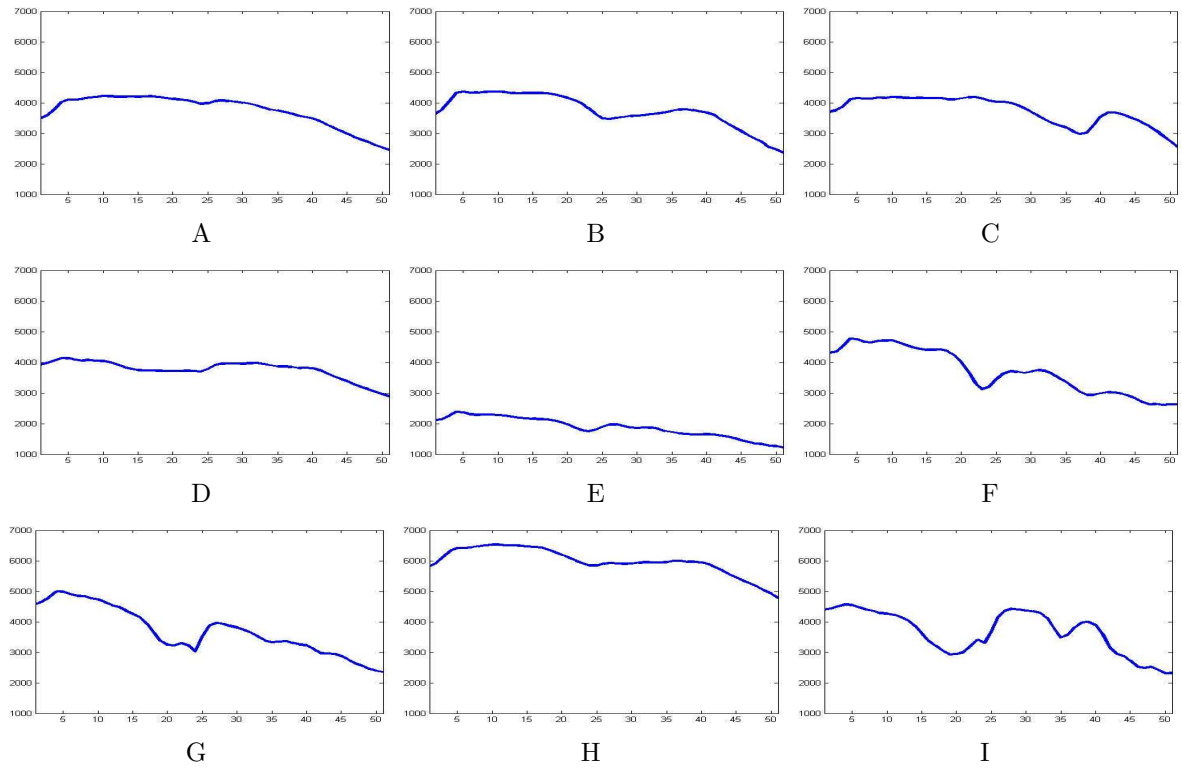


Figure 4-4. Endmembers found using SPICE on AVIRIS Cuprite hyperspectral data. μ was 0.1 for all experiments. The pruning threshold was set to 1×10^{-9}. The limits of the x-axis are 1978 to 2477 nm and the limits of the y-axis are 1000 to 7000 in units of 10,000 times the reflectance factor (Clark et al., 2004).

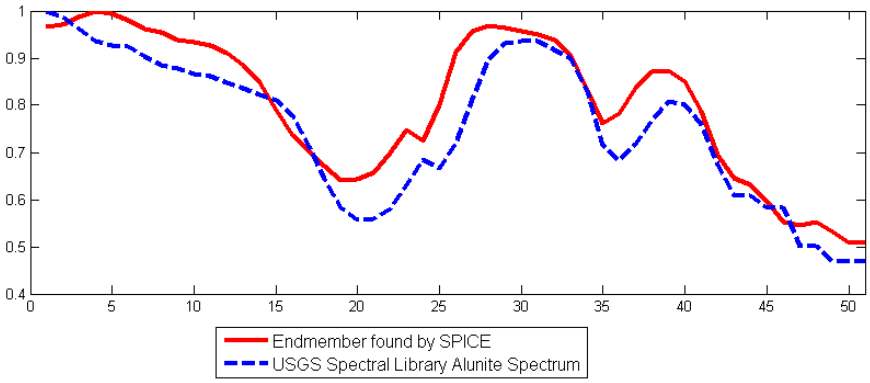


Figure 4-5. Comparison of one endmember found by SPICE and a USGS Alunite spectrum ("Alunite SUSTDA-20 W1R1Ba AREF") from the 2005 USGS spectral library.

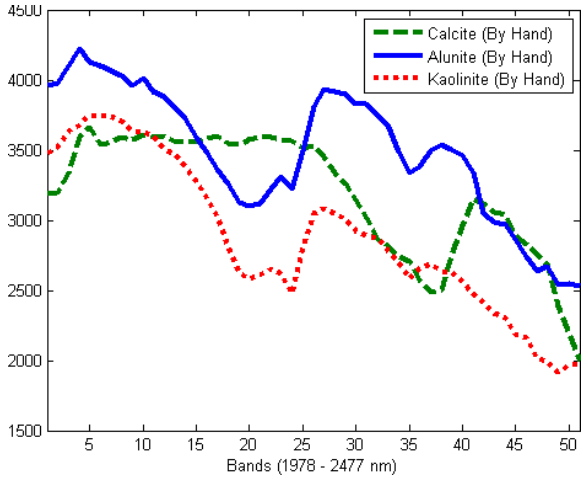


Figure 4-6. Endmembers selected from AVIRIS Cuprite data image by hand.

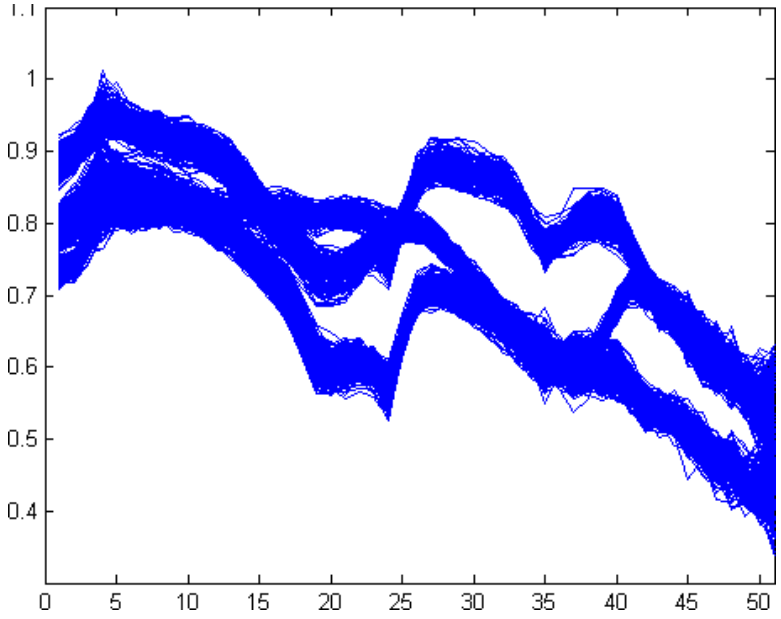


Figure 4-7. Normalized test pixels selected from AVIRIS Cuprite data.

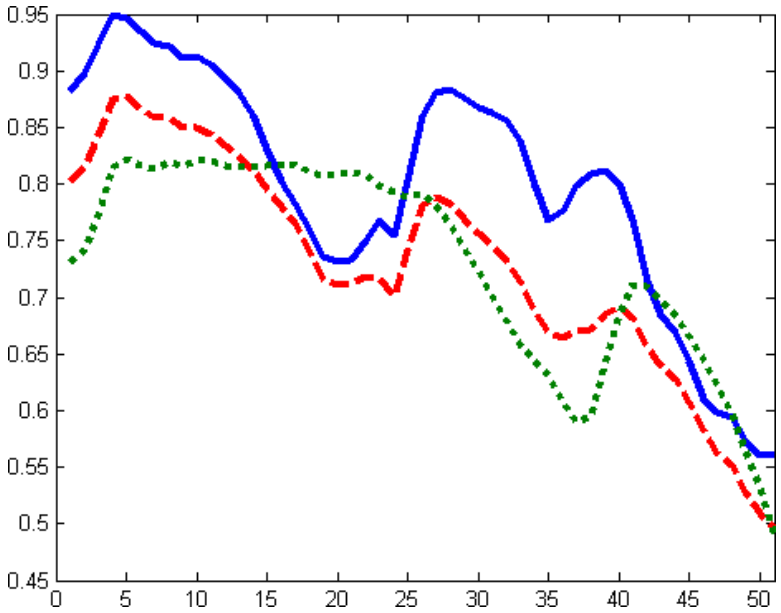
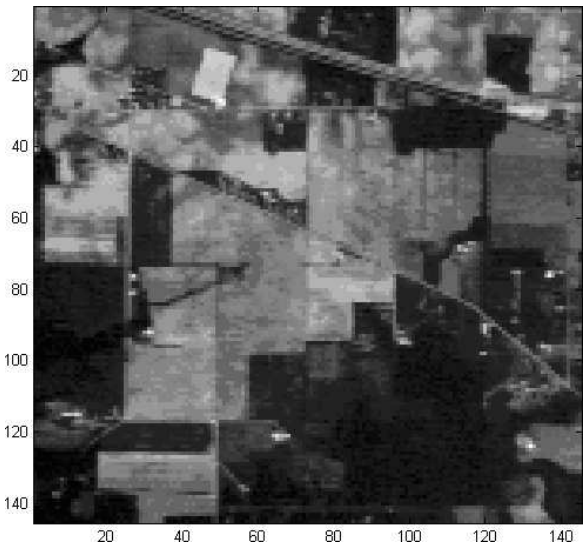
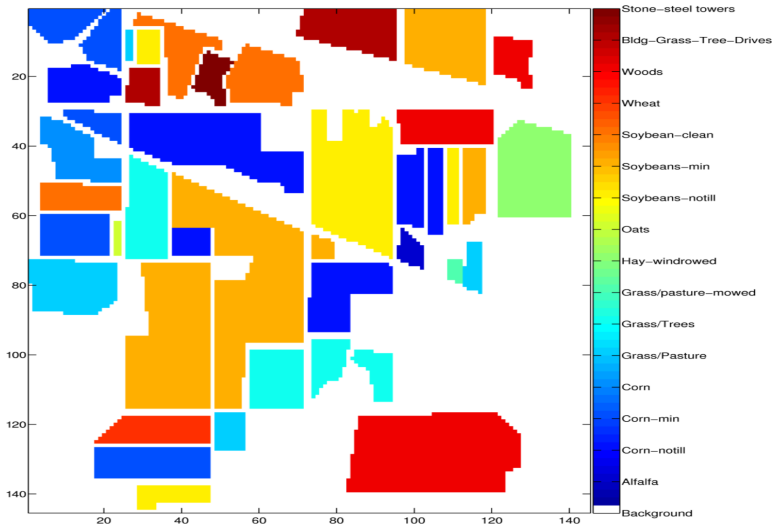


Figure 4-8. SPICE endmember results found on normalized test data selected from the AVIRIS Cuprite scene



A



B

Figure 4-9. Band 10 ($\sim$ 0.5 μm) of the AVIRIS Indian Pines data set and ground truth.

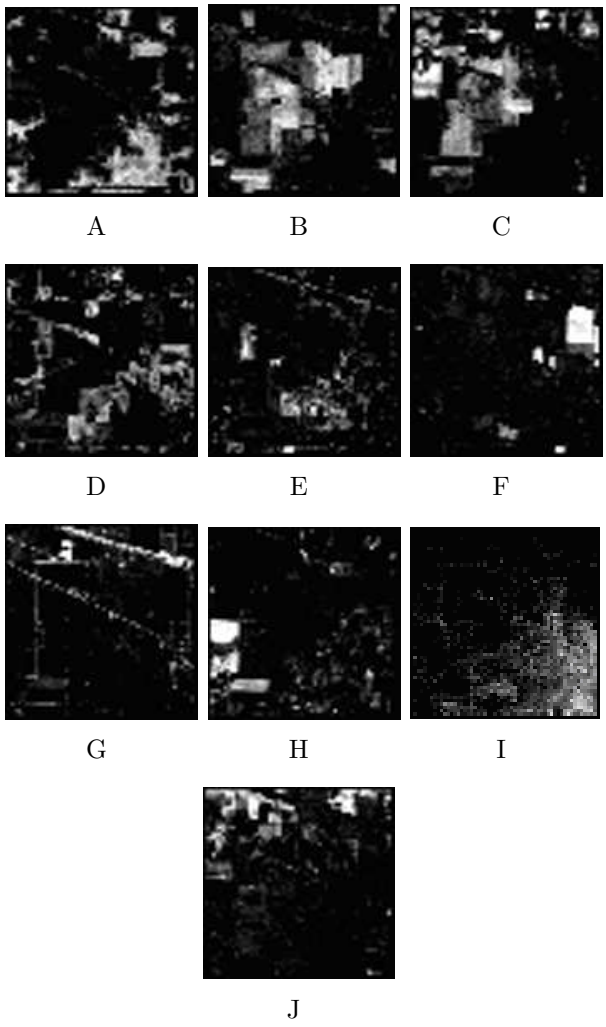


Figure 4-10. Abundance maps generated by SPICE on the Indian Pines data set.

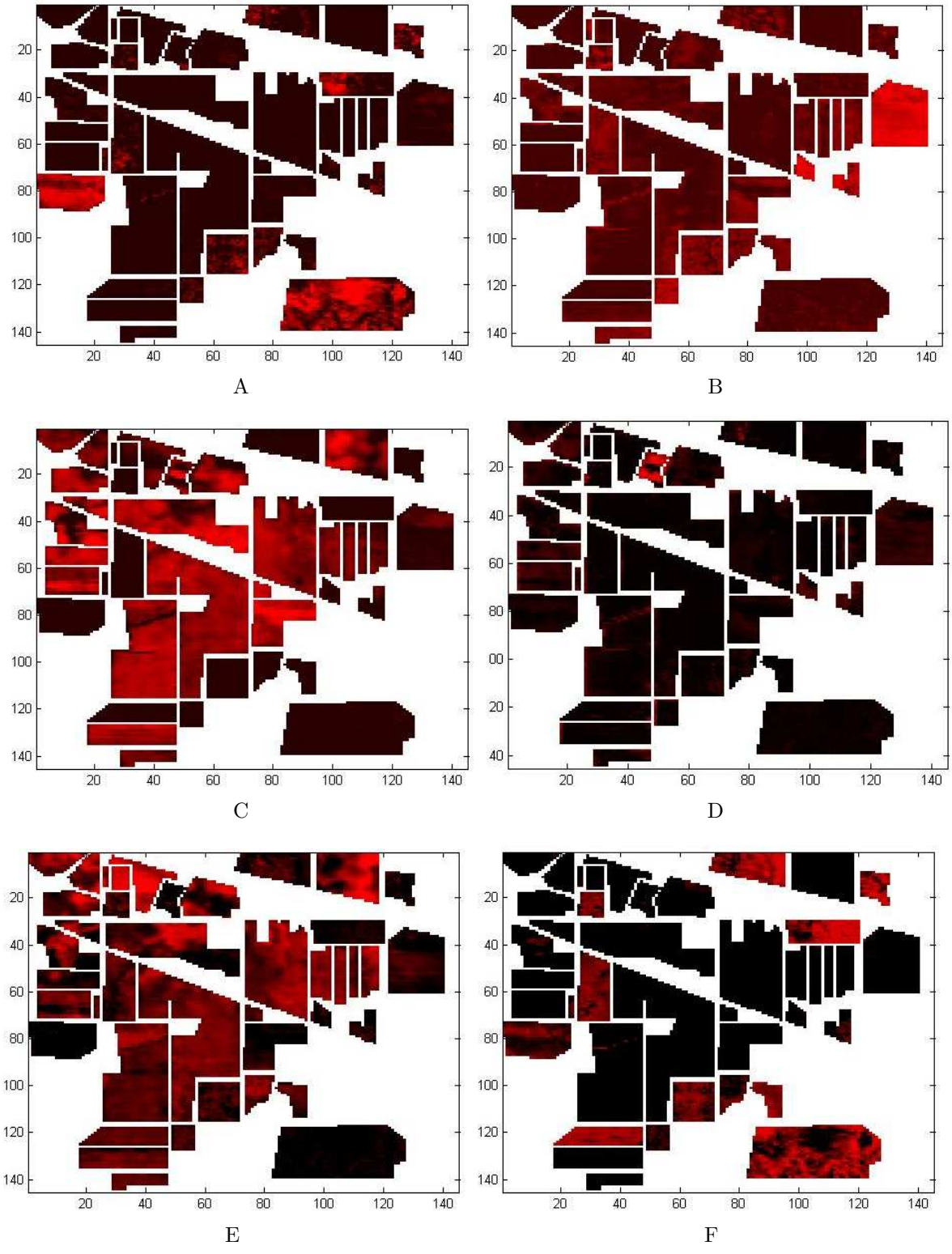


Figure 4-11. Abundance maps generated by SPICE on the labeled AVIRIS Indian Pines data set. Pixels in white correspond to unlabeled pixels. Remaining pixels range from black (abundance value of zero) to red (abundance of one).

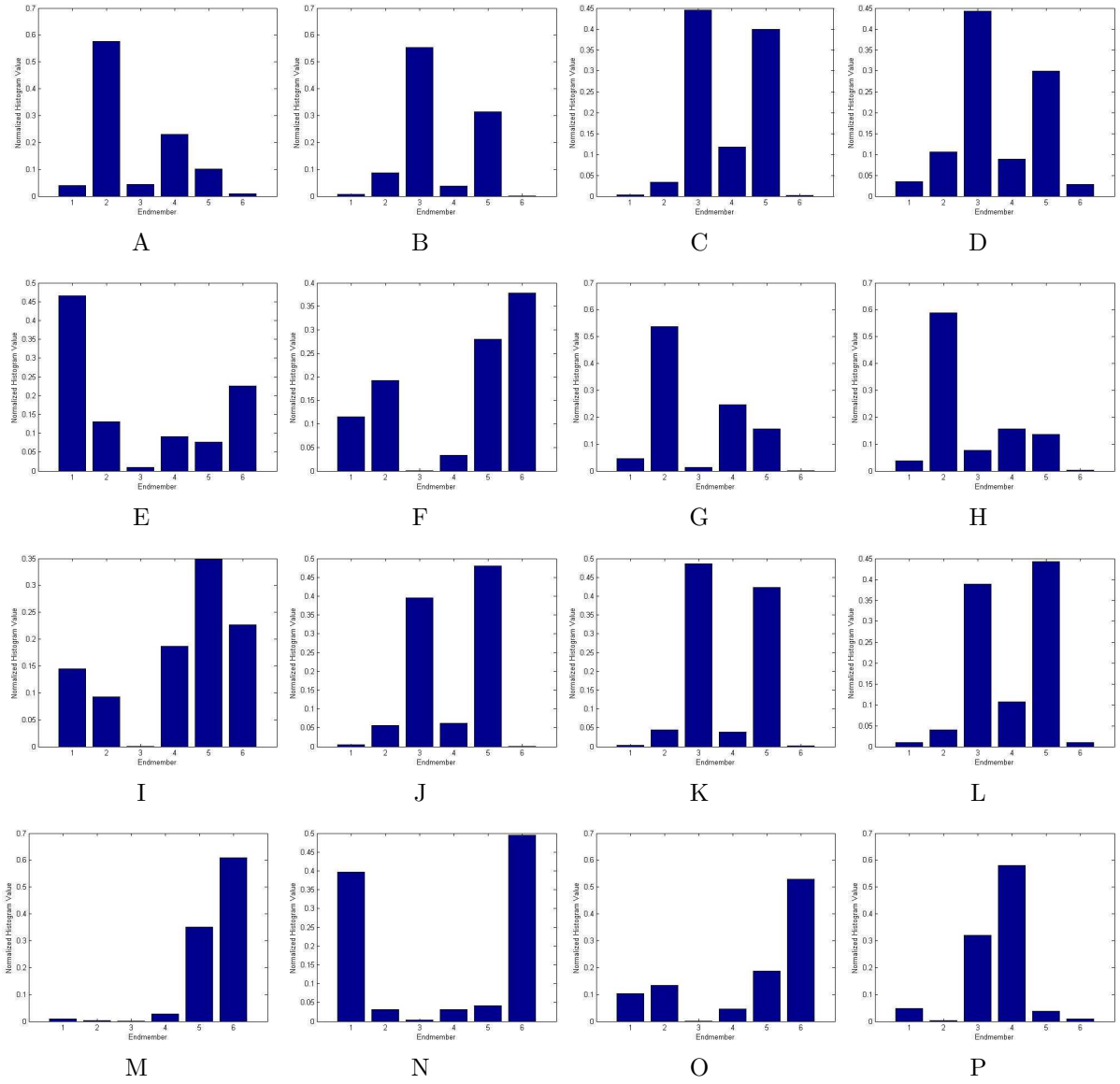


Figure 4-12. Histogram of SPICE endmember results on labeled AVIRIS Indian Pines data. Histograms show distribution of abundances values among endmembers in each ground truth class. Histograms were computed according to Equation 4–2. The sum of these histograms' Shannon's entropy values is 19.0. The histograms correspond to the following ground truth classes: (A) alfalfa, (B) corn-notill, (C) corn-min, (D) corn, (E) grass/pasture, (F) grass/trees, (G) grass/pasture-mowed, (H) hay-windrowed, (I) oats, (J) soybeans-notill, (K) soybeans-min, (L) soybean-clean, (M) wheat, (N) woods, (O) building-grass-trees-drive, and (P) stone-steel towers.

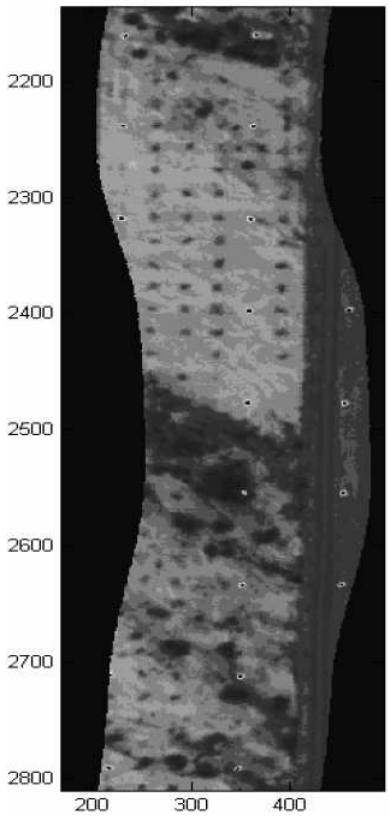


Figure 4-13. Subset at 9.19 microns of AHI hyperspectral image 2 including the overlap region

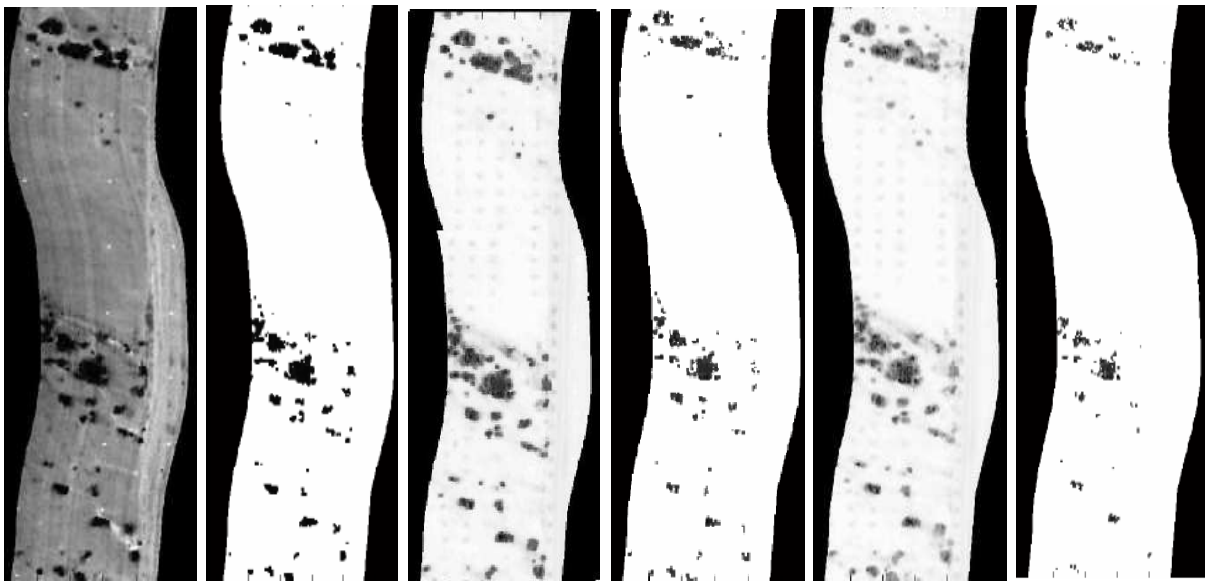


Figure 4-14. Blackbody masks created using SPICE and the clustering method. A) is the blackbody mask generated using SPICE and B) is the thresholded SPICE mask. C) is the mask generated using 4 clusters in the clustering method; D) is the thresholded version of this mask. E) is the mask generated using 5 clusters in the clustering method and F) is the thresholded version of this mask.

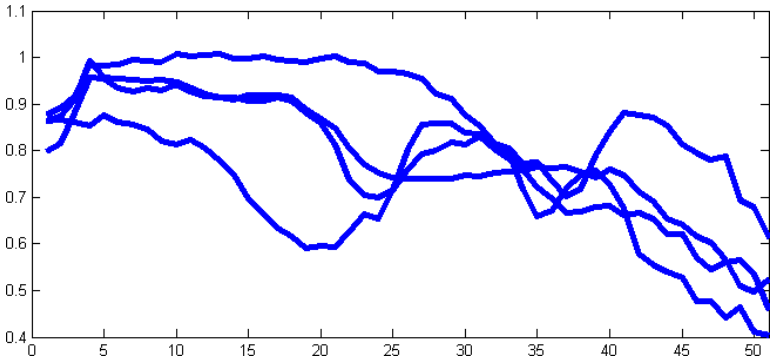


Figure 4-15. Endmembers used to generate simulated data set selected by hand from the AVIRIS Cuprite data set.

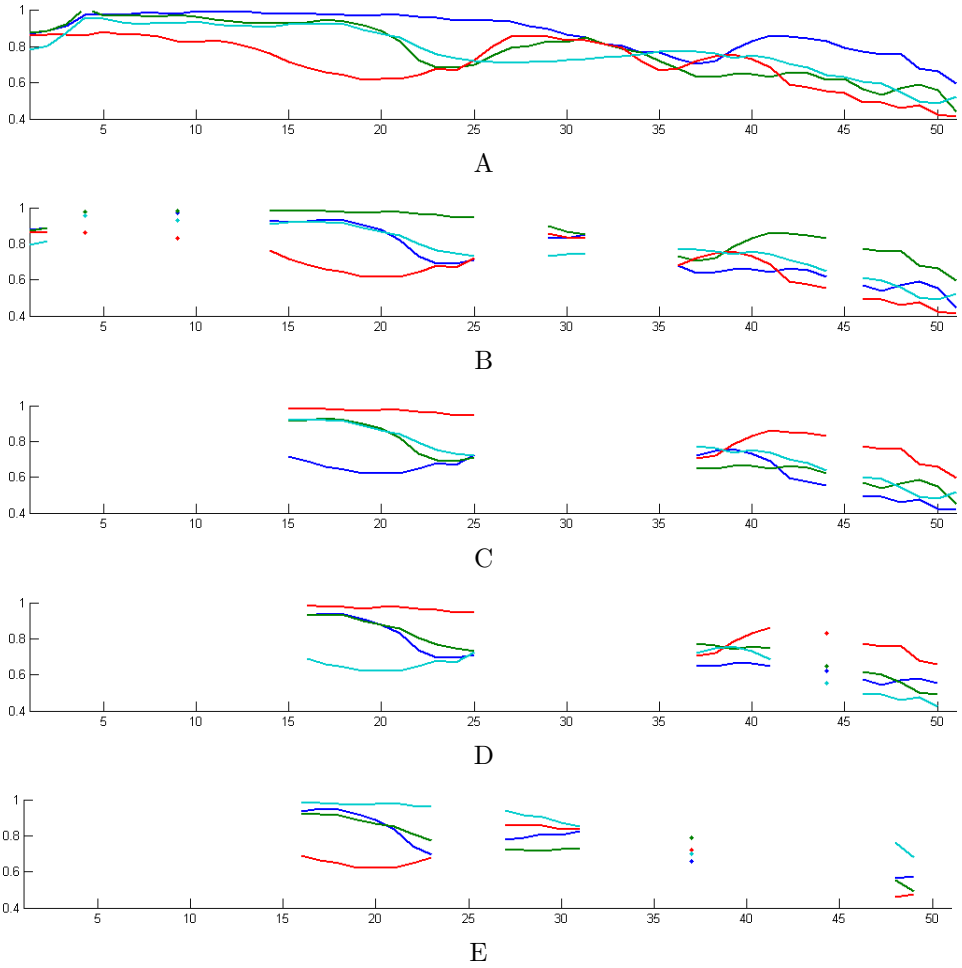


Figure 4-16. Endmembers determined using SPICE and B-SPICE with parameters $\Lambda = 0$, 0.25, 0.5, 0.75, and 1 on the simulated data set.

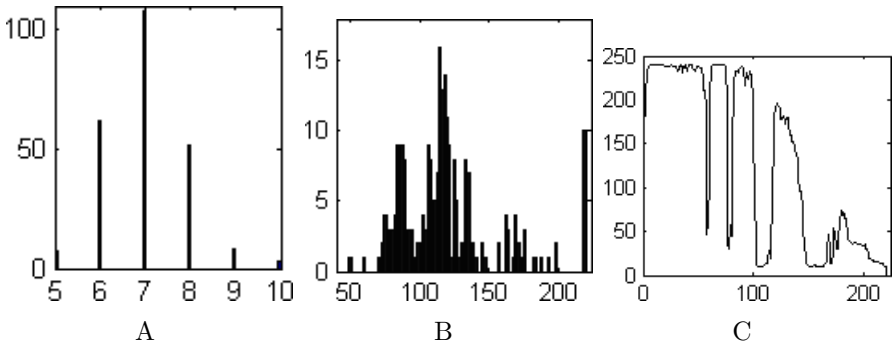


Figure 4-17. Histograms of (a) the number of endmembers (b) bands found and (c) the number of times each band is retained over 240 runs of B-SPICE using sampled parameter values.

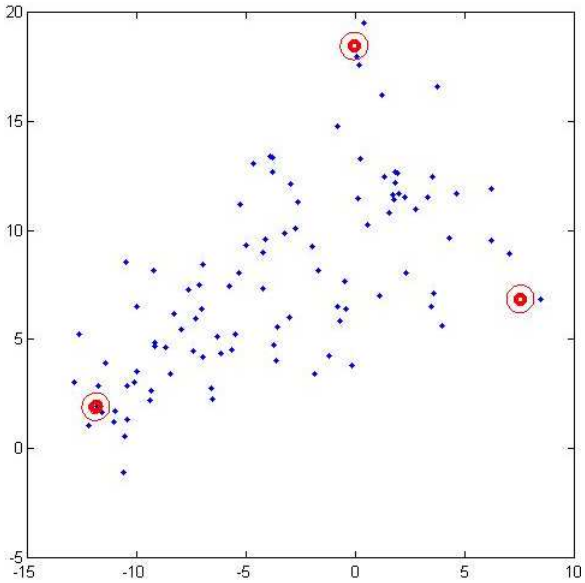


Figure 4-18. Results on two-dimensional triangle data found using ED. Blue points show the input data set. Red points are the mean endmembers of the endmember distributions found by ED. Red curves correspond to the 1st and 2nd standard deviations in each endmember distribution.

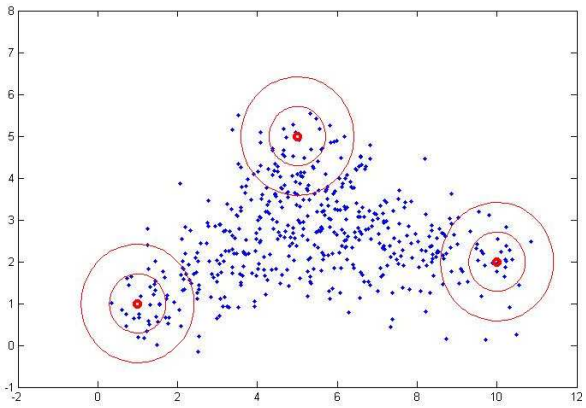


Figure 4-19. Data points generated from three endmember distributions. Blue points show the generated data set. Red points are the mean endmembers of the endmember distributions used to generate the data points. Red curves correspond to the 1st and 2nd standard deviations in each endmember distribution used to generated the data points.

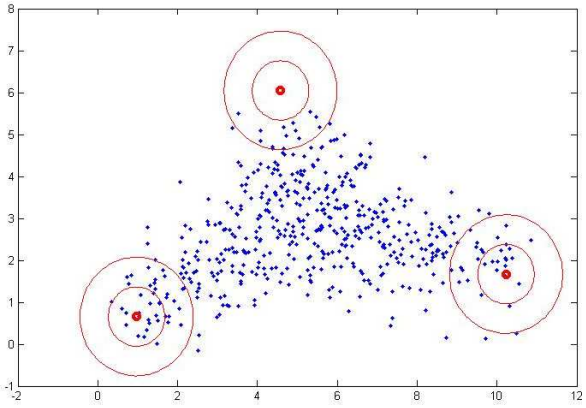


Figure 4-20. Results on two dimensional data using ED. Blue points show the generated data set. Red points are the mean endmembers of the endmember distributions found by ED. Red curves correspond to the 1st and 2nd standard deviations in each endmember distribution.

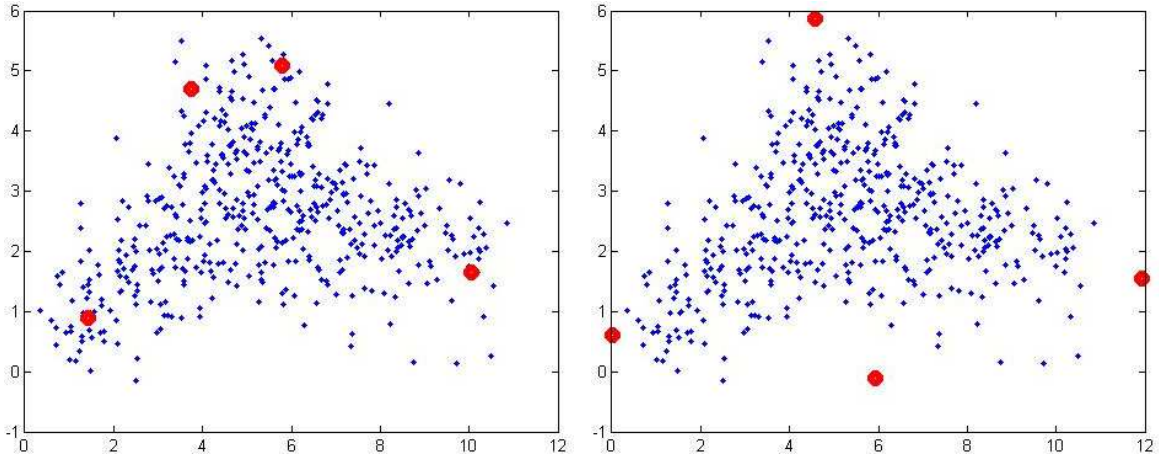


Figure 4-21. Results on two dimensional data using SPICE. Blue points show the generated data set. Red points are the endmembers found by SPICE.

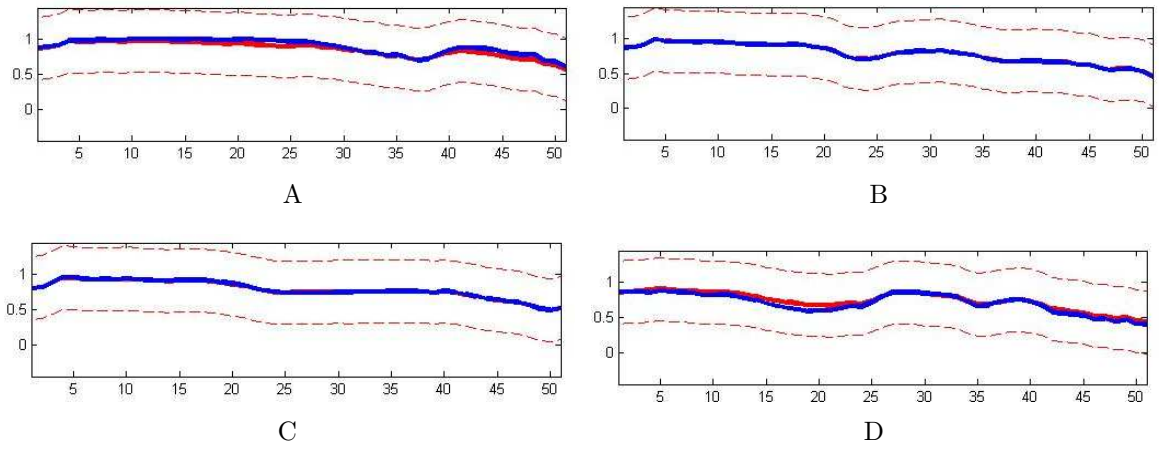


Figure 4-22. Results on simulated AVIRIS Cuprite data using ED. Solid blue curves show the true endmembers from which the data was generated. Solid red curves are the mean endmembers of the endmember distributions found by ED. Dashed red curves correspond to the 1st standard deviation in each endmember distribution.

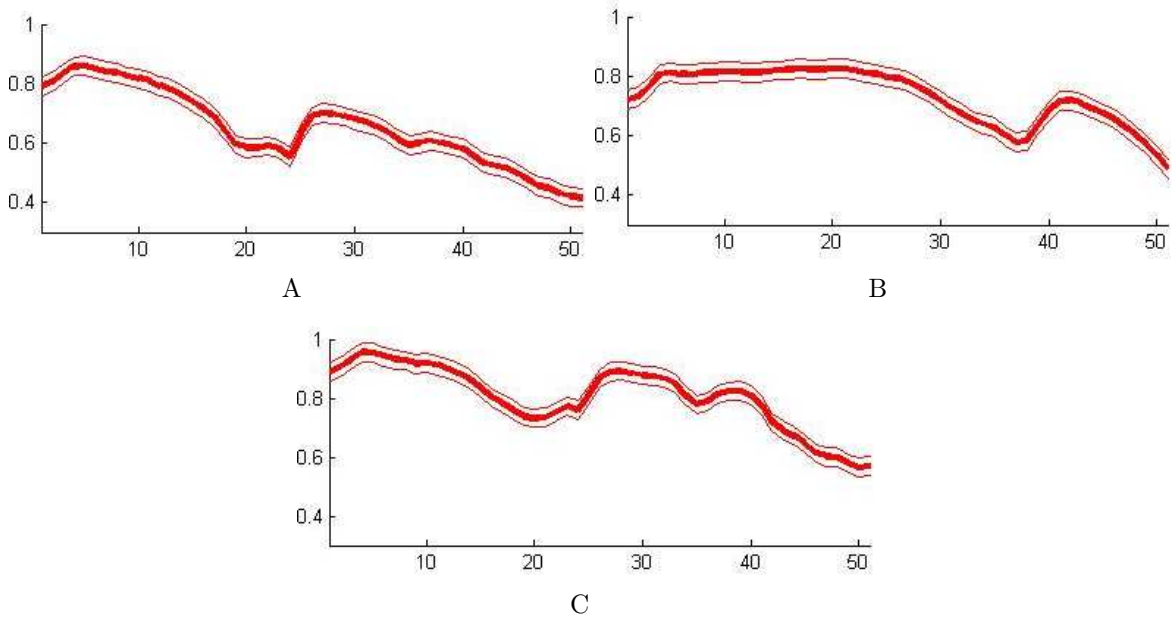


Figure 4-23. Results on a subset of AVIRIS Cuprite data found using ED. Solid red curves are the mean endmembers of the endmember distributions found by ED. Dashed red curves correspond to the 1st standard deviation in each endmember distribution.

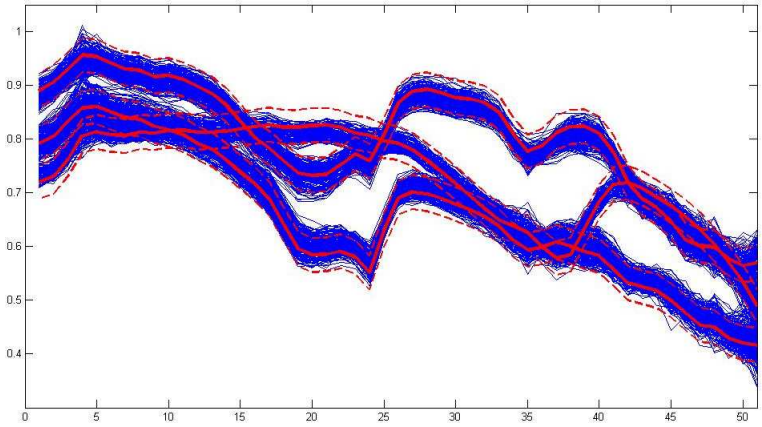


Figure 4-24. Results on a subset of AVIRIS Cuprite data found using ED. Blue curves show the input data set. Solid red curves are the mean endmembers of the endmember distributions found by ED. Dashed red curves correspond to the 1st standard deviation in each endmember distribution.

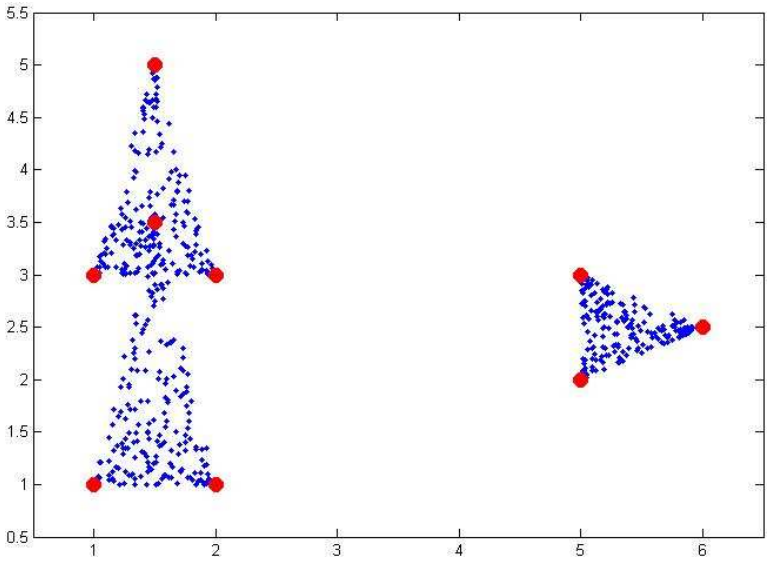


Figure 4-25. Two-dimensional data generated from three sets of endmembers. Blue points correspond to the input data set. Red points correspond to the endmembers from which the data was generated. Each triangle of data points was generated from three of the endmembers.

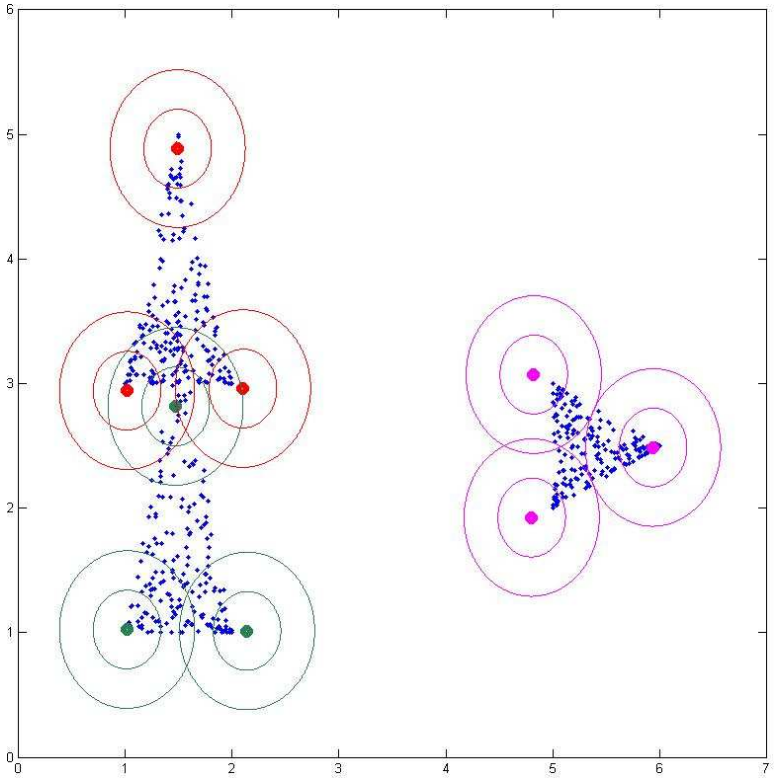


Figure 4-26. Two-dimensional data results found using PCE. Small blue points correspond to the input data set. Large points correspond to the mean endmembers for each endmember distribution. Thin curves correspond to the 1st and 2nd standard deviation curves from each endmember distribution. The color of each endmember distribution corresponds to the convex region to which it belongs.

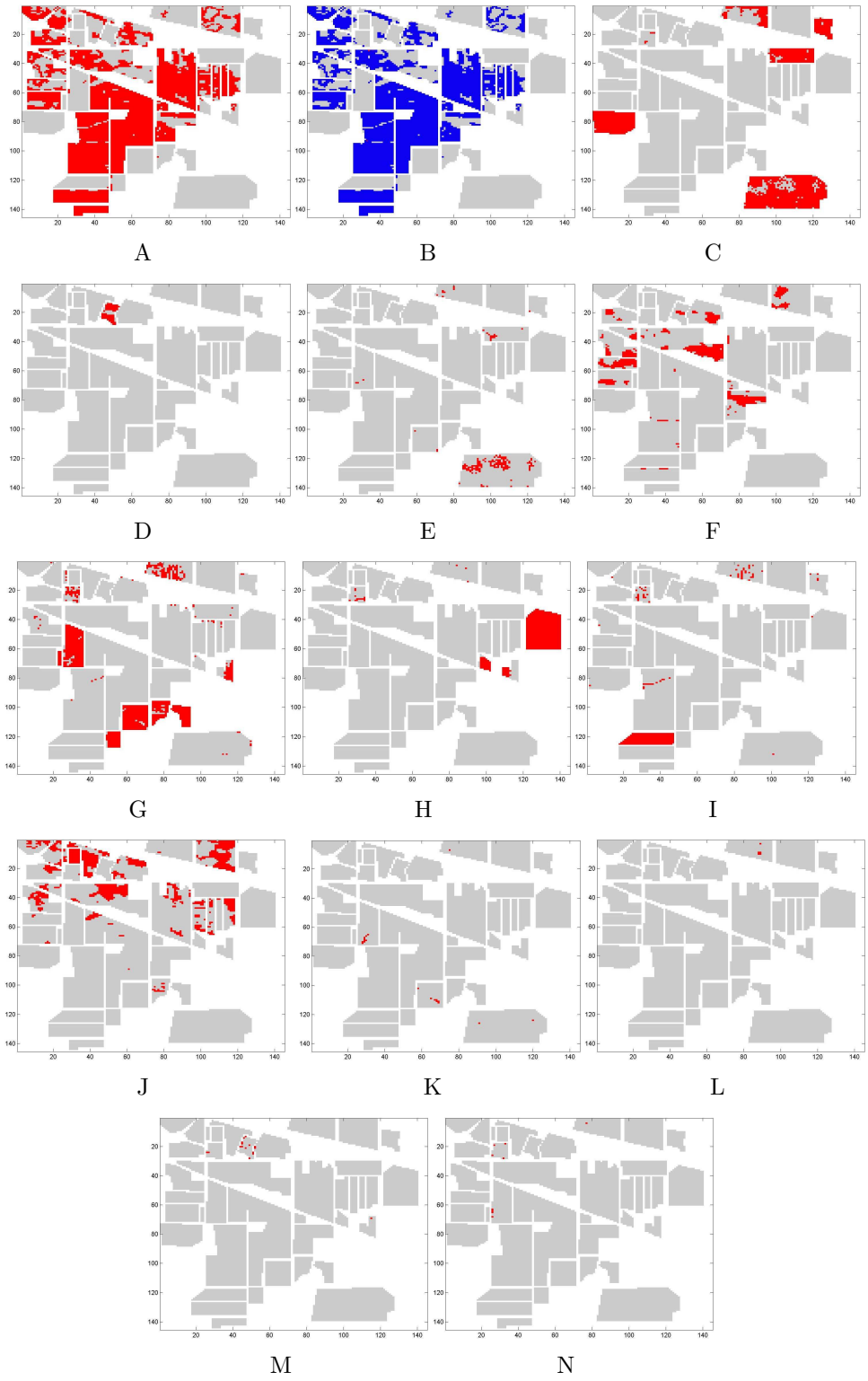


Figure 4-27. Abundance maps found using PCE on labeled PCA-reduced AVIRIS Indian Pines data. Pixels in white are unlabeled. Pixels in gray indicate pixels from another convex partition. Remaining pixels range from blue (abundance value of zero) to red (abundance value of one).

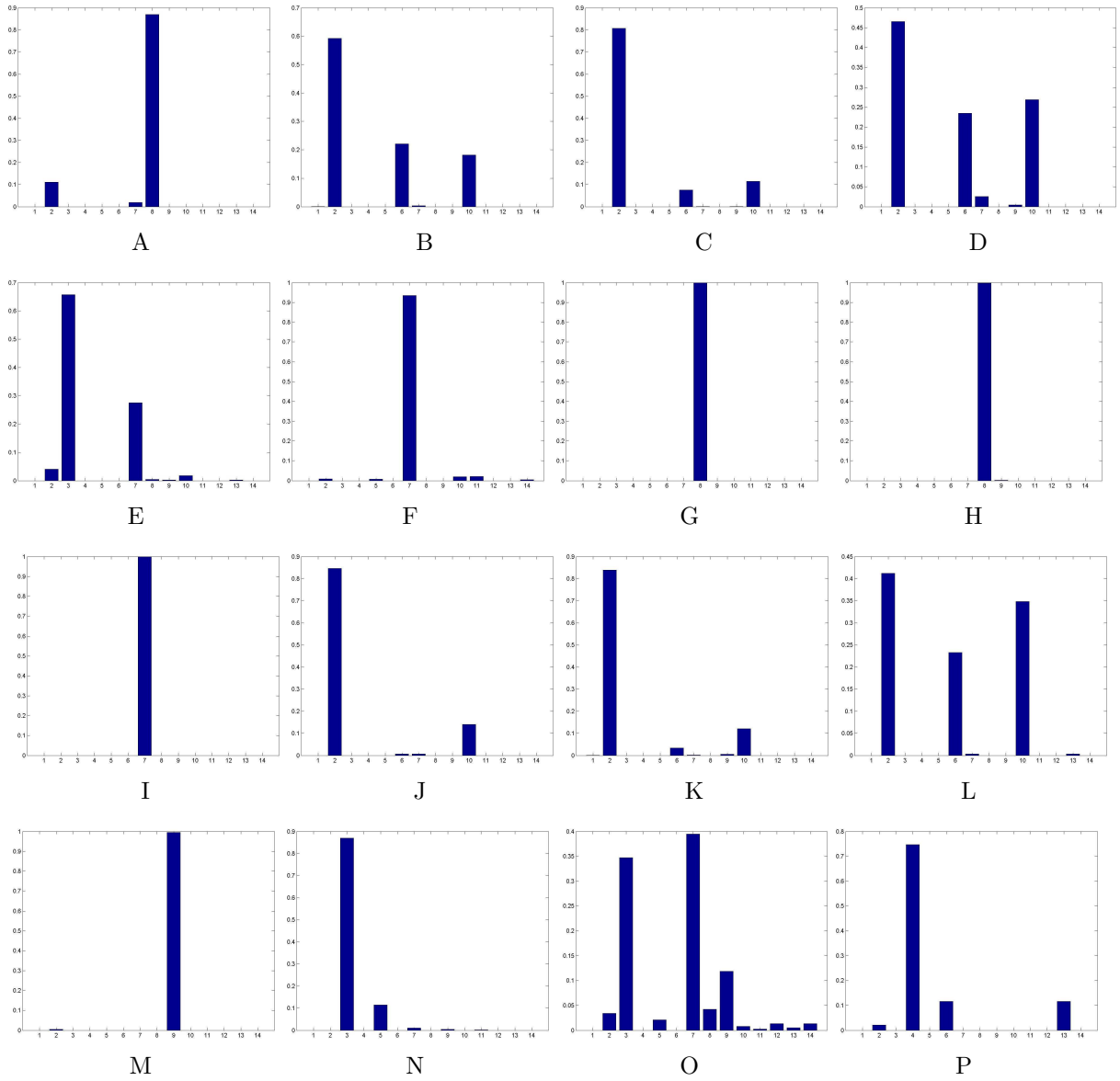


Figure 4-28. Histogram of PCE endmember results on labeled PCA-reduced AVIRIS Indian Pines data. Histograms show distribution of abundances values among endmembers in each ground truth class. Histograms were computed according to Equation 4–2. The sum of the histograms' Shannon's entropy values is 9.4. The histograms correspond to the following ground truth classes: (A) alfalfa, (B) corn-notill, (C) corn-min, (D) corn, (E) grass/pasture, (F) grass/trees, (G) grass/pasture-mowed, (H) hay-windrowed, (I) oats, (J) soybeans-notill, (K) soybeans-min, (L) soybean-clean, (M) wheat, (N) woods, (O) building-grass-trees-drive, and (P) stone-steel towers.

Figure 4-29. Abundance maps found using SPICE on labeled PCA-reduced AVIRIS Indian Pines data. Pixels in white are unlabeled. Remaining pixels range from blue (abundance value of zero) to red (abundance value of one).

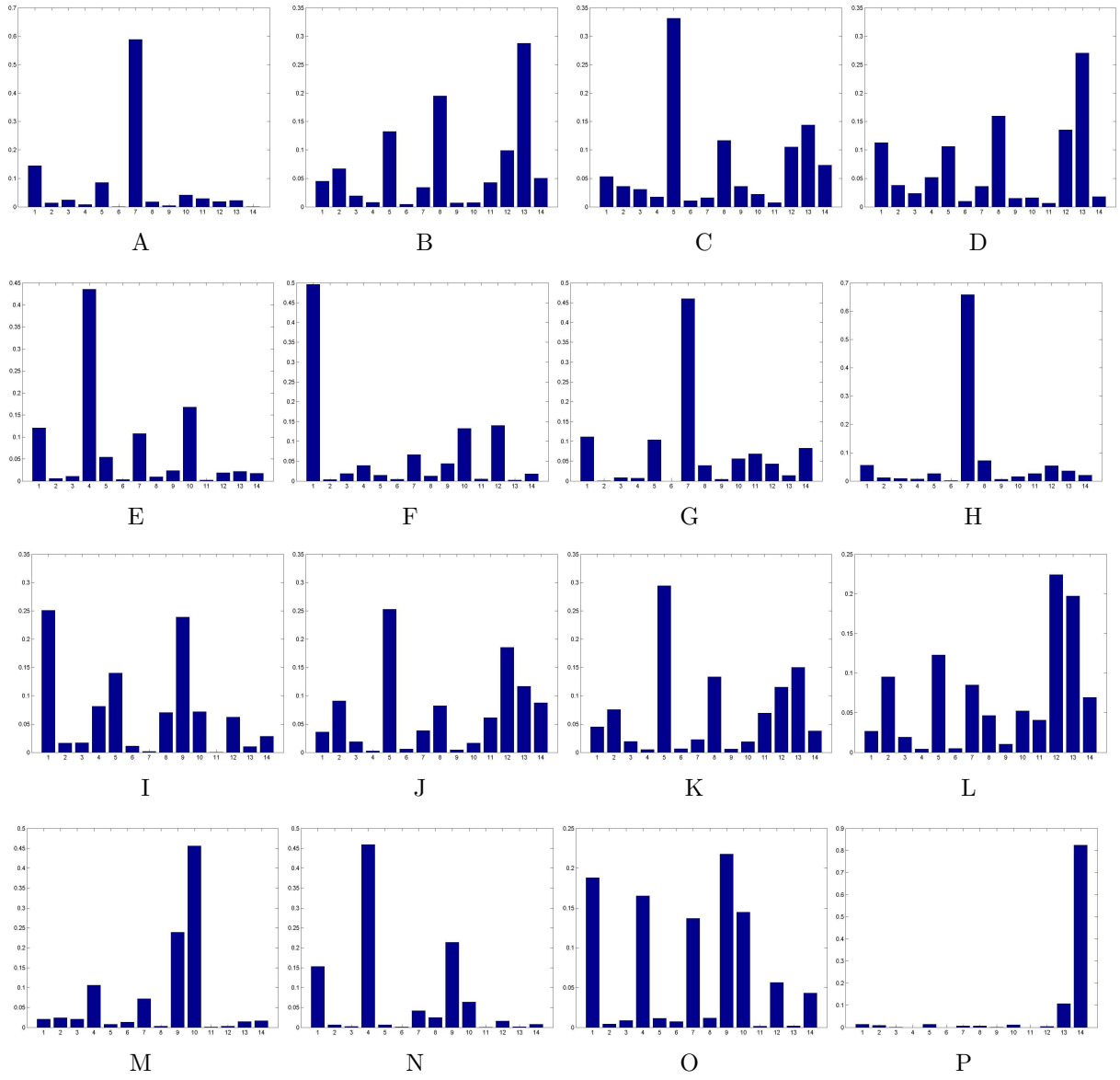


Figure 4-30. Histogram of SPICE endmember results on labeled PCA-reduced AVIRIS Indian Pines data. Histograms show distribution of abundances values among endmembers in each ground truth class. Histograms were computed according to Equation 4–2. The sum of the histograms' Shannon's entropy values is 29.2. The histograms correspond to the following ground truth classes: (A) alfalfa, (B) corn-notill, (C) corn-min, (D) corn, (E) grass/pasture, (F) grass/trees, (G) grass/pasture-mowed, (H) hay-windrowed, (I) oats, (J) soybeans-notill, (K) soybeans-min, (L) soybean-clean, (M) wheat, (N) woods, (O) building-grass-trees-drive, and (P) stone-steel towers.

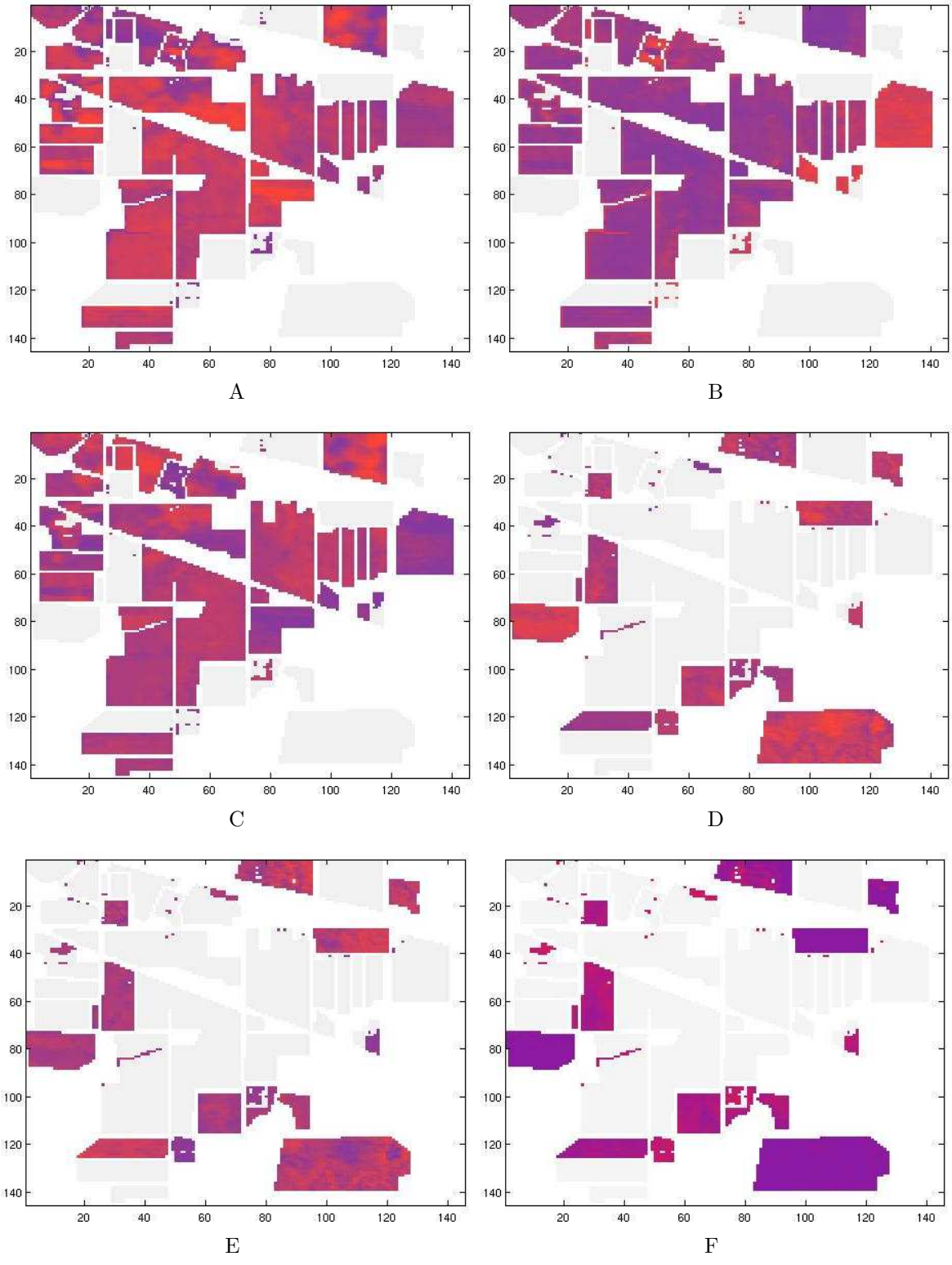


Figure 4-31. Abundance maps found using PCE on labeled AVIRIS Indian Pines data with hierarchical dimensionality reduction. Pixels in white are unlabeled. Pixels in gray indicate pixels from another convex partition. Remaining pixels range from blue (abundance value of zero) to red (abundance value of one).

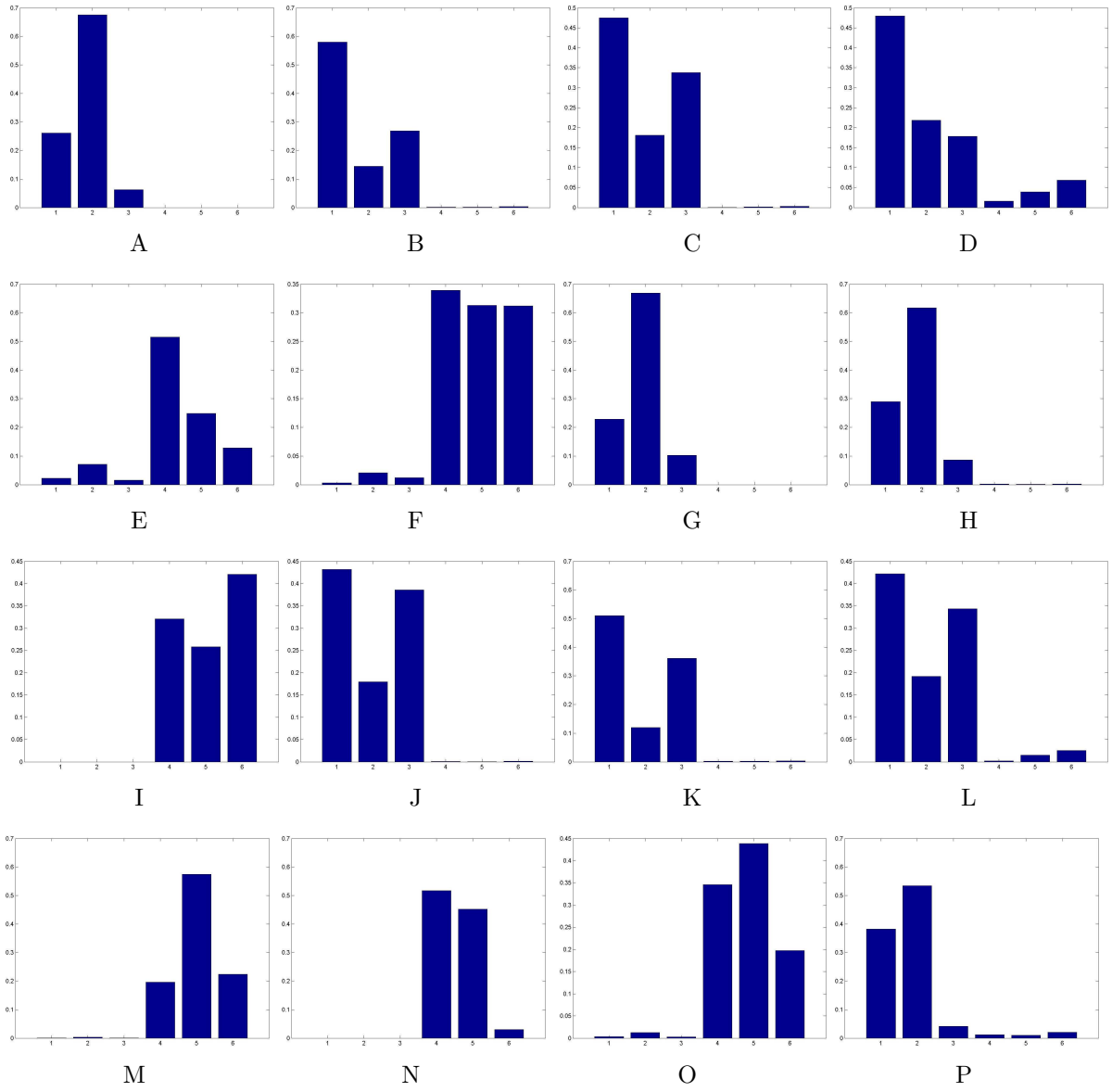


Figure 4-32. Histogram of PCE endmember results on labeled AVIRIS Indian Pines data with hierarchical dimensionality reduction. Histograms show distribution of abundances values among endmembers in each ground truth class. The sum of the histogram's Shannon's entropy values is 16.3. Histograms were computed according to Equation 4–2. The histograms correspond to the following ground truth classes: (A) alfalfa, (B) corn-notill, (C) corn-min, (D) corn, (E) grass/pasture, (F) grass/trees, (G) grass/pasture-mowed, (H) hay-windrowed, (I) oats, (J) soybeans-notill, (K) soybeans-min, (L) soybean-clean, (M) wheat, (N) woods, (O) building-grass-trees-drive, and (P) stone-steel towers.

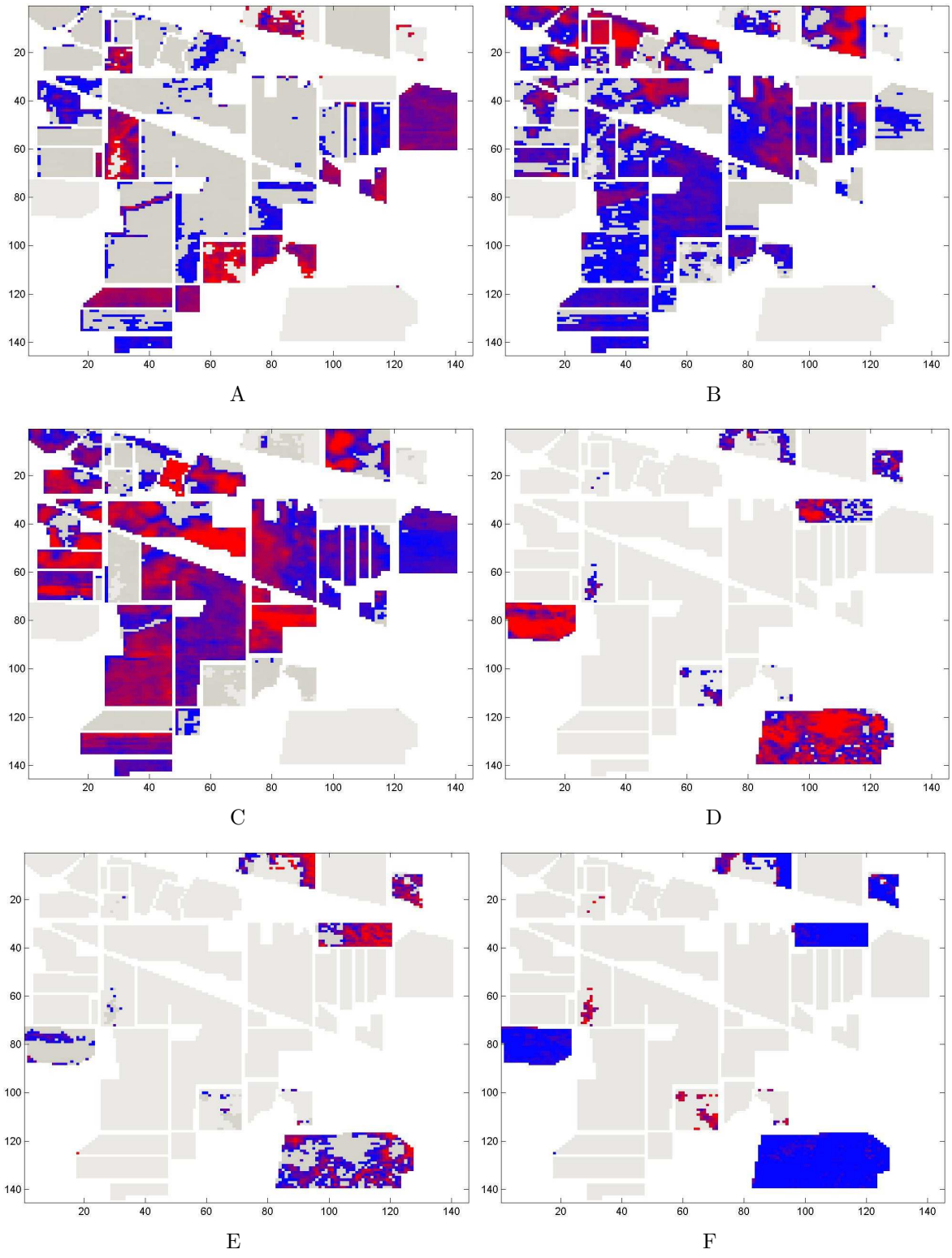


Figure 4-33. Abundance maps found using PCE on labeled AVIRIS Indian Pines data. Pixels in white are unlabeled. Pixels in gray indicate pixels from another convex partition. Remaining pixels range from blue (abundance value of zero) to red (abundance value of one).

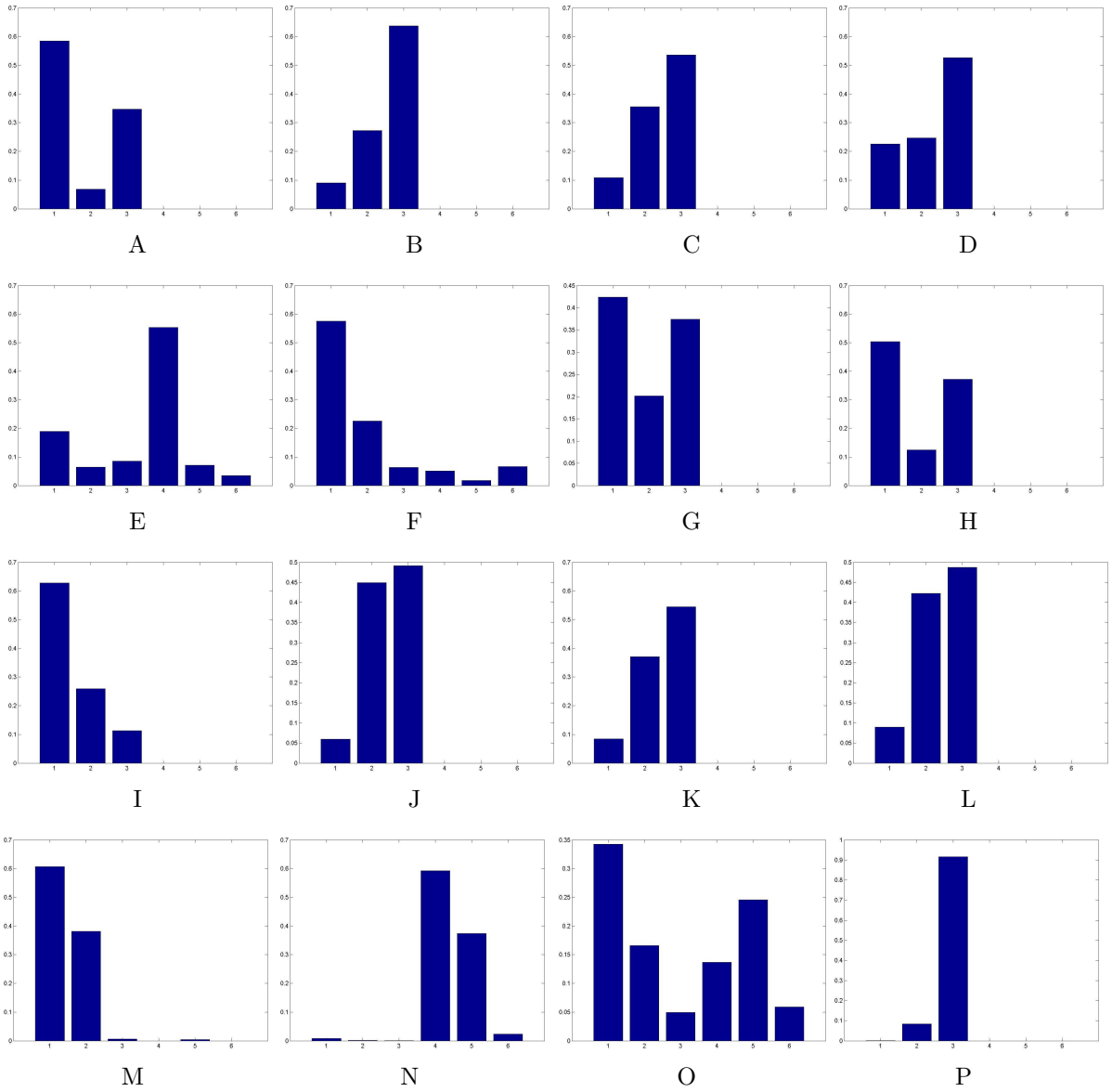


Figure 4-34. Histogram of PCE endmember results on labeled AVIRIS Indian Pines data. Histograms show distribution of abundances values among endmembers in each ground truth class. Histograms were computed according to Equation 4–2. The sum of the histogram's Shannon's entropy values is 15.4. The histograms correspond to the following ground truth classes: (A) alfalfa, (B) corn-notill, (C) corn-min, (D) corn, (E) grass/pasture, (F) grass/trees, (G) grass/pasture-mowed, (H) hay-windrowed, (I) oats, (J) soybeans-notill, (K) soybeans-min, (L) soybean-clean, (M) wheat, (N) woods, (O) building-grass-trees-drive, and (P) stone-steel towers.

CHAPTER 5
CONCLUSION

Four novel methods for hyperspectral image spectroscopy based on Bayesian methodologies were developed and tested. The Sparsity Promoting Iterated Constrained Endmembers (SPICE) algorithm incorporates sparsity promoting priors to estimate the number of endmembers while simultaneously performing endmember detection and spectral unmixing. Previously, most endmember detection algorithms required the number of endmembers in advance. The algorithm's sparsity promoting priors drive the proportion values of unneeded endmembers to zero allowing SPICE to remove those endmembers without any effect on pixel representation via endmembers and proportions.

The Band Selecting Sparsity Promoting Iterated Constrained Endmember (B-SPICE) algorithm extends SPICE to include hyperspectral band selection. Sparsity promoting priors are applied to band weights to determine the hyperspectral bands which distinguish between the endmembers in the data set. Therefore, B-SPICE autonomously determines the number of needed wavelengths. In addition, B-SPICE is able to identify the needed wavelengths, perform endmember detection and determine the number of endmembers needed, simultaneously.

The Endmember Distribution (ED) detection algorithm learns endmember distributions to incorporate spectral variability into the endmember detection model. Previously, endmember detection algorithms constrained endmembers to be single spectral vectors. By utilizing endmember distributions, several pixels of the same material with some spectral variation can all be identified as having a full abundance for the same endmember.

The Piece-wise Convex Endmember (PCE) detection algorithm used the Dirichlet process to learn the number of convex regions needed to describe an input hyperspectral scene. For each convex region, an individual set of endmember distributions and proportion values were determined. In contrast, previous endmember detection algorithms applied the same set of endmembers to every data point in a scene. Using PCE, different

portions of an input hyperspectral scene can be represented using separate sets of endmembers. This results in better suited endmembers for all of the various regions in an input image.

During development and testing of the methods, several interesting areas for future research were uncovered. Currently, the ED algorithm assumes a constant and known covariance matrix for each endmember distribution. Investigations into methods of learning appropriate covariance matrices for each endmember distribution can be done. By learning covariance matrices, endmembers distributions can be further tailored to the input data set. Also, both the B-SPICE and PCE algorithms utilize optimization schedules and many parameter values. Studies on methods to determine the appropriate optimization schedules and parameter values with regard to the input data set can be conducted.

The PCE algorithm currently assigns each data point to a single partition. Investigations into methods of allowing data points to have partial membership in several partitions can be conducted. By allowing partial memberships, overlapping clusters may be more likely to be found using PCE. The membership values could also provide additional insight into how well-suited each set of endmembers are for an individual data point.

REFERENCES

C. E. Antoniak. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics*, 2(6):1152–1174, 1974.

R. Archibald and G. Fann. Feature selection and classification of hyperspectral images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 4 (4):674–677, Oct. 2007.

AVIRIS. (2004, Sep) AVIRIS free standard data products. Jet Propulsion Laboratory, California Institute of Technology, Pasedena, CA. URL http://aviris.jpl.nasa.gov/html/aviris.freedata.html.

M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Donne, and J. F. Huntington. ICE: A statistical approach to identifying endmembers in hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 42:2085–2095, Oct. 2004.

C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

J. Boardman, F. Kruse, and R. Green. Mapping target signatures via partial unmixing of AVIRIS data. In R. Green, editor, *Summaries of the 5th Annu. JPL Airborne Geoscience Workshop*, volume 1, pages 23–26, Pasadena, CA, 1995. JPL Publ.

L. M. Bruce, C. H. Koger, and J. Li. Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 40(10):2331–2338, Oct. 2002.

G. Casella and E. I. George. Explaining the Gibbs sampler. *The American Statistician*, 46 (3):167–174, Aug. 1992.

C.-I. Chang and Q. Du. Estimation of number of spectrally distinct signal sources in hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 42(3): 608–619, Mar. 2004.

C.-I. Chang and S. Wang. Constrained band selection for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 44(6):1575–1585, June 2006.

C.-I. Chang, Q. Du, T.-L. Sun, and M. L. G. Althouse. A joint band prioritization and band decorrelation approach to band selection for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 37(6):2631–2641, Nov. 1999.

S.-S. Chiang, C.-I. Chang, and I. W. Ginsberg. Unsupervised hyperspectral image analysis using independent component analysis. In *Proceedings of the IEEE: Geoscience and Remote Sensing Symposium*, pages 3136–3138, July 2000.

S. Chib and E. Greenberg. Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(4):327–335, Nov. 1995.

R. N. Clark, A. Swayze, R. Wise, K. E. Livo, T. M. Heofen, R. F. Kokaly, and S. J. Sutley. USGS digital spectral library (splib05a), January 2004. URL http://speclab.cr.usgs.gov/spectral-lib.html.

M. D. Craig. Minimum-volume transforms for remotely sensed data. *IEEE Transactions on Geoscience and Remote Sensing*, 32(3):542–552, May 1994.

S. DeBacker, P. Kempeneers, W. Debruyn, and P. Scheunders. A band selection technique for spectral classification. *IEEE Geoscience and Remote Sensing Letters*, 2(3):319–323, July 2005.

L. Devroye. *Non-Uniform Random Variate Generation*. Springer-Verlag, 1986.

H. Du, H. Qi, X. Wang, R. Ramanath, and W. E. Snyder. Band selection using independent components analysis for hyperspectral image processing. In *Proceedings of the IEEE: 32nd Applied Imagery Pattern Recognition Workshop*, 2003.

R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience, 2 edition, Oct. 2001.

T. S. Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230, 1973.

M. A. T. Figueiredo. Adaptive sparseness for supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25:1150–1159, September 2003.

A. N. French, T. J. Schmugge, and W. P. Kustas. Discrimination of senescent vegetation using thermal emissivity contrast. *Remote Sensing of Environment*, 74:249–254, 2000.

M. Grana and M. J. Gallego. Associative morphological memories for spectral unmixing. In *European Symposium on Artificial Neural Networks*, pages 481–486, Apr. 2003.

M. Grana, P. Sussner, and G. Ritter. Associative morphological memories for endmember detection in spectral unmixing. In *The 12th IEEE International Conference on Fuzzy Systems*, volume 2, pages 1285–1290, May 2003.

M. Grana, C. Hernandez, and J. Gallego. A single individual evolutionary strategy for endmember search in hyperspectral images. *Information Sciences*, 161(3-4):181–197, Apr. 2004.

M. Grana, C. Hernandez, and A. d'Anjou. An evolutionary algorithm based on morphological associative memories for endmember selection in hyperspectral images. In M. Grana, R. Duro, A. d'Anjou, and P. P. Wang, editors, *Information Processing with Evolutionary Algorithms*, pages 45–59. Springer Berlin Heidelberg, 2005.

A. A. Green, M. Berman, P. Switzer, and M. D. Craig. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing*, 26:65–73, Jan. 1988.

B. Guo, S. R. Gunn, R. I. Damper, and J. D. B. Nelson. Band selection for hyperspectral image classification using mutual information. *IEEE Geoscience and Remote Sensing Letters*, 3(4):522–526, Oct. 2006.

T. Han, D. G. Goodenough, A. Dyk, and H. Chen. Hyperspectral feature selection for forest classification. In *Proceedings of the IEEE: Geoscience and Remote Sensing Symposium*, pages 1471–1474, 2004.

J. C. Harsanyi and C.-I. Chang. Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach. *IEEE Transactions on Geoscience and Remote Sensing*, 32(4):779–785, July 1994.

S. Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice-Hall, 2 edition, 1999.

G. Heo and P. Gader. KG-FCM: Kernel-based global fuzzy c-means clustering algorithm. Technical report, Department of Computer and Information Science and Engineering, University of Florida, 2008.

R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

R. Huang and M. He. Band selection based on feature weighting for classification of hyperspectral imagery. *IEEE Geoscience and Remote Sensing Letters*, 2(2):156–159, Apr. 2005.

A. Hyvarinen and E. Oja. Independent component analysis: Algorithms and applications. *Neural Networks*, 13:411–430, Mar. 2000.

A. Ifarraguerri and C.-I. Chang. Multispectral and hyperspectral image analysis with convex cones. *IEEE Transactions on Geoscience and Remote Sensing*, 73(2):756–770, Mar. 1999.

S. Jain and R. M. Neal. A split-merge Markov chain Monte Carlo procedure for the Dirichlet process mixture model. Technical Report 2003, University of Toronto, Toronto, ON, Canada, Jul. 2000.

P. Kealy and S. Hook. Separating temperature and emissivity in thermal infrared multispectral scanner data: Implications for recovering land surface temperatures. *IEEE Transactions on Geoscience and Remote Sensing*, 21(10):2127–2132, 2000.

N. Keshava. Best bands selection for detection in hyperspectral processing. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 5, pages 3149–3152, May 2001.

N. Keshava. Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1552–1565, 2004.

N. Keshava and J. F. Mustard. Spectral unmixing. *IEEE Signal Processing Magazine*, 19: 44–57, 2002.

B. Krishnapuram, A. J. Harternik, L. Carin, and M. A. T. Figueiredo. A Bayesian approach to joint feature selection and classifier design. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1105–1111, Sep. 2004.

S. Kumar, J. Ghosh, and M. M. Crawford. Best-bases feature extraction algorithms for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 39(7):1368–1379, July 2001.

D. Lee and H. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999.

D. Lee and H. Seung. Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems 13*, pages 556–562, 2000.

J. B. Lee, A. S. Woodyatt, and M. Berman. Enhancement of high spectral resolution remote-sensing data by a noise-adjusted principal components transform. *IEEE Transactions on Geoscience and Remote Sensing*, 28(3):295–304, May 1990.

H.-D. Lin and L. M. Bruce. Projection pursuits for dimensionality reduction of hyperspectral signals in target recognition applications. In *Proceedings of the IEEE: Geoscience and Remote Sensing Symposium*, pages 960–963, 2004.

C. Liou and K. O. Yang. Unsupervised classification of remote sensing imagery with non-negative matrix factorization. In *Proceedings of Twelfth International Conference on Neural Inf. Proc.*, pages 295–304, 2005.

P.G. Lucey, T. J. Williams, M. Mignard, J. Julian, D. Kobubun, G. Allen, D. Hampton, W. Schaff, M. J. Schlagen, E. M. Winter, W. B. Kendall, A. D. Stocker, K. A. Horton, and A. P. Bowman. Ahi: An airborne long-wage infrared hyperspectral imager. In *Proceedings of the SPIE*, volume 3431, pages 36–43, Nov 1998.

LYNXSAR. General Atomics, San Diengo, CA, URL http://www.LYNXSAR.com.

D. J. C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

D. Manolakis, D. Marden, and G. A. Shaw. Hyperspectral image processing for automatic target detection applications. *Lincoln Laboratory Journal*, 14(1):79–116, 2003.

A. Martinez-Uso, F. Pla, P. Garcia-Sevilla, and J. M. Sotoca. Automatic band selection in multispectral images using mutual information-based clustering. In *Proceedings of the 11th Iberoamerican Congress on Pattern Recognition*, pages 644–654, Cancun, Mexico, Nov. 2006.

A. Martinez-Uso, F. Pla, J. M. Sotoca, and P. Garcia-Sevilla. Clustering-based hyperspectral band selection using information measures. *IEEE Transactions on Geoscience and Remote Sensing*, 45(12):4158–4171, Dec. 2007.

A. Mendez-Vazquez and P. Gader. Sparsity promotion models for the Choquet integral. In *Proceedings of the IEEE: Symposium on Foundations of Computations Intelligence*, pages 454–459, Apr. 2007.

L. Miao and H. Qi. Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization. *IEEE Transactions on Geoscience and Remote Sensing*, 45(3):765–777, Mar. 2007.

L. Miao, H. Qi, and H. Szu. Unsupervised decomposition of mixed pixels using the maximum entropy principle. In *Proceedings of the 18th International Conference On Pattern Recognition*, volume 1, pages 1067–1070, 2006.

D. S. Myers. Hyperspectral endmember detection using morphological autoassociative memories. Masters thesis, University of Florida, Gainesville, Florida, 2005.

J. Nascimento and J. Bioucas-Dias. Blind hyperspectral unmixing. In *Proceedings of the VI Conference on Telecommunications*, pages 617–620, Peniche, Portugal, May 2007a.

J. Nascimento and J. Bioucas-Dias. Hyperspectral unmixing algorithm via dependent component analysis. In *Proceedings of IEEE: Geoscience and Remote Sensing Symposium*, Barcelona, Spain, July 2007b.

J. M. P. Nascimento and J. M. Bioucas-Dias. Does independent component analysis play a role in unmixing hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(1):175–187, Jan. 2005a.

J. M. P. Nascimento and J. M. Bioucas-Dias. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43 (4):898–910, Apr. 2005b.

R. M. Neal. Bayesian mixture modeling by Monte Carlo simulation. Technical Report CRG-TR-91-2, University of Toronto, Toronto, ON, Canada, Jun. 1991.

R. M. Neal. Markov chain sampling methods for Dirichlet process mixture models. Technical Report 9815, University of Toronto, Toronto, ON, Canada, Sep. 1998.

N. Otsu. A threshold selection method from gray-level histograms. *IEEE Trans. Sys., Man., Cyber.*, 9:62–66, 1979.

V. P. Pauca, J. Piper, and R. J. Plemmons. Nonnegative matrix factorization for spectral data analysis. *Linear Algebra Applications*, 416(1):321–331, Jul. 2005.

G. M. Petrie, P. G. Heasler, and T. A. Warner. Optimal band selection strategies for hyperspectral data sets. In *Proceedings of the IEEE: Geoscience and Remote Sensing Symposium 1998: Sensing and Managing the Environment*, pages 1582–1584, 1998.

A. Plaza, P. Martinez, R. Perez, and J. Plazas. Spatial/spectral endmember extraction by multidimensional morphological operators. *IEEE Transactions on Geoscience and Remote Sensing*, 40(9):2025–2041, September 2002.

A. Ranganathan. The Dirichlet process mixture (DPM) model. Internet draft, October 2006. URL http://www.cc.gatech.edu/~ananth/docs/dirichlet.pdf.

C. Rasmussen. The infinite Gaussian mixture model. In S. A. Solla, T. K. Leen, and K. R. Muller, editors, *Advances in Neural Information Processing Systems*, volume 12, pages 554–560. MIT Press, 2000.

M. Riedmann and E. J. Milton. Supervised band selection for optimal use of data from airborne hyperspectral sensors. In *Proceedings of the IEEE: Geoscience and Remote Sensing Symposium*, volume 3, pages 1770–1772, Jul. 2003.

G. X. Ritter and P. D. Gader. Fixed points of lattice transforms and lattice associative memories. *Advances in Imaging and Electron Physics*, 144:165–242, 2006.

G. X. Ritter and G. Urcid. Fast autonomous endmember determination using lattice algebra, 2008. Manuscript submitted for publication.

D. M. Rogge, B. Rivard, J. Zhang, A. Sanchez, J. Harris, and J. Feng. Integration of spatial-spectral information for the improved extraction of endmembers. *Remote Sensing of Environment*, 110:287–303, 2007.

S. B. Serpico and L. Bruzzone. A new search algorithm for feature selection in hyperspectral remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 39(7):1360–1367, July 2001.

Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei. Hierarchical Dirichlet processes. *Technical Report 653, Department of Statistics, UC Berkeley*, 2004.

Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei. Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101:1566–1581, 2006.

J. Theiler and K. Glocer. Sparse linear filters for detection and classification in hyperspectral imagery. In *Proceedings of the SPIE*, volume 6233, page CID:62330H, May 2006.

S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. Academic Press, 2003.

R. Tibshirani. Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society B*, 58(1):267–288, 1996.

M. Tipping. Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, 1:211–244, 2001.

T.-M. Tu. Unsupervised signature extraction and separation in hyperspectral images: A noise-adjusted fast independent components analysis approach. *Optical Engineering*, 39 (4):897–906, 2000.

T.-M. Tu, H.-C. Shyu, Y.-S. Sun, and C.-H. Lee. Determination of data dimensionality in hyperspectral imagery - pnapca. *Multidimensional Systems and Signal Processing*, 10(3): 255–273, July 1999.

T.-M. Tu, P. S. Huang, and P.-Y. Chen. Blind separation of spectral signatures in hyperspectral imagery. In *Proceedings of the IEEE: Vision, Image and Signal Processing*, volume 148, pages 217–226, Aug. 2001.

S. Venkataraman, L. M. Bruce, A. Cheriyadat, and A. Mathur. Hyperspectral dimensionality reduction via localized discriminant bases. In *Proceedings of the IEEE: Geoscience and Remote Sensing Symposium*, pages 1245–1248, July 2005.

D. D. Wackerly, W. Mendenhall, and R. L. Scheaffer. *Mathematical Statistics with Applications*. Wadsworth, 5 edition, 1996.

J. Wang and C.-I. Chang. Applications of independent component analysis in endmember extraction and abundance quantification for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 44(9):2601–2616, September 2006.

L. Wang, Y. Zhang, and Y. Gu. Unsupervised band selection method based on improved N-Findr algorithm for spectral unmixing. In *Proceedings of the 1st ISCAA*, pages 1018–1021, 2006.

M. West, P. Muller, and M. D. Escobar. Hierarchical priors and mixture models with application in regression and density estimation. In P. R. Freeman and A. F. M. Smith, editors, *Aspects of Uncertainty*, pages 363–386. John Wiley, 1994.

D. Whitley. An overview of evolutionary algorithms: Practical issues and common pitfalls. *Information and Software Technology*, 43(14):817–831, Dec. 2001.

P. Williams. Bayesian regularization and pruning using a Laplace prior. *Neural Computation*, 7:117–143, 1995.

E. M. Winter. Detection of surface mines using hyperspectral sensors. In *Proceedings of the IEEE: Geoscience and Remote Sensing Symposium*, pages 1597–1600, Anchorage, AK, September 2004.

M. E. Winter. Fast autonomous spectral end-member determination in hyperspectral data. In *Proceedings of the Thirteenth International Conference on Applied Geologic Remote Sensing*, pages 337–344, Vancouver, B.C., Canada, 1999.

H.-T. Wu, J.-F. Yang, and F.-K. Chen. Source number estimators using transformed Gerschgorin radii. *IEEE Transactions on Signal Processing*, 43(6):1325–1333, June 1995.

E. P. Xing, M. I. Jordan, and R. Sharan. Bayesian haplotype inference via the Dirichlet process. *Journal of Computational Biology*, 14:267–284, 2007.

X. Yu, S. Reed, and A. D. Stocker. Comparitive performance analysis of adaptive multispectral detectors. *IEEE Transactions on Signal Processing*, 41(8):2639–2656, August 1993.

A. Zare and P. Gader. Sparsity promoting iterated constrained endmember detection for hyperspectral imagery. *IEEE Geoscience and Remote Sensing Letters*, 4(3):446–450, July 2007a.

A. Zare and P. Gader. SPICE: A sparsity promoting iterated constrained endmember extraction algorithm with applications to landmine detection from hyperspectral imagery. In *Proceedings of the SPIE*, page CID:655319, Orlando, Fl, 2007b.

A. Zare and P. Gader. Hyperspectral band selection and endmember detection using sparsity promoting priors. *IEEE Geoscience and Remote Sensing Letters*, 5(2):256–260, Apr. 2008.

A. Zare, J. Bolton, P. Gader, and M. Schatten. Vegetation mapping for landmine detection using long wave hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 46(1):172–178, January 2008.

BIOGRAPHICAL SKETCH

Alina Zare received her Bachelor of Science degree in computer engineering from the University of Florida in 2003. She continued her studies at the University of Florida to graduate with her Master of Science and Doctor of Philosophy degrees from the department of Computer and Information Science and Engineering in 2008. Her research interests include sparsity promotion, machine learning, Bayesian methods, image analysis and pattern recognition.

RANDOM SET FRAMEWORK FOR CONTEXT-BASED CLASSIFICATION

By

JEREMY BOLTON

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2008

© 2008 Jeremy Bolton

ACKNOWLEDGMENTS

I thank my mother, Lois Bolton, father, Wade Bolton, and sister, Chelsea Bolton for their relentless love and support.

I thank my advisor, Paul Gader, for his guidance and encouragement throughout my tenure at the University of Florida. I thank my committee Paul Gader, Joseph Wilson, Gerhard Ritter, Arunava Banerjee, and Clint Slatton for their insight and guidance which has steered my research and bettered resulting contributions.

I thank my lab mates for their support and am thankful for their ability to endure my shenanigans. I thank Alina Zare, Nathan VanderKraats, Nicholas Fisher, Xuping Zhang**,** Raazia Mazhar, Wen-Hsiung Lee and Seniha Esen Yuksel for their encouragement, suggestions and aid in my research.

I thank colleagues, Jim Keller, Hishem Frigui, and Dominic Ho, for their collaboration on a variety of research projects.

I thank William Clark of Army Research Office (ARO), Russell Harmon of ARO, Miranda Schatten of Night Vision and Electronic Sensors Directorate, and Michael Cathcart of Georgia Tech, for their support of my research.

TABLE OF CONTENTS

page

ACKNOWLEDGMENTS 3

LIST OF TABLES 6

LIST OF FIGURES 7

ABSTRACT 8

CHAPTER

1 INTRODUCTION 10

Problem Statement and Motivation 10
Proposed Solution 12

2 LITERATURE REVIEW 16

Concept Drift 16
The Problem of Concept Drift 17
Concept Drift Solutions 18
Instance selection 19
Instance weighting 21
Ensemble learning 23
Applications to Hyperspectral Imagery 26
Probability Introduction 29
Topology 29
Probability Space 30
Measure 31
Standard Random Variables 32
Standard Statistical Approaches for Context Estimation 33
Random Sets 34
General Case: Random Closed Set 34
Random Set Discussion 35
Theory of Evidence 39
Point Process 40
Random Measures 45
Variational Methods 46
Set Similarity Measures 48
Random Set Applications 51
Point Process Applications 51
En Masse Context-Based Methods 53

3 TECHNICAL APPROACH 55

Mathematical Basis of the Random Set Framework 55
Possibilistic Approach 57
Development 58
Dependent Optimization 61
Independent Optimization 63
Evidential Model 63
Development 64
Optimization 64
Probabilistic Model 65
Development 65
Optimization 68
Discussion 73

4 EXPERIMENTAL RESULTS 85

KL Estimation Experiment 86
Experimental Design 86
Results 88
Synthetic Data Experiment 89
Experimental Design 90
Results 92
Hyperspectral Data Experiment 94
Experimental Design 95
Results 97
Upper and Lower Bounding Experiment 99
Experimental Design 100
Results 101

5 CONCLUSIONS 116

LIST OF REFERENCES 119

BIOGRAPHICAL SKETCH 127

LIST OF TABLES

Table page

4-1. Average inference error for each dataset using 15 test and 15 train samples.....103

4-2. Average classification error of the listed context-based classifiers on four data sets used in the Synthetic Data Experiments.....106

4-3. How classification varies with respect to the number of germ and grain pairs for data set 3 (with no outlying samples) in the Synthetic Data Experiment. 106

LIST OF FIGURES

Figure page

1-1. Spectral samples exhibiting contextual transformations. 15

1-2. Illustration of contextual transformations in a feature space.....15

3-1. Samples of Gaussian distributions drawn using randomly selected means and variances which where drawn uniformly from a specified interval.....81

3-2. Learning the representative function using update Equations.....82

3-3. Similarities and distinctions between the proposed method and standard methods.....83

4-1. Illustration of data sets one, two, and three. 103

4-2. Error analysis of the Riemann and uniform approximation methods with respect to time and number of observation samples.....104

4-3. Trials using data sets 1, 2, 3 and 4 in the Synthetic Data Experiment. 105

4-4. ROC curve for The Hyperspectral Data Experiment. Note the dashed plot is the results from the probabilistic context-based classifier using the analytical solution for KL estimation as discussed in Equation 3-40. 107

4-5. Hyperspectral Experiment ROC curve of PD versus PFA for the possibilistic, evidential probabilistic, set-based kNN, and whiten / dewhiten approaches 108

4-6. Example of a false alarm POI from The Hyperspectral Data Experiment. 109

4-7. Example of a target alarm POI, from The Hyperspectral Data Experiment.....110

4-8. Example of a target alarm POI from The Hyperspectral Data Experiment.....111

4-9. Example of a false alarm POI from The Hyperspectral Data Experiment. 112

4-10. Example of a false alarm POI from The Hyperspectral Data Experiment 113

4-11. Detection results for the possibilistic RSF classifier and results for standard Gaussian mixture classifiers equipped with variable numbers of mixture components.....114

4-12. Non-crossvalidation detection results for the possibilistic RSF classifier and the oracle classifier. 115

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

RANDOM SET FRAMEWORK FOR CONTEXT-BASED CLASSIFICATION

By

Jeremy Bolton

December 2008

Chair: Paul Gader
Major: Computer Engineering

Pattern classification is a fundamental problem in intelligent systems design. Many different probabilistic, evidential, graphical, spatial-partitioning and heuristic models have been developed to automate classification. In some applications, there are unknown, overlooked, and disregarded factors that contribute to the data distribution, such as environmental conditions, which hinder classification.

Most approaches do not account for these conditions, or factors, that may be correlated with sets of data samples. However, unknown or ignored factors may severely change the data distribution making it difficult to use standard classification techniques. Even if these variable factors are known, there may be a large number of them. Enumerating these variable factors as parameters in clustering or classification models can lead to the *curse of high dimensionality* or sparse random variable densities. Some Bayesian approaches that integrate out unknown parameters can be extremely time consuming, may require *a priori* information, and are not suited for the problem at hand. Better methods for incorporating the uncertainty due to these factors are needed.

We propose a novel *context-based* approach for classification within a random set framework. The proposed model estimates the posterior probability of a class and context given

both a sample a set of samples, as opposed to the standard method of estimating the posterior given a sample. This conditioned posterior is then expressed in terms of priors, likelihood functions and probabilities involving both a sample and a set of samples. Particular attention is focused on the problem of estimating the likelihood of a set of samples given a context. This estimation problem is framed in a novel way using random sets. Three methods are proposed for performing the estimation: possibilistic, evidential, and probabilistic. These methods are compared and contrasted with each other and with existing approaches on both synthetic data and extensive hyperspectral data sets used for minefield detection algorithm development.

Results on synthetic data sets identify the pros and cons of the possibilistic, evidential and probabilistic approaches and existing approaches. Results on hyperspectral data sets in indicate that the proposed context-based classifiers perform better than some state-of-the-art, context-based and statistical approaches.

CHAPTER 1
INTRODUCTION

Problem Statement and Motivation

When collecting data, many known and unknown factors transform the observed data distribution. In many applications, sets of samples are collected at a given time, for example, remote sensing. In remotely sensed imagery, images are taken from a remote location such as a plane. These images are essentially sets of pixels, or samples, that are collected at the same time. In this instance, many of the unknown or unspecified factors may influence all of the samples in the image, or some subset thereof, similarly. That is, all of the samples in an image subset may undergo the same transformation induced by these factors.

Optical character recognition (OCR) is another application where factors may influence the results of classification. In OCR, if a classifier could identify a font or font size of a document, the problem of character recognition may be simplified. In this problem, the font or font size is a factor, or context, which may change the appearance of the sample, or the character.

Before we fully characterize the problem at hand, we state some assumptions and define a few terms which are necessary for the problem statement. We assume that similar samples collected in similar conditions or situations will undergo similar transformations. We define a *population* as a set of samples collected under the same conditions or situation. We define the idea of *context* as the surrounding conditions or situations in which data are collected. We define *contextual factors* as the unknown or unspecified factors that transform the data's appearance. Given these definitions, we can define a *contextual transformation* as a transformation that acts on sets of samples on a context-by-context basis. We attempt to estimate a population's context using the observed population's distribution.

In a probabilistic approach, context can be viewed as hidden random variables that are correlated with the observed samples. This view implies that the observed samples are dependent on these hidden variables.

In many standard models, classification accuracy suffers due to contextual factors. If these variables are ignored, many classification methods will suffer since the sample values may be severely altered by contextual transformations. On the other hand, if their values are specified and corresponding parameters are enumerated in a model, problems such as the curse of dimensionality or sparse probability distributions may hinder classification results.

Example 1.1 Contextual transformations: In this example, we illustrate that contextual factors are present in remotely sensed *hyperspectral imagery* (HSI) collected by *airborne hyperspectral imager* (AHI). In this data, each pixel in an image has a corresponding spectral vector, or *spectral signature*, with intensity values in the *long wave infrared* (LWIR), 7.8 um to 11.02 um. Each spectral signature is usually viewed as a plot of wavelength vs. intensity. Figure 1-1A illustrates multiple spectral signatures, or *spectra*, from a target class and a non-target class indicated by a solid line and a dashed line, respectively. Two consequences of contextual transformations can hinder classification. The first problem is the obvious change in sample appearance in varying contexts, which we refer to as a *non-disguising transformation*. An algorithm must know the appearance of a target sample for identification; therefore, if a target can potentially take on multiple appearances then a classifier must be aware of all potential appearances. The second problem occurs when samples from one class, in some context, are transformed to appear as samples from another class in another context, which we refer to as *disguising transformations*. We characterize these problems separately since their solutions require different approaches.

Solutions to non-disguising transformations require knowledge of the various target class appearances. An algorithm developer could simply add model constructs or parameters to account for varying appearances. For example, a developer could add densities to a mixture model to account for multiple appearances due to multiple transformations. However, this solution will not resolve the problem of disguising transformations since samples from different classes have the same appearance. In this situation, context estimation is used to identify relevant models that were constructed for similar contexts that our test population has been observed and thereby disregarding models or parameters constructed for irrelevant contexts.

Assume we want to classify the bolded spectral signature shown in Figure 1-1A. Classification is difficult since this spectral vector has the same appearance as some target and non-target spectra from various contexts. However, if we disregarded the spectra collected in a different context, classification becomes less complicated as illustrated in Figure 1-1B.

Example 1.2 Feature space transformation: Suppose we have images of scenes containing pixels with values in $\mathcal{R}^n$. For the sake of illustration, we assume n=2 and each image X, has a continuum of pixels. Each of these pixels corresponds to a measurement of some object in the real world. We would assume that the pixel's value would depend on the object it represents in the real world, but there are contextual factors that will influence the pixels' values.

In this example, there are five images containing pixels that represent two objects in the real world, 'x' and 'o'. Some of these images were taken in different contexts thus each is affected by different influencing factors. These contextual factors transform the data collected in distinct contexts, differently. These transformations may cause sets of samples to have different spatial distributions, or shapes, in a feature or sample space as shown in Figure 1-2A.

Assume the goal is to label some samples in $X1$, denoted by '*', using some labeled samples from the other images illustrated in Figure 1-2B. If we ignore the population information, the classification problem becomes more difficult as shown in Figure 1-2C. Instead, if we emphasize, to an algorithm, datasets which appear to have been collected in a similar context, the job of classification may be simplified, as shown in Figure 1-2D. A similar spatial distribution of sets may indicate that a similar transformation has acted on the populations and have therefore been collected in similar conditions. We propose that if this contextual information is gathered and utilized correctly, classification results should improve.

Proposed Solution

The problem of variable contextual factors is similar to some existing problems such as concept drift where the idea of a target class and/or its governing distribution may change with respect to time or some hidden context. In Example **1.2** , a solution would need to include a method for determining a similar distribution, or shape, relationship between populations. A more general solution would provide a method for modeling the shape of populations from a particular context.

Standard context-based classifiers suffer from a number of limitations. Most notably, they lack the ability to solve the problem of disguising transformations, as mentioned in Example **1.2**. Many classifiers attempt to estimate context, which we propose is best identified by analysis of an entire population, by inspecting a single sample. Many existing models also suffer from restrictions, inappropriate assumptions, and the lack of ability to handle all forms of concept drift. Most standard statistical methods make the *independently identically distributed* (i.i.d.)

assumption that limits their ability to capture any information found through the analysis of the set of samples.

The proposed solution uses a random set [1]-[7] model for population context estimation. A population's context is then considered when each sample of the population is classified. This model has the ability to estimate context by inspecting the distribution of a set of samples. Populations, after undergoing contextual transformations induced by contextual factors, are compared to contextual models—modeled using random sets—in attempts to identify the context in which they were collected. Specifically, the creation of the proposed context-based classifier consists of factors for context estimation and class estimation. The classification factor will estimate the class of each sample using class models, one for each context. The context estimation factor will identify the relevance of each model based on the estimated context of the test population and subsequently weight each model's contribution by contextual relevance. The identification of context allows for more informed class estimation emphasizing models relevant to the test population's context and ignoring the irrelevant models.

Note that the proposed model implicitly acquires context of a sample set without explicitly performing any estimation of the contextual factors. A subsequent benefit to this approach is that it avoids the curse of high dimensionality and sparse densities, which are potential pitfalls of methods that would directly account for these contextual factors.

The proposed random set model allows for evidential, probabilistic, and possibilistic approaches due to the inherent versatility of the random set. Furthermore, it also has the ability to avoid the aforementioned limitations and to handle all forms of concept drift. Existing standard and state-of-the-art methods are surveyed, analyzed, and compared to the proposed approach.

Results from experiments indicate that the proposed random set model improves classification results from existing methods in the face of hidden contexts.

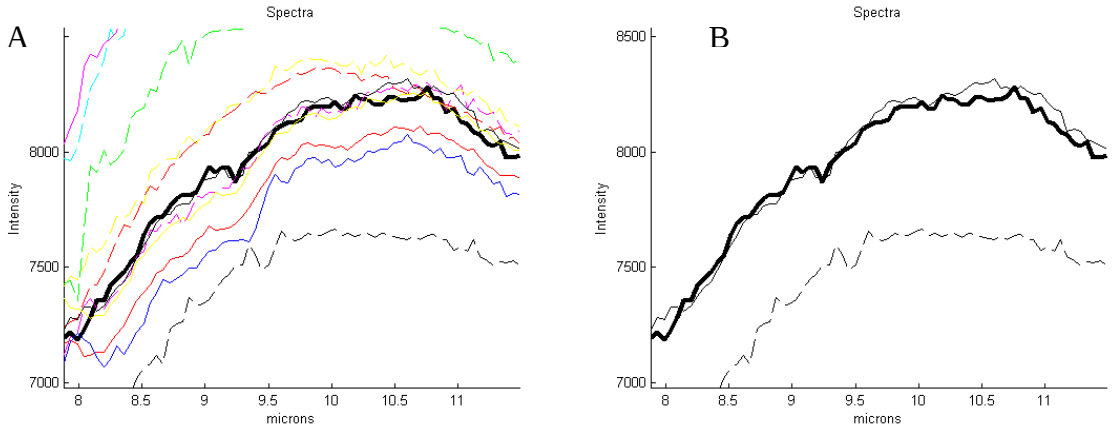


Figure 1-1. Spectral samples exhibiting contextual transformations. A) Spectra from target and non-target classes collected by AHI in multiple contexts. The target class is indicated by a solid line and a non-target class is indicated by a dashed line. B) An unlabeled sample shown in bold along with two labeled samples collected in the same context.

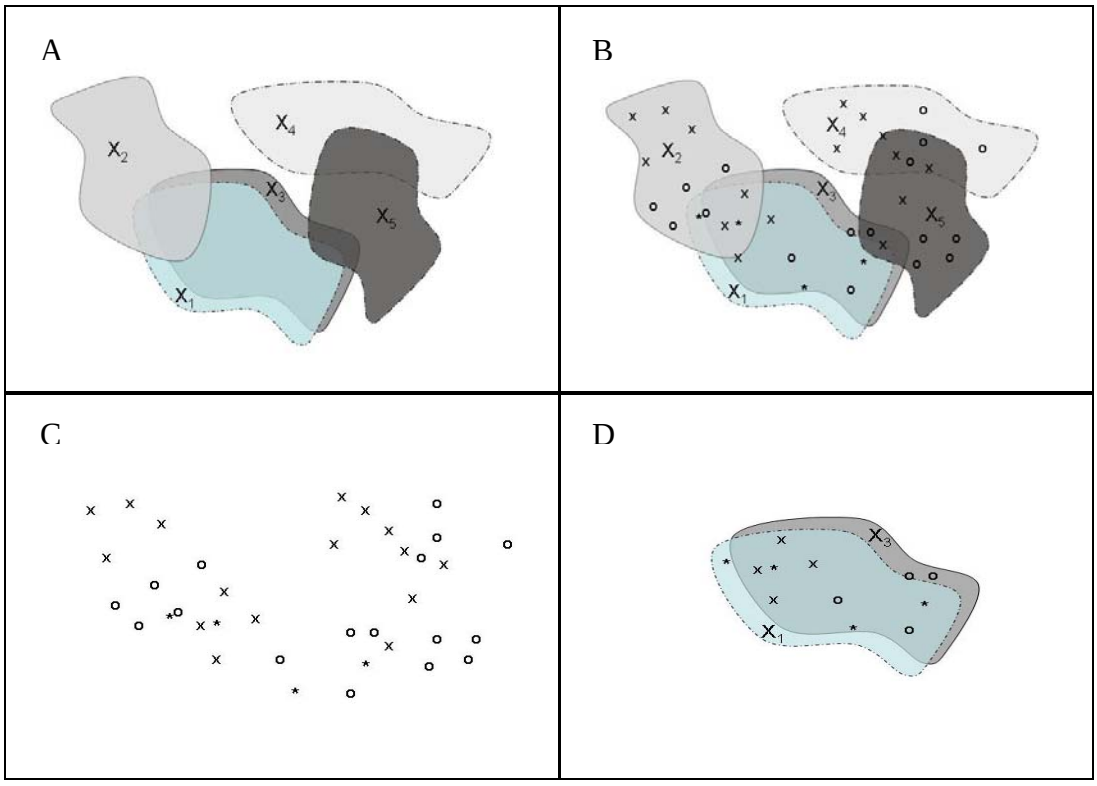


Figure 1-2. Illustration of contextual transformations in a feature space. A) Five images in some feature space that is a subset of $\mathcal{R}^2$. B) Labeled samples from each training image and unlabeled samples from the test image. C) All samples without contextual information. D) Using a similarly distributed training image to label the samples in the test image.

CHAPTER 2
LITERATURE REVIEW

The following is a review of current literature pertinent to problems and solutions arising from contextual factors. First, the problem of concept drift is detailed along with standard and state-of-the-art solutions [12]-[58]. Next, a brief review of context-based approaches with applications to hyperspectral imagery is given [59]-[67]. Next, a brief mathematical and statistical review is given to assist in the development of the proposed random set framework [1]-[11]. Standard statistical methods are reviewed and their potential uses for context-estimation are developed. Through the development we indicate that alternative methods may model the idea of context better than standard approaches. Next, the random set is defined and introduced as a method better suited for context estimation [1]-[7]. This is followed by a few examples of set similarity measures, which are reviewed to assist in set analysis [69]-[72]. Next, we review of some existing formulations and applications of random sets. Finally, we review some state-of-the-art, en masse, context-based approaches, which treat sets as unitary elements for context-estimation.

Concept Drift

The idea that samples of a class may change with respect to time is an area of recent research. We begin our discussion with a benchmark solution to this problem. One of the first algorithms developed to analyze and contend with this occurrence is STAGGER, which was developed by Schlimmer and Granger, and is based on a psychological and mathematical foundation [24]. STAGGER has 4 major steps: initialization, projection, evaluation, and refinement. In initialization, the description of a concept or class is constructed using a set of pairs consisting of logical statements, or characterizations, used to describe a class and corresponding weights used to weight the importance of each description. In this step, the

concept is specified. In projection, a Bayesian scheme is implemented to estimate the frequency of occurrences of the characterizations in subsequent samples. These probabilities are updated after the class of a new sample is determined. In this step, new samples are inspected to determine if frequency or weighting of each characterization is representative of the data. In evaluation, the effectiveness of each characterization is determined based on the number of correct and incorrect predictions for each characterization. In this step, the concept characterizations are evaluated to determine if there should be a change in these concept characterizations. In refinement, the characterizations and corresponding weights are modified based on their evaluations to improve their effectiveness as predictors.

The Problem of Concept Drift

STAGGER is one approach that contends with the change of concepts with respect to time or some hidden context. One of the more popular formulations of this problem, concept drift, has recently become an area of much research [18]-[57]. In concept drift, a concept may depend on some hidden context which is not given explicitly. Changes in the hidden context then induce changes in our target concept. This principle has been adopted by researchers in the machine learning community and has many applications in scientific research. Solutions to the problem should be able to adjust for concept drift, distinguish noise from concept drift and recognize and adjust for repeat concepts [18].

Concept drift can be divided into two categories: real and virtual. In real concept drift, the concept or idea of a target class may change. In virtual concept drift, the data distribution for a target class may change. The former is truly a *concept shift*—a change in concept—whereas the latter is simply a *sampling shift*—a change of data distribution due to some unknown context or variables. The idea of virtual concept drift is similar to our problem of hidden, population-

correlated variables, since this may lead to a change in data distribution due to some hidden context.

Concept drift can also be categorized as sudden or gradual. In sudden concept drift, the drift may be abrupt and substantial; whereas in gradual concept drift, the drift may be gradual and minimal. The problem at hand can be described as abrupt or sudden concept drift. The developed model allows for data to be collected at variable times and may not necessarily be a continuous flow of data with respect to time; in fact, the drift may be fairly substantial.

Concept Drift Solutions

There are three major approaches that are used to account for concept drift: instance selection, instance weighting and ensemble learning. In instance selection the goal is to select relevant samples from some training set for use in classifying test samples. A simple example of this approach would be windowing—using sliding windows or *k nearest neighbors* (kNN) [22], [23], [25]-[30]. Instance weighting involves weighting instances of a training set based on their relevance. Usually in instance weighting a learning algorithm is trained to appropriately weight these instances such as boosting [31]-[33], [39], [40]. In ensemble learning, a set of concept descriptions are maintained and some combination of these descriptions are used to predict current descriptions, as in STAGGER. This general approach could also be interpreted as some sort of model selection where the set concept descriptions are in fact models or algorithms whose results are to be combined based on each concept descriptions' relevance to a certain population [21], [24], [34]-[58].

In existing concept drift solutions, there are a number of restrictions, assumptions, and limitations that induce models that will not be able to account for all contextual transformations. Furthermore, almost all existing context-based solutions cannot solve the problem of disguising transformations as defined in Example 1.2. This drawback is due to the fact that context

estimation is performed by inspecting one sample, rather than the entire population. There are five major limitations or pitfalls exhibited by existing concept drift algorithms.

1. Estimates context based on a single sample (C.1)
2. Recognizes only some forms of concept drift (C.2)
3. Identifies context arbitrarily or with major assumptions (C.3)
4. Admits solutions that are not robust to outliers (C.4)
5. Assumes semi-supervised environment (C.5)

We emphasize property C.1 since this is a conceptual flaw implemented by many concept drift algorithms. This assumption presumes that the situation discussed in Example 1.2, disguising transformations, will not occur. Next, we survey standard and state-of-the-art approaches to concept drift. In the following, we parenthetically indicate where properties C.1 – C.5 are observed by the surveyed approaches. In almost all existing approaches, C.1 is present except when the approach is highly supervised and makes major assumptions for context identification.

Instance selection

In *full memory approaches*, all training samples are kept but a subset are selected to classify a given test sample. The process by which these samples are selected is the crux of instance selection approaches.

Widmer proposed the choice of a dynamic window size that is chosen based on time and classifier performance [30]. If the classifier is performing well, it is assumed that the concept has been constant for some time and a large window of samples are retained (C.2 and C.5). However, if performance decreases, it is assumed the concept is changing or has changed and the window size is shrunk (C.3 and C.4).

Klinkenberg *et al*. proposed an instance selection approach, where a variable sized window is kept over the *m* most recent training samples, assuming that the last *m* samples will be reflective of new test samples (C.2) [33], [34]. The selected window size minimizes the error of a support vector machine that is trained using the last *h* training samples. After the SVM is trained

using the last *h* samples, an upper bound on the error can be directly estimated from the SVM parameters [28], [7].

After these *m* SVMs have been trained on their last *h* samples, the training set with least error is selected. The window size is set to *h* as in Equation 2-1 and the corresponding training samples are used to classify the next test set.

$$h = \arg\min Err_{\xi\alpha}^{m}(h) \tag{2-1}$$

Here the SVM is used for an upper bound error estimate, and when its estimate increases, a change in context is assumed (C.3 and C.4).

Salganicoff proposed *Darling* which retains a selected sample until new samples are presented which occupy a similar subspace of the sample space [22]. This approach assumes context changes are directly related to the sequence of observance and that context is selected based on a single sample (C.1 and C.3).

Maloof *et al.* proposed an instance selection approach which is similar in ideology to instance weighting methods [26], [27]. In *partial-memory approaches*, each classification decision is made using some current characterization of a class and some subset of previously observed samples. The term partial-memory refers to the fact that only a subset of previously observed samples is retained to assist in classification and concept updating. Specifically in this method, the concept descriptions are updated using *selected* samples and misclassified samples [26]. Given a classifier *C*, a data set *D*, and a partial memory *P*, the update procedure consists of six major steps.

1. P={}
2. Classify *D* with *C*
3. Add misclassified samples to *P*
4. Retrain *C* using *P*
5. Select *appropriate P*
6. Repeat from step 2 when presented with new Data *D*'

Note that the classifier focuses on samples that it is misclassifying assumed to be due to concept drift (C.5 and C.4). An example of how to select an appropriate set *P* is to retain particular samples if they help form the decision boundary. One selection technique AQ-PM, which assumes a convex data set, identifies extreme points such as the points forming a covering hyperrectangle thus enclosing, or bounding, particular samples.

Instance weighting

Instance weighting approaches weight certain samples differently for the purposes of classification. A popular instance weighting scheme is boosting. A popular boosting algorithm is Adaptive Boosting, or *AdaBoost*, where misclassified samples are emphasized during parameter learning stage in a statistical manner [30]-[33]. The error term is calculated as follows:

$$\varepsilon_t = \sum_{i=1}^{n} D_t(i)[y_i \neq C_t(x_i)]. \tag{2-2}$$

In Equation 2-2, *t* is the learning iteration, x_i is sample *i*, $y_i \in \{-1,1\}$ is the class for x_i, C_t is the classifier at iteration *t*, $D_t(i)$ is the weight for sample x_i at iteration *t*, and ε_t is the average misclassification at iteration *t.* If the classifier misclassifies some samples, assumedly due to concept drift (C.3and C.5), the misclassified samples are emphasized (C.4) in the error term using the weight update formula.

$$D_{t+1}(i) = \frac{D_t(i)\exp\left(-\alpha y C_t(x_i)\right)}{Z_{t+1}} \tag{2-3}$$

This update increases the weights of misclassified samples to coerce the learning of the new concept in later iterations. Note this is similar to increasing the prior of x_i, in the statistical sense. Note if the boosting is done offline, just during training, this approach no longer exhibits property C.5, and maybe not C.4; however, it will exhibit property C.1.

Dura, Lui, Zhang, and Carin proposed neighborhood-based classifiers where a test sample's neighborhood is used for classification [35]-[38]. This approach uses and *active learning* framework which attempts to extract information from some dataset and extend it to another sample under test (C.1). Classification is performed as shown in Equation 2-4.

$$p(y_i \mid N(\mathbf{x}_i), \boldsymbol{\theta}) = \sum_{j=1}^{n} b_{ij}\, p(y_i \mid \mathbf{x}_j, \boldsymbol{\theta}), \;\; p(y_i \mid \mathbf{x}_j, \boldsymbol{\theta}) = \frac{1}{1+\exp\left(-y_i \boldsymbol{\theta}^T \mathbf{x}_j\right)}. \tag{2-4}$$

In Equation 2-4, $y_i \in \{-1,1\}$ is a class label, $\mathbf{x}_i$ is a test sample, $\mathbf{x}_j$'s are retained samples that are in the neighborhood $N(\mathbf{x}_i)$, b_{ij}'s are the weights for each neighbor, and $\boldsymbol{\theta}$ is a parameter vector. The construction of b_{ij}, the weight, and $N(\mathbf{x}_i)$, the neighborhood, are the crux of this algorithm. A few suggestions are shown in Equations 2-5 and 2-6.

$$N_t(\mathbf{x}_i) = \{(\mathbf{x}_j, b_{ij}) : b_{ij} > 0, \mathbf{x}_j \in X\} \tag{2-5}$$

$$\text{where } b_{ij}^t = \frac{\exp(-.5\frac{\left\|\mathbf{x}_i - \mathbf{x}_j\right\|}{\sigma_i^2})}{\sum_{k=1}^{n} \exp(-.5\frac{\left\|\mathbf{x}_i - \mathbf{x}_j\right\|}{\sigma_i^2})}. \tag{2-6}$$

In Equations 2-5 and 2-6, b_{ij} is the transition probability from $\mathbf{x}_i$ to $\mathbf{x}_j$ in less than *t* steps in Markov random walks [36], [37].

In some of their other proposed methods, an information theoretic approach is taken to construct $N(\mathbf{x}_i)$ based on maximizing the determinant of the Fisher information matrix [35], [38]. Note this approach also exhibits property C.1 since each sample is classified using itself and training data, not its population.

Note since the parameter $\boldsymbol{\theta}$ doesn't vary, we assume there is only one concept descriptor, which is why we consider this an instance weighting approach. We note that approach could also be implemented using an ensemble learning approach.

Ensemble learning

In ensemble learning an ensemble of concept descriptions, such as classifiers, are maintained and used in harmony for classification. A popular approach, ensemble integration, employs a weighted scheme to determine the relevance of each classifiers output given a sample [41].

$$C(x_i) = \sum_{j=1}^{J} w_{ij} C_j(x_i) \tag{2-7}$$

Here the construction of the weight w_{ij} is done to emphasize classifiers of greater contextual relevance. Equation 2-7 can be implemented in many ways such as static voting/weighting or dynamic voting/weighting [39]-[58]. In ensemble approaches, the crux of the problem is deciding how to weight each context-based model.

The popular *bagging* approach constructs *N* classifiers where each are trained using *N* corresponding training sets [43]. The training sets are constructed by randomly sampling the entire training set with replacement. Each of the sampled training sets contains *m* samples where *m* is less than the number of total training samples. The classifiers, which act on individual samples, are then combined using voting and averaging techniques (C.1 and C.3).

The random forest model is a new approach using dynamic classifier integration [44], [45], [47]. This model attempts to minimize correlation between the individual classifiers while maintaining accuracy [43], [44]. Random subspaces and/or subsets of samples are chosen and a classifier, or tree, is trained using the corresponding samples (C.3). This is repeated *N* times to create a forest of *N* trees. Most of the time the classifiers are simply partitionings of the space resulting in boolean classification. Given a test sample, classification is determined by weighting each tree's confidence using the confidences of neighboring samples in the feature space (C.1 or

C.3 and C.5 depending on implementation). The weight w_i for tree *i* is assigned using Equation 2-8.

$$w_i(x) = \frac{\sum_{j=1}^{k} 1_{OOB_i}(x_j)\phi(x, x_j)mr_i(x_j)}{\sum_{j=1}^{k} 1_{OOB_i}(x_j)\phi(x, x_j)} \tag{2-8}$$

In Equation 2-8, $mr_i(x_j) \in \{-1,1\}$ indicates whether classifier *i* has correctly classified sample *j*, ϕ is a weighting function based on distance, *k* is the size of the neighborhood, and 1_{OOB_i} is the indicator function, which indicates whether its argument is an *out-of-bag* (OOB) sample—a sample not used to train classifier *i*.

The use of OOB samples allows for unbiased estimates. We note that given some assumptions, the random forest approach is shown to perform at least as well as boosting and bagging [44].

Tsymbal *et al.* proposed an ensemble approach that maintains a set of models optimized over different time periods to handle local concept drift (C.2) [21], [39]. The models predictions are then combined, in a sense integrating over classifiers. The selection of classifier predictions is done based on a local classification error estimate performed after initial training. During testing, *k* nearest neighbors of each test sample are used to predict the local classification errors of each classifier (C.1). Using these estimated errors, each classifier's predictions are weighted and the total prediction is calculated using integration.

Kuncheva and Santana *et al.* developed an ensemble approach where contexts or training sets are constructed by clustering the training data [48], [49]. Then for each cluster, *N* classifiers are ranked such that each has a ranking in each cluster—set of samples. The weights for

combination are proportional to the classifiers correct classification. A test sample is then classified using the *k* best classifiers from the sample subspace in which it resides (C.1).

Frigui *et al.* used fuzzy clustering methods to partition a feature space into assumed contexts [52]. During classification, the models representing a context in which a test sample lies are used for classification where the classifiers are weighted by the corresponding fuzzy memberships of the test sample to the fuzzy cluster (C.1).

Harries *et al.* proposed an algorithm to learn hidden contexts called *Splice* [57], [58]. In this algorithm, a continuous dataset is partitioned, heuristically, into time intervals which supposedly represent partial contexts. Classifiers are then trained and ranked on each interval. The intervals, and classifiers, are then clustered similarly to an agglomerative clustering algorithm. If a classifier performs well on multiple contexts, the corresponding contexts and classifiers are merged and the classifiers are re-ranked based classification results. The weights are then selected similarly to the approaches proposed by Kuncheva and Santana *et al.* (C.1) [48], [49].

Santos *et al.* proposed a subsetting algorithm that randomly creates subsets of the training data (C.3) [50]. A classifier is trained on each subset, assumed to be indicative of a context, and a genetic algorithm selection scheme is used to select the best fit classifiers, where fitness is based on error rate, cardinality, and diversity. Context models are then weighted based on which subset a test sample resides (C.1).

Qi and Picard proposed a context-sensitive Bayesian learning algorithm that models each training set as a component in a mixture of Gaussians [55]. In this model each training set, or context, has a corresponding linear classifier.

$$p(y \mid x, D) = \sum_{i \in I} p(y \mid x, D_i)\, p(D_i \mid x, D) \tag{2-9}$$

In Equation 2-9, y is the class label for sample x using training dataset $D_i \in D = \{D_1,...,D_I\}$. The term $p(y \mid x, D_i)$ is estimated using the expectation propagation method [56]. Note the data set weights are chosen based solely on the sample x and not the sample and its population (C.1). Also, note that each D_i are training sets and not necessarily the population of sample x.

In the proposed random set model for context based classification, test sets are used to estimate context which alleviates property C.1, and furthermore does not induce properties C.2-C.5.

Applications to Hyperspectral Imagery

In the experiments, the proposed methods are tested using a hyperspectral dataset with apparent contextual factors. For this reason, we briefly discuss current, state-of-the-art methods used to contend with contextual factors in hyperspectral imagery. We note that some methods take different approaches or assume a different testing environment.

There are two major approaches for solutions to contextual transformations in hyperspectral data classification. The first approach relies on physical modeling using environmental information. The other uses statistical and/or mathematical methods to identify or mitigate the effects of contextual transformations. Next, we list some popular existing approaches which have shown to be successful in some testing situations.

There has been much research that uses the physical modeling of the environmental factors on measured data. Here, classifiers may use the output of physical models, for example MODTRAN, which generate the appearance of target spectra in certain environments [59], [60]. For example, the hybrid detectors developed by Broadwater use target spectra that are estimated using MODTRAN, which is given environmental information about the scene [61]. This

approach, and many like it, are shown to be very successful when environmental conditions are available.

Healy *et al.* proposed to use MODTRAN to produce spectra of various materials in various environmental conditions [62]. A vector subspace for each material is then defined by selecting an orthonormal basis for the material subspace. Confidence is then assigned to test spectra based on their distance to this subspace. This approach provides a robust and intuitive solution; however, this classification method will suffer in the presence of disguising transformations.

Kuan *et al.* proposed a projection matrix, rooted in a physics-based linear reflectance model, which in effect normalizes environmental conditions between two images [63]. This approach has shown to be successful at identifying regions of images and detecting change in co-registered imagery. This approach can learn a transformation of a set of samples; however, this approach requires a fairly large number of test sample labels be known for the construction of the transformation matrix.

Fuehrer *et al.* proposed the use of atmospheric sampling where a sample of some material is projected into some feature space based on atmospheric conditions in which it was observed [64]. Samples in this feature space may then be used to assist, using locality analysis, in identifying material and atmosphere when presented with a test image. This method has shown to be successful at classification and modeling; however, it cannot account for disguising transformations.

In these approaches, environmental conditions of a scene are assumed to be known *a priori*, or some ground truth is assumed to be known *a priori*, which may not be the case. In these other cases, different approaches must be taken.

The other tactic of existing methods uses various statistical and mathematical approaches to account for contextual transformations. Some selection, ensemble, and context-based methods attempt to identify models relevant to a test sample through context estimation. Some active learning approaches attempt to transfer knowledge to test samples.

Mayer *et al.* propose the whitening / dewhitening transformation. In this approach, transformation matrices are constructed to whiten and dewhiten spectra from an image [65]. In this approach, the whitening and dewhitening matrices are constructed to whiten the effects of environmental conditions. However, this approach requires a semi-supervised testing environment to construct the projection matrix. It also assumes that whitening of spectra will reduce or eliminate the effects of contextual factors. This assumption implies that the contextual transformation is simply a linear transformation based on a population's statistical properties, such as the mean and covariance. Mayer proposes the matched filter described in Equation 2-10.

$$MF_k = \left(x_{t,k} - \bar{x}_t\right)^T R_{22}^{-1}\left(L_t^{Transform} - \bar{x}_t\right), \quad \text{where } L_t^{Transform} - \bar{x}_t = R_{tt}^{1/2} R_{11}^{-1/2}(L_1 - \bar{x}_1). \tag{2-10}$$

In Equation 2-10, $x_{t,k}$ is a test sample, $\bar{x}_1$ is the mean of clutter samples from labeled image 1, $\bar{x}_t$ is the mean clutter estimate from the test image, L_1 is the target estimate for labeled image 1, and R_{11}, R_{tt} are the clutter covariance matrices for image 1 and the test image, respectively.

Rajan *et al.* propose an active learning approach where a classifier, or learner, attempts to acquire knowledge from a teacher about new data points that may be from an unknown distribution [66]. In this so called *KL-max* approach, the new data points and corresponding labels are chosen to maximize the KL divergence between the learned distributions and the learned distributions including the new data point and corresponding label. The labels, which are distributions, are then updated using the new data point and label. This approach could be used

for context estimation where various labels from existing classifiers are chosen based on the KL divergence; however, it estimates these labels sample-by-sample.

Many of the aforementioned existing methods either operate in different testing conditions, such as semi-supervised classification or environmental conditions are known *a priori*, or they cannot account for disguising transformations.

Probability Introduction

We now provide a brief mathematical and probabilistic review of the concepts that will be used in the proposed model. Due to the complex formulation of random sets, our review starts with the building blocks of probability and measure theory. The main purpose of the following review is the introduction of notation. For a rigorous mathematical development, see the literature [1]-[7].

Informally, a random variable is a mapping from a probability space to a measurable space. The probability space consists of a domain, family of subsets of the domain, and a governing probability distribution. To formally define random variables, we need to introduce concepts from topology and measure theory.

Topology

Definition 2.1 Topology: A topology T on a set X is a collection of subsets of X that satisfy
1. $\phi, X \in T$,
2. T is closed under finite unions and arbitrary intersections.
Such a pair, (X,T), is referred to as a topological space [10].

The set X is subsequently referred to as a topological space. Topologies are generally described by construction. Usually, a topology is said to be generated from some basis or sub-basis B.

Definition 2.2 Basis for a topology: A basis for a topology T on X is a collection $\mathcal{B}$ of subsets of X such that

1. For all $x \in X$ there exists a $B \in \mathcal{B}$ such that $x \in B$.
2. If $B_1, B_2 \in \mathcal{B}$ and $x \in B_1 \cap B_2$ then there exists a B_3 such that $x \in B_3$ and $B_3 \subseteq B_1 \cap B_2$ [10].

Definition 2.3 Subbasis for a topology: A subbasis for a topology on *X* is a collection of subsets of *X* whose union is *X*. The topology generated by a subbasis *S* is the collection *T* of all unions and finite intersections of the elements of *S* [10].

The constituent sets of a topology are the focus of this review. Therefore, we fully detail them and the idea of measurability.

Definition 2.4 Open set: Given a topological space (*X*,*T*), all sets $G \in T$ are called open sets [10].

Definition 2.5 Closed set: The complement of an open set is a closed set [10].

A major misconception is that sets are either closed or open; however, this is not the case. In fact sets in a topology can be open, closed, neither, or both. For instance in the standard topology on $\mathcal{R}$, the interval [0,1) is neither open nor closed [10]. We emphasize that this is greatly dependent on how the topology is generated. There are topologies that do not share the intuitive characteristics of the standard topology on $\mathcal{R}$.

We next define some attributes of a topological space, which help characterize important concepts. Many of these attributes such as compactness are assumed when dealing with sets, but in the following, they are formally defined for clarity.

Definition 2.6 Cover: A collection of subsets of a space *X* is said to cover *X* if the union of its elements is *X*. Furthermore, an open cover of *X* is a cover whose elements are open sets [10].

Definition 2.7 Connectedness: A topological space (*X*,*T*) is connected if there does not exist a pair of disjoint, non-empty, open subsets *U* and *V* of *X* whose union is *X* [10].

Definition 2.8 Compactness: A space is compact if every open covering of *X* contains a finite subcollection that also covers *X* [10].

Probability Space

Next, we define necessary constructs for a probability space. We then define a standard random variable which will aid in the development of the random set.

Definition 2.9 σ-Algebra: If X is a set, then a σ-algebra $\sigma(X)$ on X is a collection of subsets of X that satisfy

1. $X \in \sigma(X)$
2. $A \in \sigma(X) \Rightarrow A^C \in \sigma(X)$
3. If $\{A_n\}_{n=1}^{\infty}$ is a sequence of elements of $\sigma(X)$, then $\bigcup_{n=1}^{\infty} A_n$ is $\sigma(X)$. Furthermore, $\sigma(X)$ is closed under countable intersections [9].

Note that if $\{A_n\}$ is a finite or countably infinite collection of elements of $\sigma(X)$, then $\left(\bigcup A_n\right)^c = \bigcap A_n \in \sigma(X)$ thus a σ-algebra is also closed under countable intersections. Hence, σ-algebras are topologies since the requirements for σ-algebras subsume the requirements of topologies. Note that σ-algebras also require closure under complementation, which is not a requirement of a topology. This closure under complementation allows for an intuitive application to probabilistic analysis. A σ-algebra is a type of topology useful in the field of probability and measure theory. In fact, most probability spaces are defined using Borel σ-algebras.

Definition 2.10 Borel σ-algebra: The Borel σ-algebra on a topological space X, written $\mathcal{B}(X)$, is the smallest σ-algebra that contains the family of all open sets in X.

Elements of a Borel σ-algebra are called *Borel Sets*.

Measure

Before we introduce random variables, we explain the idea of measurability. Although the general idea of measure is fairly complex, we give a simple overview.

Definition 2.11 Measure: A measure on $\sigma(X)$ is a function $\mu : \sigma(X) \to [0, \infty)$ satisfying

1. $\mu(\phi) = 0$
2. $A \cap B = \phi \Rightarrow \mu(\{A \cup B\}) = \mu(A) + \mu(B), \forall A, B \in \sigma(X)$, if finite or

$A_n \in \sigma(X), \forall j \neq k \;\; A_j \cap A_k = \phi \Rightarrow \mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$ if infinite [9].

The elements of $\sigma(X)$ are called measurable sets [9].

Some measures have added constraints such as the probability measure.

Definition 2.12 Probability measure: A probability measure is a measure $P:\sigma(X)\to[0,1]$ with the added constraint $P(X)=1$.

We have now properly defined the probability measure which is one of three elements necessary for a probability space. The other two elements are the domain and a corresponding σ-algebra.

Definition 2.13 Measure space: A measure space is a triple $(X,\sigma(X),\mu)$ where the pair $(X,\sigma(X))$ is referred to as the measurable space, *X* is a topological space, $\sigma(X)$ is a σ-algebra on *X*, μ is a measure on $\sigma(X)$ [9].

Definition 2.14 Probability space: A probability space is a triple $(\Omega,\sigma(\Omega),P)$, where Ω is a topological space, $\sigma(\Omega)$ is a σ-algebra on Ω, and *P* is a probability measure on $\sigma(\Omega)$ [9].

Definition 2.15 Measurable function: A function $f:\sigma(X)\to\mathcal{A}$ is measurable if for any interval $A\subseteq\mathcal{A}$, $f^{-1}(A)\in\sigma(X)$ [9].

A random variable is a measurable mapping from some probability space into a measurable space.

Standard Random Variables

Random variables are the basis of statistical modeling and analysis. The use of statistical modeling and analysis is abundant in the pattern recognition and machine learning community. These tools, along with others, allow researchers to model systems and automate *intelligent* decision making.

Now that we have defined all the necessary structures, we are able to define the random variable.

Definition 2.16. Random variable: Given a probability space $(\Omega,\sigma(\Omega),P)$ and some measurable space $(X,\sigma(X))$ for some positive integer *d*, a random variable, *R*, is a measurable mapping from a probability space to a measurable space such that $\forall \mathrm{Y}\in\sigma(X), R^{-1}(\mathrm{Y})\in\sigma(\Omega)$, if the random variable is defined on the entire space [9].

We note here that in applications, many ignore this initial mapping from the probability space to the measurable space. This mapping is necessary for formal definitions; however, it is

not necessary for most applications and the cumbersome notation is disregarded. Hereafter, we may disregard this initial mapping unless its recognition is required.

Standard Statistical Approaches for Context Estimation

There are a few issues that will arise if standard statistical techniques are used for context estimation. Next, we detail some of these potential pitfalls.

In standard approaches, the probability or likelihood of multiple occurrences are calculated using a joint distribution

$$P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n \mid C) \tag{2-11}$$

where $x_1, x_2, ..., x_n$ are *n* observations and *C* is some context. A few issues that arise from this approach are as follows:

1. Estimation of the joint likelihood function may be complicated by sparsity (J.1)
2. Estimation requires the matching of observations to random variables (J.2)
3. Likelihood calculation is highly dependent on number of observations (J.3)

Issue J.1 will occur when there are a large number of random variables compared to number of observations. Issue J.2 occurs since there is a distinction made between the observations. If X_i is different from X_j, then each observation will have to be paired with a random variable. This presents a problem of matching each observation to a random variable which also results in issue J.3.

Standard random variables are used to model the outcomes of single events or trials. In some approaches, a set of observations is modeled using a standard random variable where the set of observations is interpreted as a sequence of trials from the same experiment. This approach is similar to a common assumption for simplified joint estimation, the i.i.d. assumption.

$$P(x_1, x_2, ..., x_n \mid C) = P(x_1 \mid C)P(x_2 \mid C)...P(x_n \mid C) \tag{2-12}$$

This assumption presumes that observations *x* can be fully described by one random variable. However, this simplification results in a two additional issues:

1. Estimate of the joint likelihood is not robust to outliers due to the product of sample likelihoods (J.4)
2. Contextual information concerning joint observation is reduced to a product of sample likelihoods (J.5)

Note that even with the i.i.d. assumption, issue J.3 is still present. For example, as the number of observations occurs, the likelihood of some context must decrease, which is an unintuitive result for modeling context. This result is intuitive if we are modeling a sequence of experiments. Issue J.4 occurs since we have turned joint estimation into a product of singleton likelihoods.

Random Sets

One type of random variable—the random set—has not been researched as extensively as the standard random variable in the intelligent systems community. We consider only random subsets of $\mathcal{R}^d$ in the following. First, the formal definition of the random set and some associated constructs are presented. Next, a brief inspection and discussion of the random set is presented including its relationship to belief and possibility theory. Finally, the shortcomings of standard point process models for context estimation are discussed, which provides motivation of the proposed implementations.

General Case: Random Closed Set

Assume that $\mathcal{E} \subseteq \mathcal{R}^d$ is a topological space. We will denote the family of closed subsets of $\mathcal{E}$ as $\Im$. We can define a measurable space $(\Im, \sigma(\Im))$ associated with some probability space $(\Omega, \sigma(\Omega), P)$, where all $\Im$-valued elements will be referred to as closed sets. Informally, a random set is a measurable mapping from the aforementioned probability space to the measurable space.

Note that the construction of an intuitive σ-algebra for closed set values is not as clear as the construction for real number values. For example, a measurable interval for a random variable may be [-1, 4]. This interval, or set, is constructed by accumulating all the numbers *greater than or equal to* -1 and *less than or equal to* 4. However, relationships such as *greater than* or *less than* do not linearly order sets. One σ-algebra that is used with random sets is constructed by the Hit-Miss or Fell topology, such that any observed set $X \in \mathfrak{I}$ either intersects, hits, or does not intersect, misses, some $K \in \mathcal{K}$, where $\mathcal{K}$ is the family of compact sets. The families of sets that are used as basis elements to generate the Fell topology are $\mathfrak{I}^K = \{F \in \mathfrak{I} : F \cap K = \phi, K \in \mathcal{K}\} \in \sigma(\mathfrak{I})$ and $\mathfrak{I}_G = \{F \in \mathfrak{I} : F \cap G \neq \phi, G \in \mathcal{G}\} \in \sigma(\mathfrak{I})$. The Fell topology is a standard topology on $\mathfrak{I}$.

Definition 2.17 Fell topology: The Fell topology is a topology $(\mathfrak{I}, T)$ where *T* has subbasis which consists of $\mathfrak{I}_G$ and $\mathfrak{I}^K$.

Note that the Borel σ-algebra generated by the Fell Topology on $\mathfrak{I}$ coincides with the σ-algebra generated by $\mathfrak{I}_K$ [1]. We can now formally define the random closed set.

Definition 2.18 Random closed set measurable with respect to the Fell topology: Let $\mathfrak{I}$, be a collection of all closed sets from a topological space and let $\mathcal{B}(\mathfrak{I})$ denote the σ-algebra generated by $\mathfrak{I}_K$. Given a measurable space $(\mathfrak{I}, \mathcal{B}(\mathfrak{I}))$ associated with some probability space $(\Omega, \sigma(\Omega), P)$,a measurable mapping $\Xi : \Omega \to \mathfrak{I}$ is called a random closed set measurable with respect to the Fell Topology if $\mathcal{B}(\mathfrak{I})$ [1].

Random Set Discussion

The random set is governed by its distribution $P(\mathfrak{I}_K) = P\{\Xi \in \mathfrak{I}_K\}, \mathfrak{I}_K \in \mathcal{B}(\mathfrak{I})$. Since $\mathcal{B}(\mathfrak{I})$ is generated by $\mathfrak{I}_K$, it seems reasonable to determine the measure, or probability, of some set *K* using $\mathfrak{I}_K$ where $P\{\Xi \in \mathfrak{I}_K\} = P\{\Xi \cap K \neq \phi\}$ is a well defined measure. In fact, since these sets $\mathfrak{I}_K$ for each *K* compose our Borel σ-algebra, our probability distribution is defined on these sets with corresponding values being the probability of an observed Ξ will intersect *K*.

Note that the sets in $\Im_K$ just have to have a non-empty intersection with some set value *K*. In effect, the calculation of likelihood of a random set value *K* can be viewed as calculating the measure of the sets that contain at least one similar component as the set *K*.

Definition 2.19 Capacity functional: The real-valued function, T_Ξ ,associated with Ξ, $T_\Xi(K) = P(\Xi \in F_K) = P\{\Xi \cap K \neq \phi\}, K \in \mathcal{K}$ is called the capacity functional if the following requirements are satisfied [1]:

1. $T_\Xi(\phi) = 0$
2. $0 \leq T_\Xi(K) \leq 1, K \in \mathcal{K}$
3. $K_n \downarrow K \Rightarrow T_\Xi(K_n) \downarrow T_\Xi(K)$ (upper semi-continuous)
4. $\Delta_{K_n} \ldots \Delta_{K_1} T_\Xi(K) \leq 0, \ \forall n \geq 1, \ K, K_1, ..., K_n \in \mathcal{K}$ (completely alternating/ completely $\cup$ – alternating)

where $\Delta_{K_n} \ldots \Delta_{K_1} T_\Xi(K) = -P\{\Xi \cap K = \phi, \Xi \cap K_i, i = 1, ..., n\}$.

For an extensive explanation, the reader is directed to literature [1]-[6].

The capacity functional can be viewed as an optimistic estimate of the probability of a random set. In fact, it can be shown that this measure is an upper bound the family of probability measures $\mathbf{P}_\Xi$ associated with random set Ξ, that is $T_\Xi(K) = \sup\{P(K) : P \in \mathbf{P}_\Xi\}$ [1]. This also means that the capacity functional is an upper probability. It can be shown that $T_\Xi(K)$ dominates $P(K)$, $\forall P \in \mathbf{P}_\Xi$, which means $T_\Xi(K) \geq P(K), \forall K \in \mathcal{K}, \ \forall P \in \mathbf{P}_\Xi$ [1].

To uncover other functionals associated with the random set, we dissect the set $\Im_K$ into three disjoint sets.

$$\Im_K = \{F \in \Im : F \subset K\} \cup \{F \in \Im : K \subset F\} \cup \{F \in \Im : F \cap K \neq \phi, F \not\subset K, K \not\subset F\} \qquad (2\text{-}13)$$

Since the constituent sets in Equation 2-13 are disjoint, we can divide the capacity functional into these following terms:

$$\begin{aligned} &P\{\Xi \cap K \neq \phi\} = P\{K \subset \Xi\} + P\{\Xi \subset K\} + P\{\Xi_i \cap K \neq \phi, \Xi \not\subset K, K \not\subset \Xi\} \\ &= I_\Xi(K) + C_\Xi(K) + H_\Xi(K). \end{aligned} \qquad (2\text{-}14)$$

Note that T_{Ξ} is not additive with respect to *K*, but rather partitions of $\Im_K$. For example, if $K = K_1 \cup K_2, K_1 \cap K_2 = \phi$, then $P\{\Xi \cap K_1 \neq \phi\} + P\{\Xi \cap K_2 \neq \phi\} \neq P\{\Xi \cap K \neq \phi\}$ may be possible. This is true since it may be the case that $\exists K_3$ such that $K_3 \in \Im_{K_1} \cap \Im_{K_2}$; and by definition, $K_1 \cap K_2 = \phi$ does not imply $\Im_{K_1} \cap \Im_{K_2} \neq \phi$. In fact, T_{Ξ} is a subadditive fuzzy measure on $\Im$,

$$P\{\Xi \cap K \neq \phi\} \leq P\{\Xi \cap K_1 \neq \phi\} + P\{\Xi \cap K_2 \neq \phi\}. \tag{2-15}$$

We now define the functionals developed in Equation 2-14.

Definition 2.20 Inclusion functional: The inclusion functional calculates the measure of the sets in which *K* is included—all the sets which have *K* as a subset.

$$I_{\Xi}(K) = P(\Xi \in F_{K\subset *}) = P\{K \subset \Xi\} \text{ where } F_{K\subset *} = \{F \in \Im : K \subset F\} \tag{2-16}$$

The inclusion functional can be used to describe a random set; however, it does not generally, uniquely determine the distribution of a random set due to some pathological cases. It's alternative interpretation is its relation to the capacity functional of $\Xi^{\mathbf{C}}$ [1].

$$I_{\Xi}(K) = P(\Xi^{\mathbf{C}} \cap K = \phi) = 1 - T_{\Xi^{\mathbf{C}}}(K). \tag{2-17}$$

Definition 2.21 Containment functional: The containment functional which calculates the measure of the sets which are contained in *K*.

$$C_{\Xi}(K) = P(\Xi \in F_{*\subset K}) = P\{\Xi \subset K\} \tag{2-18}$$

where $F_{*\subset K} = \{F \in \Im : F \subset K\}$.

It can be shown that the containment functional is completely intersection monotone making it the dual of the capacity functional [1]. It can be shown that the following relationship exists between the capacity and containment functionals:

$$C_{\Xi}(K) = P\{X \subset K\} = 1 - T_{\Xi}(K^{\mathbf{C}}) \tag{2-19}$$

This relation also gives an intuitive explanation as to why the containment functional also determines the distribution of a random set, if defined on the open sets. This dual relationship shared between the capacity and containment functionals is similar to the relationship shared

between belief and plausibility functions. Belief functions are used extensively in evidential reasoning and are discussed in the Theory of Evidence section [8].

For the purposes of the random set, the containment functionals superadditivity property can be viewed as a pessimistic estimate of a random set value. The containment functional uses a containment requirement for the probabilistic frame of reference, meaning it uses sets that are contained in *K* to calculate probability. In other words, this value is the probability that only the elements of *K* will be generated, whereas, the capacity functional requires only the existence of one similar element. In fact, it can be shown that the containment functional is a lower probability

$$C_{\Xi}(K) = \inf\{P(K) : P \in \mathbf{P}_{\Xi}\} \text{ [1].} \tag{2-20}$$

This implies that $C_{\Xi_i}(K)$ is dominated by $P(K)$, $\forall P \in \mathbf{P}_{\Xi}$, $\forall K \in \mathcal{K}$. All probability measures on a random set are wedged in between these bounds, that is

$$C_{\Xi}(K) \le P(K) \le I_{\Xi}(K) + C_{\Xi}(K) + H_{\Xi}(K) = T_{\Xi}(K) \ \forall P \in \mathbf{P}_{\Xi}, \forall K \in \mathcal{K}. \tag{2-21}$$

This is intuitive since the capacity functional is the probability that the random set will hit a given set, whereas the containment functional is the probability that the random set is fully contained within the given set.

Definition 2.22 Hit and miss functional: The hit and miss functional calculates the measure of sets that intersect the set *K*, but have no inclusion or containment relationship.

$$H_{\Xi}(K) = P(\Xi \in F_{*\not\subset K, K\not\subset *}) = P\{\Xi_i \cap K \neq \phi, \Xi \not\subset K, K \not\subset \Xi\} \tag{2-22}$$

where $F_{*\not\subset K, K\not\subset *} = \{F \in \Im : F \not\subset K, K \not\subset F\}$

The hit and miss functional is not used in the literature. It simply identifies sets that have a non-empty intersection with a set *K*, non-containment relationship with a set *K*, and non-inclusion relationship with a set *K*. Its use alone for the purposes of probability assignment would not be intuitive.

The inclusion and containment functionals identify the sets *above* or *below K* in the lattice of subsets of $\Im_K$, that is these functionals identify the sets that can be linearly ordered with respect to *K* by inclusion and containment. On the other hand, the hit and miss functional considers all sets at the same level as *K* on the lattice, and are not comparable using inclusion and containment.

Theory of Evidence

We briefly discuss the relationship between random sets and the *Theory of Evidence*, as developed by Dempster and Shafer.

Definition 2.23 Belief function: A function $BEL: 2^X \rightarrow [0,1]$ is a belief function on some space *X* if the following constraints are satisfied

1. $BEL(\phi) = 0$
2. $BEL(X) = 1$
3. *BEL* is completely monotone [1], [8].

Definition 2.24 Plausibility functions: The dual of the belief function, the plausibility function has the expected dual form

$$PL(A) = 1 - BEL(A^C). \tag{2-23}$$

Just as the capacity functional is an optimistic estimation of the probability of a set outcome, the plausibility function is an optimistic estimation of the probability of an occurrence of an element in *A*. Belief functions are completely determined by their mass functions.

Definition 2.25 Mass functions: A function $m: 2^X \rightarrow [0,1]$ is a mass function if $m(\phi) = 0$ and $\sum_{A \subseteq X} m(A) = 1$.

Note that the containment functional of a random closed set is a belief function, which can also be described by its corresponding mass function.

$$BEL(A) = \sum_{B \subseteq A} m(B) = P_\Xi\{\Xi \subseteq A\} = C_\Xi(A); \tag{2-24}$$

Whereas, a general belief function is a containment functional only if some continuity conditions are met [1].

Note that $\left(2^X, \sigma(2^X), m\right)$ forms a probability space, where *m* is a probability on sets $A \in 2^X$. Furthermore, the corresponding belief function resembles a cumulative distribution function on 2^X using containment relationship to accumulate measure.

The purpose of distributing mass, *m*, to subsets of outcomes rather than simply the outcomes themselves in evidential reasoning is an attempt to model uncertainty. Rather than merely having the ability to state the probability of each outcome, the mass function can assign probability of an outcome occurring in a set without explicitly expressing the probability of its constituents [8].

Point Process

General random set models are seldom used in the machine learning community. This is interesting since random variables and statistical models are ubiquitous in the same community. One reason for this is that the general random set has no simple or even established parametric form or simple methods for estimation. Specific types of random sets, such as point processes, do have simple parametric forms which allow for optimization and estimation; however as will be discussed, they are rarely used to model sets of occurrences.

Next, we define some popular parametric forms of the point process and discuss their pros and cons. We conclude that most parametric forms of the point process are restricted to behave as standard random variables. They do not take advantage of the information attained from the co-occurrence, or observation, of a set of samples, but rather treat these samples as independent occurrences.

Definition 2.26 Counting measure: Assume $\mathcal{E} \subseteq \mathcal{R}^d$ is a topological space. A measure μ on a family of Borel sets $\mathcal{B}(\mathcal{E})$ is called a counting measure if it takes only non-negative integer values, that is $\mu : \mathcal{B}(\mathcal{E}) \to \{0,1,2,...\}$ [4].

A counting measure is locally finite if the measure is finite on bounded subsets of $\mathcal{E}$. Therefore, a locally finite counting measure has a finite number of points in its support in any compact set [4].

Definition 2.27 Point process: A point process $\Phi : \Omega \to N$ is a random closed set with associated probability space, $(\Omega, \sigma(\Omega), P)$, and a measurable space $(N, \mathcal{B}(N))$ where N is the family of all sets φ of points in $\mathcal{E}$ if φ is locally finite (each bounded subset of $\mathcal{E}$ must contain only a finite number of points of φ)[4].

Less formally, a point process is a random choice of $\varphi \in N$ governed by P. In practice, point processes are considered to be random sets of discrete points or as random measures which count the number of points within bounded regions. Random measures are further discussed in the Random Measure section. Since a point process is a random set, the same principles and theorems that apply to random sets apply to point processes.

Since point processes are locally finite, their capacity functional are expressed as follows:

$$T_{\Phi}(K) = \mathbf{P}(\Phi \cap K \neq \phi) = \mathbf{P}(|\Phi \cap K| \neq 0) = \mathbf{P}(\Phi(K) \neq 0), \quad \text{(2-25)}$$

where $\Phi(K) = |\Phi \cap K|$.

Since we know the intersections will have a finite number of elements, we can model these probabilities as counting probabilities [4].

Definition 2.28 Intensity measure: The intensity measure Λ of Φ is the mean value of $\Phi(K)$, defined as $\Lambda(K) = E[\Phi(K)]$, where Λ is simply a random variable with probability space $(\mathcal{K}, \sigma(\mathcal{K}), \mu)$ and measurable space $(\mathcal{R}^{+}, \sigma(\mathcal{R}^{+}))$. Simply, $\Lambda(K)$ is the mean number of points of a realization of Φ in K [1], [4].

In many applications, point processes are modeled in terms of intensity measures to provide for a simpler functional model. It provides for an intuitive idea of intensity and allows for a simple parametric form. The following are examples of a few popular parameterizations: random point, binomial point process, Poisson point process and the Gibbs point process.

Definition 2.29 Random point: A random point is a point process ξ with singleton outcomes. The capacity functional of this random point can be estimated
$P(\{\xi\} \cap K \neq \phi) = P(|\{\xi\} \cap K| \neq 0) = P(\xi \in K)$ [4]. (2-26)

Assume that is our random point is uniformly distributed in some compact set $K \subset \boldsymbol{\mathcal{E}}$. Let ν be the Lesbegue measure on $\boldsymbol{\mathcal{E}}$ that corresponds to length, area, or volume, depending on the dimension of $\boldsymbol{\mathcal{E}}$. Note this measure represents the uniform distribution on the space $\boldsymbol{\mathcal{E}}$. For each subset *A* of *K* we could then define the point process distribution, corresponding to the random point as follows:

$$P(\xi \in A) = \frac{\nu(A)}{\nu(K)} \tag{2-27}$$

This is essentially a standard random variable which should be clear from Equation 2-27. This formulation is simply a ratio of the measure of *A* and the total measure, the measure of *K*. This seems reasonable for the probability of a uniformly distributed random point to fall in volume *A* to assume this value.

Definition 2.30 Binomial point process: A binomial point process with *n* points is *n* independent uniformly distributed random points $\xi_1, \xi_2, ..., \xi_n$ which are distributed over the same compact set $K \subset \boldsymbol{\mathcal{E}}$. This binomial point process, written $\Phi_{W^{(n)}}$ is governed by the following joint distribution

$$P(\xi_1 \in A_1, \xi_2 \in A_2, ..., \xi_n \in A_n) = \prod_{i=1}^{n} P(\xi_i \in A_i) = \frac{\prod_{i=1}^{n} \nu(A_i)}{\nu(K)^n} \tag{2-28}$$

For each subset *A* of *K*. Since ν is a Lebesgue measure, there are three inherent properties of the binomial point process.

1. $\Phi_{W^{(n)}}(\phi) = 0$
2. $\Phi_{W^{(n)}}(K) = n$
3. $\Phi_{W^{(n)}}(A_1 \cup A_2) = \Phi_{W^{(n)}}(A_1) + \Phi_{W^{(n)}}(A_2), A_1 \cap A_2 = \phi$ [4].

The above formulation of random points is indicative of the i.i.d. assumption. The above formulation treats each element of a random set, as being independent of each other. This assumption retards the random sets ability to maintain co-occurrence information about the

samples, and furthermore, behaves similarly to the standard random variable with the i.i.d. assumption.

The aptly named binomial point process has an expected value, $E\left[\Phi_{W^{(n)}}(A)\right]$, modeled by a binomial distribution with parameters *n* and $p = P(\xi \in A)$ [4]. The mean of a binomial distribution is simply the product of its parameters *n* and *p*, yielding

$$E\left[\Phi_{W^{(n)}}(A)\right] = np = \frac{n\nu(A)}{\nu(K)}. \tag{2-29}$$

This means that the intensity—mean number of points per unit volume—is given by

$$\lambda = \frac{n\nu(A)}{\nu(K)}\frac{1}{\nu(A)} = \frac{E\left[\Phi_{W^{(n)}}(K)\right]}{\nu(K)}. \tag{2-30}$$

Although each of the points is distributed uniformly about the sample space in a binomial point process, the number of points contained in subsets of *K*, are not independent, since this distribution is defined for a fixed number of points *n*. If we were to construct $\Phi_{W^{(n)}}$ in terms of the number of points per subset as in [4] , the distribution would be more descriptive.

$$P(\Phi_{W^{(n)}}(A_1) = n_1, ..., \Phi_{W^{(n)}}(A_k) = n_k) \tag{2-31}$$
where $n_1 + n_2 + ... + n_k = n$ and $k = 1,2,...$

Example 2.1 Dependence on number of samples: It is clear that the number of points contained in subsets of *K* are dependent due to the fact that $n_1 + n_2 + ... + n_k = n$. If we know that $\Phi_{W^{(n)}}(A_1) = n_1$, then we also know that $\Phi_{W^{(n)}}(K \setminus A_1) = n - n_1$ [4].

We reiterate that the binomial point process treats its outcomes as the product standard random variables with the i.i.d. assumption and it is highly dependent on the number of points in a given area *A*.

Definition 2.31 Poisson point process: Let Λ be a locally finite measure on a topological space $(\mathcal{E}, \mathcal{B}(\mathcal{E}))$. The Poisson point process Π_Λ with intensity measure Λ is a random subset of $\mathcal{E}$ that satisfies the following constraints
1. For each bounded subset *K* of $\mathcal{E}$, the random variable $|\Pi_\Lambda \cap K|$ has a Poisson distribution

with mean $\Lambda(K)$.

2. Random variables $|\Pi_\Lambda \cap K|$ are independent for disjoint *K* [4].

The corresponding capacity functional takes the form

$$T_{\Pi_\Lambda}(K) = P\{\Pi_\Lambda \cap K \neq \phi\} = 1 - \exp(-\Lambda(K)) \text{ [1], [4].} \tag{2-32}$$

The first constraint suggests that $\Lambda(K)$ is parameterized by λ, the parameter of the Poisson distribution. This parameterization is usually of the form $\Lambda(K) = \lambda\nu(K)$, where ν is a measure, usually Lebesgue, of the set value *K* for all $K \in \mathcal{K}$. The second constraint imposes *independent scattering*, the number of points in disjoint Borel sets are independent. Note that this second constraint implies that there is *no interaction between points in a pattern*—elements in a set [4]. This parameterization would therefore be limiting for context estimation.

The last point process model that is discussed is the Gibbs point process which has roots in statistical physics. They are motivated by Gibbs distributions which describe equilibriums states of closed physical systems. In Gibbs theory, likelihoods of configurations are modeled assuming that the higher the probability of a system of objects, the lower the potential energy of the system [4]. This ideology is modeled in their definition.

Definition 2.32 Gibbs point process: A point process Ψ is a Gibbs Point Process with exactly *n* points if its capacity function is governed by the probability density function defined in Equation 2-33.

$$f(K) = \frac{\exp(-U(K))}{Z} \tag{2-33}$$

Hence the distribution is calculated in the standard fashion.

$$P(\Phi \in K) = \int\cdots\int_K f(x_1,...,x_n)dx_1...dx_n \tag{2-34}$$

In Equation 2-33, the function $U : \mathcal{R}^{nd} \rightarrow \mathcal{R}$ is the energy function and *Z* is the partition function. Note in Equation 2-34, order of integration is irrelevant since $K = \{x_1,...,x_n\}$ [4].

In practice, the energy function is chosen to be a sum of interaction potentials

$$U(K) = \sum_{A \subseteq K} V(A). \tag{2-35}$$

Frequently, V is assumed to have small values for large subsets of K. This assumption leads to the use of a *pair potential* function

$$U(K) = \sum_{i=1}^{n}\sum_{j=1}^{n}\theta\left(\left\|x_i - x_j\right\|\right). \tag{2-36}$$

The Gibbs point process can also be formulated for varying numbers of points n. This is called the *grand canonical ensemble* and assumes n is random [4]. Let $\mathcal{K}_n$ be the family of sets with n points. Then we can define $\mathcal{K} = \bigcup_{n=0}^{\infty}\mathcal{K}_n$ [4].

We can now define a density on $\mathcal{K}$.

$$f(K) = ca^n \exp(-U(K)). \tag{2-37}$$

where c and a are the appropriate normalization factors [4].

Random Measures

Random measures associated with random sets are generalizations of counting measures. As a random counting measure is a function on a point process, a random measure, associated with random sets, is a function on a random set.

Definition 2.33 Random measure: Assume $\mu : \mathcal{F} \to [0,\infty)$ is a fixed measure and Ξ is a random closed set with respect to the Fell Topology. Then $M_{\Xi,\mu}(F) = \mu(F \cap \Xi)$ is a random measure which maps from some probability space $(\Omega, \sigma(\Omega), P)$ to a measurable space $(\mathcal{M}, \mathcal{B}(\mathcal{M}))$ where $\mathcal{M}$ is the family of all locally finite measures on $\mathcal{F}$ and $\mathcal{B}(\mathcal{M})$ is generated by $\{M \in \mathcal{M} : M(F) > t\}$ for every $F \in \mathcal{F}$ and $t > 0$ [1].

For each instance X of Ξ, we have a corresponding instance $M_{X,\mu}$ of random measure $M_{\Xi,\mu}$, specifically a measure taking on a non-negative value for each set F. Note that throughout the literature, the measure μ is assumed to be additive and thus it has all corresponding characteristics. If we restricted $\mu : \mathcal{F} \to [0,1]$, it can define a probability measure on $\mathcal{R}^d$,

namely $P_{\Xi,\mu}(F) = M_{\Xi,\mu}(F)$. Therefore each instance of a random set Ξ has a corresponding measure $P_{X,\mu}$ [1].

$$P_{X,\mu} = \frac{\mu(F \cap X)}{\mu(X)}, \forall F \in \Im \tag{2-38}$$

To avoid cumbersome notation, we may omit μ and refer to $P_{X,\mu}$ as P_X when there is no ambiguity. This construction can be generalized by a taking a measurable random function $\varsigma(x), x \in \mathcal{E}$. We can then define a random measure as in Equation 2-39.

$$M_{\Xi,\mu}(F) = \int_{\Xi \cap F} \varsigma(x) d\mu(x) \tag{2-39}$$

Then we can construct a measure P_X as in Equation 2-40 [1].

$$P_X = \frac{\int_{X \cap F} \varsigma_X(x) d\mu(x)}{\int_X \varsigma_X(x) d\mu(x)}, \forall F \in \Im \tag{2-40}$$

We have therefore defined a mapping from X to $P_{X,\mu}$. Note in this construction we assume a dependence of ς on Ξ, denoted by ς_X.

Note, we have also defined a family of measures P_Ξ associated with random set Ξ. The random measure could be viewed as a distribution on distributions, or a measure on measures, which is related to variational approaches for approximate inference.

Variational Methods

The use of variational methods for approximate inference has become a popular classification method in the machine learning community. We give a brief description in order to identify its relationship to random sets, or more specifically, random measures. The goal of variational approaches is to determine the posterior $P(Z \mid X)$, of latent variables Z given observed data X where Z are typically class labels and parameters of distributions for the

elements of *X*. This approach is typically preferred over standard methods when the latent variable space is large, the expectations with respect to the posterior are intractable, or the integrations required are intractable or have no closed form representation [97].

Variational inference approximation balances the pros and cons of typical estimation approaches such as EM and other more computationally intensive methods such as stochastic techniques [97]. EM approaches suffer from the aforementioned problems; whereas stochastic methods such as Markov Chain Monte Carlo (MCMC) methods can generate exact results, but not in finite time [97].

In standard approaches such as EM, parameters are estimated by inspecting a small portion of the parameter space, which may make it more likely to settle in local optima rather than the global. MCMC methods attempt to construct the true distribution over all the possible values of the parameters using sampling methods. This approach allows for a globally optimal choice of parameter values or allows for the integration over all possible values. However, these approaches are only guaranteed as the sampling tends to infinity, but they may be useful when the sample space allows for a tractable solution [97].

In variational methods for approximate inference, function learning is the objective and typically hyperparameters, prior distributions on a function's parameters, are used to model a family of function values. It can be shown that the optimization of the log likelihood of the set of observations *X* can be separated into two terms:

$$\ln p(X) = L(q) + KL(q \parallel p) \tag{2-41}$$

where $L(q) = \int q(Z) \ln\left(\frac{p(X,Z)}{q(Z)}\right) dZ$ and $KL(q \parallel p) = -\int q(Z) \ln\left(\frac{p(Z \mid X)}{q(Z)}\right) dZ$.

It can also be shown that we can maximize the lower bound *L(q)* by minimizing the KL divergence between $q(Z)$ and $P(Z \mid X)$. Therefore, this is approach is a variational method, as

$p(Z|X)$ is estimated by optimizing the log likelihood with respect to the function q. Given the use of hyperparameters the optimization with respect to q is called a *free form* estimate, that is, q is only restricted by the parameterization of the hyperparameters. Therefore this expression can be seen as the optimization of a functional with respect to a function,

$$H[q] = \ln p(X). \tag{2-42}$$

The parameter distributions are typically formulated for simple integration, such that the parameters can be integrated out for the purposes of inference, usually classification. That is, the parameters are never estimated explicitly.

In summary, variational learning estimates a function through the use of observed data and parameter distributions governed by hyperparameters. These parameter distributions, which are distributions on distributions, are similar to the idea of random measures. However, as discussed in the Technical Approach, the purpose of the random measure within the random set framework is different from the use of hyperparameters in variational inference.

Before we discuss random set applications, it is necessary to review some measures, metrics and divergences defined on sets or measures.

Set Similarity Measures

In data sample analysis, it is necessary to have some sort of similarity measure for the purposes of comparing and contrasting the samples. If we are performing contextual analysis it seems appropriate to have a similarity measure to compare and contrast sets. The following is a brief review of standard and modern set similarity measures.

One way to analyze the similarity of measures would be to use a distribution similarity measure or divergence. Popular examples are the *Kullback-Leibler* (KL) divergence, which was

informally introduced in the previous section, and Chernoff divergence. The well-known KL divergence between distributions P_0 and P_1 is computed as follows:

$$KL(P_1 \,||\, P_0) = \int p_1(x) \log\left(\frac{p_1(x)}{p_0(x)}\right) dx \tag{2-43}$$

The Chernoff divergence is computed as follows:

$$C(P_0, P_1) = \max_{0 \le t \le 1} \left[-\log \mu(t) \right] \tag{2-44}$$
$$\text{where } \mu(t) = \int [p_0(x)]^{1-t} [p_1(x)]^t dx .$$

Upon inspection, both of these divergences seem to quantify the idea of similarity of measures based on the underlying distribution of mass.

Another common approach is the use of compressed distribution similarity measures. Common histogram measures are the L_1 and weight L_2 measures.

$$d_{L1}(H, K) = \sum_i |h_i - k_i| \tag{2-45}$$

$$d^2_{L2}(H, K) = (\mathbf{h} - \mathbf{k})^t \mathbf{A} (\mathbf{h} - \mathbf{k}) \tag{2-46}$$

In Equations 2-45 and 2-46, **A** is a weight matrix; *H* and *K* represent histograms, weighted clusters, or feature subsets of two discrete sets. Although popular, these similarity measures give rise to problems in robustness. For example, when computing the differences in histogram bins, Equations 2-45 and 2-46 do not account for neighboring bins.

A common similarity measure used in topological spaces is the Hausdorff metric. This metric computes the difference between two sets by finding the maximum difference of the minimum point-wise differences.

$$d^H(X, Y) = \max\left(\sup_{x_i \in X} \inf_{y_i \in Y} \|x_i - y_i\|, \sup_{y_i \in Y} \inf_{x_i \in X} \|x_i - y_i\| \right). \tag{2-47}$$

Although this similarity measure is indeed a metric, it seems to lack robustness. For example, two point sets having all constituents the same, less one outlier, would still be assigned a high difference value.

Another recently researched approach is the *earth mover distance* (EMD) [70], [71]. The idea behind the EMD is to calculate the minimum work needed to transform a discrete set *X* into a discrete set *Y* given some constraints. This minimization is done using linear programming. In fact, this distance calculation is a reformulation of the well known transportation problem. In this framework, one of the sets is considered a supplier and one a consumer where each supplier has a supply quantity x_i and each consumer has a demand quantity y_i. Given a shipping cost c_{ij} for each supplier / consumer pair, c_{ij}, the goal is to find the optimal flow of goods, f^*_{ij}, such that the cost is minimal. Using the optimal flow, EMD is calculated as follows:

$$EMD(X,Y) = \frac{\sum_{i\in I}\sum_{j\in J} c_{i,j} f^*_{i,j}}{\sum_{i\in I}\sum_{j\in J} f^*_{i,j}} = \frac{\sum_{i\in I}\sum_{j\in J} c_{i,j} f^*_{i,j}}{\sum_{j\in J} y_j} \tag{2-48}$$

where $\mathbf{f}^* = \arg\min_{\mathbf{f}} \sum_{i\in I}\sum_{j\in J} c_{i,j} f_{i,j}$

subject to

$f_{i,j} \geq 0, \quad \forall i \in I, j \in J,$

$\sum_{i\in I} f_{i,j} = y_j$, and

$\sum_{j\in J} f_{i,j} \leq x_i.$

Note the above formulation requires that each consumer be completely satisfied. For the purposes of set similarity measures, the idea of flow is simply the matching of similar points in the set. The difference between these points is then computed using the cost, which if formulated accordingly, can be a difference measure of these points. Also note that if the numbers of points

are different in the sets *X* and *Y*, then we can assign fraction values to the supplies and demands to allow for fractional point matching.

Houissa *et al.* proposed an algorithm that uses EMD as a metric for the comparison of images for image retrieval from a data base [72]. This is novel approach of using a set metric to analyze the similarity of two sets. In fact, the use of the aforementioned set metrics and divergences is fairly common in the machine learning community.

Random Set Applications

Next, we review current uses of random sets and en masse approaches in the machine learning and pattern recognition communities. The most widely used formulation of the random set is by far the point process [74]-[96].

Point Process Applications

Popular applications of point processes in machine learning and pattern analysis arenas include, but are not limited to, the following: event prediction [89],[90],[92], object recognition / tracking [74], [79]-[83], and particle modeling [4],[85],[93],[94]. Although we do not detail particle modeling, we explicitly mention it since many forms of the point process have deep roots in statistical physics, and therefore, many point process models relate to physics-based concepts. In many fields of physics, one studies the interaction between groups of particles.

In machine learning, these groups of particles are treated as sets of samples distributed by a point process. One of the more popular applications of point processes is event prediction. In this application the point process domain is the real line, typically time, and the particles are events. Other applications include sample clustering. In most applications, the point process is used similarly to standard random variables with standard probabilistic techniques.

There are no known, to the authors, applications of point processes that include the comparison of sets of samples, which is odd since they are random sets. We review some past

and current research involving the use of point processes in a manner relevant to context estimation.

Linnett *et al.* have used Poisson point processes to model segments of images for texture-based classification [84]. In this approach, samples from a same class are considered the same context. Each image is discretized and each pixel with similar gray values is bin grouped into similar point processes. A Bayesian posterior is then calculated estimating the class of each segment. Note that in this approach, the point process is used as a standard clustering algorithm, grouping samples from the same class together.

Stoica *et al.* proposed the Candy model which models road segments, in remotely sensed imagery, as a marked Poisson point process for roadway network extraction [74]. Each line segment is considered a point, or center, with marks such as width, length, and orientation. The interaction of the segments is governed by a Gibbs point process whose energy function contains a data term and a line segment interaction term. The segment interaction term penalizes short line segments. Segments are then merged based on an MCMC sampling method which adds points to segments, deletes points from segments, and merges segments. In later work, they incorporated Gibbs point processes within this model [80].

Descombes, *et al.* used a point process to model segments of images within the Candy model framework [81]. They improved their model by adding a prior density on the line segments. The prior is modeled as a point process, referred to as the Potts model, where the energy function is calculated based on the number of points in a clique in a segment, such that smaller segments are penalized.

Other work, such as extensions of the Candy model, continues their research of the point process for image analysis [82]. They improved their object process which is used to model the

target line networks in remotely sensed images by adding an additional term in its governing density to account for interactions with other object processes.

The point process is used by Savery and Cloutier to model clusters of red blood cells and correlate their orientation with other attributes of the blood [85]. In this paper, the point process is used to model different red blood cell configurations in the presence of backscattering noise. An energy function is used to assign a value to each configuration of blood cells; this function is placed inside an exponential function to estimate the likelihood of each configuration. An MCMC method is then employed to estimate the true configuration of the red blood cells.

En Masse Context-Based Methods

We refer to methods that treat a set of samples as a singleton unit as en masse approaches. These approaches use the same ideology as the random set and attempt to perform inference or analysis using the set.

Dougherty *et al.* proposed a set-based kNN algorithm is proposed to contend with data sets that may be distributed differently with respect to time [12]. In this approach, the idea of context is maintained by using each training set as a set prototype. The algorithm is able to contend with contextual factors and even disguising transformations. In this approach, the *k* nearest neighbors, neighboring training sets, of the test set are identified. Here context is identified through a similarity measure, specifically the Hausdorff metric, between the test *set*, and a prototype set. Classification of the individual samples is performed using the labels of the *k* nearest samples from the *k* nearest sets. Although this approach is improved over other context-based methods and solutions to concept drift, it suffers from a lack of robustness due to the use of the Hausdorff metric.

Bolton and Gader applied set-based kNN to remotely sensed data for target classification [15]. Contextual factors were apparent in this data set. The application of set-based kNN

improved classification results by correctly identifying the contexts using sets of samples; however, the resiliency of the Hausdorff metric was questionable.

Dougherty *et al*. motivated a statistical approach, an extension of set-based kNN, to identify population correlated factors for improved classification [12], [13], [14]. Dougherty *et al*. provided a very theoretical approach which was suggestive of Poisson point processes [12].

We extend Dougherty's theoretical approach and provide a general random set framework for context based classification which permits possibilistic, probabilistic and evidential implementations.

CHAPTER 3
TECHNICAL APPROACH

We propose a context-based approach for classification posed within a random set framework. The incorporation of random sets equips a classification algorithm with the ability to contend with hidden context changes. The goal of the proposed algorithm is, given an input sample set, or population, identify the population's context and classify the individual input samples.

We propose two models for context estimation and provide analogous inference and optimization strategies. The first model is similar to the germ and grain model which is commonly used in point process simulation [4]. We develop possibilistic and evidential approaches within this model and detail some optimization strategies. The second model utilizes random measures. We propose an unnormalized likelihood function which provides for a probabilistic estimate of context within this model. Finally, we provide a discussion to identify the similarities and differences of the proposed random measure model and standard statistical methods.

Mathematical Basis of the Random Set Framework

Assume a topological space $\mathcal{E} = \mathcal{R}^d$ with samples $x \in \mathcal{E}$. Let $\{\Xi_1,...,\Xi_I\}$ be random sets with respect to the Fell topology. Each Ξ_i is used to model a distinct context *i*, where we assume $\{\Xi_1,...,\Xi_I\}$ to be exhaustive. Assume a sample set *X*, test or train, containing a finite number of observations $X = \{x_1, x_2,..., x_n\}$ from some random set. Let $Y : \mathcal{E} \to \mathcal{Z}^+$ be a label function that maps each *x* to a given label $y \in \{1,2,..,l\} \subseteq \mathcal{Z}^+$, where $\mathcal{Z}^+$ denotes the positive integers.

Standard techniques estimate $P(y \mid x)$ for classification. If we believe that *x* was measured or observed in the presence of contextual factors, we can assume that our label function depends

on the context. If Y not independent of some context Ξ in which x was observed, the posterior estimate can be formulated as follows:

$$P(y \mid x, X) = \sum_{i=1}^{I} P(y, \Xi_i \mid x, X) . \tag{3-1}$$

Equation 3-1 is interpreted as calculating the probability that sample x has class label y and was generated in context i. In Equation 3-1 the posterior is marginalized over each potential context i. For reasons developed throughout Chapters 1 and 2, context identification is performed by indentifying contextual transformations; therefore, the observed population X is used for context estimation. Using Bayes' rule and making some independence assumptions, we arrive at Equation 3-2.

$$P(y \mid x, X) = \sum_{i=1}^{I} \frac{P(x, X \mid y, \Xi_i) P(y, \Xi_i)}{P(x, X)} \propto \sum_{i=1}^{I} P(x \mid y, \Xi_i) P(X \mid \Xi_i) P(y \mid \Xi_i) P(\Xi_i) \tag{3-2}$$

In Equation 3-2, we assume x is independent of X given its context and label. We also assume X is independent of y given the context. Equation 3-2 provides a random set framework for context based classification.

The factors in Equation 3-2 have intuitive meanings. The factor $P(x \mid y, \Xi_i)$ can be interpreted as the probability or likelihood that x was collected in context i and is of class y. A suitable implementation would be I classifiers, such that when each is presented with a sample x, could identify it as having class label y in its corresponding context i.

The result of classification within a particular context i, $P(x \mid y, \Xi_i)$, is weighted by the term $P(X \mid \Xi_i)$ which can be interpreted as the probability of observing X in context i. The result is an intuitive weighting scheme that weights each classifier's output based on contextual relevance to the test population.

The $P(y,\Xi_i)$ factor is interpreted as a prior likelihood of observing some class and context. Depending on the implementation, this term may be better estimated using, $P(y,\Xi_i) = P(y \mid \Xi_i)P(\Xi_i)$, where $P(y \mid \Xi_i)$ is the probability of class *y* given context *i* and $P(\Xi_i)$ is the prior probability of context *i*. Note that $P(x \mid y,\Xi_i)$ and $P(X \mid \Xi_i)$ are terms of great interest as they embody the context-based approach and will be further discussed and analyzed.

Estimating $P(x \mid y)$ has been researched for years using various models and estimation techniques. The estimation of $P(X \mid \Xi)$ and $P(x \mid y,\Xi)$ has not been researched quite as thoroughly, especially $P(X \mid \Xi)$. It seems proper that the values $P(X \mid \Xi)$ should be estimated using determining functionals of Ξ. The random set model provides for considerable flexibility since these probabilities can be estimated using evidential, probabilistic, or possibilistic techniques.

The proposed generalized, context-based framework may have different interpretations and a potential myriad of implementations. We develop two models for the estimation of $P(X \mid \Xi)$ within the proposed framework. A germ and grain model is specified and accompanied by possibilistic and evidential approaches for the estimation of $P(X \mid \Xi)$. Then a random measure model is specified and a probabilistic approach is developed for the estimation of $P(X \mid \Xi)$.

Possibilistic Approach

In this possibilistic approach, $P(X \mid \Xi)$ is estimated using the capacity functional.

$$P(X \mid \Xi) = P_{\Xi}(X) = T_{\Xi}(X) \tag{3-3}$$

For the initial development of this model we will let *Y* be a random set. Classification of the samples from *X* can be defined as partitioning the set such that subsets of *X* are assigned some

class label *y*. This first model can be considered a preliminary or intermediate model. The classifier in each context is modeled using the constructs which are modeling the context, that is, Y is a random subset of each Ξ. This possibilistic implementation provides for a simple and efficient parametric model which allows for direct analysis of the driving terms in Equation 3-2 and concurrent optimization of the classifier and contextual parameters. Optimization techniques for classifiers that do not share parameters with the germ and grain model are also provided.

Development

Note that in this initial model we use $P(\{x\} \mid Y, \Xi)$ instead of $P(x \mid Y, \Xi)$. This slight modification is due to the fact that the classifier in this initial implementation is modeled by random set constructs. Therefore the samples must be formally defined as singleton sets. However, this is not always the case and the notation $P(x \mid Y, \Xi)$ should be used, when a standard statistical classifier is used.

For the purposes of analysis, we focus on the terms $P(X \mid \Xi)$ and $P(\{x\} \mid Y, \Xi)$. These terms drive the context-based classifier so their isolation will aid in analysis. We assume the prior probabilities of all contexts $P(\Xi_i)$ are equal and that the probabilities of the class given the context $P(Y \mid \Xi_i)$ are equal. Given this we have

$$P(Y \mid \{x\}, X) \propto \sum_{i=1}^{I} P(\{x\} \mid Y, \Xi_i) P(X \mid \Xi_i) \tag{3-4}$$

We develop a model similar to that of the germ and grain model [4], [5], [16], that is, the random set is modeled as a union of random hyperspheres. This model provides a simple yet versatile parametric model to allow for the estimation of the terms in Equation 3-4. The germs are the random hypersphere centers and the grains refer to the size or volume of the hypersphere, which

is directly related to the radii. If random set Ξ_i follows a germ and grain model, it is defined by Equation 3-5, where ξ_{ij} are the germs and Ξ_{ij} are the grains.

$$\Xi_i = \bigcup_{j=1}^{n_i} (\{\xi_{ij}\} + \Xi_{ij}) \tag{3-5}$$

In Equation 3-5, n_i is the number of grains used to model context *i*. In our model we assume each grain is governed by a random radius r_{ij} that is exponentially distributed.

$$p(r_{ij}) = \lambda_{ij} \exp(-\lambda_{ij} r_{ij}) \tag{3-6}$$

This implies that the probability that $\{x\}$ hits a grain, $P(\{x\} \mid \Xi_{ij})$, can be estimated as follows

$$P(\{x\} \mid \Xi_{ij}) = T_{\Xi_{ij}}(\{x\}) = P(r_{ij} > \|x - \xi_{ij}\|). \tag{3-7}$$

Substituting the probability density in Equation 3-6 into Equation 3-7 yields

$$P(\{x\} \mid \Xi_{ij}) = 1 - P(r_{ij} \le \|x - \xi_{ij}\|)) = \exp(-\lambda_{ij} \|x - \xi_{ij}\|). \tag{3-8}$$

Equation 3-8 is used to model the constituent grains Ξ_{ij} and subsequently used to model Ξ_i and Y. The capacity functional of Ξ_{ij}, $P(\{x\} \mid \Xi_{ij})$, is subsequently used to estimate the capacity functional of Ξ_i.

$$P(X \mid \Xi_i) = P(\Xi_i \in \Im_X) = T_{\Xi_i}(X) \tag{3-9}$$

In this model, the calculation of $P(X \mid \Xi_i)$, follows from the calculation of the capacity functional of the constituent grains.

$$P(X \mid \Xi_i) = 1 - \prod_{j=1}^{n_i} \left(1 - T_{\Xi_{ij}}(X)\right) \tag{3-10}$$

Equation 3-10 states that the probability that *X* hits Ξ_i is the same as the probability that *X* does not miss all $\Xi_{ij}, \forall j = 1,...,n_i$. Given our model, we can calculate $T_{\Xi_{ij}}(X)$ using Equation 3-11.

$$T_{\Xi_{ij}}(X) = \max_{x \in X} T_{\Xi_{ij}}(\{x\}). \tag{3-11}$$

The proof is discussed in the Lemma 3-1.

Lemma 3-1. Let Ξ be a random set taking on set values in $\Im$ and having a probability distribution P_Ξ on $\mathcal{B}(\Im)$ and corresponding capacity functional T_Ξ. If we restrict the elements of $\Im$ to be a random disc or hypersphere then $T_\Xi(X) = \max_{x \in X} T_\Xi(\{x\})$ if X is finite or $T_\Xi(X) = \sup_{x \in X} T_\Xi(\{x\})$ if X is infinite.

Proof. We show if $T_{\Xi_{ij}}(\{x_1\}) \geq T_{\Xi_{ij}}(\{x_2\})$ then $P(\Xi_{ij} \cap \{x_1\} \neq \phi) = P(\Xi_{ij} \cap \{x_1, x_2\} \neq \phi)$, which can we inductively show implies $T_\Xi(X) = \max_{x \in X} T_\Xi(\{x\})$.

Base Case: First assume *without loss of generality* (WLOG) that $T_{\Xi_{ij}}(\{x_1\}) \geq T_{\Xi_{ij}}(\{x_2\})$. If random hypersphere is determined by a random radius, then $P(r \geq d(x_1, c)) \geq P(r \geq d(x_2, c))$, where d is some metric, r is the radius of the hypersphere and c is the hypersphere center. This implies that $d(x_1, c)) \leq d(x_2, c)$ if r is governed by a distribution that is monotonic with respect to distance, such as the exponential distribution. This is due to the fact that the probability of intersection is a function of distance only. This implies that each hypersphere that $\{x_2\}$ hits, $\{x_1\}$ must hits. So in this model we can assume

$$T_{\Xi_{ij}}(\{x_1\}) > T_{\Xi_{ij}}(\{x_2\}) \Rightarrow P(\Im_{x_1}) > P(\Im_{x_2}) \Rightarrow \forall K, K \in \Im_{x_2} \Rightarrow K \in \Im_{x_1} \tag{3-12}$$

Equation 3-12 implies that $P(\Xi_{ij} \cap \{x_1\} \neq \phi) = P(\Xi_{ij} \cap \{x_1, x_2\} \neq \phi)$.

Induction Step: Now assume $T_\Xi(K) = \max_{x \in K} T_\Xi(\{x\})$.We show that $P(\Xi_{ij} \cap (K \cup \{x_1\}) \neq \phi) = \max\left(\max_{x \in K} T_\Xi(\{x\}), T_\Xi(\{x_1\})\right)$. We know that there exists some $\hat{x} = \arg\max_{x \in K} T_\Xi(\{x\})$ and therefore $\hat{x} = \arg\min_{x \in K} d(x, c))$, where ties are arbitrarily broken. There are two cases. First assume $d(x_1, c)) \leq d(\hat{x}, c)$, which implies that $T_{\Xi_{ij}}(\{x_1\}) \geq T_{\Xi_{ij}}(K)$. Using the same argument in the Base Case, that is, every hypersphere that hits K, must hit $\{x_1\}$. In the other case, if $T_\Xi(\{x_1\}) < T_\Xi(K)$, then by the same logic every hypersphere that hits $\{x_1\}$, must hit K. Therefore, $T_\Xi(X) = \max_{x \in K \cup x_1} T_\Xi(\{x\})$ and given the Base Case, is true for all sets of countable size. Thus

$$T_{\Xi_{ij}}(X) = T_{\Xi_{ij}}\left(\bigcup_{x \in X} \{x\}\right) = \max_{x \in X} T_{\Xi_{ij}}(\{x\}). \tag{3-13}$$

Q.E.D.

For classification purposes, assume that some subset of the grains represent some class Y, which are identified in some index set C_y.

$$Y = \bigcup_{j:(i,j)\in C_y} (\{\xi_{ij}\} + \Xi_{ij}) \tag{3-14}$$

If we assume that the measure of the random hypersphere overlap in each context, $P(\{x\} \mid \Xi_{ij}, \Xi_{ik}),\ j \neq k$ is negligible, then the term $P(\{x\} \mid Y, \Xi_i)$ can be estimated as follows:

$$P(x \mid Y, \Xi_i) \propto \sum_{j:(i,j)\in C_y} P(\{x\} \mid \Xi_{ij}) . \tag{3-15}$$

The assumption in Equation 3-15 admits simplified update equations during the optimization stage.

Dependent Optimization

In this development, we propose an optimization method that assumes parametric dependence of the classifying and context estimating factors. Optimization of the parameters λ_{ij} is performed using a *minimum classification error* (MCE) objective [86], [87] and [88]. The objective is to maximize the difference between correct and incorrect classification. Equation 3-16 is used as an MCE objective function. Each parameter is updated in an iterative fashion using gradient descent. For optimization purposes, let $X_i \in \mathbf{X} = \{X_1, ..., X_I\}$ be training sets that represent different contexts.

$$D(x, X, \lambda_{ij}) = \begin{cases} -\left(\begin{array}{l} \sum_{j:(i,j)\in C_y} P(\{x\} \mid \Xi_{ij}) P(X \mid \Xi_i) - \\ \sum_{(m,k):(m,k)\notin C_y} P(\{x\} \mid \Xi_{mk}) P(X \mid \Xi_m) \end{array} \right), x \in y \\ \left(\begin{array}{l} \sum_{j:(i,j)\in C_y} P(\{x\} \mid \Xi_{ij}) P(X \mid \Xi_i) - \\ \sum_{(m,k):(m,k)\notin C_y} P(\{x\} \mid \Xi_{mk}) P(X \mid \Xi_m) \end{array} \right), x \notin y \end{cases} \tag{3-16}$$

In Equation 3-16, the second terms sum over context, grain pairs that model a class other than C_y, where C_y is the class modeled by parameter λ_{ij}. This objective can be interpreted as an

optimization of λ_{ij} with respect to observations from the context and class it represents as long as it doesn't hinder the classification of observations from other classes in any context.

For stability and quick convergence, a loss function is used.

$$l(x, X, \lambda_{ij}) = \frac{1}{1+\exp(-D(x, X, \lambda_{ij}))} \tag{3-17}$$

The total loss is then defined by Equation 3-18.

$$L(\lambda_{ij}) = \sum_{X \in \mathbf{X}} \sum_{x \in X} l(x, X, \lambda_{ij}) \tag{3-18}$$

We have the following gradient descent update formula where *t* represents the iteration number and α is the learning rate.

$$\lambda_{ij}^{t+1} = \lambda_{ij}^{t} - \alpha \frac{dL}{d\lambda_{ij}} \tag{3-19}$$

$$\text{where} \frac{dL}{d\lambda_{ij}} = \sum_{X \in \mathbf{X}} \sum_{x \in X} l(x, X, \lambda_{ij})(1 - l(x, X, \lambda_{ij})) \frac{dD}{d\lambda_{ij}}, \tag{3-20}$$

and

$$\begin{aligned}
&\frac{dD}{d\lambda_{ij}} = \left(-\left\|x - \xi_{ij}\right\|\right)\exp(-\lambda_{ij}\left\|x - \xi_{ij}\right\|)P(X \mid \Xi_i) + \\
&\left(\sum_{m \in C} \exp(-\lambda_{im}\left\|x - \xi_{im}\right\|)\right)\left(-\prod_{m \neq j} 1 - \exp(-\lambda_{im}\left\|x^{im} - \xi_{im}\right\|)\right) \times \\
&\left(-\left\|x^{ij} - \xi_{ij}\right\|\right)\left(-\exp(-\lambda_{ij}\left\|x^{ij} - \xi_{ij}\right\|)\right) - \\
&\left(\sum_{m \notin C} \exp(-\lambda_{im}\left\|x - \xi_{im}\right\|)\right)\left(-\prod_{m \neq j} 1 - \exp(-\lambda_{im}\left\|x^{im} - \xi_{im}\right\|)\right) \times \\
&\left(-\left\|x^{ij} - \xi_{ij}\right\|\right)\left(-\exp(-\lambda_{ij}\left\|x^{ij} - \xi_{ij}\right\|)\right)
\end{aligned} \tag{3-21}$$

where $x^{ij} = \arg\max_{x}(P(\{x\} \mid \Xi_{ij}))$.

The germs are not optimized in the experiments. However, similar gradient descent methods could be employed.

The proposed updates indicated by Equations 3-18, 3-19 and 3-20 have the added benefit of concurrently updating classification and contextual parameters since both are implemented as

the same structures. Next, we provide a general optimization strategy using the germ and grain model with a possibilistic estimate. That is, we optimize the contextual parameters based on their ability to correctly estimate context.

Independent Optimization

We estimate the contextual parameters using the following MCE objective.

$$D(\lambda_{ij}) = \sum_{X \in \Xi_i} P(X \mid \Xi_i) - \sum_{X \notin \Xi_i} P(X \mid \Xi_i) \tag{3-22}$$

The objective in Equation 3-22 is to maximize the difference between correct and incorrect context estimation. Using a similar gradient descent strategy, we arrive at Equation 3-23.

$$\begin{aligned} \frac{dD}{d\lambda_{ij}} = & \sum_{X \in \Xi_i} \left(-\prod_{m \neq j} 1 - \exp(-\lambda_{im} \left\| x^{im} - \xi_{im} \right\|) \right) \left(- \left\| x^{ij} - \xi_{ij} \right\| \right) \left(- \exp(-\lambda_{ij} \left\| x^{ij} - \xi_{ij} \right\|) \right) - \\ & \sum_{X \notin \Xi_i} \left(-\prod_{m \neq j} 1 - \exp(-\lambda_{im} \left\| x^{im} - \xi_{im} \right\|) \right) \left(- \left\| x^{ij} - \xi_{ij} \right\| \right) \left(- \exp(-\lambda_{ij} \left\| x^{ij} - \xi_{ij} \right\|) \right) \end{aligned} \tag{3-23}$$

Equation 3-23 provides for efficient optimization of the contextual parameter λ_{ij}, based on maximizing the separation between correct and incorrect contextual identification.

Evidential Model

In the possibilistic approach, we estimate $P(X \mid \Xi_i)$ using the capacity functional. In the evidential approach we use the inclusion functional to estimate the term $P(X \mid \Xi_i)$. There are two major reasons why we have chosen the inclusion functional for evidential modeling rather than the containment functional. First, we have a continuous model with discrete observations. This means the probability of containment would be zero for essentially all possible discrete observations *X*. Second, the inclusion functional is more intuitive for set-valued random elements, whereas, containment, similar to the idea of belief, is intuitive for modeling uncertainty with singleton random elements.

Development

We develop the evidential approach using the germ and grain model and assume the radii are exponentially distributed. Given these assumptions, we calculate the probability of inclusion given one random hypersphere as follows:

$$P\{X \subset \Xi_{ij}\} = P(\{F : X \subset F\}) = \exp(-\lambda_{ij} \left\| \hat{x}^{ij} - \xi_{ij} \right\|), \quad \text{where } \hat{x}^{ij} = \arg\min_{x}(P(\{x\} \mid \Xi_{ij})). \tag{3-24}$$

For calculation of inclusion, note that we use $\hat{x}^{ij}$ rather than x^{ij}. As x^{ij} is the closest $x \in X$ to germ ξ_{ij} and determines a non-empty intersection relationship of X and Ξ_{ij}, $\hat{x}^{ij}$ is the furthest $x \in X$ to germ ξ_{ij} and determines an inclusion relationship of X and Ξ_{ij}.

This probability can be accumulated across the constituent random hyperspheres using the same ideology taken during the calculation of the capacity functional in Equation 3-10. Therefore we calculate the probability of inclusion of random set Ξ_i across the constituent hyperspheres using Equation 3-25.

$$P(X \mid \Xi_i) = P\{X \subset \Xi_i\} = P(\{F : X \subset F\}) = 1 - \prod_{j=1}^{n_i} \left(1 - \exp(-\lambda_{ij} \left\| \hat{x}^{ij} - \xi_{ij} \right\|)\right). \tag{3-25}$$

Equation 2-25 states that the probability that a random set Ξ_i includes a set *X* is equal to the probability that each of the constituent random hyperspheres Ξ_{ij}, *does not* have a *non-inclusion* relationship with *X*.

Optimization

Using the objective defined in Equation 3-22, the parameters can be optimized using gradient descent as defined in Equation 3-19. For the optimization of λ_{ij}, we substitute Equation 3-26 into Equation 3-19.

$$\frac{dD}{d\lambda_{ij}} = \sum_{X \in \Xi_i} \left(-\prod_{\substack{m=1 \\ m \neq j}}^{n_i} \left(1 - \exp(-\lambda_{im} \left\| \hat{x}^{im} - \xi_{im} \right\|)\right) \right) \left(-\exp(-\lambda_{ij} \left\| \hat{x}^{ij} - \xi_{ij} \right\|) \right) \left(- \left\| \hat{x}^{ij} - \xi_{ij} \right\| \right) -$$

$$\sum_{X \notin \Xi_i} \left(-\prod_{\substack{m=1 \\ m \neq j}}^{n_i} \left(1 - \exp(-\lambda_{im} \left\| \hat{x}^{im} - \xi_{im} \right\|)\right) \right) \left(-\exp(-\lambda_{ij} \left\| \hat{x}^{ij} - \xi_{ij} \right\|) \right) \left(- \left\| \hat{x}^{ij} - \xi_{ij} \right\| \right). \quad (3\text{-}26)$$

Note we have performed optimization independent of the classifier which is assumed to be independent of λ_{ij}. Depending on the classifier utilized, similar optimization techniques could be used for its parameters.

Probabilistic Model

In the probabilistic approach, we model context using a class of functions on random sets called random measures. That is, for each observed set we construct a corresponding measure. We perform analysis in this space of measures rather than in closed subsets of $\mathcal{E}$, or $\Im$, as in previous models, in hopes of extracting supplementary information to that found during analysis in $\Im$.

Development

Recall in Equation 2-33, a likelihood function was derived for a Gibbs point process using an energy function U which was used to assign likelihood based on the configuration of points in some set *X*. We have noted that different forms of *U* yield different issues and may imply certain constraints on a point process.

We now define an unnormalized likelihood function using an energy functional which calculates the energy of a particular configuration by analyzing an observed function or measure. The goal is to permit a tractable contextual estimate, as opposed to an energy function as in Equation 2-35. Furthermore we desire the ability to analyze the *shape* of a function across $\mathcal{E}$ rather than inspecting pairs of elements in $\mathcal{E}$ as in Equation 2-36. Also, we define the likelihood

function such that it can be parameterized to recognize different random measures, whereas Gibbs point processes are typically used to calculate probability using the energy of a closed system and not necessarily distinct random measures' characterizations.

Since we are analyzing functions, we use the KL divergence on functions. We note that other measures or divergences on functions may be used as well. We define the energy functional for random measure M_{Ξ} as

$$U_{\Xi}(P_X) \equiv KL(P_X \parallel Q_{\Xi}). \tag{3-27}$$

We refer to Q_{Ξ} as the *representative measure* for random measure M_{Ξ} and it can be thought of as a parametric representation of Ξ. We can now define the unnormalized likelihood functional for random measure M_{Ξ} as

$$p_{M_{\Xi}}(P_X) \equiv \exp\left(-KL(P_X \parallel Q_{\Xi})\right). \tag{3-28}$$

Note that this likelihood compares how measure is distributed between the function P_X and Q_{Ξ}. Hereafter, we denote Q_{Ξ} by Q or Q_i for a particular context *i*. If the distribution of mass in P_X becomes more similar to that in Q, a higher likelihood is assigned to P_X, using the KL divergence to assess similarity. Therefore, an intuitive value for Q would be the measure that minimizes the sum over the KL divergences of observed samples $D = \{P_{X_1}, P_{X_2}, ..., P_{X_n}\}$ from M_{Ξ_i},

$$Q \equiv \underset{R \in \mathcal{M}}{\arg\inf}\left(\sum_{j=1}^{N} KL\left(P_{X_j} \parallel R\right)\right). \tag{3-29}$$

Hereafter, we denote the densities corresponding to measures *Q* and P_X as *q* and ν_X respectively, and assume they exist. The likelihood function defined in Equation 3-28 is used for contextual estimation given the random set framework for context-based classification.

Specifically, we use the likelihood on random measures to calculate the contextual estimation term.

$$P(X \mid \Xi_i) \propto \exp\{-KL(P_X \parallel Q_i)\} \tag{3-30}$$

In Equation 3-30, Q_i is the representative measure for context *i* and P_X is the measure corresponding to observed set *X*. We use the KL divergence to compare distributions using their corresponding densities ν_X and q_i, to determine the likelihood of context *i*. Therefore, we can calculate or approximate Equation 3-30 using Equations 3-31 or 3-32, respectively.

$$P(X \mid \Xi_i) = P(\nu_X \mid q_i) = \exp\left\{-\int \nu_X(x) \log\left(\frac{\nu_X(x)}{q_i(x)}\right) dx\right\} \tag{3-31}$$

$$P(X \mid \Xi_i) = P(\nu_X \mid q_i) \approx \exp\left\{-\sum_{x \in A} \nu_X(x) \log\left(\frac{\nu_X(x)}{q_i(x)}\right) \Delta x\right\} \tag{3-32}$$

In Equation 3-32, $A \subseteq \mathcal{E}$ is used to estimate the KL divergence. The choice of *A* is further detailed in the Discussion section.

The choice between Equations 3-31 and 3-32 depends on the formulation of parameter q_i, specifically, whether an analytical representation of the KL divergence exists or whether it is convenient for parameter estimation given an assumed parametric form of the model.

The density q is the parameter for $P(\nu_X \mid q)$, which itself may be parameterized for convenience, for example, $q \sim N(\mu, \Sigma)$ or $q \sim Exp(\lambda)$. We note that estimation may benefit if density q is modeled using a more complex distribution such as a Gaussian mixture; however, this may lead to difficulty in computation and may complicate parameter learning [98].

In the probabilistic approach, we need to construct the ν_X given some observed set *X*. One possible construction would be to use a simple Lesbegue or uniform measure over the discrete points in *X*.

Example 3.1 Discrete measure: Assume $X = \{x_1, x_2, ..., x_n\}$. Then we could construct our measure M_X using a cardinality based measure μ_c such that

$$P_X(F) = \frac{\mu_c(F \cap X)}{\mu_c(X)} = \frac{|F \cap X|}{|X|}. \tag{3-33}$$

We note that this measure meets the requirements outlined in the definition of a random measure; however, it is discontinuous, not smooth, which may lead to optimization issues. Furthermore, as we will see during the construction, issues may arise if M_X has a limited support. Therefore, it is beneficial to provide a parametric measure to provide a smooth measure with a large support.

If we use the generalized development of the random measure, and therefore the general construction of an instance of a random measure as in Equation 2-40, we can develop a parametric measure that is continuous and has a large support, given some assumptions.

Example 3.2 Continuous parametric measure: Assume $X = \{x_1, x_2, ..., x_n\}$ are a finite number of observations from some infinite set $\mathcal{X} \subseteq \mathcal{R}^d$. If we assume that elements in *X* are similarly distributed to this continuous set in space $\mathcal{R}^d$ we could estimate the measure on this set using parameters calculated from *X*, and define a measure $M_X : \mathcal{F} \to [0,1]$, by

$$P_X(F) = \frac{\sum_{X \cap F} \varsigma_X(x)\Delta x}{\sum_X \varsigma_X(x)\Delta x} = \frac{\sum_{X \cap F} N(x \mid \mu_X, \Sigma_X)\Delta x}{\sum_X N(x \mid \mu_X, \Sigma_X)\Delta x} \approx \frac{\int_{x \in F \cap \mathcal{R}^d} N(x \mid \mu_X, \Sigma_X)dx}{\int_{x \in \mathcal{R}^d} N(x \mid \mu_X, \Sigma_X)dx}. \tag{3-34}$$

We estimate the center of mass μ_X and covariance function Σ_X of the set $\mathcal{X}$ using the set of observed finite samples in *X* and use these estimates for the parameters of the Gaussian density. We have therefore constructed an example of a measure given an observed sample *X*, which is continuous, has a large support, and has a parametric form.

Other parametric forms of ν_X could be developed through many existing methods. If we assume a complicated parametric form for ν_X, some methods that might be used to estimate ν_X, such as the standard EM algorithm, may be subject to initialization conditions and therefore will not strictly satisfy Equation 2-40.

Optimization

Next, we develop optimization strategies and example model implementations that would use Equation 3-31 or 3-32. The developed probabilistic model allows for closed form solutions for optimization given certain model assumptions and appropriate objective functions. Roughly

speaking, the optimization of parameter q, using parametric representations of ν_X, proceeds in two main steps. During the first step, parameters of the densities ν_X are estimated for each $X \in \mathbf{X}$, using standard methods such as EM or ML estimates. The result is a set of densities, and therefore measures, $\{\nu_{X_1}, \nu_{X_2}, ..., \nu_{X_n}\}$. In the second step, representative measure q_i is estimated for each random set Ξ_i by maximizing a likelihood function that is a product of factors involving context dependent classification factor $p(x \mid y, \Xi_i)$, context estimation factor $P(X \mid \Xi_i) = P(\nu_X \mid q_i)$, and prior $P(\Xi_i)$, with respect to function q_i. We focus on the maximization of $P(\nu_X \mid q_i)$ since the classification factors of each context can be estimated using standard techniques. Note that factor $P(\nu_X \mid q_i)$ treats ν_X as the samples rather than *x* as in standard methods.

Specifically in the first optimization example, we assume a form of ν_X and q such that the integral in Equation 3-31 can be calculated analytically. We take an EM approach for optimization; specifically we take an expectation over the contextual parameters given each ν_X constructed from observation set *X*. We assume $q \sim N(\mu, \Sigma)$ and $\nu_X \sim N(\mu, \Sigma)$. Initially each ν_X is constructed from the observed samples from the corresponding $x \in X$. Once each ν_X is constructed, the individual elements of the sets $x \in X$ are no longer referenced in the optimization process.

We begin by defining our objective and corresponding log likelihood function given our initial independence assumptions of the random set framework arriving at

$$L(\Theta) = \log\left(\prod_{X \in \mathbf{X}} P(X \mid \Xi_i) P(\Xi_i) \prod_{x \in X} p(x \mid y, \Xi_i) \right). \tag{3-35}$$

Next, we take an expectation over the contextual parameters given our observed populations,

$$\underset{\Xi|X}{E}\left[L(\Theta)\right] = \sum_{X\in\mathbf{X}}\sum_{i=1}^{I}\left[\log\left(P(X\mid\Xi_i)\right)+\log\left(P(\Xi_i)\right)+\sum_{x\in X}p(x\mid y,\Xi_i)\right]P(\Xi_i\mid X)\,. \tag{3-36}$$

We disregard the classification term for now, as this type of optimization is ubiquitous throughout the literature, and therefore we focus on the contextual terms.

$$R(\Theta) = \sum_{X\in\mathbf{X}}\sum_{i=1}^{I}\left[\log\left(P(X\mid\Xi_i)\right)+\log\left(P(\Xi_i)\right)\right]P(\Xi_i\mid X) \tag{3-37}$$

Using Equation 3-31, we get

$$R(\Theta) = \sum_{X\in\mathbf{X}}\sum_{i=1}^{I}\left[\log\left(\exp\{-KL(\nu_X\parallel q_i)\}\right)+\log\left(P(\Xi_i)\right)\right]P(\Xi_i\mid X)\,. \tag{3-38}$$

After some algebra we arrive at

$$R(\Theta) = \sum_{X\in\mathbf{X}}\sum_{i=1}^{I}\left[-\int\nu_X(x)\log\left(\frac{\nu_X(x)}{q_i(x)}\right)dx+\log\left(P(\Xi_i)\right)\right]P(\Xi_i\mid X). \tag{3-39}$$

Analytically integrating and ignoring a constant [98], we arrive at

$$R(\Theta) = \sum_{X\in\mathbf{X}}\sum_{i=1}^{I}\left[-.5\left(\log\left(\frac{\left|\Sigma_{q_i}\right|}{\left|\Sigma_{\nu_X}\right|}\right)+Tr\left[\Sigma_{q_i}^{-1}\Sigma_{\nu_X}\right]+(\mu_{\nu_X}-\mu_{q_i})^T\Sigma_{q_i}^{-1}(\mu_{\nu_X}-\mu_{q_i})\right)+\log\left(P(\Xi_i)\right)\right]P(\Xi_i\mid X) \tag{3-40}$$

We then perform the maximization step by differentiating Equation 3-40 with respect to the parameters. At this point we note that many closed form representations can be found for the KL divergence of distributions other than the Gaussian, such as the exponential distribution. Setting the result of the differentiation of Equation 3-40 to zero and solving for parameters μ_{q_i}, Σ_{q_i}, and $P(\Xi_i)$, results in update Equations 3-41, 3-42, and 3-43, respectively.

$$\mu_{q_i} = \frac{\sum_{X \in \mathbf{X}} \mu_{\nu_X} P(\Xi_i \mid X)}{\sum_{X \in \mathbf{X}} P(\Xi_i \mid X)} \tag{3-41}$$

$$\Sigma_{q_i} = \frac{\sum_{X \in \mathbf{X}} \left[\Sigma_{\nu_X} + (\mu_{\nu_X} - \mu_{q_i})(\mu_{\nu_X} - \mu_{q_i})^T\right] P(\Xi_i \mid X)}{\sum_{X \in \mathbf{X}} P(\Xi_i \mid X)} \tag{3-42}$$

$$P(\Xi_i) = \frac{\sum_{X \in \mathbf{X}} P(X \mid \Xi_i)}{\sum_{i=1}^{I} \sum_{X \in \mathbf{X}} P(X \mid \Xi_i)} \tag{3-43}$$

Finally, we use Bayes' rule to solve for $P(\Xi_i \mid X)$.

$$P(\Xi_i \mid X) \propto P(X \mid \Xi_i) P(\Xi_i) \tag{3-44}$$

Recall, $P(X \mid \Xi_i)$ is given by Equation 3-31.

However as previously mentioned, if a more complex distribution is assumed for the model or the sample ν_X, the KL divergence may not have a closed form representation. We now develop an optimization strategy for this case.

Assume the representative measure is a Gaussian mixture, $q_i \sim \sum_{j=1}^{n_i} N(\mu_{ij}, \Sigma_{ij})$, which does not permit a closed form solution. We note there are numerical / statistical methods that can be used to help estimate the KL divergence [98]; however, the optimization of the parameters in q_i would become an issue if those techniques were used.

For development of this optimization technique, we skip to Equation 3-37 and substitute in Equation 3-32 arriving at

$$R(\Theta) = \sum_{X \in \mathbf{X}} \sum_{i=1}^{I} \left[-\sum_{x \in A} \left[\nu_X(x) \log\left(\frac{\nu_X(x)}{q_i(x)} \right) \Delta x \right] + \log\left(P(\Xi_i)\right) \right] P(\Xi_i \mid X). \tag{3-45}$$

Upon inspection, we see that optimization with respect q_i is analogous to minimizing the KL divergences between each ν_X and q_i. If we assume q_i is a Gaussian mixture, with some algebra, we arrive at

$$R(\Theta) = \sum_{X \in \mathbf{X}} \sum_{i=1}^{I} \left[-\sum_{x \in A} \left[\nu_X(x) \log\left(\nu_X(x)\right) \Delta x - \nu_X(x) \log\left(\sum_{j=1}^{J} \pi_{ij} N(x \mid \mu_{ij}, \Sigma_{ij}) \right) \Delta x \right] + \log\left(P(\Xi_i)\right) \right] P(\Xi_i \mid X). \tag{3-46}$$

After performing the maximization step for parameter μ_{ij} we can get a closed form solution assuming Equation 3-48 is independent of μ_{ij}.

$$\mu_{ij} = \frac{\sum_{X \in \mathbf{X}} \left[\sum_{x \in A} \left[x[\Delta x \nu_X(x)] \gamma_{xij} \right] \right] P(\Xi_i \mid X)}{\sum_{X \in \mathbf{X}} \left[\sum_{x \in A} \left[\Delta x \nu_X(x) \gamma_{xij} \right] \right] P(\Xi_i \mid X)} \tag{3-47}$$

$$\text{where } \gamma_{xij} = \frac{\pi_{ij} N(x \mid \mu_{ij}, \Sigma_{ij})}{\left(\sum_{j=1}^{J} \pi_{ij} N(x \mid \mu_{ij}, \Sigma_{ij}) \right)} \propto p(\mu_{ij} \mid x) \tag{3-48}$$

While updating the parameters, we assume γ_{xij} is independent of the other parameters, which is a common assumption in machine learning [97]. In fact, this result is a similar to the result attained using a standard EM approach, taking the expectation over each component given the individual samples using $p(\mu_{ij} \mid x)$ [97]. The other parameters are solved similarly,

$$\Sigma_{ij} = \frac{\sum_{X \in \mathbf{X}} \left[\sum_{x \in A} (x - \mu_{ij})(x - \mu_{ij})^T \left[[\Delta x \nu_X(x)] \gamma_{xij} \right] \right] P(\Xi_i \mid X)}{\sum_{X \in \mathbf{X}} \left[\sum_{x \in A} \left[\Delta x \nu_X(x) \gamma_{xij} \right] \right] P(\Xi_i \mid X)} \tag{3-49}$$

$$\text{and } \pi_{ij} = \frac{\sum_{X \in \mathbf{X}} \left[\sum_{x \in A} \left[\Delta x \nu_X(x) \gamma_{xij} \right] \right] P(\Xi_i \mid X)}{\sum_{X \in \mathbf{X}} \left[\sum_{x \in A} \left[\Delta x \nu_X(x) \right] \right] P(\Xi_i \mid X)}. \tag{3-50}$$

Optimization is again performed in sequence with parameter γ_{xij} being calculated last in each epoch.

To properly calculate the factor Δx in the update Equations 3-47, 3-49 and 3-50, we use the standard approximation of the Riemann integral. If *x* is multidimensional, $x \in \mathbb{R}^d$, construction of Δx involves creating incremental volumes ΔV by constructing a hypergrid or hyper-rectangles. Hereafter, we refer to ΔV as Δx. One intuitive method of constructing the set *A* would be to construct samples by taking all N^d combinations of the *N* samples in *X* in each dimension *d*. However, if samples *x* are multidimensional, then construction of Δx may be intractable. If a smaller *A* was constructed, the Riemann approximation may decrease in accuracy.

We propose an efficient estimation of the KL divergence that assumes Δx is constant and that the samples that comprise *A* are uniformly sampled from some hyperrectangle created from observations of the distributions ν_X and q_i. This approximation, which is similar to Markov Chain Monte Carlo (MCMC) integration, is intuitive since if the samples are, in fact, uniformly distributed, Δx should be constant. In the experiments, we analyze the error using synthetic and real data sets.

Discussion

There are many interesting results of this derivation. For clarification, we first provide a few examples in order to flesh out some of these details. Next we discuss certain similarities and distinctions between the proposed method, standard methods and variational methods. In particular we compare optimization and inference results of the proposed method to standard statistical methods. Lastly, we compare the proposed method to typical variational methods for approximate inference.

We noted earlier that in the construction of the proposed likelihood function we hoped to gain some versatility over standard probabilistic approach that assume i.i.d. However, some approaches that could be employed for the construction of ν_X, may implicitly assume that the singleton elements of *X* are i.i.d. However, we note these effects do not necessarily trickle up to inference at the measure level. After we introduce the optimization methods, which helps identify some characteristics of the proposed approach, we illustrate some of the similarities and differences between using standard statistical approaches which assume i.i.d. and the proposed method.

Example 3.3 Construction of ν_X : Equation 3-32 can be rewritten

$$P(\nu_X \mid q_i) = \prod_{x \in A} \exp\left\{ -\nu_X(x) \log\left(\frac{\nu_X(x)}{q_i(x)} \right) \Delta x \right\}. \tag{3-51}$$

Note that that ν_X is a function of our observation set, *X* and therefore each term in Equation 3-33 is dependent on the set *X*. Note, the use of samples $x \in A$ is simply to estimate the KL divergence, that is, the only reason to use the underlying space is to sample the values of ν_X and q_i. In fact, the samples in *A* do not even need to be elements of the observation set *X*.

Since the likelihood function can be factorized as in Equation 3-33, we could interpret the resulting product as stating that each value $\nu_X(x)$ is distributed by standard random variable $M_{\Xi_i}(x)$ which is determined by random set Ξ_i and is represented by $q_i(x)$ given representative function q_i. Note that ν_X is a function of the set *X* and that each corresponding value $\nu_X(x)$ is drawn from a *distinct* random variable $M_{\Xi_i}(x)$ at each *x* in the domain of M_{Ξ_i}, as illustrated in Figure 3-1.

So in effect, a random measure is a continuum of random variables on some subset of $\mathcal{E}$, one for each element in the domain of M_Ξ, namely $M_\Xi(x)$. As mentioned, each random variable $M_\Xi(x)$ has a corresponding parameter $q(x)$. Note the parametric form of the representative function q, has allowed us to maintain a continuum of random variables in a concise manner, but at the cost of versatile forms of $q(x)$ and therefore $M_\Xi(x)$. That is, $M_\Xi(x_1)$ is a random variable that maps into $\mathcal{R}$ whose distribution is intrinsically governed by

the distribution of the random set Ξ. We shall refer to the value $q(x_1)$ as the representative value for random variable $M_\Xi(x_1)$.

Example 3.4 Random variable $M_\Xi(x)$: If we wanted to minimize the KL divergence of two probability measures, the two functions must coincide. Assume we wanted to minimize the sum of KL divergences between *q* and samples ν_X. At the optima each representative value $q(x)$ is the representative function's value at *x* which minimizes this sum of the KL divergences, given the constraint that the representative function must be probability measure. Assume we collect *N* samples from Ξ and have *N* corresponding measures. Note that given for each instance $\nu_{X_j}(x), j = 1,...,N$ is an instance of random variable $M_\Xi(x)$ for a fixed *x*. If we minimize the expression $\sum_j KL\left(P_{X_j} \| Q\right)$ with respect to each value $q(x)$ at a fixed *x*, using Equation 3-32 and subject to the constraint that *Q* must be a probability measure, that is $\sum_{x\in A} q(x)\Delta x = 1$, we arrive at

$$q(x) = \frac{\sum_j \nu_{X_j}(x)}{\sum_{x\in A}\sum_j \nu_{X_j}(x)\Delta x} = \frac{\sum_j \nu_{X_j}(x)}{\sum_j \sum_{x\in A} \nu_{X_j}(x)\Delta x} = \frac{\sum_j \nu_{X_j}(x)}{\sum_j 1} = \frac{\sum_j \nu_{X_j}(x)}{N}. \tag{3-52}$$

Note that $q(x)$ is the arithmetic mean of $\nu_{X_j}(x), j = 1,...,N$. This means the *representative value is the mean value for $M_\Xi(x)$ for each x, and therefore minimizes the squared Euclidean distance* between samples $P_X(x)$ from random variable $M_\Xi(x)$, as illustrated in Figure 3-1B.

One result of using a parametric form for the representative function is that the representative value $q(x)$ may no longer be the exact mean of random variable $M_\Xi(x)$ due to the particular constraints, for example if it is assumed Gaussian distributed. However, the assumption of a parametric model is important, otherwise, we would need a random variable for each point in the domain of M_Ξ which does not permit a tractable solution, unless a very simple domain is assumed. As found throughout machine learning techniques, there is a tradeoff between data fidelity and tractability.

Example 3.5 Representative value: Given a set of observed Gaussian measures constructed by selecting the mean and covariance from a uniform interval, assume we wish to construct *Q* using the update Equations 3-47, 3-49 and 3-50. Note this implies we are assuming that *q* is Gaussian. The resulting representative values are not necessarily the arithmetic means of samples from $\nu_\Xi(x)$, as illustrated in Figure 3-1B. Although, the update equations are optimal

assuming Gaussian, there are not necessarily optimal over all possible distributions due to this extra constraint.

The first optimization technique proposed that uses Equation 3-31 to calculate the KL divergence, learns the parameters of *q* using some parameters of our observed distribution ν_X. However, in the second optimization technique proposed, the parameters of *q* are learned using the underlying space, samples in $\mathcal{R}^d$. We note there is some similarity between these update equations and those that are developed in standard EM algorithms such as Equation 3-53.

$$\mu_{ij} = \frac{\sum_{x \in A} x \gamma_{xij}}{\sum_{x \in A} \gamma_{xij}}. \tag{3-53}$$

However, we note that in the proposed update equations, there is a discrete expectation over random measures, not simply an expectation over standard random variables. We also note that when samples are clustered, the set *A* is typically the data. However, in the proposed approach, the samples in set *A* are not directly important, as long as their use permits a good estimate of the KL divergence.

The major difference in the update formulas is the factor $\nu_X(x)\Delta x$. Note that in the KL divergence we integrate with respect to our sample ν_X, which is also a density. In the discrete approximation, the factor $\nu_X(x)\Delta x$ is used instead. One interpretation is that we are taking the expected value of the difference between ν_X and q_i. This interpretation shows, that during optimization, we are trying to minimize the difference between samples ν_X and representative measure q_i.

Another interpretation is the representative function q_i is being coerced into a form similar to the samples ν_X vicariously through its parameters μ and σ using samples *x* in *A* and weights

$\nu_X(x)\Delta x$, as illustrated in Figure 3-2. This coercion is performed through the parameters, for example, the means μ, which reside in the same space as *x*. For this reason, the samples *x* are included in the update equations. However in Equation 3-47, the factor $\nu_X(x)\Delta x$ weights each sample *x* by its corresponding measure in the distribution, ν_X. In fact, μ is optimized such that $q(x)$ is similar to $\nu_X(x)$, not necessarily to maximize $q(x)$ with respect to the samples *x* as is the case with standard statistical optimization.

However, there are similarities in standard statistical optimization and the proposed method. In standard statistical methods, the learned posteriors / likelihoods are optimized while assuming i.i.d. In the proposed method, the representative function is optimized using observed measures which may have been constructed using similar optimization techniques that are used in standard statistical methods. In the developed approach, the observed measures are essentially likelihood functions optimized with respect to each observed set, and therefore most likely assume i.i.d. during their construction. We illustrate situations when the proposed methods result in similar and different optimizations than standard methods.

Example 3.6 Optimization similarities and differences: Assume we have multiple observation sets $D = \{X_1, X_2, ..., X_N\}$ observed in the same context and we wish to optimize likelihood functions for context estimation. We optimize a standard likelihood function, which assumes i.i.d., using the EM algorithm while training on the dataset $X = \bigcup_{i=1}^{N} X_i$. We will also learn the proposed random measure likelihood function by optimizing the representative function given observed measures $\{\nu_{X_1}, \nu_{X_2}, ..., \nu_{X_N}\}$ using the method in Equations 3-47, 3-49, and 3-50. We will construct the observed measures using the standard EM algorithm for Gaussian mixtures.

Results are illustrated and further detailed in Figure 3-3. Note that the resulting likelihood from EM optimization results in a measure learned from the set of all singleton samples, whereas, the learned representative function is a measure which was learned from a set of measures. If the distribution of *X* is similar to that of each X_i with respect to the number of samples in the distribution, the representative measure learned will be similar to the likelihood learn using standard methods. This is because all information can be detailed without any set information; however, if the distributions are different with respect to the number of samples, the learned

measures will be different. This result is illustrated in Figure 3-3. This distinction is a direct result of the proposed methods ability treat each set as a unitary element.

We have identified some fundamental differences between the proposed method and standard techniques. Note there are some similarities and differences when performing inference using standard techniques and the proposed technique. In many cases the calculation of likelihood is different during inference; however in some cases, the result of inference—determination of the most probable context—is similar. In fact, if the representative measure q is the actual learned likelihood of the standard method, that is $p(x \mid y) = q(x) \; \forall x$, then the result of inference will be the same. This shared similarity between the two approaches is again shared if the distribution of X is similar to that in each X_i.

Example 3.7 Inference similarities and differences: The random measure approach assigns high likelihood to sets, or random measure instances, that have a similar distribution throughout the domain; whereas standard approaches assign high likelihood as long as each observed singleton sample appears in a place of high likelihood. This difference is illustrated in Figure 3-3C, which continues from Example 3.6. Note that although this is a fundamental difference, the result of context estimation may be similar using both approaches depending on the observed measures construction and the results of optimization.

During the optimization of the proposed likelihood function, the representative function is learned. This is similar to variational methods where functions are learned by optimizing objective functionals.

Example 3.8 Comparison with standard inference using variational methods: Given an observed set X, we want to determine if it was observed in context i. Using the proposed method on random measures, we would first construct ν_X. Next, we could determine the unnormalized likelihood of some context using $p(\nu_X \mid M_{\Xi_i}) = \exp\big(-KL(P_X \mid\mid Q_i)\big)$. Whereas, with a standard variational method, or most standard methods of inference given a joint observation set, the initial observation set is explicitly assumed i.i.d., during optimization and inference. For example, the standard initial assumption made in variational inference given a set X is

$$p(X \mid Z) = \left(\frac{\tau}{2\pi}\right)^{N/2} \exp\left\{-\frac{\tau}{2}\sum_{n=1}^{N}\left(x_n - \mu\right)^2\right\} \tag{3-54}$$

where $Z = \{\mu, \tau\}$ [97]. Therefore, the estimate of the posterior $p(Z|X)$ also is i.i.d.

In Example 3-9, we compared the proposed method to standard variational methods; however, we ignored the use of hyperparameters. The hyperparameters would be better suited for contextual inference since they govern distributions on distributions and inference could be performed on observed measures. We explore the viability of using the subsequent hyperparameters for a means of context estimation.

Example 3.9 Context estimation using variational methods: For the construction of the hyperparameters, assume Equation 3-54. We then model parameters μ and τ using a normal and gamma distribution, respectively.

$$p(\mu \mid \tau) = N\left(\mu \mid \mu_i, (\lambda_i \tau)^{-1}\right)$$
$$p(\tau) = Gamma\left(\tau \mid \alpha_i, \beta_i\right) \quad (3\text{-}55)$$

It can be shown that the parameter μ_i is updated using

$$\mu_i^{t+1} = \frac{\lambda_i \mu_i^t + N\bar{x}}{\lambda_i + N}. \quad (3\text{-}56)$$

We note this is similar to update Equation 3-41, save the expectation over sets used by the proposed method. Therefore it cannot treat set values as unitary elements and will differ from the proposed method similarly to standard statistical methods, as illustrated in Examples 3-6 and 3-7.

Again, note that Equation 3-56 is somewhat similar to the optimization of the random set, where the Gaussian is the measure resulting from the update Equations 3-41, 3-42, and 3-43. The difference here is that there is a prior distribution on the parameters of some family of distributions. This simplifies computation to some degree as the random element is reduced to being a standard random variable in $\mathcal{R}^d$. Note that this is an *atypical use of the intermediate constructs of standard variational inference;* however, this potential use fits the problem of developing a likelihood on functions given a simple model.

Example 3.10 Context estimation using a mixture of Gaussian hyperparameters: In Example 3.9, we constructed a hyperparameter given a single Gaussian measure constructed from an observation set *X*. We can similarly construct a mixture of Gaussians given an observation set *X*. Given a set of observed parameters $\{\boldsymbol{\mu}, \boldsymbol{\Lambda}\}$, developed from some observed set *X*, we can estimate the likelihood of some context *i* given some trained parameters $\{\mathbf{m}_i, \boldsymbol{\beta}_i, \mathbf{W}_i, \mathbf{V}_i\}$, learned given an assumed Gaussian-Wishart prior governing the mean and precision of each Gaussian component

$$p(\boldsymbol{\mu}, \boldsymbol{\Lambda} \mid \mathbf{m}_i, \boldsymbol{\beta}_i, \mathbf{W}_i, \mathbf{V}_i) = \prod_{j=1}^{J} N\left(\mu_j \mid m_{ij}, (\beta_{ij}\Lambda_j^{-1})\right) W\left(\Lambda_j \mid W_{ij}, V_{ij}\right). \quad (3\text{-}57)$$

This development by Bishop [89] has surfaced a few inherent issues that accompany this approach. First there is the assumption that the hyperparameters is factorizable, which was previously mentioned and may or may not be that constraining dependent on application area. However, the fact that we are now performing inference in the parameter space, rather than the space of measures has lead to other issues. Note that μ_j and m_{ij} are both indexed by *j*, although they both are elements of sets $\boldsymbol{\mu}$ and $\mathbf{m}_i$, respectively. This implies that in order to properly calculate the likelihood, there must be the same number of observed samples as there are Gaussian-Wishart priors and that the observations and distribution components must be matched.

These issues are a direct result of the hyperparameters being intermediate constructs. These constructs have one purpose, which is to model one set of observations. In fact, they are not meant to be used directly since their only use is to integrate out intermediate parameters. That is why these standard variational learning should not be used for context estimation.

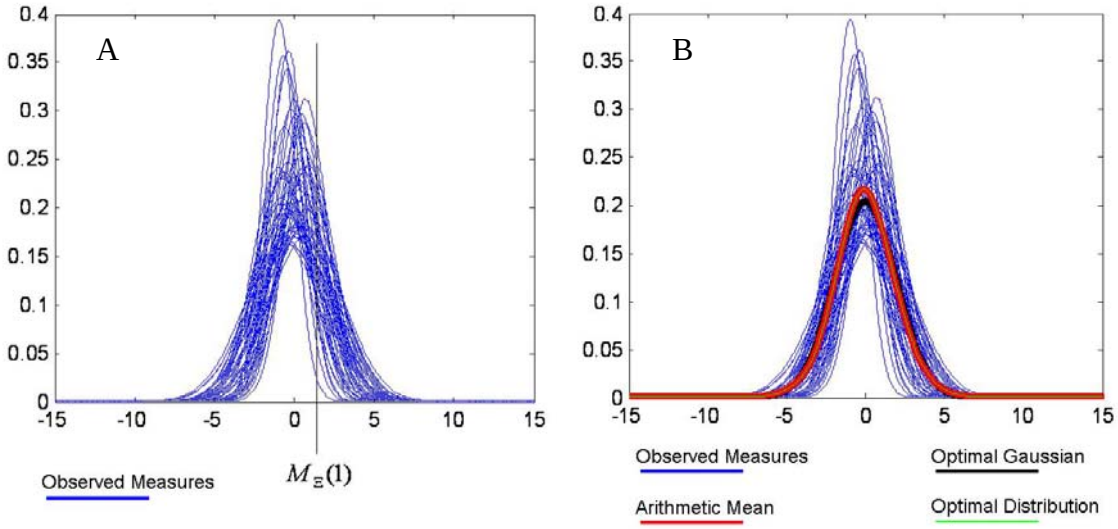


Figure 3-1. Samples of Gaussian distributions drawn using randomly selected means and variances which where drawn uniformly from a specified interval. A) Fifty sample measures are plotted. The resulting value at each point x, is a random variable. For example, random variable $M_{\Xi}(1)$ has corresponding samples that that lie on the line x=1. B) The arithmetic mean, optimal Gaussian and optimal Distribution are shown given the 50 Gaussian samples. The corresponding KL divergence values are 88.5, 91.2 and 88.5, respectively. The arithmetic mean is the optimal distribution; they coincide.

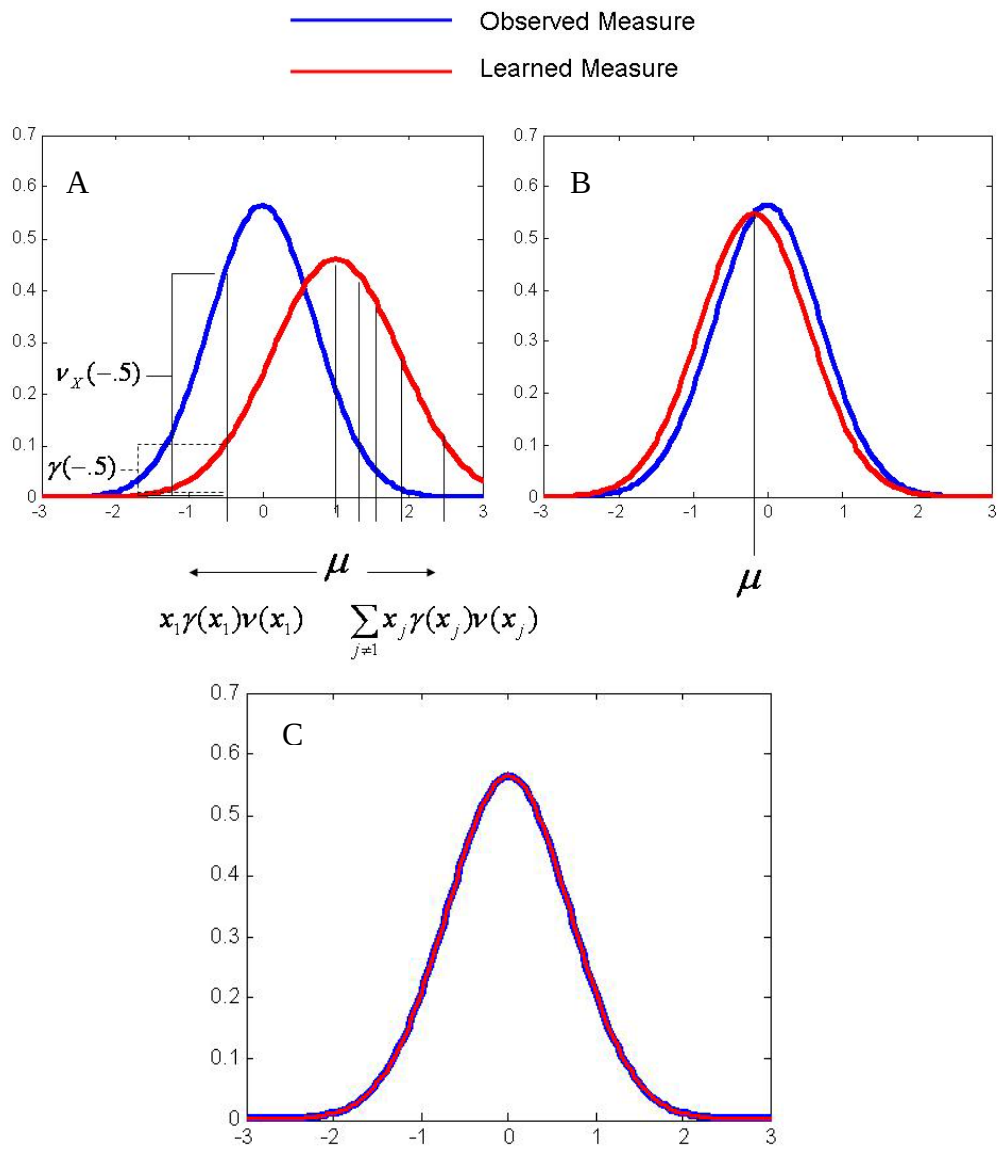


Figure 3-2. Learning the representative function using update Equations 3-47, 3-49 and 3-50 given set $A = \{-.5,1,1.3,1.5,1.9,2.5\}$. These plots illustrate the fact that the proposed method learns the function ν_X and does not fit the learned parameters to the individual samples in *A*. A) The observed measure and the initialized learned measure *q*. In standard learning techniques, optimization of the parameter μ would occur when it was the mean value of the samples in *A*, 1.28. However, the proposed objective is optimized when the correct function is learned. Parameter μ is coerced toward point $x_1 = -.5$, since $\nu_X(-.5)$ is large compared to the other samples in *A*. B) After a couple of iterations, μ becomes -.33. It should be clear that optimization coincides with function matching rather than fitting the function to the samples in *A*. C) If we use the set *A* which is a uniform sampling of 61 points in the range [-3,3], we get a better estimate of the KL divergence and the learned measure coincides with the observed measure.

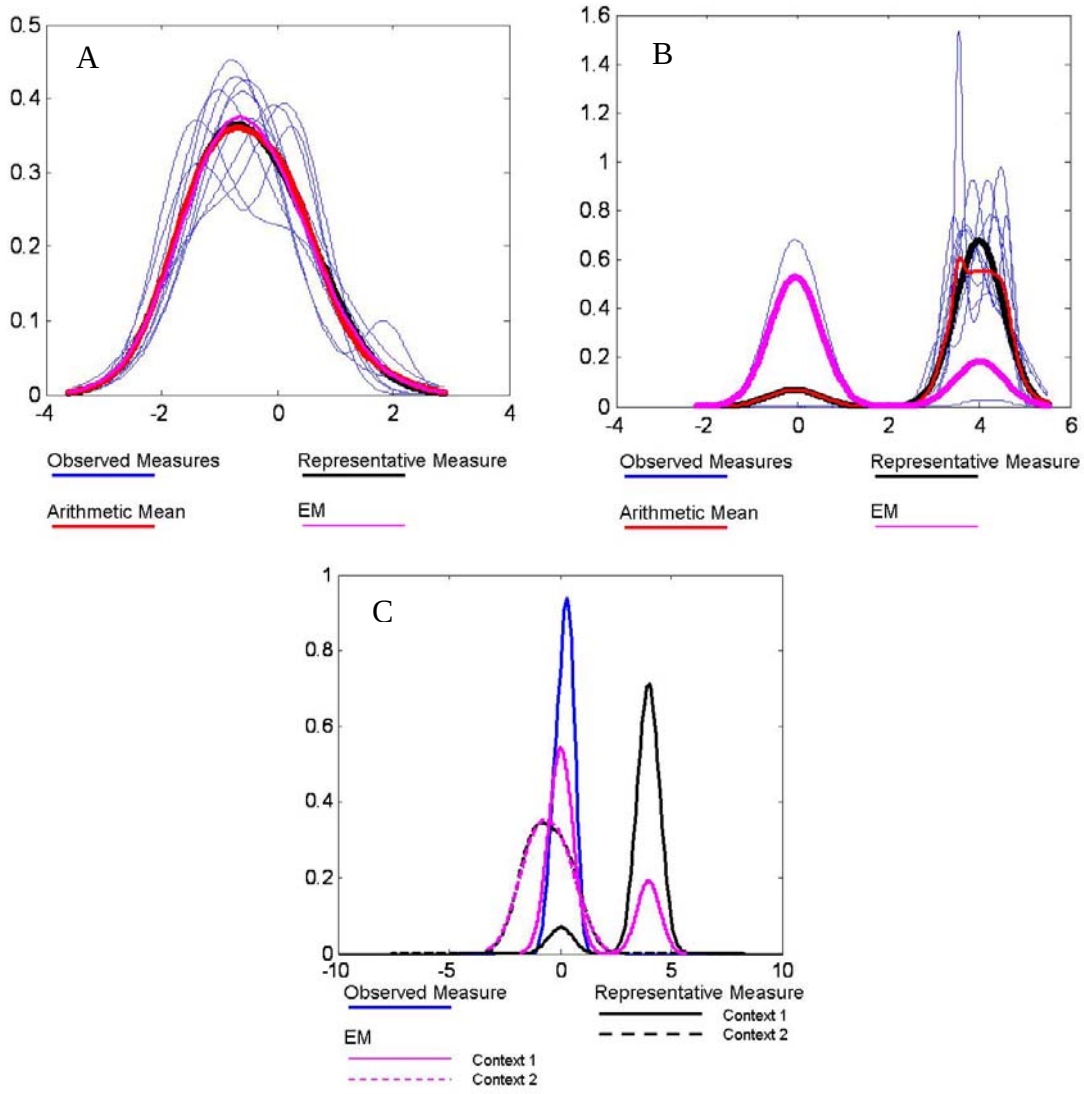


Figure 3-3. Similarities and distinctions between the proposed method and standard methods. A) The resulting EM likelihood and representative measure when optimized with respect to 10 observed sets (observed in context 2) each with a similar distribution as their union. B) The learned EM likelihood and representative measure when presented with 10 observed sets (observed in context 1) where one set has a distinct distribution compared to the union. The proposed method assumes each measure is a single sample and does not weight the one set with a different distribution any differently than the other measures. However, the standard method looks at the distribution of the singleton samples. We have constructed the set with a different distribution to have a large number of singleton samples (comparatively), to emphasize this ideological difference. C) When presented with a test set, the contextual estimates vary greatly between the standard approach and the proposed approach. Using the standard approach, context 1 is the most probable (100% to 0%); whereas, using the proposed random measure approach context 2 is the most probable (83% to 17%). Using standard i.i.d. joint estimation, the likelihood of samples lying under the

observed measure will have a greater likelihood in context 1 since the likelihood estimate for context 1 has greater likelihood values (as opposed to the likelihood for context 2) in the corresponding domain (approximately [-1, 1]). However, when comparing the representative measures for each context to the observed measure, the representative for context 2 is more similar to the observed measure.

CHAPTER 4
EXPERIMENTAL RESULTS

The three methods for context estimation developed within the random set framework (RSF) were tested using synthetic and real datasets. Four major experiments were performed. In the first experiment, we analyzed the use of different KL divergence approximation methods for estimating context in the proposed random measure model. In the second experiment, each of the three methods was tested using synthetic datasets created to imitate data in the presence of contextual factors. We compared synthetic data classification results of the proposed RSF approaches to that of set-based kNN [12], [15] and the whitening / dewhitening transform [65]. The main purpose of the experiments using synthetic data sets is to identify situational pros and cons of each of the approaches. Each method's ability to identify correct context is evaluated through its classification results since the ultimate goal is classification. Hence, we may refer to our results as context estimation results, but show the classification error on a sample basis. In the third experiment, the proposed methods are applied to an extensive hyperspectral data set collected by AHI for the purposes of landmine detection. This data set exhibits the effects of contextual factors. The purpose of the experiments using real data sets is to show the applicability of the proposed random set framework to real-world problems. We compared the hyperspectral data classification results to that of set-based kNN and the whitening / dewhitening transform. In the final experiment, the possibilistic approach is compared to a similar classifier that does not use contextual information, and it is also compared to an oracle classifier that always selects the correct context for the purposes of context-based classification. These comparisons compare the possibilistic approach to, informally, its lower and upper bounds.

KL Estimation Experiment

Experiment 1 demonstrates the efficacy of five different constructions of the set *A* for KL estimation in the proposed random measure model. Recall that if Equation 3-32 is used for KL estimation, the set *A* must be constructed such that it admits a tractable calculation but not at the expense of correctness. Therefore, we varied both the construction and size of *A* and analyzed how each affects its ability to estimate context.

Experimental Design

We compared the results of context estimation using three synthetic datasets. In the experiments each training set is constructed randomly by sampling from a Gaussian mixture with two components. Three Gaussian mixtures are used to simulate three distinct contexts. Fifteen samples are generated from each component in each context. This experimental design attempts to simulate a two class problem within each of three contexts. Ten training populations are constructed from each of the three Gaussian mixtures to simulate sets of samples observed in 3 distinct contexts. Observed measures are then created using Equation 2-40 and assuming ς is a Gaussian distribution; training is performed using Equations 3-47, 3-49, and 3-50. A test population is then generated from one of the Gaussian mixtures, which is randomly selected, and its corresponding observed measure is created assuming it is Gaussian. The representative function in the random measure model is learned from the 10 training measures and used to estimate the correct context of the test measure.

Experiments were performed using three data sets where each data set, from one to three, represents an increasingly difficult context estimate problem due to highly overlapping contexts. The data sets are Gaussian sample sets, so all experiments are repeated 50 times. Examples of each dataset are illustrated in Figure 4-1.

Each of the random measure models under test uses Equation 3-32 to estimate the KL divergence and performs contextual estimation using the random measure likelihood function as in Equation 3-32. The five methods used to construct the set *A* are as follows. In the Riemann test method, *A* is composed of N^d samples constructed by taking all combinations of test sample values in all dimensions and a Riemann integral is approximated. In Riemann test and train method, *A* is a constructed as in the Riemann test, but the samples are constructed using testing and training samples. In the naïve test method, *A* contains the observed test samples and Δx is assumed constant. In the naïve test and train method, *A* contains the observed test samples and the observed training samples and Δx is assumed constant. In the uniform MCMC method, *A* is the result of sampling a uniform distribution from within the hyperrectangle covering the train and test samples and Δx is assumed constant. Note the Riemann test and Riemann test and train methods are the same during the training phase, but differ during testing. The same is true for naïve test and naïve test and train methods. We point this out since during training only the training samples are used by all of the methods.

The Riemann approaches approximate the Riemann integral, which is a fairly standard approach. However it may be intractable for high dimensional data. Using the observed samples to partition the space into these grids would require an exponential number of elements in *A* with respect to the number of observed samples.

In the naïve test approach, *A* is simply the observed test samples and Δx is assumed constant. In the naïve test and train approach, *A* is simply the union of the test and training samples of the particular context which is to be inferred. We note these approaches are very tractable but we hypothesize that they will not be good estimates of the KL divergence.

The uniform MCMC method constructs a hyperrectangle that covers the testing and training sets using simple min and max operations. Then a fixed number of samples, the same number of samples in the observed set X in this experiment, are uniformly sampled from the covering hyperrectangle and Δx is constant for the approximation. The intuition behind this approach is that if the samples are truly uniform, Δx should be similar for each sample. The hypothesis is that this method will balance tractability and correctness.

Fifty experiments are run on each of the three data sets. For each method the representative measure is assumed to be Gaussian. The resulting contextual estimation results are compared to those attained by the random measure model using the analytical KL solution. The error of the methods under test is the average difference between themselves and the analytical solution, which is assumed to be the correct estimate. We also compare the contextual estimation error as a function of the number of observed samples. The hypothesis is that as the number of samples increases, the KL estimates will improve.

Results

The results of context estimation are shown in Table 4-1. The Riemann approaches have the least error total for all three data sets. Uniform MCMC had a low error and performed slightly better than the Riemann test and train method for datasets 2 and 3. The naïve methods had the most error for each data set, and the naïve test method had the maximum error, 8.7%, on data set 2.

Interestingly, Riemann test, which only uses the test samples for estimation purposes, performs better than Riemann test and train which uses both test and train samples. This is due to our Riemann approximation. Due to the construction of A, Riemann test and train, will have considerably more elements in the set A. Although more elements may mean higher granularity and potentially a better estimate of Δx, it has also exacerbated error in estimation. We used the

upper bound estimate to approximate the integral, which means KL estimates are slightly high for each Δx. Therefore, if we have considerably more Δx we may have more error, even with the better granularity.

Given the error estimates, the uniform MCMC seems to perform similarly to the Riemann estimates. However, it takes much less time to calculate than the Riemann approaches. Figure 4-2A shows a plot of context estimation error versus the number of samples in the initial observation set. For the Riemann approaches, there are exponentially many points that are added to correctly partition the space like a grid. On the other hand, the uniform MCMC approach performs uniform sampling and constructs *A* to have the same number of samples that are in the observation set.

Figure 4-2B shows the computation time needed for the Riemann test and train and uniform MCMC methods versus the size of the observed set. Although the Riemann approaches perform slightly better at integral estimation, uniform MCMC does comparably well and needs a very small amount of relative computation time. The runtime for the Riemann approach is exponential with respect to the number of observed samples, whereas the uniform approximation has a linear relationship as shown in Figure 4-2C.

Synthetic Data Experiment

The classification ability of the methods is under test in this experiment. Again synthetic data is created to simulate the effects of contextual factors. Four data sets are constructed such that each exposes a pro and/or con for each of the proposed methods. Each of the four data sets are illustrated in Figure 4-3 which helps to visualize the experimental setup and the purpose for each of the carefully constructed datasets. We also experiment with the whitening/dewhitening transform and set-based kNN to expose their pros and cons and for comparison purposes.

Experimental Design

Again samples are randomly generated from a Gaussian mixture with two components where samples from each component are assumed to be from the same class. Again, there are three contexts which allows for clear, less cluttered, analysis. Ten training populations are constructed from each of the three Gaussian mixtures to simulate sets of samples observed in three distinct contexts.

The contextual parameters, λ_{ij}, for the possibilistic and evidential models are optimized as described in Equations 3-23 and 3-26, respectively. In the probabilistic approach, using random measures, the observed measures are created using Equation 2-40, assuming they are Gaussian. The representative functions of the random measure likelihood functions were learned using the EM algorithm in Equations 3-47, 3-49 and 3-50, in a supervised manner. That is, each model's representative function was optimized using only the samples from the corresponding context.

We performed 50 trials on each data set; in each trial, a test set was generated randomly from one of the Gaussian mixtures associated with one of the contexts. For the random measure model, the corresponding measure was created using the standard EM algorithm assuming a Gaussian mixture of two components. The proposed evidential, probabilistic and possibilistic methods were equipped with Gaussian mixtures optimized separately using the standard EM algorithm. The contextual components were optimized separately as previously discussed.

The set-based kNN algorithm assigned, to each test sample, the label of the closest training sample in the closest set, that is, $k = 1$. The whiten / dewhiten transform was calculated as described in Equation 2-10 for each training image. The resulting confidence value was simply averaged over the training sets, since this algorithm does not provide for context estimation or relevance weighting.

Data set 1 is a fairly simple data set which should allow for simple context estimation and, within each context, simple classification. An example of data generated under data set 1 is shown in Figure 4-3. There are some disguising transformations present; however, the hypothesis is that most of the classifiers will perform well since context estimation is fairly simple in this data set.

In data set 2, we orient the Gaussians such that samples from class ‘x’ are relatively the same as compared to the samples from class ‘o’ in each of the three contexts. This data set was constructed to highlight the fact that the whitening / dewhitening transform assumes similar orientation of classes throughout each context. Therefore, the hypothesis is that the whitening/dewhitening transform will perform well on this data set. Each of the other methods should perform well since there remains only a slight presence of disguising transformations, and context estimation is therefore simple.

In data set 3, we introduce the presence of an outlier in the test set. The hypothesis is that the possibilistic approach should remain a good classifier since it has shown to be robust [16]. The evidential estimate will be affected by the outlier since it is a pessimistic approach. The probabilistic approach may be slightly affected if the observed measure is skewed toward the outlier. Set-based kNN will be affected by the outlier due to the use of the Hausdorff metric. The whitening / dewhitening transform may be affected since the outlier may drastically influence the whitening process.

In data set 4, we introduce multiple outliers which are placed relatively near to the observed samples. This data set is constructed to alter the observed measure and therefore, our hypothesis is that the probabilistic approach will be highly affected, along with the evidential approach and set-based kNN. The possibilistic approach should be unaffected by the outliers.

Classification results from the whitening / dewhitening transform will be drastically changed if the outliers greatly skew the whitening process.

Lastly, we analyzed the classification results of the evidential and possibilistic approaches, on dataset 3, when the number of germ / grain pairs was varied.

Results

The average classification errors are presented for each classifier for each dataset in Table 4-2. In data set 1, each method performed with under a 10% error and the best method, the evidential model, performed best with a 4.1% error. The whitening / dewhitening transform performed the worst since it relies on the fact that each class is relatively oriented in the same manner throughout each context, which is not the case (slightly) in data set1. The possibilistic approach performed the worst out of the proposed methods. Upon inspection, it fails to correctly identify context when an observed sample falls near a germ of an incorrect context. This is illustrated in Figure 4-3A. In the trial illustrated in Figure 4-3A, Context 3 is the most probable which is incorrectly estimated due to the close proximity of one of the test samples to the germ for context 3. The evidential and probabilistic models performed similarly, well. Set-based kNN performed slightly worse, which is attributable to the use of a nearest neighbor classifier as opposed to a Bayesian classifier.

In data set 2, the whitening / dewhitening transform results improved as expected. The evidential and probabilistic models performed equally as well. Set-based kNN and the possibilistic model performed relatively similarly.

In data set 3, the presence of an outlier drastically affected the classification results of the evidential model and set-based kNN. This data set is illustrated in Figure 4-3C. Both of the metrics used by these methods are pessimistic and are therefore affected by outliers. The

possibilistic and probabilistic approaches remained unaffected. Similarly, the whitening / dewhitening transform produced similar results as those found in data set 1.

In data set 4, the evidential model and set-based kNN remained highly affected by the presence of outliers. This data set is illustrated in Figure 4-3D. The incorporation of multiple outliers also affected the results of the probabilistic approach. The presence of multiple outliers was enough to greatly affect the observed measure and therefore tainted the context estimation. The possibilistic approach remained unaffected by the outliers. The whitening / dewhitening transform performed relatively well, although the samples from each class were not relatively oriented the same in each context. However, we note that in each context samples from class ‘o’ were to the right of samples from class ‘x’ in each class.

Table 4-3 shows the classification error, on data set 4, of the evidential and possibilistic models as the number of germ grain pairs varied. It also shows the classification error of the probabilistic approach for a baseline comparison. Overall the classification error decreases as the number of germ and grain pairs increases. This result is expected since more germ / grain pairs should allow for more detailed shape characterization.

Conversely, in standard techniques, the optimization of a statistical classifier using probability density functions may be subject to overtraining, especially if the number of densities used is increased or the number of densities is large compared to the number of training samples. In fact, if a probability density function is optimized with respect to a small number of samples the density will become focused on the few samples thus closing some abstract decision boundary tightly around said samples, causing overtraining. The overtraining during optimization corresponds to increasing the likelihood of samples in the correct probability density.

In the germ and grain model, the probability of sets of intersecting random hyperspheres increases as the hyperspheres grow. Therefore overtraining, in the aforementioned sense, is not an issue. However, optimization in the germ and grain model may cause the random radius to diverge, seemingly, the opposite of overtraining. Appropriate MCE optimization techniques, as developed here, must be implemented to prevent divergence.

However, classification error will increase with an increase in the number of germ / grain pairs, if the increase in pairs induces one of the situations outlined in Figure 4-3A, for the possibilistic approach, or Figure 4-3C and Figure 4-3D, for the evidential approach.

Hyperspectral Data Experiment

The classifiers under test were applied to remotely sensed, hyperspectral imagery collected from AHI [101], [102]. AHI was flown over an arid site at various times in the years 2002, 2003 and 2005. Eight AHI images which covered approximately 145,000m^2 were collected at altitudes of 300m and 600m with spatial resolutions of 10cm and 15cm, respectively. Each image contains 20 spectral bands after trimming and binning, ranging over LWIR wavelengths 7.88um – 9.92um. Ground truth was provided by Radzelovege *et al.* [100]. The maximum error was estimated to be less than one meter.

The scenes consisted mainly of targets, soil, dirt lanes and senescent vegetation. There are 4 types of targets. Targets of type 1 are plastic mines buried 10.2cm deep, targets of type 2 are metal mines buried 10.2cm deep, targets of type 3 are metal mines flush with the ground and targets of type 4 are circular areas of loosened soil, referred to as holes, with diameters less than one meter in length.

Since the imager was flown over the course of 4 years at various times of day, it is reasonable to assume that environmental conditions were variable. In fact, the presence of

contextual transformations including disguising transformations was confirmed, as shown in Figure 1-1.

Experimental Design

Labeled data sets were constructed from the imagery such that all samples from each data set were assumed to be observed in the same context. Training set construction was done manually since the ground truth error was large enough to prevent automation of this task. We note that the spectral signatures of all target types were similar enough to group into the same class given this data set. Each training set consisted of 10 samples from one of four classes: target, soil type 1, soil type 2, and vegetation. Therefore each training set, whose samples are assumed to be observed in the same context, consisted of 40 samples total. There were eight training populations, one from each image used to model the context of each image. Each context was modeled using four germ and grain pairs. The contextual parameters, λ_{ij}, were optimized using Equations 2-23 and 2-26 for the possibilistic and evidential approaches, respectively. Again, the probabilistic approach was trained using the EM algorithm in a supervised manner, that is, each model was optimized using only the samples from the corresponding context to be modeled. Gradient descent optimization for the evidential and possibilistic approaches was terminated after 200 iterations or sooner if the change was minimal. The germs were set to the results of *k*-means clustering of the samples of each class for each context. The learning rate for gradient descent optimization was set to 0.1.

The three context-based classifiers within the random set framework were equipped with Bayesian classifiers implemented as a mixture of Gaussians containing two components. The classification parameters were learned using the well-known EM algorithm in a supervised manner. Specifically, optimization for the mixture components modeling a particular class was

performed using only samples from the corresponding class in the corresponding context. Diagonal loading of the covariance matrices was done to mitigate the effects of low sample numbers and high dimensionality.

Set-based kNN was equipped with a simple classifier that was the inverse distance of the test sample to the closest representative of the target class, in the closest training set and k =1. This classifier permits gray level confidences, which allows for comparison to the other algorithms in the ROC curve.

The whitening/dewhitening transform was calculated as described in Equation 2-10 for each training image. The resulting confidence value was simply averaged over the training images, since this algorithm does not provide for context estimation or relevance weighting.

Test sets, or populations, were constructed from subsets of the imagery. The well-known RX algorithm [99] was run by Ed Winter from Technical Research Associates Inc. on the imagery as a pre-screener, or anomaly detector, to collect *points of interest* (POIs). There are 4,591 POIs and 1161 actual targets in the entire dataset. Sets of samples surrounding each POI in a 9x9 pixel window were collected to form test sets. This implies there is a total of 4,591 test sets each set consisting of 81 spectral signatures. Note that each test set is assumed to be a population, which means it is assumed that each sample in the set is observed in the same context. For this dataset, the 9x9 pixel window is large enough to encompass a target and background samples, but small enough to ensure that all samples have been observed in the same context.

Each sample in the test set is classified target or non-target by each of the classifiers. The probability of target is calculated for each sample within a test set and each POI is assigned a probability of target detection by taking the mean probability of target over the center samples

within a 3x3 window, since this is the standard size of a target. We note that the prescreener was not able to identify all targets in the scene, and the maximum *probability of detection* (PD) for the classifiers is 75% or 867 targets.

Cross validation is implemented at the image level, that is, spectra from a test image are not used for training purposes while said image is under test. Note that this testing procedure assumes that there exists a training population from an image other than the test image that contains samples observed in a context similar to those in the test image. We note that this may not be a valid assumption, and may make classification very difficult; however, this testing procedure mimics the testing conditions of real-world application, that is, the exact context and labels of some of the spectra from a test image may not be known *a priori*.

Classification results of all target types are presented in one *receiver operating characteristic* (ROC) curve which is shown by PD versus *false alarm rate* (FAR). We note that previous research has indicated that a minefield can be minimally detected when the PD is greater than 50% and the FAR is less than 10^{-2} FA/m^2 and is successfully detected when the PD is greater than 50% and the FAR is less than 10^{-3} FA/m^2 [100].

Results

ROC curves for each algorithm are shown in Figure 4-4. All methods performed well achieving greater than 50% PD at relatively low FARs. Note the Probabilistic RSF approach was run using the uniform sampling technique for KL estimation and using the analytical integral, assuming Gaussian. The analytical approach performed best, although it assumed Gaussian, whereas, the uniform sampling method used a Gaussian mixture with four components. Although the uniform sampling allows for a more versatile modeling scheme, the analytical calculation of the KL divergence seemed more important than versatility for correct context estimation. Due to the high dimensionality and sparsity of the data, the KL estimate using uniform sampling suffers.

ROC curves with error bars are shown in Figure 4-5. In Figure 4-5, each mine encounter is treated as a binomial distribution and the error bars illustrate a confidence window of 95%. Note the PDs are normalized to 100% for binomial estimation, and there is good separation of the possibilistic and evidential approaches with 95% confidence indicating a non-random result.

All context-based approaches performed better than the whiten / dewhiten transform save the probabilistic approach using uniform sampling. This result is expected since these approaches are able to identify relevant contexts and use this information to correctly classify samples that have undergone contextual transformations. However, the whiten / dewhiten transform performed relatively well, which indicates that some of the classification issues induced by contextual transformations can be mitigated by means of whitening the data. Figure 4-7 shows a correctly classified POI, where each context-based approach identifies a relevant context, context 3 or 4, and consequently classifies the POI correctly.

Each of the RSF classifiers performed better than the set-based kNN classifier. This is due to their ability to identify relevant contexts in a probabilistic manner rather than a nearest neighbor manner as indicated in Figure 4-10 and Figure 4-9. This is also due to the nearest neighbor approach in the classifier as indicated in Figure 4-6 and Figure 4-9. This is due to the fact that nearest neighbor approaches do not directly incorporate the idea of probability, or weights, and therefore assign confidence based on some fixed number of samples, in this case k=1. We note in previous experiments, k=1 provided the best results for set-based kNN [17].

In Figure 4-8, a POI is incorrectly classified by the possibilistic approach. In this case, the possibilistic approach was the only method to identify context 4 as a relevant context, and consequently misclassified this POI. The situation that occurred in the synthetic data experiment that is shown in Figure 4-3A has occurred, that is, a sample from the test set has come into close

proximity of a germ / grain pair modeling context 4, which has caused the possibilistic approach to choose this context as most likely as opposed to context 3. Although the possibilistic, or optimistic approach, has caused the possibilistic approach to misclassify the POI in Figure 4-8, it also allows for resiliency in the face of outliers. An instance where the possibilistic approach chose a context different from all other approaches is shown in Figure 4-10. This POI was correctly classified by the possibilistic approach and the chance of FA was lessened by all of the probabilistic approaches as they were able to identify two relevant contexts, one which provides correct classification.

The evidential approach performed best, achieving highest PDs at almost all FARs. This result is similar to that found in the synthetic data experiment, save the situation illustrated in Figure 4-3D. The evidential approach provides a good contextual model as the inclusion functional provides an intuitive model for shape characterization.

The probabilistic approach performed well in the synthetic data experiment balancing shape characterization and robustness. However, its results in high dimensional data were inconsistent. Providing enough samples using the uniform sampling method in high dimensions was not practical, and using the analytical integral provided better results. However, the Gaussian assumption limited its shape characterization which influenced its classification results.

Upper and Lower Bounding Experiment

In this experiment we compared the proposed possibilistic context-based classifier to a standard Bayesian classifier, a non-context based classification method. We also compared the results of the possibilistic classifier to results from a context-based oracle classifier that always chooses the correct context. This comparison provides an idea of an upper bound and a lower bound for the proposed method, where the standard classifier is a lower bound since it makes no

use of contextual information and the oracle classifier is the upper bound since it makes the best use of contextual information.

Experimental Design

The experimental setup was similar to that in the hyperspectral experiment. Eight training sets were constructed each representing a set of samples, both target and non-target, observed in some context. Each training set consists of samples collected from an image, where eight contexts are modeled using samples from the eight distinct images. Note that in this training set there are 20 samples from each class in each context. Also in this experiment more spectral bands are used, that is, each image contains 40 spectral bands after trimming and binning, ranging over LWIR wavelengths 7.88um - 9.92um. Again, the possibilistic classifier is equipped with a mixture of Gaussians for sample classification. The oracle uses the same classifiers as the possibilistic approach, however, it always chooses the correct Gaussian mixture.

Classification results of the standard Bayesian classifier are compared to that of the possibilistic RSF classifier. The hypothesis is that both classifiers can account for non-disguising transformations, however, a standard Bayesian classifier cannot account for disguising transformations, whereas the possibilistic classifier can.

The number of mixture components used in the standard Bayesian classifier is varied to illustrate how its ability to classify in the presence of non-disguising transformations relates to the number of mixture components. The hypothesis is that as the number of components increases, the results should improve since it will be better equipped to handle non-disguising transformations. However, regardless of the number of mixture components, the standard Bayesian classifier cannot handle disguising transformations and its results should not best those of the possibilistic classifier, assuming context estimation is performed correctly.

The possibilistic classifier was equipped with two mixture components per class per context for a total of 56 components since for each test set there were seven contexts available each with four classes each containing two mixture components. We compared the results to those of a standard Bayesian classifier with three, seven and 14 mixture components per class.

For comparison to the upper bound, the testing procedure will remain the same, except the classifier trained on the test image will be available to the classifiers during testing; therefore, cross validation is no longer being performed. The results of the possibilistic classifier will be compared to the oracle classifier. The oracle classifier is equipped with similar Gaussian mixtures as the possibilistic RSF classifier; however, it always uses the Gaussian mixture that was trained on the test image. The results of this classifier can be seen as an upper bound of the classification results within this framework. Therefore, it provides a means to assess the ability of the context estimation methodology used in the RSF classifier, namely the optimistic germ and grain model.

Results

The use of possibilistic context estimation within the RSF significantly improved classification results. Probability of detection is improved at all FARs and is improved as much as 10 percentage points. False alarm rates are decreased at all PDs and are reduced as much as 50% at PDs of $4x10^{-3}$ FAs/m^2 through $8x10^{-3}$ FAs/m^2.

Classification results of the standard Bayesian classifier became better as the number of mixture components increased. The increase of mixture components equipped the standard classifier with the ability to account for non-disguising transformations. When the number of components was less than the number of contexts, the standard classifier performed poorly. This is expected as it could not account for all of the non-disguising transformations. However, its performance improved as the number of mixture components became greater than or equal to the

number of contexts. The results also indicate that the RSF classifier was able to account for disguising transformations with an improvement in classification when compared to the standard classifier with the same number of overall mixture components.

The RSF Bayesian classifier performed similarly to the oracle RSF Bayesian classifier indicating that using the random set framework is an excellent method for context estimation. In fact, the RSF Bayesian classifier using the germ and grain model weighted the context which was chosen by the oracle as the most likely context 66% of the time, and furthermore, weighted that context as one of the two most likely contexts 86% of the time. However, we note there is room for improvement which can be noticed at low FARs.

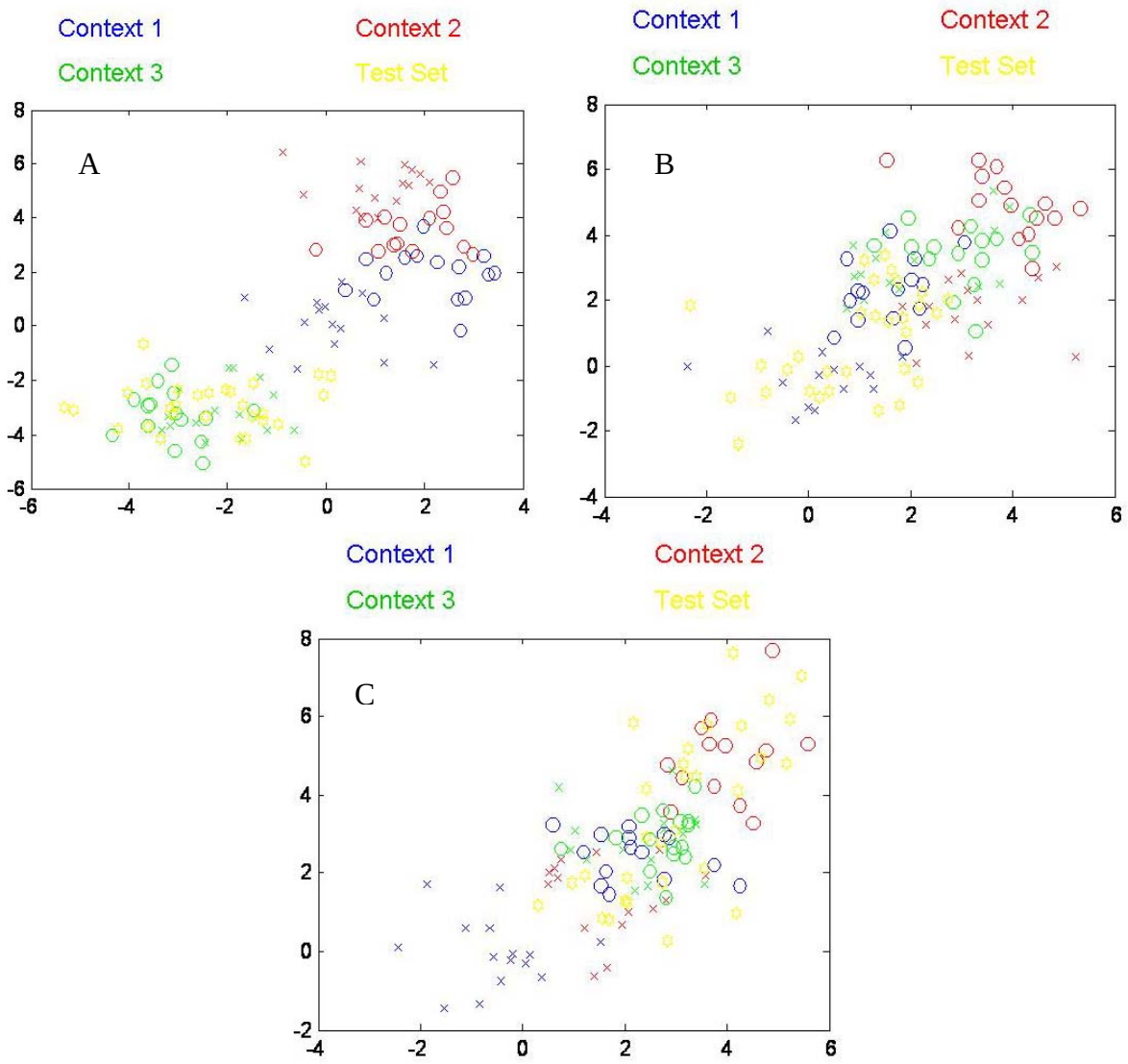


Figure 4-1. Illustration of data sets one, two, and three. A) Samples from a distinct context are shown in distinct colors. Distinct class is shown using a distinct symbol. This is the easiest data set since each context is fairly separable. B) In data set 2, context 3 is overlapped highly by both context 1 and context 2. C) In data set 3, context 1 is completely overlapped by context 2 and context 3.

Table 4-1. Average inference error for each dataset using 15 test and 15 train samples.

KL Estimation	Data Set 1	Data Set 2	Data Set 3
Riemann Test	.0094	.0390	.0522
Riemann Test and Train	.0114	.0638	.0642
Naïve Test	.0750	.0870	.0722
Naïve Test and Train	.0128	.0639	.0683
Uniform MCMC	.0094	.0562	.0581

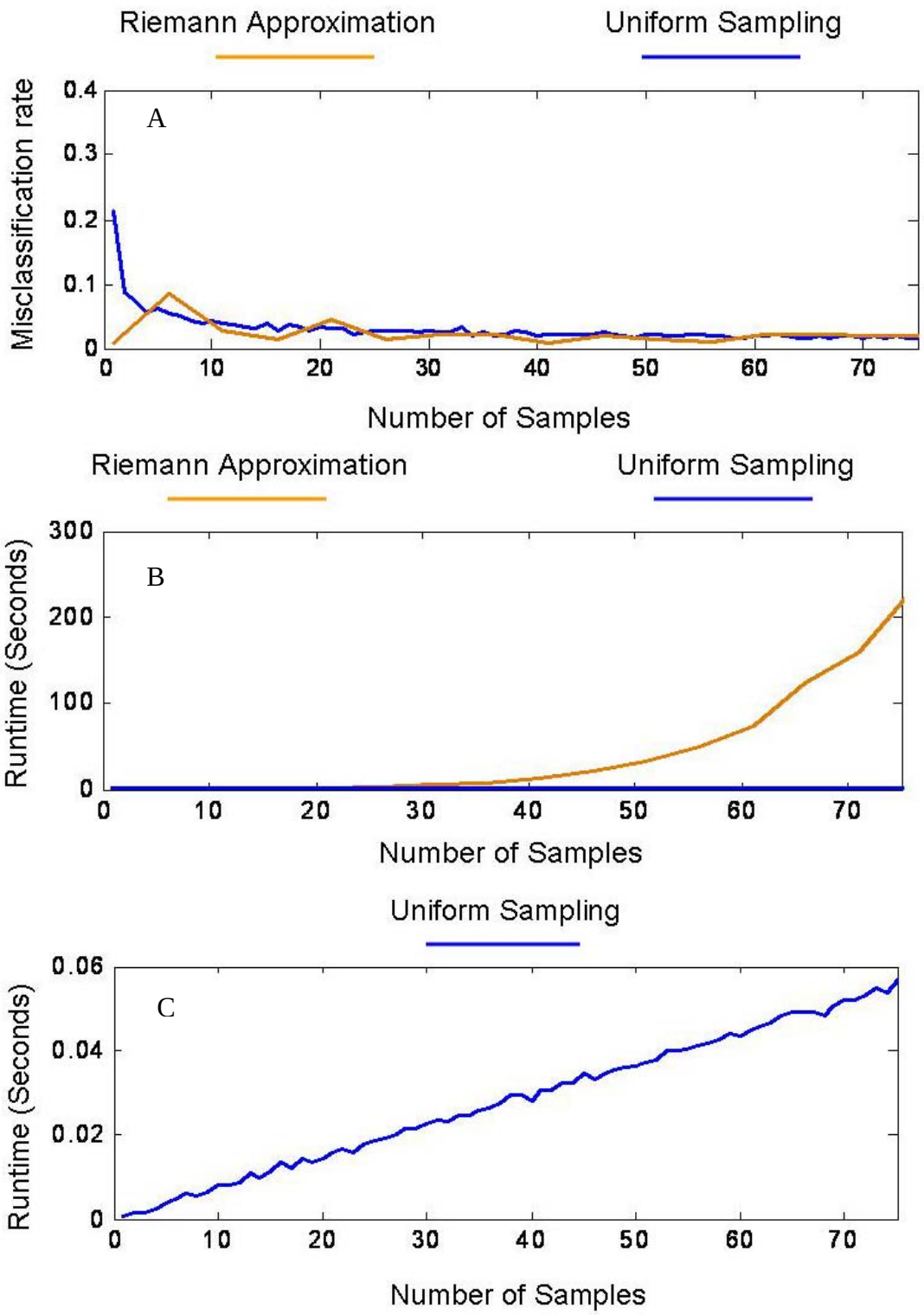


Figure 4-2. Error analysis of the Riemann and uniform approximation methods with respect to time and number of observation samples. A) Plot of context misclassification rate versus the number of samples in the observed set. B) Plot of runtime versus the number of samples in the observations set. C) Close of the plot of runtime versus number of observation samples for the uniform approximation method.

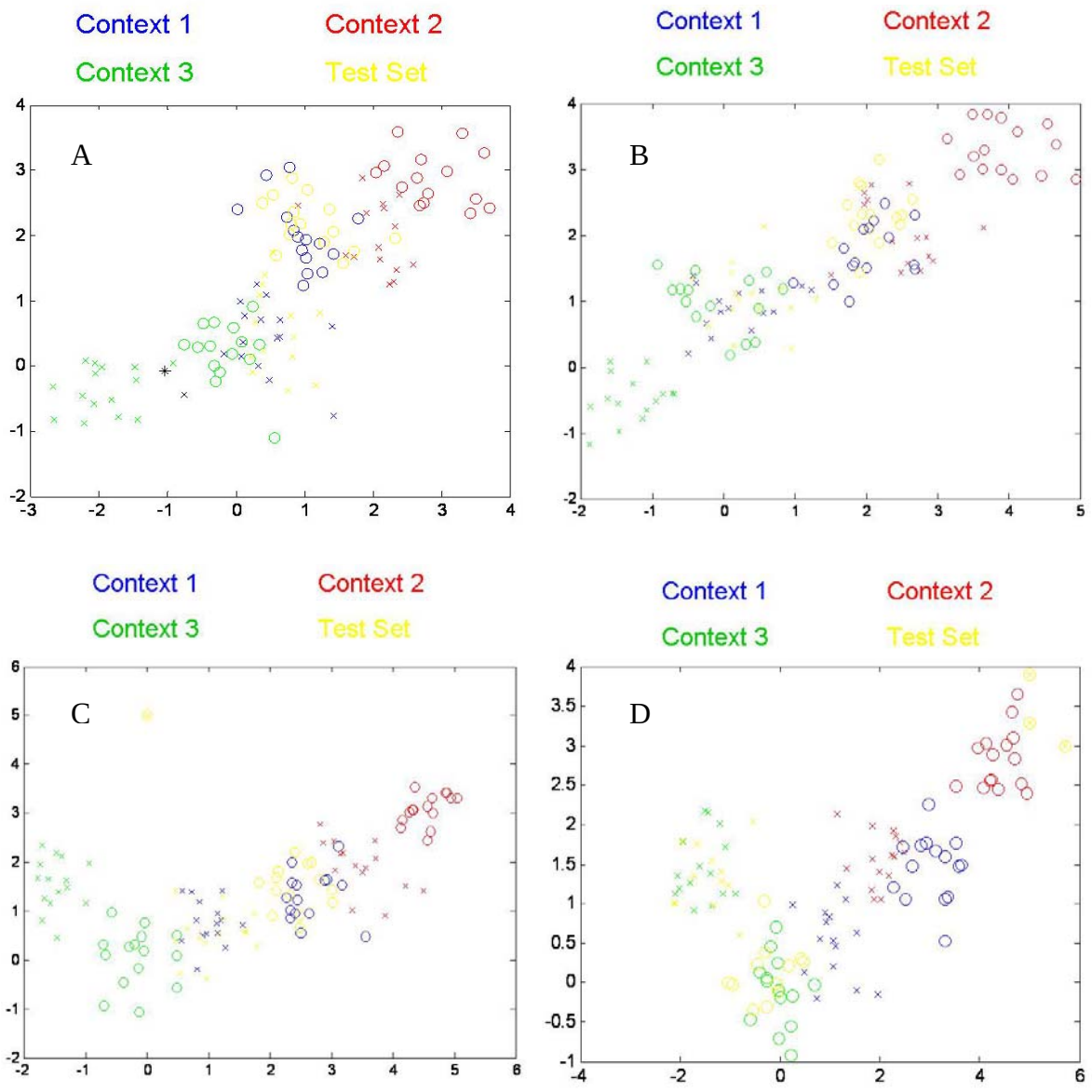


Figure 4-3. Trials using data sets 1, 2, 3 and 4 in the synthetic data experiment. A) Illustration of a trial on data set 1 from the synthetic data experiment where the possibilistic model fails to correctly identify context. Here the germ from context 3 is indicated with a black '*'. Note there is a sample from the test set indicated by a black 'x', which lies very near to the grain. This increases the probability of context 3. B) Trial example of data set 2 in the synthetic data experiment. Samples from either class are oriented relatively the same in each of the 3 contexts. C) Trial example of data set 3 in the synthetic data experiment. Each test set in each of the 50 trials has two outlying samples at [0, 5]. D) Trial example of data set 4 in the synthetic data experiment. Each test set has 6 outliers located near [5, 3.5].

Table 4-2. Average classification error of the listed context-based classifiers on four data sets used in the Synthetic Data Experiments.

Context Classifiers	Data Set 1	Data Set 2	Data Set 3	Data Set 4
Evidential Model	.0413	.0273	.2073	.2500
Probabilistic Model	.0427	.0280	.0667	.2562
Possibilistic Model	.0647	.0480	.0693	.0542
Set-Based kNN	.0560	.0373	.2647	.2520
Whiten/De-Whiten	.0993	.0220	.1033	.0792

Table 4-3. How classification varies with respect to the number of germ and grain pairs for data set 3 (with no outlying samples) in the Synthetic Data Experiment.

Context Classifiers	1 Pair/Context	2 Pair/Context	3 Pair/Context	4 Pair/Context
Evidential Model	.0447	.0487	.0367	.0373
Probabilistic Model	.0453	.0473	.0400	.0336
Possibilistic Model	.0553	.0460	.0453	.0460

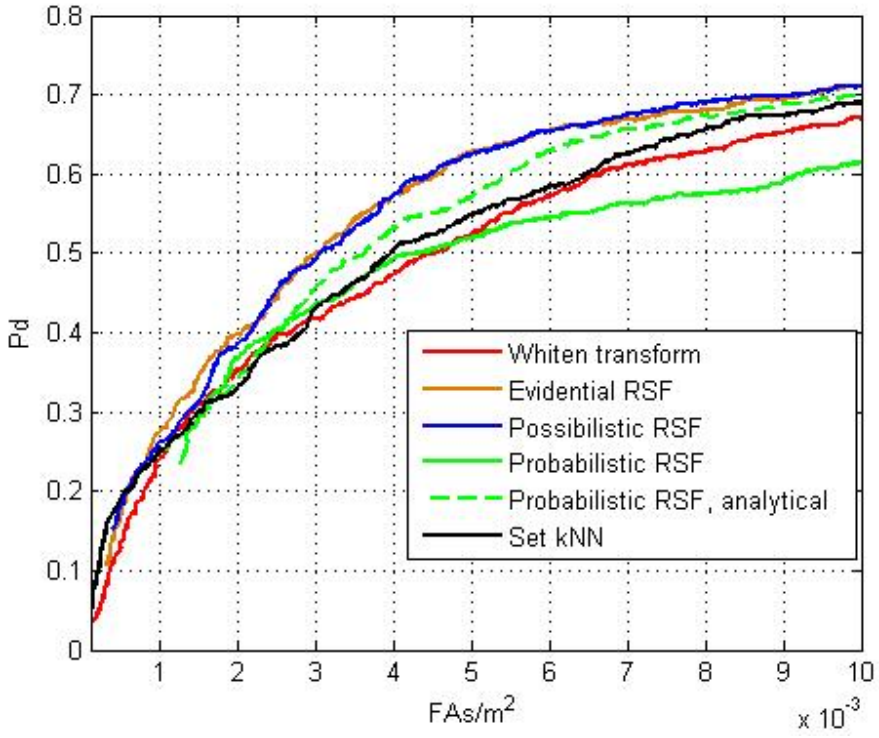


Figure 4-4. ROC curve for The Hyperspectral Data Experiment. Note the dashed plot is the results from the probabilistic context-based classifier using the analytical solution for KL estimation as discussed in Equation 3-40.

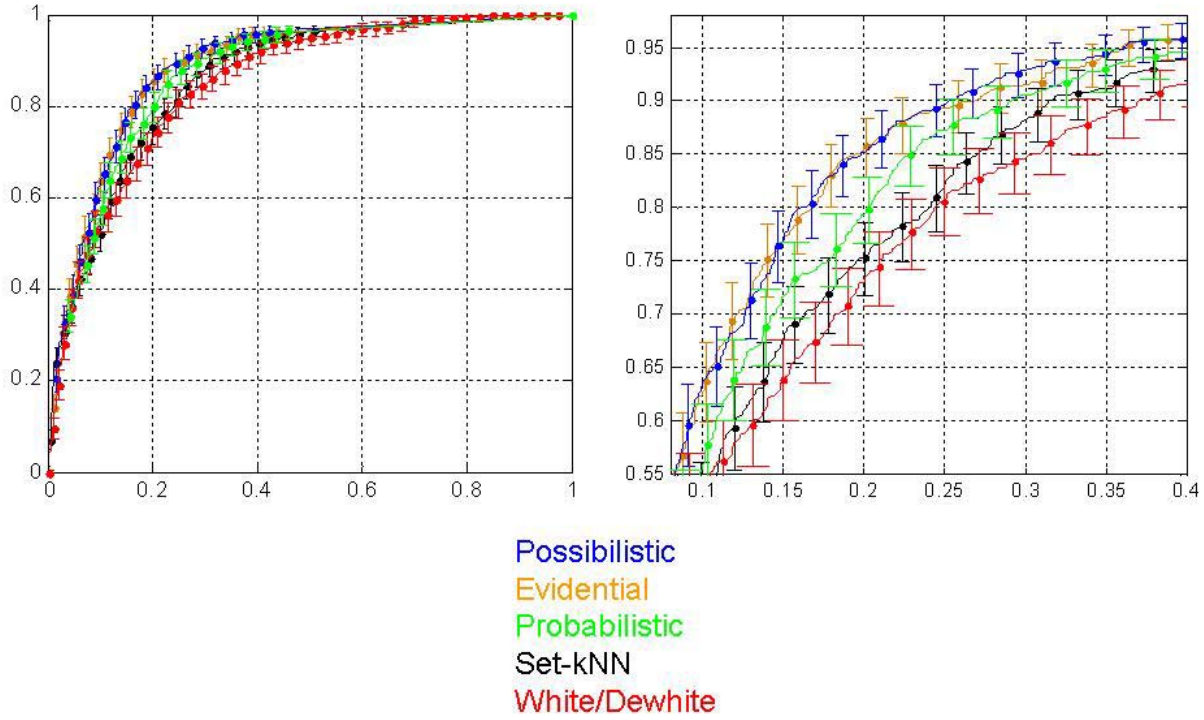


Figure 4-5. Hyperspectral Experiment ROC curve of PD versus PFA for the possibilistic, evidential probabilistic, set-based kNN, and whiten / dewhiten approaches. Error bars show the 95% confidence range assuming each encounter is a binomial experiment. For this reason, PDs are normalized to include only targets that were observed by the algorithms under test, and do not include targets missed by the prescreener.

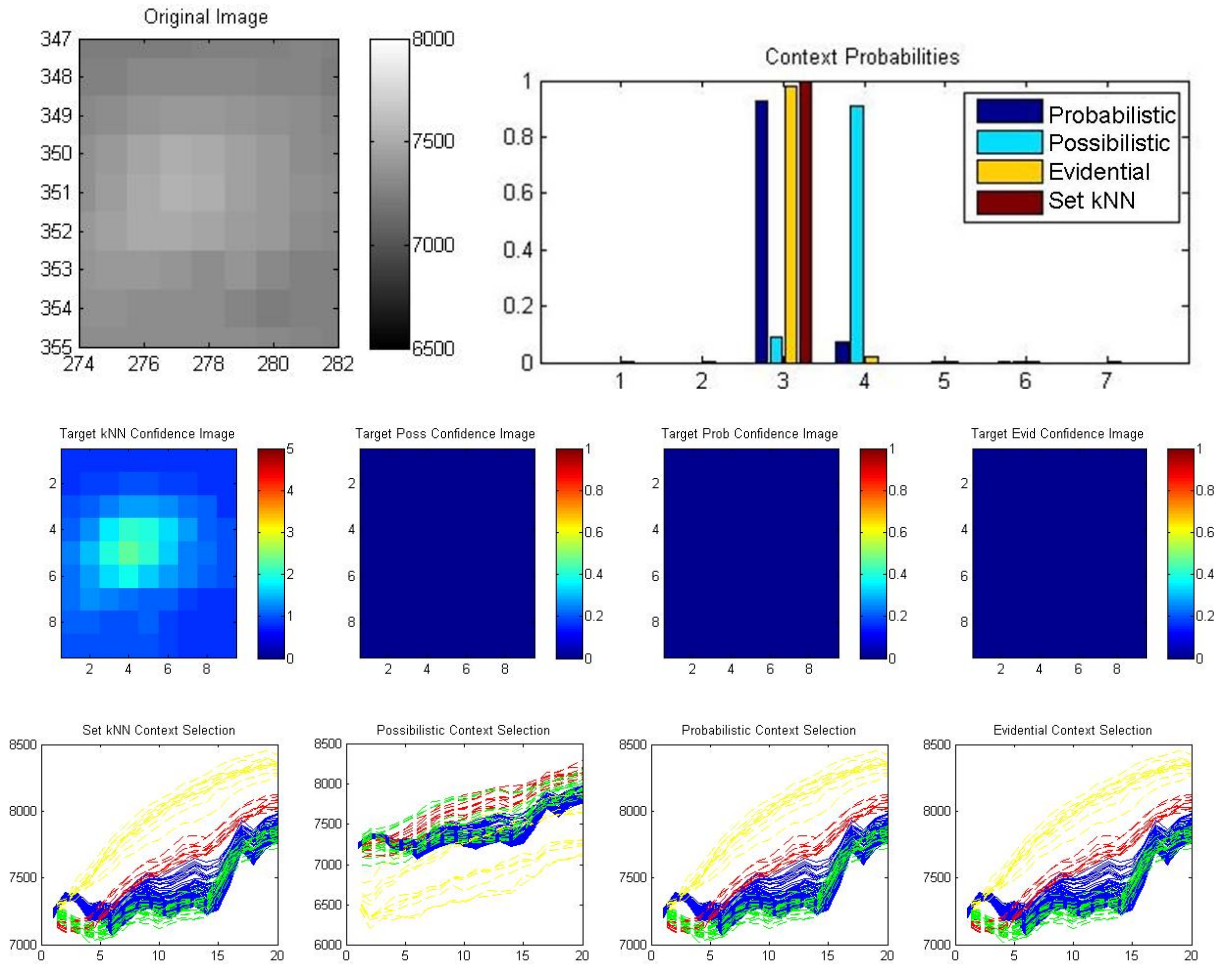


Figure 4-6. Example of a false alarm POI from The Hyperspectral Data Experiment. A snippet of the original AHI image at wavelength 8.9um is shown in the upper left where the prescreener alarmed. The second row are the confidence images of set-base kNN, possibilistic, probabilistic, and evidential approaches, from left to right. Their contextual estimates of the potential seven contexts are shown in the bar chart in the upper right. Note there are seven potential contexts and not eight since are performing crossvalidation at the image level. Under the confidence images, in the bottom row, are the spectral plots of the test population, shown in blue. Also shown in these plots are the spectra used to create the contextual models of the context that the corresponding approach selected as most probable. These training spectral are color coded by class, where red, green, and yellow correspond to target, soil types, and vegetation, respectively.

Note in this example set-based kNN submits a marginal confidence, due to its use of a nearest neighbor based classifier and choice of context 3. the probabilistic and evidential approaches select context 3 as well; but their classifier makes use of covariance which allows for correct classification. The possibilistic approach chose a context which correctly identifies the spectra as soil.

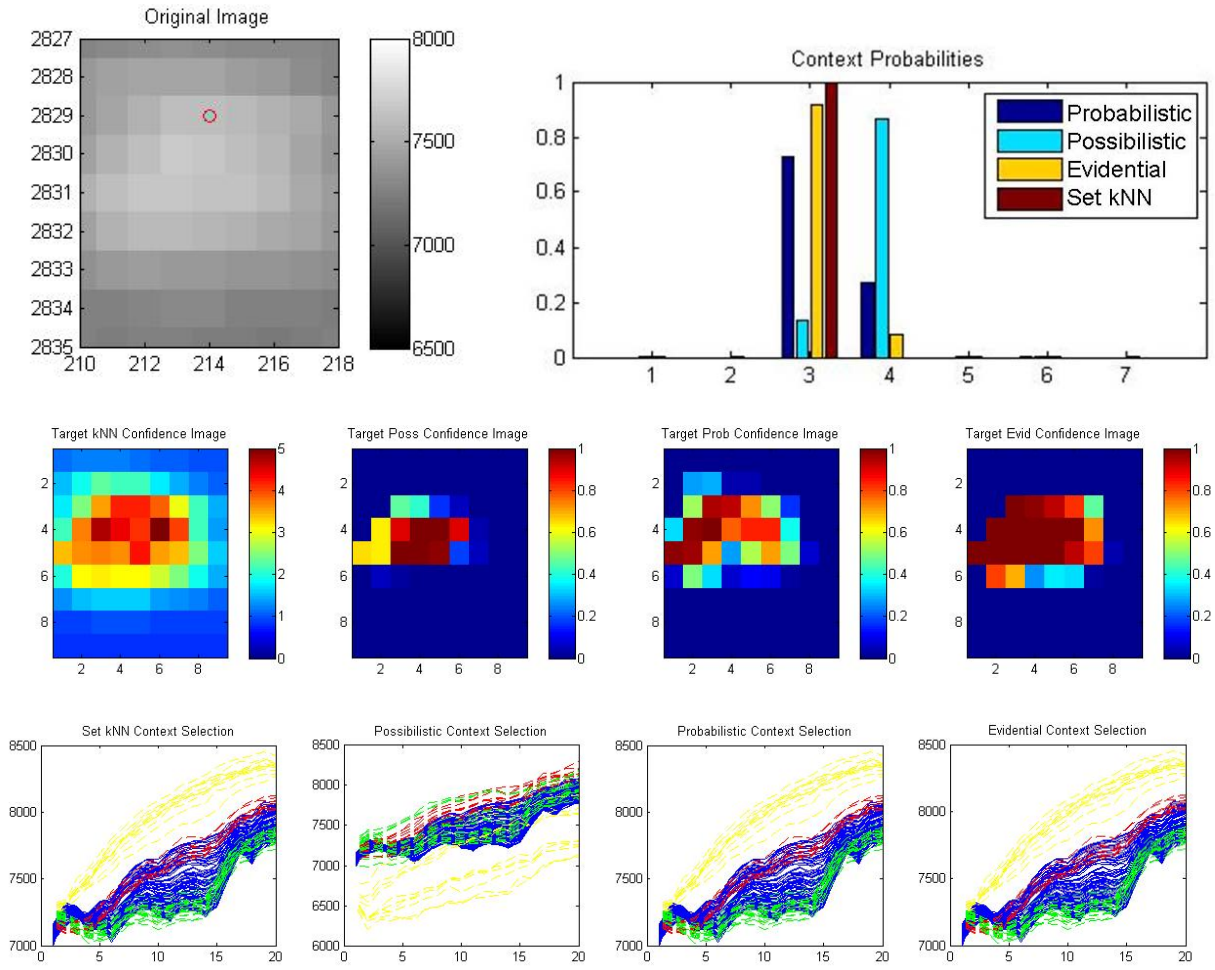


Figure 4-7. Example of a target alarm POI, from The Hyperspectral Data Experiment. A snippet of the original AHI image at wavelength 8.9um is shown in the upper left where the prescreener alarmed. The red circle indicates that this is a target. The second row are the confidence images of set-base kNN, possibilistic, probabilistic, and evidential approaches, from left to right. Their contextual estimates of the potential seven contexts are shown in the bar chart in the upper right. Under the confidence images, in the bottom row, are the spectral plots of the test population, shown in blue. Also shown in these plots are the spectra used to create the contextual models of the context that the corresponding approach selected as most probable. These training spectral are color coded by class, where red, green, and yellow correspond to target, soil types, and vegetation, respectively.

Note in this example each algorithm correctly identifies this POI as a target. Note the possibilistic approach has selected context 4 as most probable, whereas, the other three methods selected context 3. In this instance, the choice between context 3 and 4 does not change the classification results since the test spectra are similar to the target prototypes in both contexts.

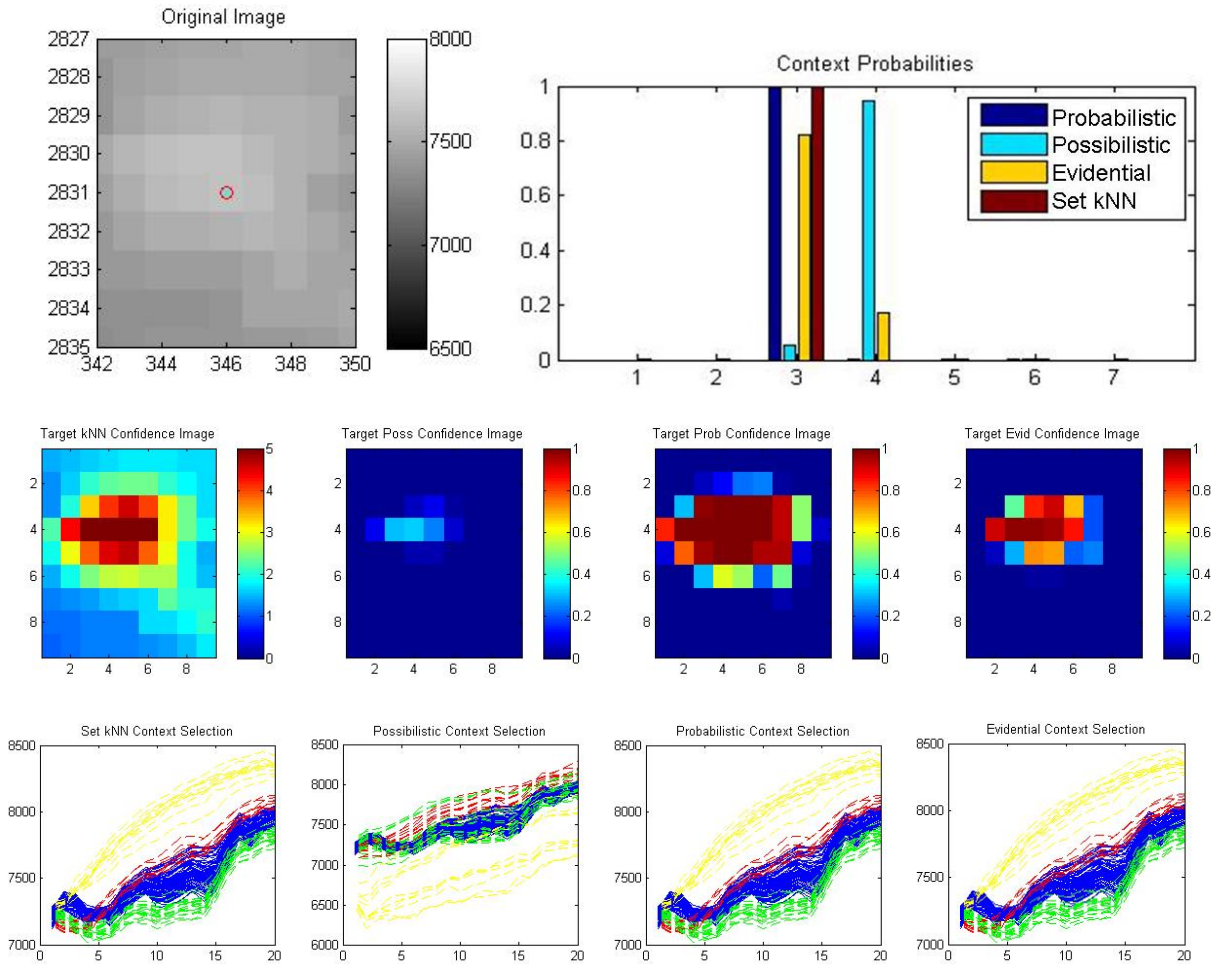


Figure 4-8. Example of a target alarm POI from The Hyperspectral Data Experiment. A snippet of the original AHI image at wavelength 8.9um is shown in the upper left where the prescreener alarmed. The red circle indicates that this is a target. The second row are the confidence images of set-base kNN, possibilistic, probabilistic, and evidential approaches, from left to right. Their contextual estimates of the potential seven contexts are shown in the bar chart in the upper right. Note there are seven potential contexts and not eight since are performing crossvalidation at the image level. Under the confidence images, in the bottom row, are the spectral plots of the test population, shown in blue. Also shown in these plots are the spectra used to create the contextual models of the context that the corresponding approach selected as most probable. These training spectral are color coded by class, where red, green, and yellow correspond to target, soil types, and vegetation, respectively.

Note in this example the possibilistic approach selects context 4, which results in incorrect classification. Also note that the evidential approach partially weights context 4, thus its confidence is not as high as set-based kNN and the probabilistic approach.

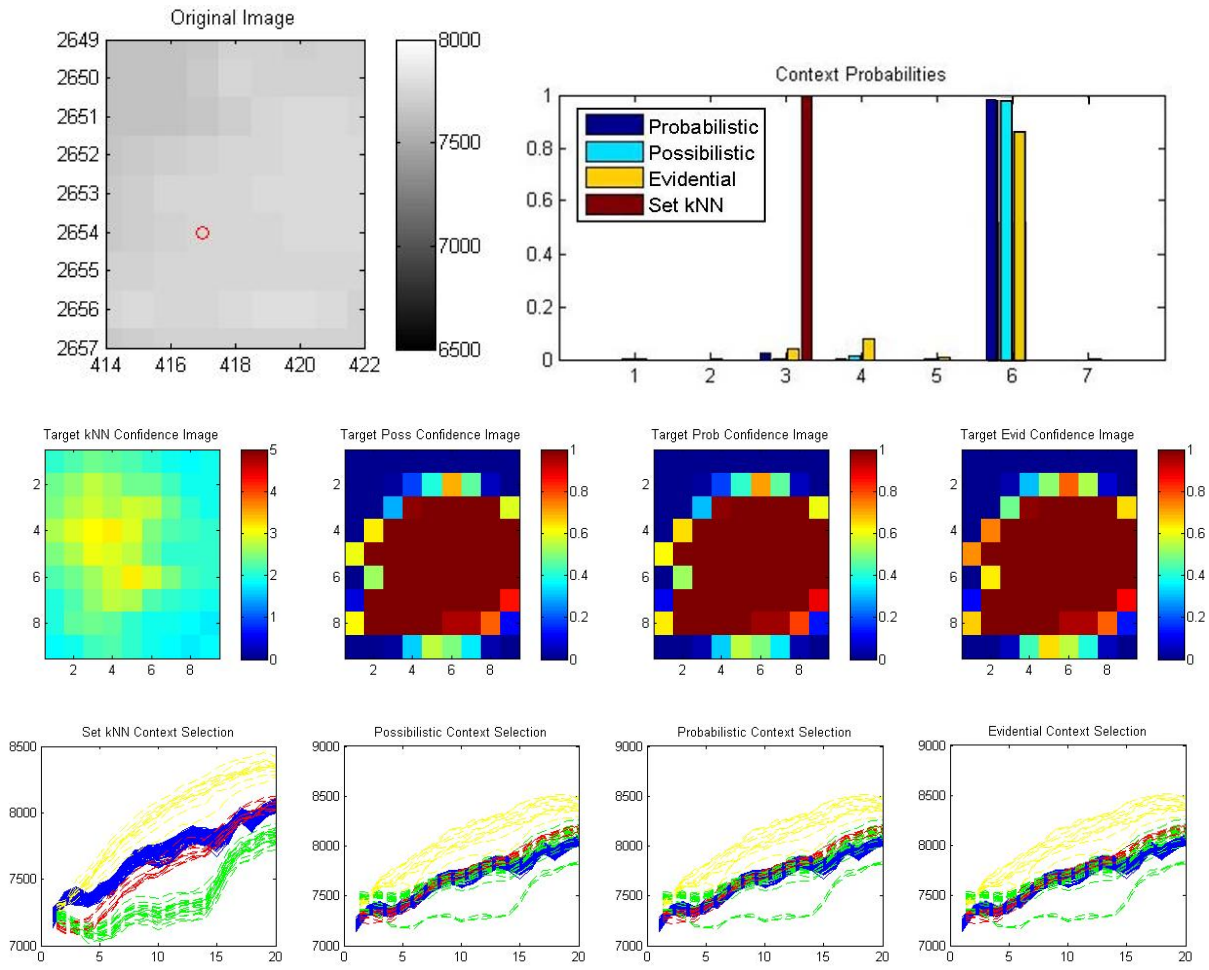


Figure 4-9. Example of a false alarm POI from The Hyperspectral Data Experiment. A snippet of the original AHI image at wavelength 8.9um is shown in the upper left where the prescreener alarmed. The red circle indicates that this is a target. The second row are the confidence images of set-base kNN, possibilistic, probabilistic, and evidential approaches, from left to right. Their contextual estimates of the potential seven contexts are shown in the bar chart in the upper right. Under the confidence images, in the bottom row, are the spectral plots of the test population, shown in blue. Also shown in these plots are the spectra used to create the contextual models of the context that the corresponding approach selected as most probable. These training spectral are color coded by class, where red, green, and yellow correspond to target, soil types, and vegetation, respectively.

Note in this example set-based kNN submits a marginal confidence rather than a high confidence due to its selection of context 3. Note the population spectra for set-based kNN selection fall in between prototypes for class target and vegetation, providing for a marginal confidence. The other 3 classifiers selected context 6 which provides for correct classification. Note samples from the target class in context 6 are extremely similar to the test samples, indicating a correct selection.

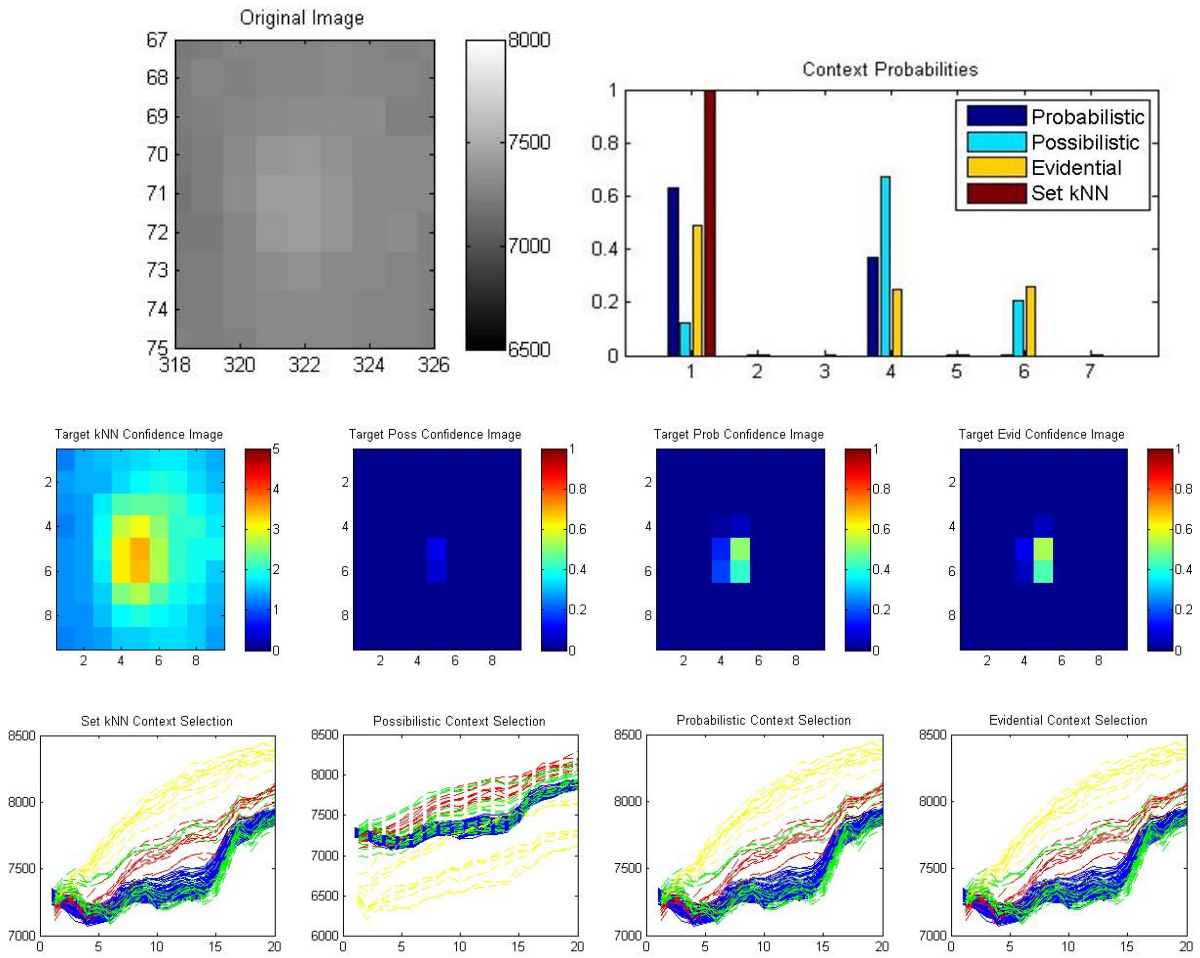


Figure 4-10. Example of a false alarm POI from The Hyperspectral Data Experiment. A snippet of the original AHI image at wavelength 8.9um is shown in the upper left where the prescreener alarmed. The second row are the confidence images of set-base kNN, possibilistic, probabilistic, and evidential approaches, from left to right. Their contextual estimates of the potential seven contexts are shown in the bar chart in the upper right. Under the confidence images, in the bottom row, are the spectral plots of the test population, shown in blue. Also shown in these plots are the spectra used to create the contextual models of the context that the corresponding approach selected as most probable. These training spectral are color coded by class, where red, green, and yellow correspond to target, soil types, and vegetation, respectively.

Note in this example set-based kNN submits high confidence due to its selection of context 1. Note the probabilistic and evidential approaches submit marginal confidences as they selected context 1. But their confidence is only marginal since they only partially selected context 1. Note the possibilistic approach selected context 4, and was able to correctly classify this POI as a false alarm.

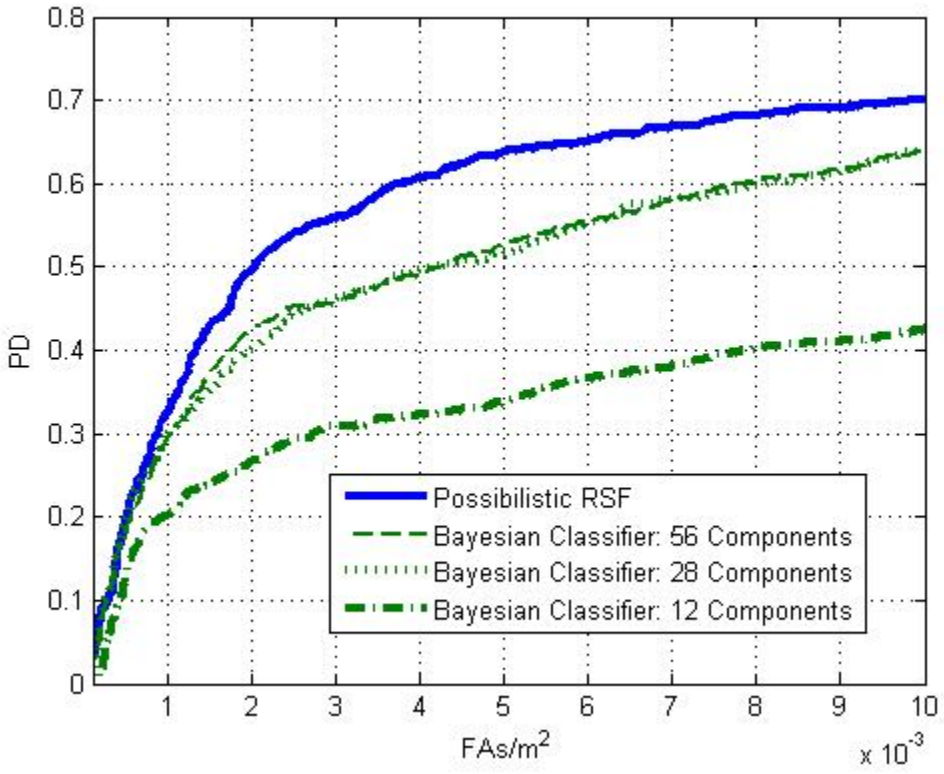


Figure 4-11. Detection results for the possibilistic RSF classifier and results for standard Gaussian mixture classifiers equipped with variable numbers of mixture components.

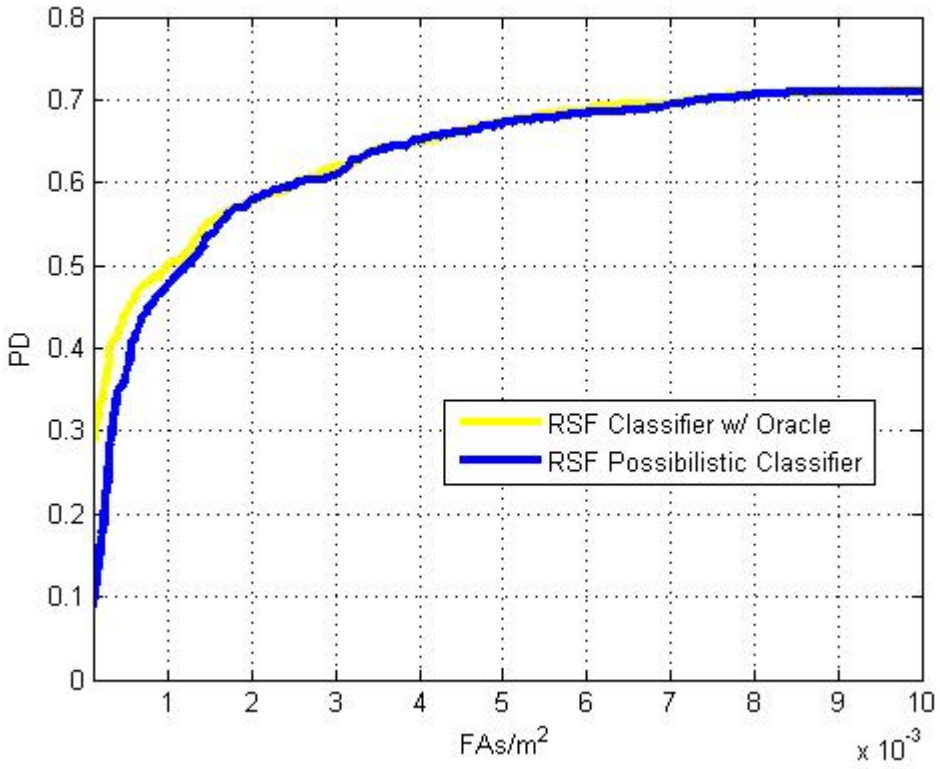


Figure 4-12. Non-crossvalidation detection results for the possibilistic RSF classifier and the oracle classifier.

CHAPTER 5
CONCLUSIONS

We developed a generalized framework for context-based classification using the theory of random sets. The resulting context-based classifier estimates the posterior of a sample, using the sample and a set—its population. Contextual transformations are identified by population analysis, and the resulting contextual estimate provides an appropriate weight of relevance to context specific classifiers. The random set framework provides the tools necessary to perform classification in the presence of contextual factors. Furthermore, it has the ability to contend with disguising transformations, which is not the case for standard classification procedures. Experimental results have shown the random set models' abilities to correctly identify context in various situations, and have shown applicability to real-world problems, improving classification results over state-of-the-art classifiers: set-based kNN and the whiten / dewhiten transform.

In the synthetic experiments, pros and cons of each approach where highlighted. The possibilistic approach was shown to be a robust classifier, resilient to outliers, but at the cost of optimism. The evidential approach has the ability to characterize shape, but at the cost of robustness. The probabilistic approach balanced these two pros and cons, allowing for some characterization of shape and some resilience. Each of these RSF classifiers was superior to set-based kNN, which is not resilient to outliers, but provides an intuitive, nearest neighbor, set comparison procedure. The whiten / dewhiten transform assumes a consistent orientation of target subspaces with respect to background subspaces, and given this assumption, provides a whitening solution. This approach can be considered a context-based method, but makes strict assumptions which the other methods do not. Therefore, the whiten / dewhiten transform performed well when said assumptions are true and performs poorly when they are not.

Each of the methods under test was able to minimally detect a minefield using a extensive hyperspectral dataset. The evidential and possibilistic methods performed best due to their resilience and shape characterization capabilities, reducing FARs by up to 25% over set-based kNN. The probabilistic model suffers partially due to its attempt to construct a representative measure given a low number samples and high dimensionality. Set-based kNN was bested by the RSF classifiers due to its lack of ability to assign gray-level weights of contextual relevance. The whiten / dewhiten transform performs worst indicating that, although some of the contextual transformations can be mitigated through the use of whitening, all of them could not.

In the final experiment, the possibilistic approach performs similarly to its upper bound, and outperformed a similar classifier that made no use of contextual information. This indicates that the possibilistic approach makes good use of contextual information which translates to improved classification results.

Each algorithm has different computational complexity. Although set-based kNN does not require training, the set-based comparison provides for a testing computation time bounded by $O(pdTN^2)$, where N is the bounding number of samples in a training or test set, d is the dimensionality of the samples, p is the number of testing populations, and T is the number of training sets. Note for each population p we must calculate the pairwise distances between the test set and all T training sets. Whereas, the RSF classifiers require a training period, but testing computation time is bounded by $O(pcdN + md^3)$, where c is determined by the fixed number of constructs, such as a germ and grain pair or a likelihood function, used to model C contexts and m is the number of constructs needed to model M classifiers. For each population, we must compare each sample to each contextual construct. Note for each Bayesian classifier we must invert a covariance matrix that is dxd; however, the use of a Gaussian classifier is not necessary

within the RSF framework. The whiten / dewhiten transform has a training period, and requires extensive testing computation time bounded by $O(pmd^3 + Nm)$. Note, for each population we must calculate and invert a covariance matrix.

Future work will include the extensive experimentation of the methods developed for the optimization and experimentation methods used by the RSF classifiers. An example of research in optimization strategies would include the investigation of the use of EM for unsupervised learning of contexts within the hyperspectral data set. This could provide for interesting findings of sub-contexts or subpopulations, within each image. Examples of future research in experimental methods would be performing experiments where the size of the populations varied. Larger populations may provide for a better estimate of context.

Extended research may include the development of a non-additive random measure. This development may provide the capability to characterize complex relationships between sets of samples, similar to a belief function. We also note that during the development of the representative function, it was determined that the point-wise average of the observed measures minimized the KL between the representative function and the observed measures, this may provide for an interesting development of posterior estimation, and relation to variational methods.

Future work should include the application of the RSF classifiers to *unexploded ordnance* (UXO) datasets. These data sets are subject to problems similar to those faced in remote sensing data, including contextual factors. The use of contextual estimation should improve classification, or target identification, for these applications as well.

LIST OF REFERENCES

[1] I. Molchanov, *Probability and Its Applications: Theory of Random Sets*. London: Springer-Verlag, 2005.

[2] J. Goutsias, R. Mahler, and H Nguyen, *Random Sets: Theory and Applications.* New York: Springer-Verlag, 1997.

[3] T. Norberg, "Convergence and Existence of Random Set Distributions" *Annals of Probability,* Vol. 12, No. 3, pp. 726-732, 1983.

[4] D. Stoyan, W. Kendall, and J. Mecke, *Stochastic Geometry and Its Applications: Second Edition.* West Sussex: John Wiley & Sons, 1995.

[5] D. Stoyan, "Random Sets: Models and Statistics", *International Statistical Review,* Vol. 66, No. 1, pp.1-27, 1998.

[6] N. Cressie and G. Laslett, "Random Set Theory and Problems of Modeling", *SIAM Review*, Vol. 29, No. 4, pp. 557-574, December 1987.

[7] D. Hug, G. Last and W. Weil, "A Survey on Contact Distributions", *Morphology of Condensed Matter: Physics and Geometry of Spatially Complex Systems*, Springer, 2002.

[8] G. Shafer, *A Mathematical Theory of Evidence*. Princeton: Princeton University Press, 1976.

[9] M. Capinski and E. Kopp, *Measure, Integral and Probability*, New York: Springer, 1999.

[10] J. Munkres, *Topology: Second Edition.* Upper Saddle River: Prentice Hall, 2000.

[11] V. Vapnik, *Statistical Learning Theory*, New York: John Wiley & Sons, 1998.

[12] E. Dougherty and M. Brun, "A Probabilistic Theory of Clustering", *Pattern Recognition*, Vol.37, pp. 917-925, 2004.

[13] E. Dougherty, J Barrera, M. Brun, S. Kim, R. Cesar, Y. Chen, M. Bittner, and J. Trent, "Inference from Clustering with Application to Gene-Expression Microarrays", *Journal of Computational Biology*, Vol. 9, No. 1, pp. 105-126, 2002.

[14] M. Brun, C. Sima, J. Hua, J. Lowey, B. Carroll, E. Suh, and E. Dougherty, "Model-Based Evaluation of Clustering Validation Measures", *Pattern Recognition*, Vol. 40, pp. 807-824, 2007.

[15] J. Bolton and P. Gader, "Application of Set-Based Clustering to Landmine Detection with Hyperspectral Imagery", *IEEE Proceedings Geoscience and Remote Sensing*, Barcelona, July 2007.

[16] J. Bolton and P. Gader, "Random Set Model for Context-Based Classification", *IEEE World Congress on Computational Intelligence FUZZ*, Hong Kong, June 2008 (Accepted).

[17] J. Bolton and P. Gader, "Application of Context-Based Classifier to Hyperspectral Imagery for Mine Detection", *SPIE Defense and Security*, Orlando, March 2008 (Accepted).

[18] Tsymbal A, "The problem of concept drift: definitions and related work", *Technical Report TCD-CS-2004-15*, Department of Computer Science, Trinity College Dublin, Ireland, 2004.

[19] M. Bentley, "Environment and Context", *The American Journal of Psychology*, Vol. 39, No. ¼. pp. 52-61, December 1927.

[20] R. Rescoria, "Probability of Shock in the Presence and Absence of CS in Fear Conditioning", *Journal of Comparative and Physiological Psychology*, 66, pp.1-5, 1968.

[21] A. Tsymbal, M. Pechenizkiy, P. Cunningham, and S. Puuronen, "Handling Local Concept Drift with Dynamic Integration of Classifiers: Domain of Antibiotic Resistance in Nosocomial Infections", *Proceedings of IEEE Symposium Computer-based Medical Systems*, 2006.

[22] M. Salganicoff, "Density Adaptive Learning and Forgetting", *Technical Report No. IRCS-93-50*, University of Pennsylvania Institute for Research in Cognitive Science, 1993.

[23] D. Widyantoro and J. Yen, "Relevant Data Expansion for Learning Concept Drift from Sparsely Labeled Data", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17, No. 3, pp. 401-412, March 2005.

[24] J. Schlimmer and R. Granger Jr., "Incremental Learning from Noisy Data", *Machine Learning*, Vol. 1, pp. 317-354, 1986.

[25] G. Widmer, "Learning in the Presence of Concept Drift and Hidden Contexts", *Machine Learning*. Vol. 23, pp. 69-101, 1996.

[26] M. Maloof and R. Michalski, "Learning Evolving Concepts Using Partial-Memory Approach", *Working Note AAAI Fall Symposium on Active Learning*, Boston, pp. 70-73, November 1995.

[27] M. Maloof and R. Michalski, "Selecting Examples for Partial Memory Learning", *Machine Learning*, Vol. 41, pp. 27-52, 2000.

[28] R. Klinkenberg, "Using Labeled and Unlabeled Data to Learn Drifting Concepts", *Workshop Notes on Learning from Temporal and Spatial Data*, Menlo Park, pp. 16-24, 2001.

[29] R. Klinkenberg and T. Joachims, "Detecting Concept Drift with Support Vector Machines", *Proceedings of the 17th Intl. Conf. on Machine Learning*, PP. 487-494, July 2000.

[30] R. Schapire, "The Strength of Weak Learnability", *Machine Learning*, Vol. 5, pp. 197-227, 1990.

[31] R. Schapire, Y. Feund, P. Bartlett, and W. Lee, "Boosting the Margin: A new Explanation for the Effectiveness of Voting Methods", *Annals of Statistics*, Vol. 26, No. 5, pp. 1651-1686.

[32] R. Schapire, "The Boosting Approach to Machine Learning: An Overview, "*MSRI Workshop on Nonlinear Estimation and Classification*", 2002.

[33] Freund, Y. and R. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting*", Proceedings of Computational Learning Theory: Second European Conference*, Barcelona, 2005.

[34] E. Bauer and R. Kohavi, "An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants", *Machine Learning*, Vol. 36, pp. 105-139, 1999.

[35] E. Dura, Y. Zhang, X. Liao, G. Dobeck, and L. Carin, "Active Learning for Detection of Mine-Like Objects in Side-Scan Sonar Imagery", *IEEE Journal of Oceanic Engineering*, Vol. 30, No. 2, April 2005.

[36] Q. Liu, X. Liao, and L. Carin, "Detection of Unexploded Ordnance via Efficient Semi-Supervised and Active Learning", *IEEE Transactions Geoscience and Remote Sensing*, (submitted).

[37] Q. Lui, X. Liao, and L. Carin, "Semi-Supervised Multi Task Learning", *Neural Information and Processing Systems (NIPS)*, 2007.

[38] Y. Zhang, X. Liao, and L. Carin, "Detection of Buried Targets Via Active Selection of Labeled Data: Application to Sensing Subsurface UXO", *IEEE Transactions Geoscience and Remote Sensing*, Vol. 42, No. 11, November 2004.

[39] A. Tsymbal and S. Puuronen, "Bagging and Boosting with Dynamic Integration of Classifiers", *Proc. Principles of Data Mining and Knowledge Discovery*, PKDD, 2000.

[40] M. Skurichina and R. Duin, "Bagging, Boosting and the Random Subspace Method for Linear Classifiers", *Pattern Analysis and Applications*, Vol. 5, pp. 121-135, 2002.

[41] N. Rooney, D. Patterson, S. Anand and A. Tysmbal, "Dynamic Integration of Regression Models", *Proceedings 5th Annual Multiple Classifier Systems Workshop*, Cagliari, Italy, 2004.

[42] N. Rooney, D. Patterson, A. Tsymbal and S. Anand, "Random Subspacing for Regression Ensembles", *Proceedings 17th Intl. Florida Artificial Intelligence Research Society*, 2004.

[43] L. Breiman, "Bagging Predictors", *Machine Learning*, Vol. 24, pp. 123-140, 1996.

[44] A. Tsymbal, M. Pechenizkiy, and Padraig Cunningham, "Dynamic Integration with Random Forests", *Machine Learning*: *EMCL*, LNAI 4212, pp. 801-808, 2006.

[45] R. Schapire, "Random Forests", *Machine Learning* Vol. 45, pp. 5-32, 2001.

[46] T. Ho, "The Random Subspace Method for Constructing Decision Forests", *IEEE Transactions PAMI*, Vol. 20, No. 8, pp. 832-844.

[47] T. Ho, "Random Decision Forest", *Proceedings 3rd Intl. Conf. on Document Analysis and Recognition*, pp. 278-282, Montreal, August 1995.

[48] L. Kuncheva, "Switching Between Selection and Fusion in Combining Classifiers", *IEEE Transactions Systems, Man, Cybernetics*, Vol. 32 No. 2, April 2002.

[49] A. Santana, R. Soares, Anne Canuto, and M. Souto, "A Dynamic Classifier Selection Method to Build Ensembles Using Accuracy and Diversity", *IEEE Proc. 9th Annual Symp. On Neural Networks*, Brazil, 2006.

[50] E. Santos, R. Sabourin, and P. Maupin, "Single and Multi-Objective Genetic Algorithms for the Selection of Ensemble of Classifiers", *Proceedings Intl. Joint Conf. on Neural Networks*, Vancouver, July 2006.

[51] F. Destempes, J. Angers and M. Mignotte, "Fusion of Hidden Markov Random Field Models and Its Bayesian Estimation", *IEEE Transactions Image Processing*, Vol. 15, No. 10, October 2006.

[52] H. Frigui, L. Zhang, P. Gader, D. Ho, "Context-Dependent Fusion for Landmine Detection with Ground-Penetrating Radar", *Proceedings of SPIE*, Orlando, 2007.

[53] R. Cossu, S. Chaudhuri and L. Bruzzone, "A Context-Sensitive Bayesian Technique for the Partially Supervised Classification of Multitemporal Images", *IEEE Transactions Geoscience and Remote Sensing*, Vol. 2, No. 3, July 2005.

[54] I. Taha and J. Ghosh, "Symbolic Interpretation of Artificial Neural Networks", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 11, No. 3, May 1999.

[55] Y. Qi and R. Picard, "Context-Sensitive Bayesian Classifiers and Application to Mouse Pressure Pattern Classification" *Proceedings on IEEE Pattern Recognition*, Vol. 3, pp. 448-451, August 2002.

[56] T. Minka, "A Family of Algorithms for Approximate Bayesian Inference", Dissertation Submitted to Department of Electrical Engineering and Computer Science MIT, 2001.

[57] M. Harries and C. Sammut, "Extracting Hidden Context", *Machine Learning*, Vol. 32, pp. 101-126, 1998.

[58] M. Harries and K. Horn, "Learning Stable Concepts in Domains with Hidden Changes in Context", *Proceedings 13th ICML, Workshop on Learning in Context Sensitive Domains*, 1996.

[59] A. Berk, L. Bernstein, and D. Robertson, "MODTRAN: Moderate Resolution Model for LOWTRAN 7. Rep. AFGL-TR-83-0187, 261, [Available from Airforce Geophysical Laboratory, Hanscom Air Force Base, MA 01731], 1983.

[60] A. Berk *et. al.*, "MODTRAN4 Radiative Transfer Modeling for Atmosphere Correction", *Proceedings of SPIE Optical Spectroscopic Techniques and Instrumentation for Atmospheric and Space Research*, Vol. 3756, July 1999.

[61] J. Broadwater and R. Challappa, "Hybrid Detectors for Subpixel Targets", *IEEE Transactions Pattern Analysis and Machine Intelligence*, Vol. 29, No. 11, pp. 1891-1903, November 2007.

[62] G. Healy and D. Slater, "Models and Methods for Automated Material Identification in Hyperspectral Imagery Acquired Under Unknown Illumination and Atmospheric Conditions", IEEE Transactions Geoscience and Remote Sensing, Vol. 37, No. 6, November 1999.

[63] C. Kuan and G. Healey, "Modeling distribution changes for hyperspectral image analysis", *Optical Engineering*, Vol. 46, No. 11, 117201, pp. 1-8, November 2007.

[64] P. Fuehrer, G. Healey, B. Rauch and D. Slater, "Atmospheric Radiance Interpolation for the Modeling of Hyperspectral Data", *Proceedings. of SPIE Algorithms and Technologies ofr Multispectral, Hyperspectral, and Ultraspectral Imagery XIV*, Vol. 6966, No. 69661O-1, pp. 1-12.

[65] R. Mayer, F. Bucholtz and D. Scribner, "Object Detection by Using 'Whitening/Dewhitening' to Transform Target Signatures in Multitemporal Hyperspectral and Multispectral Imagery", *IEEE Transactions. Geoscience and Remote Sensing*, Vol. 41, No. 5, pp. 1136-1142, May 2003.

[66] S. Rajan, J. Ghosh and M. Crawford, "An Active Learning Approach to Hyperspectral Data Classification", *IEEE Transactions Geoscience and Remote Sensing*, Vol. 46, No. 4, April 2008.

[67] S. Rajan, J. Ghosh and M. Crawford, "An Active Learning Approach to Hyperspectral Data Classification"**,** *IEEE Transactions. Geoscience and Remote Sensing*, Vol. 46, No. 4, pp. 1231-1242, April 2008.

[68] R. Xu and D. Wunsch, "Survey of Clustering Algorithms", *IEEE Transactions on Neural Networks*, No. 3, Vol. 16, pp. 645-678, May 2005.

[69] H. Luo, F. Kong, K. Zhang, and L. He, "A Clustering Algorithm Based on Mathematical Morphology", *IEEE Proceedings Intelligent Control and Automation*, Dalian, pp. 6064-6067, June 2006.

[70] Y. Rubner, L. Guibas, and C. Tomasi, "A Metric for Distributions with Applications to Image Databases", *Proceedings International Conference on Computer Vision*, pp. 59-66, Bombay, 1998.

[71] Y. Rubner, L. Guibas, and C. Tomasi, "The Earth Mover's Distance, Multi-Dimensional Scaling, and Color-Based Image Retrieval", *Proceedings ARPA Image Understanding Workshop*, pp. 661-668, May 1997.

[72] H. Houissa, N. Boujemaa and H. Frigui, “Adaptive Visual Regions Category with Sets of Point of Interest”, *Lecture Notes in Computer Science*, Vol. 4261, pp. 485-493, 2006.

[73] S. Theodoridis and K. Koutroumbas, *Pattern Recognition: Second Edition*, San Diego: Elsevier, 2003.

[74] X. Descombes and J. Zerubia, “Marked Point Process in Image Analysis”, *IEEE Signal Processing Magazine*, pp. 77-84, September 2002.

[75] A. Baddeley and J. Moller, “Nearest-Neighbor Markov Point Processes and Random Sets”, *International Statistical Review*, Vol. 57, No. 2, pp. 89-121, August 1989.

[76] J. M. Billiot, J. F. Coeurjolly, and R Drouilet, “Maximum Pseudo-Likelihood Estimator for Nearest-Neighbors Gibbs Point Processes”, *arXiv:math/0601065v1*, January 2006.

[77] J. Gubner and W. B. Chang, “Nonparametric Estimation of Interaction Functions for Two-Type Pairwise Interaction Point Processes”, *Proceedings IEEE Acoustics, Speech, and Signal Processing*, Vol. 6, pp. 3981-3984, May 2001.

[78] P. Fishman and D. Snyder, “The Statistical Analysis of Space-Time Point Processes”, *IEEE Transactions Information Theory*, Vol. 22, No. 3, May 1976.

[79] R. Stoica, X. Descombes, and J. Zerubia, “Road Extraction in Remote Sensed Images Using Stochastic Geometry Framework”, *Proceedings Intl. Workshop Bayesian Inference and Maximum Entropy Methods*, France, 2000.

[80] R. Stoica, X. Descombes and J. Zerubia, “A Gibbs Point Process for Road Extraction from Remotely Sensed Images”, *Intl. Journal of Computer Vision*, Vol. 57, No. 2, pp. 121-136, 2004.

[81] X. Descombes and J. Zerubia, “Marked Point Process in Image Analysis”, *IEEE Magazine Signal Processing*, pp. 77-84, September 2002.

[82] C. Lacoste, X. Descombes and J. Zerubia, “Point Processes for Unsupervised Line Network Extraction in Remote Sensing”, *IEEE Transactions PAMI*, Vol. 27, No. 10, pp. 1568-1579, October 2005.

[83] M. Ortner, X. Descombes and J. Zerbia, “Point Processes of Segments and Rectangles for Building Extraction from Digital Elevation Models”, *Proceedings IEEE ICASSP*, 2006.

[84] L. Linnett, D. Carmichael and S. Clarke, “Texture Classification Using a Spatial Pont Process Model”, *Proceedings IEEE Vis. Image Signal Processing*, Vol. 142, No. 1, February 1995.

[85] D. Savery and G. Cloutier, “Monte Carlo Simulation of Ultrasound Backscattering by Aggregating Red Blood Cells”, *Proceedings IEEE Ultrasonics Symposium*, 2001.

[86] B.-H. Juang, W. Chou, and C.-H. Lee, "Minimum classification error rate methods for speech recognition," *IEEE Transactions Speech Audio Process.*, Vol. 5, pp. 257–266, May 1997.

[87] S. Katagiri, B.-H. Juang, and C.-H. Lee, "Pattern recognition using a family of design algorithms based upon the generalized probabilistic descent method," *Proceedings IEEE*, Vol. 86, pp. 2345–2372, November 1998.

[88] M. G. Rahim, B.-H. Juang, and C.-H. Lee, "Discriminative utterance verification for connected digit recognition," *IEEE Transactions Speech Audio Process.*, Vol. 5, pp. 266–277, May 1997.

[89] A. Ergun, R. Barbieri, U. Eden, M. Wilson and E. Brown, "Construction of Point Processes Adaptive Filter Algorithms for Neural Systems Using Sequential Monte Carlo Methods", *IEEE Transactions Biomedical Engineering*, Vol. 54, Mo. 3, March 2007.

[90] B. Picinbono, "Time Intervals and Counting in Point Processes*"*, *IEEE Transactions Information Theory*, Vol. 50, No. 6, pp. 1336-1340, June 2004.

[91] V. Solo, "High Dimensional Point Process System Identification: PCA and Dynamic Index Models", *Proceedings of 45th IEEE Conf. Decision and Control*, San Diego, 2006.

[92] R. Sunaresan and S. Verdu, "Capacity of Queues via Point Process Channels", *IEEE Transactions Information Theory*, Vol. 52, No. 6, June 2006.

[93] J. Gubner and W. Chang, "Nonparametric Estimation of Interaction Functions for Two-Type pairwise Interaction Point Processes", *Proceedings IEEE ICASSP*, 2001.

[94] P. Diggle, T. Fiksel, P. Grabarnik, Y. Ogata, D. Stoyan and M. Tanemura, "On Parameter Estimation for Pairwise Interaction Point Processes", *International Statistical Review*, Vol. 62, No. 1, pp. 99-117, April 1994.

[95] Y. Ogata and M. Tanemura, "Likelihood Analysis of Spatial Point Processes", *Journal of the Royal Statistical Society, Series B*, Vol. 46, No. 3, pp. 496-518, 1984.

[96] V. Isham, "Multiple Markov Point Processes: Some Approximations", *Proceedings of the Royal Society of London, Series A*, Vol. 391, No.1800, pp. 39-53, January 1984.

[97] C. Bishop, *Pattern Recognition and Machine Learning*: Springer. New York, 2006.

[98] J. Chen, J. Hershey, P. Olsen and E. Yashin, "Accelerated Monte Carlo for Kullback-Leibler Divergence between Gaussian Mixture Models", *Proceedings IEEE Intl. Conference on Acoustics, Speech and Signal Processing*, pp. 4553 – 4556, 2008.

[99] X. Yu, I. S. Reed, and A. D. Stocker. "Comparative Performance Analysis of Adaptive Multispectral Detectors," *IEEE Transactions Signal Processing*, Vol. 41, No. 8, August 1993, pp. 2639-2656.

[100] W. Radzelovage and G. Maksymonko, “Lessons Learned from a Multi-Mode Infrared Airborne Minefield Detection System”, *Proceedings of the Infrared Information Symposia*, 3rd NATO-IRIS Joint Symposium, Vol. 43 No.3, July 1999, pp. 343-364.

[101] P. Lucey, T. Williams, M. Winter and E. Winter, “Two Years of Operations of AHI: an LWIR Hyperspectral Imagery”, *Proceedings SPIE*, Vol. 4030, pp. 31-40, 2003.

[102] P. Lucey, T. Williams, M. Mignard, J. Jullian, D. Kobubon, G. Allen, D. Hampton, W. Schaff, M. Schlangen, E. Winter, W. Kendall, A. Stocker, K. Horton, A. Bowman, “AHI: an Airborne Long-Wave Infrared Hyperspectral Imager”, *Proceedings SPIE*, Vol. 3431, pp. 36-43, 1998.

BIOGRAPHICAL SKETCH

Jeremy Bolton received the Bachelor of Science degree in computer engineering from the University of Florida, Gainesville, in May 2003. He received his Master of Engineering and Doctor of Philosophy from the University of Florida in December 2008.

Currently, he is a research assistant in the Computational Science and Intelligence Lab in the Computer and Information Sciences and Engineering Department at the University of Florida. Research includes the development of algorithms, methodologies, and models with a solid mathematical and/or statistical base with applications to landmine detection. Current and previous research applies these methods to a variety of data including hyperspectral, multispectral, radar, and infrared.

Jeremy Bolton is a member of IEEE Computational Intelligence Society, IEEE Geoscience and Remote Sensing Society and Society of Photographic Instrumentation Engineers.

ADAPTIVE SAMPLING BY HISTOGRAM EQUALIZATION: THEORY, ALGORITHMS, AND APPLICATIONS

A DISSERTATION
SUBMITTED TO THE SCHOOL OF ARTS AND SCIENCES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY

BY

OLADIPO O. FADIRAN

DEPARTMENT OF ENGINEERING, SYSTEM SCIENCE PROGRAM

ATLANTA, GEORGIA

MAY 2007

© 2007

OLADIPO O. FADIRAN

All Rights Reserved

CLARK ATLANTA UNIVERSITY THESIS OR DISSERTATION DEPOSITED IN THE ROBERT W. WOODRUFF LIBRARY

STATEMENT OF UNDERSTANDING

In presenting this dissertation as a partial fulfillment of the requirements for an advanced degree from Clark Atlanta University, I agree that the Robert W. Woodruff Library shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote from, to copy from, or to publish this dissertation may be granted by the author or, in his absence, the Dean of the School of Arts and Sciences at Clark Atlanta University. Such quoting, copying, or publication must be solely for scholarly purposes and must not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential gain will not be allowed without written permission from the author.

_____ _____

Signature of Author Date

NOTICE TO BORROWERS

All dissertations and theses deposited in the Robert W. Woodruff Library must be used only in accordance with the stipulations prescribed by the author in the preceding statement.

The author of this dissertation is:

Name: Oladipo O. Fadiran

Street Address: C.T.S.P.S., 223 James P. Brawley Dr. SW.

City, State, and Zip: Atlanta, GA, 30314.

The director of this thesis/dissertation is:

Professor: Dr. Péter Molnár

Department: Computer and Information Science

School: Arts and Sciences Clark Atlanta University

Office Telephone: 404-880-6419

Users of this thesis/dissertation not regularly enrolled as students of the Atlanta University Center are required to attest accordance of the preceding stipulations by signing below. Libraries borrowing this thesis/dissertation for use of patrons are required to see that each user records here the information requested.

NAME OF USER	ADDRESS	DATE	TYPE OF USE

ABSTRACT

SYSTEMS SCIENCE (ENGINEERING CONCENTRATION)

FADIRAN, OLADIPO O. B.ENGR. UNIVERSITY OF ILORIN, 1997

M.Sc. UNIVERSITY OF CAPE TOWN, 2001

ADAPTIVE SAMPLING BY HISTOGRAM EQUALIZATION:

THEORY, ALGORITHMS, AND APPLICATIONS

Advisor: Dr. Péter Molnár

Dissertation dated May 2007

We present the investigation of a novel, progressive, adaptive sampling scheme. This scheme is based on the distribution of already obtained samples.

Even spaced sampling of a function with varying slopes or degrees of complexity yields relatively fewer samples from the regions of higher slopes. Hence, a distribution of these samples will exhibit a relatively lower representation of the function values from regions of higher complexity. When compared to even spaced sampling, a scheme that attempts to progressively equalize the histogram of the function values results in a higher concentration of samples in regions of higher complexity. This is a more efficient distribution of sample points, hence the term adaptive sampling. This conjecture is confirmed by numerous examples.

Compared to existing adaptive sampling schemes, our approach has the unique ability to efficiently obtain expensive samples from a space with no prior knowledge of the relative levels of variation or complexity in the sampled function. This is a requirement in numerous scientific computing applications.

Three models are employed to achieve the equalization in the distribution of sampled function values: (1) an active-walker model, containing elements of the random walk theory, and the motion of Brownian particles, (2) an ant model, based on the simulation of the behavior of ants in search of resources, and (3) an evolutionary algorithm model. Their performances are compared on objective basis such as entropy measure of information, and the Nyquist-Shannon minimum sampling rate for band-limited signals.

The development of this adaptive sampling scheme was informed by a need to efficiently synthesize hyperspectral images used in place of real images. The performance of the adaptive sampling scheme as an aid to the image synthesis process is evaluated. The synthesized images are used in the development of a measure of clutter in hyperspectral images. This process is described, and the results are presented.

ACKNOWLEDGMENTS

I appreciate the immense contributions of Dr. Péter Molnár, my advisor, and Dr. Lance Kaplan, my co-advisor. They provided direction, encouragement, and constructive criticisms, all in the right proportions. It has been great working with you. I am also grateful to the other members of my dissertation committee, namely, Dr. George Japaridze, Dr. William Seffens, and Dr. BhuDev Sharma. Their various expertise formed an invaluable pool of resources that was always available to me.

Also worthy of mention is Dr. Ayodele Abatan, who along with my co-advisor facilitated my enrollment in the Ph.D program. The Systems Science program coordinator, Dr. Lebone Moeti, was always accessible and quick to address any issues. I am grateful to Dr. Alfred Msezane and the other members of the Center for Theoretical Physics for creating a stimulating environment for research work.

I would also like to express my gratitude for the financial support that I received from the United States Army research laboratory. My research work was fully funded through the ”Science of Land Target Signatures” project underwritten by this agency.

I am grateful to family members and friends who contributed to making this happen. I am grateful to God for putting you in my life.

To Temitope(iyebiye mi), Babayosimi, and *Atinuke*. This would not have been possible without your unwavering support through thick and thin. I am truly blessed to have you.

TABLE OF CONTENTS

ACKNOWLEDGMENTS ii

LIST OF FIGURES vii

LIST OF TABLES xi

Chapter Page

1 INTRODUCTION 1

1.1 Background and Motivation 1

1.1.1 Adaptive Sampling by Histogram Equalization (ASHE) . . 1

1.1.2 Model for Quantifying Clutter in Hyperspectral Images . . 6

1.2 Dissertation Outline 8

2 ADAPTIVE SAMPLING BY HISTOGRAM EQUALIZATION (ASHE) ALGORITHM 10

2.1 Review of Adaptive Sampling Algorithms 10

2.2 Theoretical Basis 13

2.3 Illustration of ASHE 16

2.4 Application Areas 20

2.4.1 Data Synthesis 20

2.4.2 Design of Experiments 21

2.4.3 Surface Reconstruction from Expensive Samples 21

2.4.4 Progressive Transmission/Rendering 22

3 MODELS UTILIZED IN IMPLEMENTING ASHE 23

3.1 Active Walker Model 23

3.1.1 ASHE Implementation using Active Walkers 25

3.2 Ant Model 29

TABLE OF CONTENTS

Chapter Page

3.2.1 ASHE Implementation using the Ant Model 31

3.3 Evolutionary Algorithm Model 34

3.3.1 ASHE Implementation using the Evolutionary Algorithm Model 34

4 PERFORMANCE AND SENSITIVITY ANALYSIS OF MODELS . . 38

4.1 Measure Based on Frequency Content 39

4.2 Measure Based on Entropy Measure of Information 44

4.3 Analysis of the Active Walker Model 44

4.4 Analysis of the Ant Model 56

4.5 Analysis of the Evolutionary Algorithm (EA) Model 73

4.6 Further Analysis of the Active Walker Model 84

4.7 Conclusions 85

5 NATURE, USES AND SYNTHESIS OF HYPERSPECTRAL IMAGES 89

5.1 Nature of Hyperspectral Images 89

5.2 Synthesizing Hyperspectral Images 91

5.2.1 The Digital Imaging and Remote Sensing Image Generation model (DIRSIG) 92

5.3 Image Synthesis with DIRSIG 93

6 EFFICIENT HYPERSPECTRAL IMAGE SYNTHESIS USING ASHE 100

6.1 ASHE based Image Synthesis 101

6.1.1 Imaged Scene 102

6.1.2 Input Factors to DIRSIG 103

TABLE OF CONTENTS

Chapter Page

6.1.3 Baseline ATR Performance . . . 103

6.1.4 Image Synthesis . . . 105

6.2 Experiments . . . 105

7 DEVELOPMENT OF A MEASURE OF CLUTTER FOR HYPERSPECTRAL IMAGES . . . 110

7.1 Clutter Complexity Measure . . . 111

7.2 Clutter Complexity Measure for Single Hyperspectral Bands using Real Data . . . 112

7.2.1 Baseline ATR Performance . . . 113

7.2.2 Multiple-feature CCM . . . 115

7.2.3 Single-feature CCM . . . 116

7.2.4 Single-Band CCM Experiments . . . 117

7.3 Clutter Complexity Measure for Hyperspectral Images . . . 121

7.3.1 Baseline ATR Performance . . . 121

7.3.2 Image Clutter Metrics . . . 123

7.3.3 Determining Significant Metrics . . . 124

7.3.4 Hyperspectral Images CCM Experiments . . . 129

8 CONCLUSIONS . . . 138

8.1 Summary of contributions . . . 139

8.2 Suggestions for further work . . . 142

A Clutter Complexity Metrics . . . 144

A.1 Single-band clutter metrics . . . 144

A.1.1 Global standard deviation . . . 144

TABLE OF CONTENTS

Chapter Page

A.1.2 Schmieder Weathersby 144

A.1.3 Homogeneity 145

A.1.4 Energy 145

A.1.5 Entropy 145

A.1.6 Target Interference Ratio 145

A.1.7 Outlier/Edge 146

A.1.8 FBM Hurst Parameter 146

A.1.9 Metrics c and p derived from Gabor filtered images [9] . . 147

A.2 Metric derived from band information content 148

A.2.1 Band correlation 148

A.3 Anomaly detectors 148

A.3.1 Dot product 148

A.3.2 Kullback-Leibler 148

A.4 Metrics derived from the Gray Level Co-occurrence Matrix 149

LIST OF FIGURES

Figure		Page
1.1	Basis of Adaptive Sampling by Histogram Equalization (ASHE)	4
2.1	Illustration of basis for the ASHE algorithm	14
2.2	Illustration of Performance of ASHE, in sampling 2-dimensional function	18
2.3	Illustration of Performance of ASHE, in sampling 2-dimensional image	19
3.1	Simple symmetric random walk on $\mathbb{Z}^2$	24
3.2	Illustration of 2 possible paths for ants to reach a resource. With path 'a' shorter than 'b'	30
3.3	Illustration of standard genetic algorithm recombination process.	35
4.1	Images with single spatial frequencies, and their magnitude images in the frequency domain	40
4.2	Cross-sections of images in frequency domain	41
4.3	Frequency content based measure of function complexity	43
4.4	Entropy based measure of function complexity	45
4.5	Two test functions: smooth, and rapidly varying	46
4.6	Intermediate steps from sampling a smooth test function with the active walker model	49
4.7	Intermediate steps from sampling a rapidly varying test function with the active walker model	50

LIST OF FIGURES

Figure Page

4.8 Intermediate steps from sampling a smooth test function with the active walker model, resulting in good performance. 57
4.9 Intermediate steps from sampling a smooth test function with the active walker model, resulting in poor performance. 58
4.10 Intermediate steps from sampling a rapidly varying test function with the active walker model, resulting in good performance. 59
4.11 Intermediate steps from sampling a rapidly varying test function with the active walker model, resulting in poor performance. 60
4.12 Model to simulate reduction of pheromone concentration as a function of the distance away from source. Concentration $\propto 1/d^2$. 61
4.13 Pheromone concentration/probability change from intermediate steps in the sampling of a smooth function using the ant model 62
4.14 Pheromone concentration/probability change from intermediate steps in the sampling of a rapidly varying function using the ant model 63
4.15 Intermediate stages from sampling a smooth test function with the ant model, resulting in good performance 69
4.16 All stages from sampling a smooth test function with the ant model, resulting in poor performance 70
4.17 Intermediate stages from sampling a rapidly varying test function with the ant model, resulting in good performance 71
4.18 All stages from sampling a smooth test function with the ant model, resulting in poor performance 72

LIST OF FIGURES

Figure Page

4.19 Intermediate stages from sampling a smooth test function with the Evolutionary Algorithm model, resulting in good performance. 79
4.20 All stages from sampling a smooth test function with the Evolutionary Algorithm model, resulting in poor performance. 80
4.21 Intermediate stages from sampling a rapidly varying test function with the Evolutionary Algorithm model, resulting in good performance. . . . 81
4.22 All stages from sampling a rapidly varying test function with the Evolutionary Algorithm model, resulting in poor performance. 82
4.23 *Slice* from 3-dimensional Sinc test function. (a) top view, (b) side view. 84

5.1 Example hyperspectral image, and material signatures. 90
5.2 Combination of factors used to generate images in synthesized database. Total of 216 images. 95
5.3 Example hyperspectral image bands from the same scene 96
5.4 Example hyperspectral images bands from different scenes 97
5.5 Example hyperspectral scenes showing levels of clutter. 98
5.6 Example spectral signatures of materials in synthesized scenes 99

6.1 A single band from the hyperspectral image of the urban scene used for synthesizing multiple images. 102
6.2 Distribution of baseline ATR performance values for different sampling methods 108

LIST OF FIGURES

Figure Page

7.1 RX detection in FLIR images: (a) Original image (b) Image of RX statistic (c) Resulting detection image. 114
7.2 Scatter plot of clutter complexity measure and false alarm count: (a) Weighted sum of multiple features (b) Single feature. 117
7.3 FLIR band classification by clutter complexity measure 118
7.4 Example of band selection based on wavelength to ensure a uniform distribution of the choice of hyperspectral bands. 119
7.5 Performance of clutter complexity measures represented as average false alarm count versus number of bands integrated into the RX detector. . . 120
7.6 Target detection in HSI using the Adaptive Coherence Estimator (ACE) detector 122
7.7 Gray Level Co-occurrence Matrices from hyperspectral images 125
7.8 Gabor filtered band from hyperspectral image 126
7.9 Frequency of selection of clutter metrics for different training image sets 131
7.10 Distribution of CC values resulting in the average values shown in Table7.4(a). The indicated percentages, and the actual number of images that they represent are also shown in the same table. 136

LIST OF TABLES

Table		Page
2.1	Comparison of the Requirements for the Proposed ASHE Algorithm to other Adaptive Sampling Methods.	12
4.1	Performance of active walker model in sampling the smooth function, sorted by the frequency based measure.	51
4.2	Performance of active walker model in sampling the smooth function, sorted by the entropy based measure.	52
4.3	Performance of active walker model in sampling the rapidly varying function, sorted by the frequency based measure.	53
4.4	Performance of active walker model in sampling the rapidly varying function, sorted by the entropy based measure.	54
4.5	Performance of the ant model in sampling the smooth function, sorted by the frequency based measure.	65
4.6	Performance of the ant model in sampling the smooth function, sorted by the entropy based measure.	66
4.7	Performance of the ant model in sampling the rapidly varying function, sorted by the frequency based measure.	67
4.8	Performance of the ant model in sampling the rapidly varying function, sorted by the entropy based measure.	68

LIST OF TABLES

Table Page

4.9 Performance of the Evolutionary Algorithm model in sampling the smooth function, sorted by the frequency based measure. 75

4.10 Performance of the Evolutionary Algorithm model in sampling the smooth function, sorted by the Entropy based measure. 76

4.11 Performance of the Evolutionary Algorithm model in sampling the rapidly varying function, sorted by the frequency based measure. 77

4.12 Performance of the Evolutionary Algorithm model in sampling the rapidly varying function, sorted by the entropy based measure. 78

4.13 A summary of the results from the analysis of the three sampling models. It shows the ranges of input factors that result in good performance, and averages of the entropy based performance values. 83

4.14 Performance of Active Walker model in sampling the 3-dimensional Sinc function, sorted by the frequency based measure. 86

4.15 Performance of Active Walker model in sampling the 3-dimensional Sinc function, sorted by the entropy based measure. 87

7.1 List of some of the image statistical features used in deriving the clutter complexity measure. 115

7.2 Summarized list of clutter metrics used in deriving the clutter complexity measure for hyperspectral images. 127

7.3 Averaged correlation coefficients obtained between the clutter measure and false alarm rates using different training sample sizes. 132

7.4 Averaged correlation coefficients obtained between the clutter measure and false alarm rates using different training sample sizes. 134

CHAPTER 1

INTRODUCTION

The main contribution of this work is the development of a novel, progressive, adaptive sampling method, based on the distribution of already obtained samples. This algorithm is shown to be efficient for sampling a space for which there is no prior information on the global, and relative levels of local variation of the function in question, and the cost of obtaining each sample is prohibitive. This is a requirement in numerous scientific computing applications.

We present results of applying the developed algorithm in the efficient synthesis of hyperspectral images. These images are used in the development of a framework for quantifying clutter in hyperspectral images.

1.1 Background and Motivation

1.1.1 Adaptive Sampling by Histogram Equalization (ASHE)

The reconstruction of most continuous functions from a finite number of sample points results in errors. Since there is always a constraint on the number of samples that can be obtained, the aim of efficient sampling schemes is to minimize the inherent errors that result from reconstructing a continuous function from the finite discrete samples.

An alternate approach to the sampling question is based on the Nyquist-Shannon minimum sampling theory. This shows that a sampling rate N of at least twice the highest frequency component f_s in a signal is required in order to unambiguously reconstruct the

signal from its samples. Thus, the required sampling rate is

$$N \geq 2 \times f_s \, . \tag{1.1}$$

That is, a higher sampling rate is required to unambiguously reconstruct a function with higher frequency components. Reconstructing a continuous signal from finite samples is equivalent to representing finite frequency components of the signal. Error-free reconstruction of a continuous function that is not bandlimited, requires the representation of infinite frequency components [43, 56, 67].

Some sampling algorithms focus more on reducing the *effect* of this inherent error, as is the case in image processing, where structured, and thus more apparent artifacts like aliasing are converted to noise [13, 15, 60]. Others attempt to reduce the actual error by distributing the limited samples more efficiently [52, 53, 64, 71]. A third group of algorithms combines efficient distribution of the samples with the reduction of the effect of the error, as can be found in the adaptive variant of the algorithm discussed here [22]. Algorithms that distribute samples efficiently usually harness the nonstationary nature of the function[1] to be sampled [65]. That is, samples are distributed based on local statistics. This information may be required prior to the sampling process [22], or continuously made available and updated during the sampling process, using a progressive sampling approach [64, 71].

For actual error reduction, the problem of allocating sample points efficiently becomes trivial if there is prior knowledge of levels of variation or local frequency components in a function. The samples are simply allocated based on the different levels of variation. This means that the regions of rapid variation or higher frequencies are allo-

[1] Statistical properties of nonstationary functions, such as the mean change over time

cated relatively more samples. In many cases, there is no *a priori* information on the global, or relative levels of local variation of the function being sampled. Without such knowledge, sample points are usually placed randomly, spaced evenly, or some other variant of these arrangements to avoid the artifact resulting from even spacing [13, 15]. These approaches are however, inefficient for sampling a nonstationary function.

One approach to solving this problem is the progressive intensification of sampling in a local region based on some information content criterion as in the work on ray-tracing [52, 71]. Variable sampling rates may also be achieved using variants of Markov-chain Monte-Carlo (MCMC) methods adapted for this purpose [16, 41]. Local sampling rates may also be pre-determined based on prior information on the local complexities in the function to be sampled. An example of this is found in the adaptive form of the farthest point algorithm [22]. These and similar methods however, require at least one of the following: *a priori* knowledge on the global, or relative levels of local variation of the function to be sampled [21, 22], computation to determine local information content [52, 71, 88], or an acceptance/rejection step in the progressive sampling process [16, 41]. These requirements make these methods infeasible for sampling in many applications.

We present a progressive adaptive sampling algorithm, in which the subsequent sample locations are determined based on the distribution of already collected samples. The algorithm is based on the thesis that even spaced sampling of a function with varying degrees of complexity results in a distribution of samples with relatively lower representation of values from regions of higher complexity. A simple illustration of this can be seen in Figure 1.1. Since the slope in part I of the function is higher than that in part II, more samples are collected per unit length of the function in part II. This is reflected in the distribution of the function values. It is more efficient to concentrate more samples in the region of higher slope or complexity. This results in a reduction of the dominance of

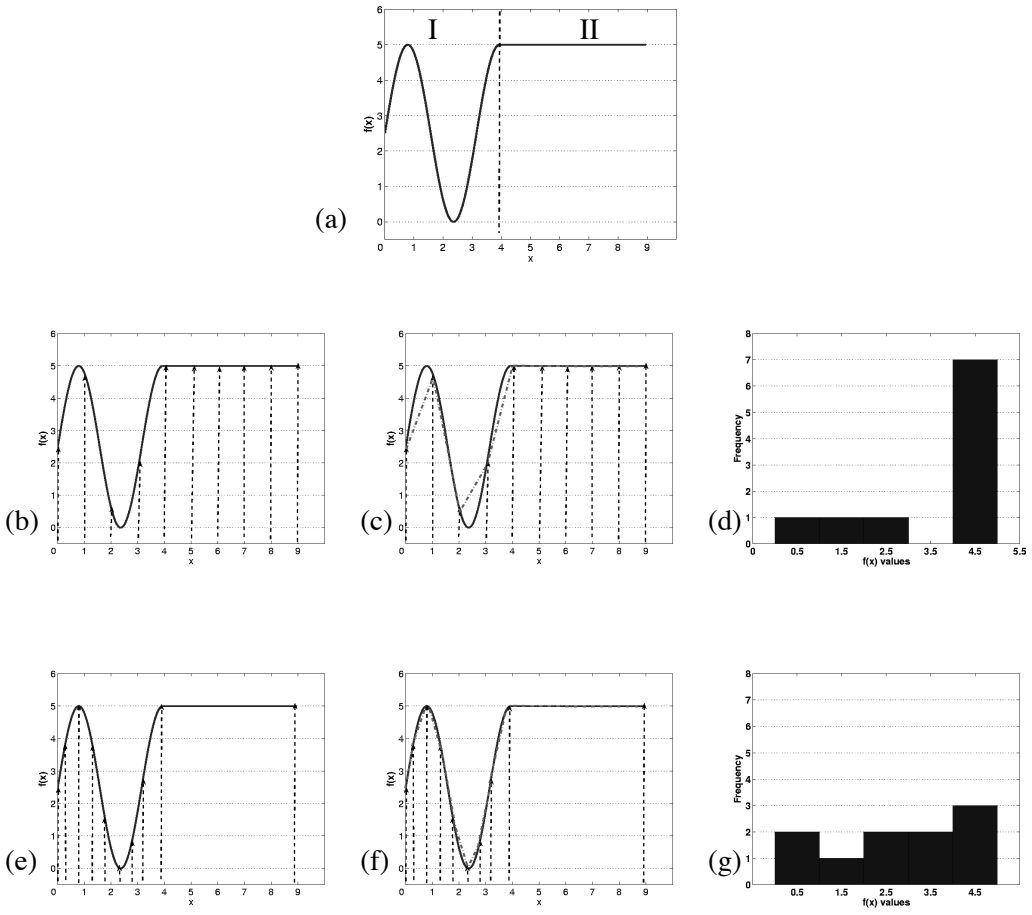


Figure 1.1: This figure illustrates the basis of adaptive sampling by histogram equalization. Figure (a) shows the function to be sampled, (b) shows an evenly spaced set of 10 samples with its linear reconstruction depicted in (c). Figure (d) is the histogram of resulting function values from evenly spaced samples. Figures (e)-(g) show the corresponding results for adaptively placed samples. The error of a linear reconstruction based on the sampled points, represented by the dotted lines, is significantly greater for the even spaced sampling, (c), than for the adaptively placed samples, (f).

the function values from the regions of lower slope, and relatively more samples in the other region. That is, a distribution of function values that tends more towards a uniform distribution. The improved efficiency in sample distribution is evidenced by the function reconstruction based on the samples, indicated by the dotted lines. The concept of samples per unit length, which this example is based on, can be easily extended to higher dimensions.

A sampling scheme that progressively attempts to equalize the histogram of these function values results in a relatively higher concentration of samples in regions of complexity. This results in a more efficient distribution of sample points, hence the adaptive sampling. This conjecture is confirmed by numerous examples shown in Chapter 3 of this dissertation. We call the algorithm Adaptive Sampling by Histogram Equalization (ASHE). The algorithm is not subject to the limitations of the adaptive algorithms mentioned earlier, and only requires that it is possible to obtain the value of the function at each sampled point. No prior knowledge of the local or global levels of variations in the function is required. Also, the only extra computational overhead required by this algorithm is the computation of a histogram at each stage of the sampling procedure. Finally, there is no acceptance/rejection step in the progressive sampling procedure, every obtained sample is kept. This makes the procedure particularly useful for obtaining expensive samples[2].

Three models are employed to achieve the progressive equalization in the distribution of sampled function values. These are:

1. an active-walker model [46, 48, 49], with basis in both the random walk theory [72], and the motion of Brownian particles [75],

[2]Situations for which obtaining each sample is prohibitive in cost, time or some other resource. Good examples are ab initio computations in the physical sciences.

2. an ant model, based on the simulation of the behavior of ants in search of resources [18, 19, 87], and

3. an evolutionary algorithm [61] model.

These are evaluated on their ability to achieve our objective of efficient sample distribution. They are also compared on the basis of ease of implementation. Appropriate models for specific applications are identified based on the analysis of these results.

1.1.2 Model for Quantifying Clutter in Hyperspectral Images

A specific application of the adaptive sampling scheme reported in this work is the efficient synthesis of images. In fact, the development of the adaptive sampling algorithm was informed by the need for such a scheme in this application. The synthesized images are used in our framework for modeling clutter in images. Models of targets and clutter, aid in the understanding of images in general [85, 94]. Targets are considered to be objects of interest in a particular image. We define clutter as any factor in the image that may increase the difficulty for an Automatic Target Recognition (ATR) algorithm in detecting or identifying a target in a scene [23, 25, 27].

Our objective is to obtain a measure of the amount of clutter in an image that will be an indication of the inherent difficulty for an ATR to find a target. This measure will form bounds on the performance of any ATR, such that a high value of this measure will indicate that an ATR will produce a high false alarm (FA) rate. A low value may, however, not result in a low FA rate. This depends on the exact nature of the ATR. Such a measure could serve as a basis for evaluating, and comparing ATRs on an objective basis. It could also serve as a basis for measuring image quality, independent of a particular target detection algorithm or scheme [66].

Our approach to obtaining a clutter measure in these images is to compute a set of statistical image features that are significant for, and monotonically related to ATR performance. In addition, these features have to be algorithmically uncomplicated to implement [27, 66]. The measure of clutter is then obtained as an aggregation of these features that correlates best with baseline ATR performance. The process of combining these features to yield the required result is obtained through a training process on a subset of available image data. Once established, this is generalized over the complete dataset [23, 25, 27].

This training process requires image data in numbers that are statistically significant. There is limited availability of these in the public domain. This limitation can be overcome by synthesizing the desired images. Tools for such image synthesis require inputs like object and scene geometry, object material properties, atmospheric conditions, and illuminating sources [10, 76]. These factors are then accounted for in the ray-tracing process that produces the final image. A database of images can then be produced by synthesizing images with a varying combination of these inputs.

To ensure that the result of the clutter analysis of images from such a synthetic database is general, and representative of real images, two basic requirements should be met. These are the fidelity of each image, and representation in all categories of ATR difficulty in the database. The former requirement is beyond the scope of this work. To achieve the second requirement, an image is modeled as a point on a multidimensional surface, whereby each dimension represents an input parameter to the image synthesis software. Thus, each image results from synthesizing with a combination of input parameters, with each being a possible source of variation with respect to ATR performance. The aim is to sample this surface in order that the resulting images show adequate statistical representation for all categories of ATR difficulty. The prohibitive cost of image

synthesis places a limitation on the number of samples that can be produced [10, 76]. There is also no prior knowledge of how the ATR performance varies with changes in the input parameters used to synthesize the images. Relatively more dense sampling in the regions of this space with higher variability with respect to ATR performance results in a more diverse set of images, and vice-versa [24, 26]. We sample this surface using the ASHE algorithm, and investigate the improvement in performance with respect to diversity in the synthesized images.

1.2 Dissertation Outline

In Chapter 2, we review some existing adaptive sampling schemes and summarize some of their limitations, especially those addressed by our proposed sampling scheme. We then establish the premise on which the Adaptive Sampling by Histogram Equalization (ASHE) algorithm is based. This is done through graphical and analytical methods. The improvement of sampling by the ASHE algorithm over random or even spaced sampling is demonstrated through examples illustrating its performance.

Three stochastic optimization models employed in implementing the ASHE algorithm are discussed in Chapter 3. The underlying theory and heuristics that these models are based on are discussed. We then describe the specific adaptation of the general forms for our particular application.

A performance and sensitivity analysis of the three models is carried out in Chapter 4. We establish two performance measures based on the entropy measure of information [34, 82], and the Nyquist-Shannon minimum sampling rate for band-limited signals [43, 67]. These give an indication of the level of variation or complexity in a function. The measure is computed as the correlation between these two measures separately, and the sample density distribution obtained by employing each model. Hereby,

a higher correlation indicates better performance. We conduct a sensitivity analysis of each model to determine the change in performance for various input factors to each model. This analysis is similar to previous work in ant models [2]. The results obtained from these, serve as good indicators of the appropriateness of each model for particular applications.

The second part of the dissertation, comprising of Chapters 5 - 7, presents results from an application that utilizes synthesized hyperspectral images. An important step in this application is aided by the ASHE algorithm. Chapter 5 gives a background on the nature and uses of hyperspectral images. The need for image synthesis is stated, and the process is described.

The need for an adaptive sampling scheme in the image synthesis process is identified in Chapter 6. The properties of the ASHE algorithm are reiterated to justify its choice for this application. Images are then synthesized based on the ASHE algorithm, and the recorded performance improvement is evaluated on objective basis.

A framework for modeling clutter in hyperspectral images is detailed in Chapter 7. The process of quantifying clutter using both real and synthesized images is then described. Numerous experiments are carried out to investigate this framework.

Chapter 8 concludes with a summary of the findings, and major contributions of the dissertation. Suggestions for future work are also made.

CHAPTER 2

ADAPTIVE SAMPLING BY HISTOGRAM EQUALIZATION (ASHE) ALGORITHM

In this chapter, we propose a novel, progressive, adaptive sampling scheme, based on the distribution of obtained samples. We conduct a brief review of some existing adaptive sampling schemes, and compare them to the proposed algorithm. This is in order to highlight the limitations of the existing algorithms that are addressed by Adaptive Sampling by Histogram Equalization (ASHE). Next, we layout the theoretical basis for the adaptive sampling algorithm. We employ analytical and graphical methods to illustrate why, and how the ASHE algorithm works. Examples are presented to illustrate the performance of the algorithm. Finally, we identify possible practical applications areas of the algorithm.

2.1 Review of Adaptive Sampling Algorithms

Processes like the reconstruction of continuous signals from finite samples, and numerical integration of a continuous signal will generally result in error. This is because they attempt to represent a continuum with a discrete space. These errors can be reduced by using relatively more samples in intervals where the sampled values vary the most. This is the aim of adaptive sampling. That is, the efficient distribution of a finite number of samples in a manner that reflects the varying levels of rate of change, or complexity in a sampled function, in order to minimize errors. Even spaced sampling of a function with nonstationary statistics is inefficient. Adaptive sampling places relatively more samples in regions of higher variance in samples.

An approach to solving this problem is the progressive intensification of sampling in a local region based on some information content criterion as in these works on ray-tracing found in [64, 71]. These algorithms utilize a refinement scheme to determine where to increase the sample density. For example, the variance of sample values in a region is computed, and the sampling density is increased in that region until a threshold is reached [52]. Other measures, such as contrast have also been used as the basis for further refinement [60]. Variable sampling rates may also be achieved using variants of Markov Chain Monte-Carlo (MCMC) methods adapted for this purpose as discussed in [41, 16]. Local sampling rates may also be pre-determined based on prior information on the local complexities in the function to be sampled. An example of this is found in the adaptive form of the farthest point algorithm discussed in [22]. The gradient based sampling algorithms increase sampling density in a region where the slope in the measured quantity exceeds a set threshold. These, and similar methods however, require at least one of the following: *a priori* knowledge on the global, and relative level of local variation of the function to be sampled [22], computation to determine local information content [71, 64, 60], an acceptance/rejection step in the progressive sampling process [41, 16], or a large number of samples to converge [64, 60, 41, 16]. These make them inappropriate or even infeasible for sampling in many applications.

Our developed algorithm only requires that it is possible to obtain the value of the function at each sampled point. No prior knowledge of the local or global levels of variations in the function is needed. This information is not available in many cases. In fact, the process of efficient distribution of sample points becomes apparent when this is available. Also, the only extra computational overhead required by this algorithm is the computation of a histogram at each stage of the sampling procedure. Finally, there is no acceptance or rejection step in the progressive sampling procedure, every obtained sam-

Table 2.1: Comparison of the Requirements for the Proposed ASHE Algorithm to other Adaptive Sampling Methods.

Requirements	**ASHE algorithm**	**Entropy Based**	**Adaptive Farthest Point**	**Monte-Carlo methods**	**Gradient Based**
Info. on Global fn.	N	N	Y	Y	N
Info. on Local fn. Variation	N	N	Y	N	N
Local/Regional Computation	N	Y	N	N	Y
Acceptance/Rejection Step	N	N	N	Y	N
Large No. Samples Needed	N	Y	N	Y	Y

Entropy based algorithms do further sampling based on local information content, while gradient based algorithms sample based on local gradients. The table entries represent $N-$ Not required, and $Y-$ Required.

ple is kept. This makes the procedure particularly useful for obtaining expensive samples. Table 2.1 shows a comparison of the ASHE sampling scheme to similar algorithms. The advantages of ASHE over the other algorithms are highlighted.

The described adaptive sampling algorithm is particularly useful in the following situations:

- No *a priori* information on the global, and relative levels of local variation of the function to be sampled is available.
- Obtaining samples is prohibitively expensive.

2.2 Theoretical Basis

Consider a monotonically increasing, non-linear function of $x \in \mathbb{R}^+$:

$$y = f(x) = Cx^n \ , \tag{2.1}$$

where C is a positive constant, and $n > 1$. An indication of the rate of change or level of complexity in the function is the derivative. This is given by:

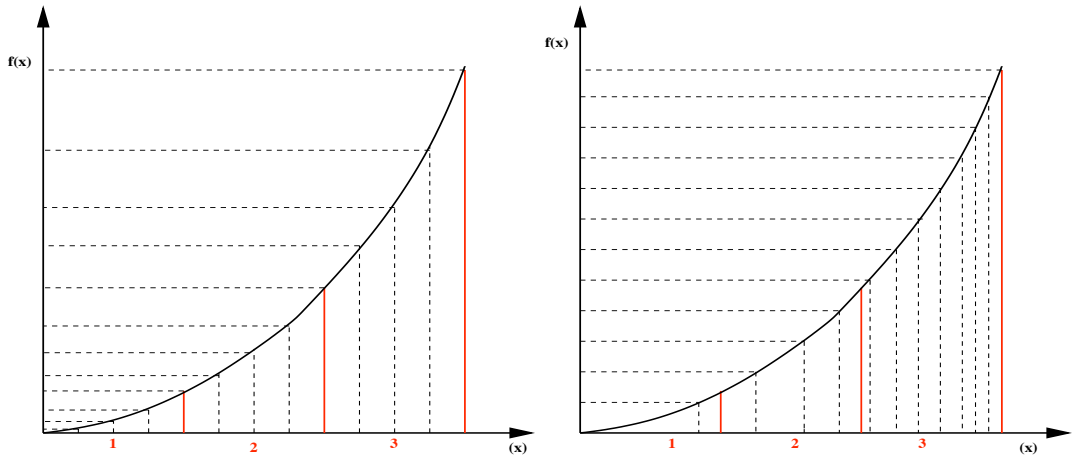
$$f^{'}(x) = nCx^{n-1} \ . \tag{2.2}$$

Based on the assumptions made about the function,

$$x_i > x_j \Longrightarrow f^{'}(x_i) > f^{'}(x_i). \tag{2.3}$$

Hence, an adaptive sampling algorithm will attempt to place relatively more samples as x increases. If there is prior knowledge on the form of the function, an optimal sample distribution may be obtained based on these derivatives. However, without such prior knowledge, the obvious solution is to sample the function with even spaced samples in x. This results in the same sample density for the different regions. A more efficient scheme will result in a higher sample density as the values of x and $f^{'}(x)$ increase.

We propose an algorithm that increases the relative sampling density as the level of complexity increases. This algorithm is based on the distribution of the samples from the co-domain. Consider an example of the described function with, even space sampling in x, as shown in Figure 2.1(a), and even spaced sampling in the co-domain $f(x)$, shown in Figure 2.1(b). The functions are divided into three equal intervals in x. An equal number of 12 samples are used in both cases. Some of the samples are at the same locations



(a) (b)

Figure 2.1: Illustration of basis for the ASHE algorithm in sampling a monotonic increasing function $x \in \mathbb{R}^+$ of the form $f(x) = Cx^n$, where C is a positive constant, and $n > 1$: (a) even spaced sampling in the domain, (b) ASHE algorithm, even-spaced sampling in the co-domain, resulting in more efficient sample density in the domain.

as the dividing red lines. As shown in Figure 2.1(a), even spaced sampling in x yields the obvious result of the same sample rate, even with increase in function complexity indicated by the increasing slope. That is, the three defined intervals, with different rates of change in sample values, have the same sample rate. Based on our discussions on error reduction, and adaptive sampling, this is inefficient.

Suppose we sample progressively, in order to equalize the distribution of the obtained samples. This will result in equal representation of samples in the co-domain. Hence, an equalized distribution of samples. This is shown in Figure 2.1(b). Projecting the even sampling in the co-domain $f(x)$ back to the domain x, shows a different sample rate for each of the regions. The relative increase in sampling density is proportional to the complexity in each region of the function, this is indicated by the slope. In contrast to other sampling schemes, the focus of the ASHE algorithm is on the co-domain, instead

of the domain.

In summary, the ASHE algorithm produces an adaptive sampling density in the domain, by varying the sample rates in *proportion* to the relative rate of change in the sampled function. This is an improvement over even spaced sampling, which produces the same sampling densities for regions containing different rates of change. However, the optimal *proportion* to determine the relative sample rate that minimizes errors will be specific to each function. Referring back to the derivative of the general form in (2.2), such optimal proportion will be a function of C, and n. We restate the fact that these are not known *a priori*. Also, any practical phenomenon to be sampled will consist of a complex combination of the type of function used in the illustration. A complete analytical consideration will have to consider these complex system. Note, however, that the extension of ASHE to such a complex system is valid. Samples are distributed in proportion to the relative levels of variation in the system. Finally, there is no prior knowledge of the exact divisions in the co-domain. This forms the basis of the histogram to be equalized. The foregoing precludes a rigorous mathematical consideration of the concept. We, however, conduct further analysis in a manner similar to that of other heuristic methods. An example of such analysis is found in [2]. These are usually performance and sensitivity analysis to determine factors that yield the best result from these algorithms, for a given class of applications. Many heuristics have been employed in solving practical problems for which obtaining an optimal solution is computationally prohibitive, or even infeasible. Some examples of such practical applications include routing for vehicles and in telecommunication networks [69, 11], and scheduling in industrial organizations [33]. Our analysis of the ASHE algorithm is reported in Chapter 4.

2.3 Illustration of ASHE

We illustrate the performance of the ASHE algorithm by comparing it to evenly spaced, and randomly spaced samples. The comparison is based on the quality of the functions that are reconstructed from sample points. Details of the actual implementation of the ASHE algorithm to generate these samples are presented later in Chapter 3. We experiment with two 2-dimensional functions, and reconstruct the function from their samples by the 4NN[1] nearest neighbor algorithm [43]. We emphasize the point that these examples are solely for the purpose of illustrating the algorithm, and not necessarily practical application areas. The areas of possible practical application will be discussed in the next section.

Figure 2.2 shows the comparison of the performance of ASHE to evenly spaced and randomly placed sample points. Note the clustering of the sample points in the regions of the functions with relatively higher local slopes when ASHE is used for sampling. Note also how the resulting normalized histograms compare to the Uniform distribution. Evenly spaced and randomly placed samples result in histograms with underrepresentation of the function values in regions of higher complexity, and a dominance of the function values in the regions of lower complexity. The histogram of the function values obtained by adaptive sampling however, shows a tendency towards the Uniform distribution. An objective measure of the histogram comparison is a sum-squared difference between the histograms and a normalized Uniform distribution with the same number of bins. The lower this value, the closer the histogram is to the Uniform distribution. The function reconstruction quality, indicated by the Peak Signal to Noise Ratio (PSNR) values, is highest when the function is sampled adaptively using the ASHE algorithm. A total of 100 experiments are conducted, and the indicated PSNR and deviation values are

[1] A sample point is reconstructed as the mean of the four nearest existing samples.

averages.

In our second experiment, we consider the image shown in Figure 2.3. This image represents a 2-dimensional function, whereby, the pixel grayscale values are the function values at each pixel location. A higher sampling rate is required in the parts of the image with dissimilar pixels because of the higher complexity. The image background is bland, and requires relatively fewer sample points. The image is of size $512 * 512 = 262,144,\ 8$ bits/pixels, and it is sampled at $16,384$ pixel locations indicating a ratio of $16:1$. The results of sampling adaptively based on ASHE are shown in the same figure. The performance is compared to the other sampling schemes as in the previous experiment. Note the efficient distribution of samples by the ASHE scheme as indicated by the cluster of sample points in the regions of high complexity - the face in the image. This has the required effect of a reconstructed image with better quality compared to the other sampling methods. Two numerical measures of image quality are used as a basis for comparison: the frequently used PSNR, and another measure of image quality called Structural Similarity (SSIM). This has been shown in [92] to be a better indicator of image quality than the PSNR. Both measures show that the image reconstructed from the adaptively sampled points based on the ASHE algorithm has the best quality. These numerical measures of image quality are supported by the better representation of the facial features in the image reconstructed from adaptively sampled points. The relationship between sample point positions and the resulting histogram of function values, in this case pixel grayscale values, is the same as those in the previous examples. That is, the histogram of the sampled grayscale values using adaptive sampling tends closer towards a Uniform distribution than the other sampling methods. A total of 100 experiments are conducted. The indicated PSNR, SSIM, and deviation values are averages of these experiments.

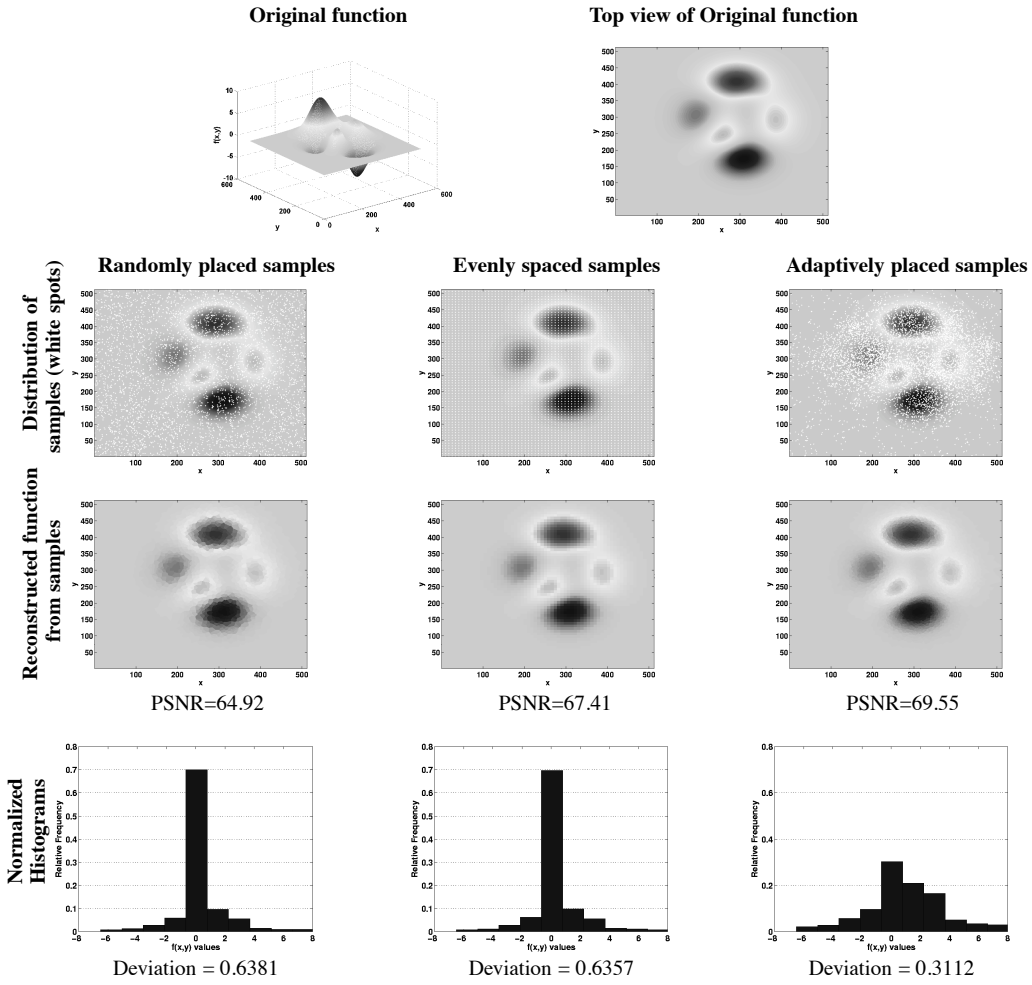


Figure 2.2: Performance comparison of the adaptive sampling algorithm to randomly placed, and evenly spaced samples for sampling a 2-dimensional function with varying slope. There are 512 values for both x, and y, yielding 262144 values. The function is sampled at 4096 locations, and reconstructed using the nearest neighbor method. Hereby, the Peak Signal to Noise Ratio (PSNR) is chosen as the objective measure of the quality of the reconstructed function. The Deviation values are defined as the mean squared deviation of the histograms from a normalized uniform distribution with the same number of histogram bins.

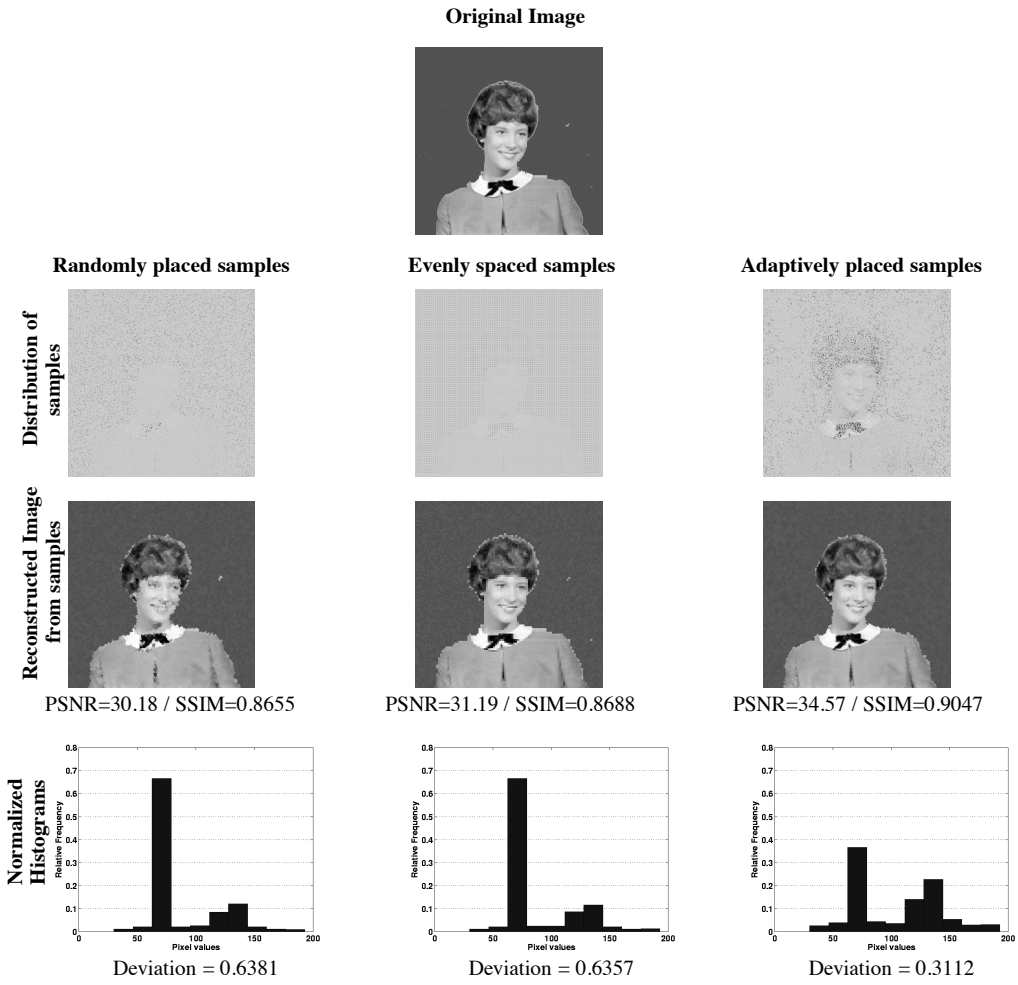


Figure 2.3: Performance comparison of the ASHE algorithm to randomly placed and, evenly spaced samples for sampling a 2-dimensional grayscale image. The original $512 * 512 = 262144$ pixels are sampled in 16384 locations, a ratio of $16:1$. The images are reconstructed using the nearest neighbor method. PSNR, and structural similarity (SSIM) are objective measures of the reconstructed image quality. The Deviation values are the mean squared deviation of the histograms from a normalized Uniform distribution with the same number of histogram bins.

2.4 Application Areas

As stated earlier, the examples in the previous section only serve as illustrations of the concept, and are not practical application areas. Based on the advantages of the ASHE algorithm stated in Section 2.1, we have identified some practical areas of applications. The following discussions are fairly generic, it should be straightforward to adapt the algorithm to specific problems.

2.4.1 Data Synthesis

Many forms of data analysis require adequate, and statistically representative population of the dataset in question to be able to make reliable inferences, and draw general conclusions. The problem in many fields of study is that the amount of available data does not fulfill the these stated requirements. Cost, time, and other limitations on resources may be prohibitive to the collection of the data. This problem has been solved in many instances by generating synthesized data. The requirement of statistical representation is usually that of maximum diversity in the dataset i.e., an equal representation of all possible members of a population. Maximum diversity is required in sample data in order to ensure that results from such are representative of the entire domain. Analysis of such data can then lead to inferences and conclusions that take all possible output scenarios into account. As demonstrated in Section 2.2, sampling to maximize diversity i.e., equalized distribution of obtained samples results in more efficient sampling. An image synthesis application that employs the ASHE algorithm is the subject of Chapters 5 - 7 of this dissertation.

2.4.2 Design of Experiments

Experimental results are usually functions of various factors. For example, a chemical reaction or biological process may depend on such factors as temperature, pressure, presence of catalyst or other reagents. It is usually required to determine the results of such experiments over a range of factors. It may be expensive or impractical to perform these experiments over all possible ranges and combinations of these factors. The results of such experiments or processes can be modeled as a multi-dimensional function with each dimension being one of the factors. Usually, there is no *a priori* information on the global, and relative levels of local variation of the outputs from these experiments. Regions of change due to a factor or combination of factors are usually of interest in these experiments. With the constraint on the number of experiments, it will be beneficial to perform more experiments in regions of this multi-dimensional space where there is relatively more change in the experimental results. This space can then be progressively sampled using the ASHE algorithm, whereby each subsequent sample location, that is, combination of factors for which the experiment is performed is determined by the current distribution of already obtained samples.

2.4.3 Surface Reconstruction from Expensive Samples

Computational Physics, Chemistry and Biology involve studies that require the computation of surfaces representing various phenomena. An example of this is the computation of the potential surface of a molecule in a particular electron state using first principles, that is, ab initio computations. These surfaces are usually multi-dimensional, and are constructed from the computed phenomenon values at various sample points in the space. In many cases, obtaining these values at each sampled point is computationally expensive. Also, no *a priori* information on the global or local variation on this surface is

available, only the ability to compute the surfaces' value at each sampled point. To minimize the surface reconstruction error from the points at which the surface values have been calculated, or to adequately represent regions of transition, it is required to compute relatively more values in regions where there is more change in the surface values. The ASHE algorithm can be used to determine the sample points where the surface values are to be computed by ensuring efficient variable sample rates.

2.4.4 Progressive Transmission/Rendering

Transmission of image data on a limited bandwidth channel can be effectively achieved by progressively sampling the image using the ASHE algorithm. This results in the more important information from the image being transmitted earlier. With this approach, truncating the data stream will only lead to the loss of the less important part of the information stream needed for reconstruction. The approach yields similar results in image rendering, with the more important region being rendered earlier such that a profile of the image is quickly represented. Similar work has been done with a different approach in [42].

CHAPTER 3

MODELS UTILIZED IN IMPLEMENTING ASHE

In this chapter, we discuss the three models employed in implementing the function values equalization described in the Adaptive Sampling by Histogram Equalization (ASHE) algorithm. All the described models are generally utilized in many areas of optimization, especially for problems in which directly obtaining optimal solutions is infeasible due to the computational cost. For these problems, the models are used to obtain variables that minimize or maximize a function. We apply variants of these models to obtain an efficient sample distribution as described in the ASHE algorithm. Though somewhat different, our problem may also be seen as an optimization problem, in which we intend to maximize the efficiency of the sample distribution. This efficiency is defined based on an objective measure. The set of sample points obtained by these algorithms constitutes a set of solutions.

The underlying concept that each model is based on is presented. Any variations or modification of the general form for our specific purpose is stated, and justified. General examples are presented to aid in the understanding of these models. The specific details of implementing the ASHE algorithm with each model are then described. Examples to illustrate the ASHE implementation with each model are, however, delayed until the discussion on their performance and sensitivity analysis in Chapter 4.

3.1 Active Walker Model

The active walker model can be explained both in the framework of the motion of Brownian particles, and as a variant of the random walk [72]. The simple Brownian motion

will not result in any structure required to model the systems to be studied. Therefore, Brownian particles "*with the ability to generate self-consistent fields, which in turn influence their subsequent movement, physical and chemical behaviors*" are introduced [75]. These are called *active Brownian particles*. The term *active walker* was first introduced in this work [29], in which a discrete approximation of the motion of these active particles was used to model a complex system. The active walker model has been used to simulate, and analyze numerous complex systems in both the physical and life sciences [48, 49, 46, 79, 78, 37, 38].

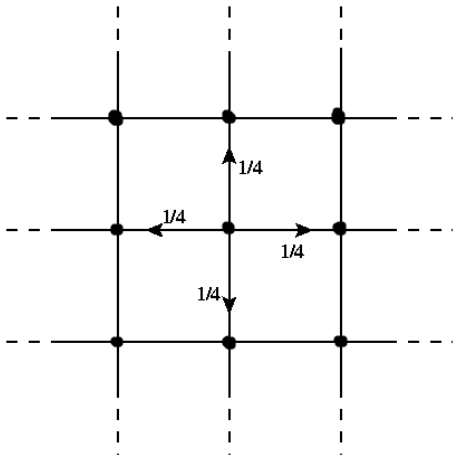


Figure 3.1: Simple symmetric random walk on $\mathbb{Z}^2$

The random walk approach is, however, the more appropriate of the two frameworks to explain our specific use of the active walker model. Consider the simple, symmetric random walk on $\mathbb{Z}^2$ shown in Figure 3.1 [72]. From the starting central position, the random walk can be seen as a specific case of a Markov-chain [63], in which the transition probabilities are given as follows:

$$p_{ij} = \begin{cases} \frac{1}{4} & \text{if } \|i-j\| = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

Each step taken in a random walk is discrete, and of equal sizes. The chain starts with an equal probability of moving into any of the vacant positions. The generalization of this to a higher-dimensional space is trivial. In their most general form, active walkers are pseudo-random walkers with the following properties:

- They take discrete, but not necessarily equal step sizes.
- The direction of their movement may be either random or biased.
- Their step sizes and direction of movement may depend either on local information contained in their current location, or global information in the walking space. They may also depend on a combination of these.
- In the case of multiple active walkers, the behavior of each walker may depend on peer interaction.
- Multiple active walkers can not occupy the same location at the same time.

All the variable properties, such as the movement, are governed by defined fitness criteria. For example, in a situation where moving charged particles are simulated, the distance and direction moved by an active particle may depend on the charge carried by the particle and those in its vicinity [79]. The result is a pseudo-random walk, which is biased based on the fitness criteria. The adaptive or biased random walk approach has been used in solving optimization problems [12].

3.1.1 ASHE Implementation using Active Walkers

For our specific application, simulated active walkers are employed to implement the ASHE algorithm by placing sample points in the location of the walker in the sample space. Initial samples are obtained by placing even spaced active walkers in the space.

A histogram is formed from the function values obtained from these initial locations. We establish a fitness criterion based on the state of the histogram. Whereby, after each sample addition, the normalized histogram of samples is updated, and compared to a normalized Uniform distribution with the same number of bins. The comparison of the two histograms results in a fitness criterion FC which is given by:

$$FC = \sqrt{\sum_{i=1}^{n} (\hat{h} - h_i)^2} \,, \tag{3.2}$$

where n is the number of bins in the histograms, h_i are the relative frequencies from the sample distribution, and $\hat{h} = \frac{1}{n}$ is one of the equal valued relative frequencies from the Uniform distribution. FC has a lower bound of zero. A decrease in the value of the FC indicates that the newly added sample moved the histogram of samples closer to the normalized Uniform distribution. The active walker that obtained the sample then moves a *short* step in order to sample more in its current vicinity. An increase in the value of the FC due to the addition of a sample, indicates a deviation of the distribution of samples from the Uniform distribution. The active walker that obtains the sample resulting in the deviation is made to sample in a location different from its current vicinity by taking a *long* step. The definitions of the terms "short", and "long" steps will be addressed in detail in Chapter 4, while discussing the sensitivity analysis. The distance between the current and subsequent locations of a walker is determined as the resultant of vector lengths along each dimension. The direction of each vector is randomly chosen, independently, resulting in a random direction for the resultant vector. This process is then continued to progressively sample the space until the required number of samples is obtained. Multiple active walkers are usually employed to ensure that the entire sample space is covered. A variant of the self-avoidance mechanism is also included to

ensure that a location is not sampled multiple times [72]. In summary, the active workers employed to implement ASHE have the following specific behaviors:

- The position of an active walker represents a sampled location.
- There is no cost associated with the distance moved by an active walker.
- A self-avoidance mechanism is implemented to ensure that a location is not sampled multiple times.
- The next location of an active walker is dependent on its current location, and the step size adapting criterion. The adapting or fitness criterion is the change in state of the distribution of function values.
- Their direction of movement is random.

Algorithm 1 shows the implementation of ASHE using the active walker model.

Algorithm 1 . Active walker model implementation of the ASHE algorithm

Initial definitions:

Objective function, e.g. function value
Variables/factors the objective function is dependent on
Range and possible values that all variables/factors can take

Sampling initialization:

Obtain initial randomly located samples using active walkers
Compute/obtain the objective function values from initial sample points
Compute normalized histogram from initial sampled function values
Compute Overall Fitness Criterion OFC

while Sample points $\leq$ required no. of samples **do**
for all Active walkers **do**
Obtain new sample point
if Location has already been sampled
Obtain closest unsampled location
(random choice if multiple unsampled locations exist at same distance)
end if
Add new sample from active walker to existing samples
Compute new normalized histogram after single addition, and
Compute New Fitness Criterion NFC
if $NFC < OFC$
Single walker takes *short* step size in random direction
else Single walker takes *long* step size in random direction
end if
end for
Compute new overall normalized histogram
Compute OFC
end while

3.2 Ant Model

This model is based on the behavior of ants in search of resources, usually food. Many insect species deposit a substance called pheromones when walking to or from food sources [40, 32]. The role played by this mechanism on their ability to efficiently search for food has been studied [31, 14]. Many of these insects, for example ants, possess little or no sense of sight, and communicate primarily through their sense of smell. Consider the simple case of an ant colony and a resource, say food, as shown in Figure 3.2. The ants can reach the food by either of the two paths, with one being much longer than the other. The ants deposit pheromones as they traverse these paths. Assume an initial random access of the paths, and also assume that the effect of deposited pheromone spreads, and fades with time, due to a diffusion process. The pheromone update over time, is generally modeled as:

$$\tau_{ij} \longleftarrow (1-p).\tau_{ij} + \sum_{k=1}^{m} \Delta\tau_{ij}^{k} \,, \tag{3.3}$$

where i and j are endpoints of the path, p is the evaporation rate, m is the number of ants, and $\Delta\tau_{ij}^{k}$ is the quantity of pheromone laid on the path (i, j) by ant k. It is expected that the pheromone concentration on the shorter path will be refreshed more often, therefore, the maintained concentration level will be higher. This will attract more of the ants into using this path to get to the food. Hence, the optimal route to the resource is established. Probabilistic models of these kind of behavior have been developed, and they show that the initial equal probability of taking either path is updated to increase the probability of the shorter path [31]. In summary, the colony of ants communicate indirectly to reinforce a good solution by modifying their environment through positive feedback.

This kind of communication is referred to as *stigmergy* [39]. When compared to other

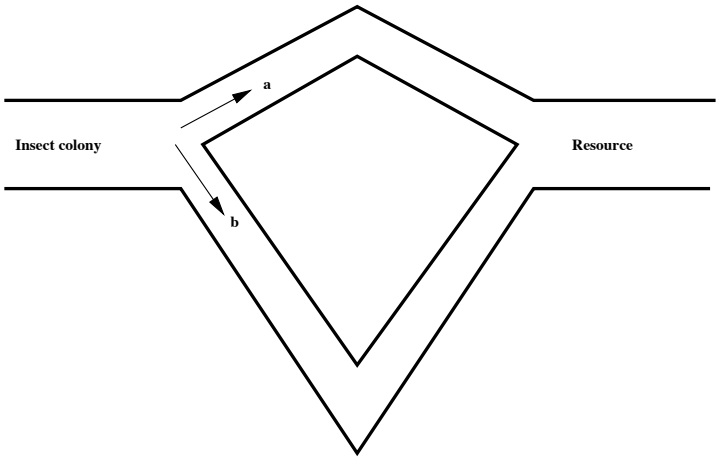


Figure 3.2: Illustration of 2 possible paths for ants to reach a resource. With path 'a' shorter than 'b'

forms of communications, it is noted to have two unique characteristics [17]. These are:

- It is indirect, and non-symbolic, insects communicate by modifying their environment.
- It is local, that is, the information can only be accessed by insects that visit the vicinity of the pheromone *footprint*.

Algorithms based on the ant model have been successfully used in solving numerous optimization problems [19, 86, 55]. These fall under the generic name of Ant Colony Optimization (ACO) algorithms. Some variants of this are the Ant System (AS) [20], and the Ant Colony System (ACS) [18]. A good review of the progress in this area of study can be found in [17]. A basic assumption made by all these algorithms is that the ants live in an environment where time is discrete.

3.2.1 ASHE Implementation using the Ant Model

In implementing the ASHE algorithm, the described 1-dimensional path model is extended to multiple dimensions. Whereby, ants *forage* in the space to be sampled, and samples are obtained from their current locations. At the start of the sampling process, each sample location is allocated equal probability of being foraged. The probabilities are modeled as pheromone concentrations. The sampling is done in discrete time intervals. The *resource* is the obtained sample. Whether a positive feedback is sent by an ant depends on the change in the state of the distribution of the already obtained samples. The same fitness criterion FC defined in (3.2) for the active walker model is used. In this case, a positive feedback is only generated if the obtained sample moves the updated distribution closer to the Uniform distribution. The ant modifies its environment by depositing pheromones in the multidimensional vicinity of where it obtains a *good* sample, indicated by a reduction in the value of FC. The effect of this deposit is a relative increase in the probabilities associated with the sample locations in the vicinity of this sample. The questions concerning the amount of increase, and the extent of the spread will be addressed in the performance and sensitivity considerations in Chapter 4. The increase is highest in the locations nearest to where the sample was obtained, and tapers down at a non-linear rate. All locations that are already sampled are allocated a zero probability to avoid multiple sampling of the same location. At each discrete time step, the probabilities associated with the unsampled locations are reduced by the same factor. This is to simulate the process of the pheromone evaporation with each time step. Subsequently, the ant samples in its vicinity in a manner that reflects the probabilities associated with the sample locations. That is, where there are multiple non-zero probability locations of the same distance, it samples the location with the highest probability. Otherwise, it samples in the only non-zero probability location in its vicinity. The defini-

tion of vicinity will be clarified in the discussions on sensitivity analysis. It is expected, that the effect of the feedback created by an ant generally goes farther than its movement in any one time step. The ant model implementation of the ASHE algorithm is shown in Algorithm 2.

Algorithm 2 . Ant model implementation of the ASHE algorithm

Initial definitions:

Objective function, e.g. function value
Variables/factors the objective function is dependent on
Range and possible values that all variables/factors can take
Associate equal probabilities with all sample locations

Sampling initialization:

Obtain initial samples from location of randomly placed ants
Compute/obtain the objective function values from initial sample points
Compute normalized histogram from initial sampled function values
Compute Overall Fitness Criterion OFC

while Sample points $\leq$ required no. of samples **do**
  for all ants **do**
    Obtain new sample point from non-zero probability location in *vicinity*
      if Multiple non-zero locations exists
        Obtain sample from location with highest associated probability
      end if
    Add new sample from ant to existing samples
    Set probability associated with sampled location to zero
    Compute new normalized histogram after single addition, and
    Compute New Fitness Criterion NFC
      if $NFC < OFC$
        Increase probabilities associated with locations around sample
        by values that sum up to 1 in the surrounding sample locations
      end if
  end for
    Multiply probabilities associated with sample locations by constant < 1
    Normalize probabilities based on non-zero locations in the space
    Compute new overall normalized histogram
    Compute OFC
end while

3.3 Evolutionary Algorithm Model

The class of evolutionary algorithms (EAs) includes genetic algorithms [61], genetic programming [47], evolutionary strategies [30], and evolutionary programming [28]. Common to all the variants of this class of algorithms are elements of the principles of natural biological evolution. They operate on populations based on the principles of survival of the fittest. Population members deemed to be best suited for surviving in a particular environment form the basis for creating the next generation. Evolutionary algorithms model this natural process by applying principles like recombination, mutation, and migration. A good introduction to evolutionary algorithms [3] addresses these basic concepts.

EAs have been used variously to solve search and optimization problems [59, 84]. In general, they consider a population of possible solutions, and remove the poor solutions based on some fitness criterion. The surviving population members then form the basis for producing a new generation. The new generation is produced primarily by combining surviving members. The rationale for this is that combining elements from fit members will result in even fitter members. A mutation process is also used to generate new members. This is an occasional perturbation that results in a new member, whose properties are not completely accounted for by any existing member. The population size may be kept constant or varied over the generations. This death, survival, and mutation process continues until an acceptable solution is obtained.

3.3.1 ASHE Implementation using the Evolutionary Algorithm Model

In implementing the ASHE algorithm using an evolutionary algorithm approach, we start the sampling process in even spaced locations. The coordinates of these locations serve

as the population in the first generation. Again, the fitness criterion is the same as in the other models, and stated in (3.2). That is, the change in the FC value due to the addition of a sample from a location determines whether it is considered fit, or not. Based on the prior discussions, a sample from a location that results in the reduction of the FC value is considered fit. The recombination process used in the standard genetic algorithms will not necessarily produce a new fit member. To illustrate this point, consider two fit members of a current generation. For simplicity, let us assume an even number of dimensions in the space to be sampled, say two. The population members consist of an ordered pair of integers. Recombination between two fit members will yield new offspring that do not necessarily have any relation to the parents. For our purpose, recombination between two fit members may yield new locations that have no bearing on the original location. This is shown in Figure 3.3.

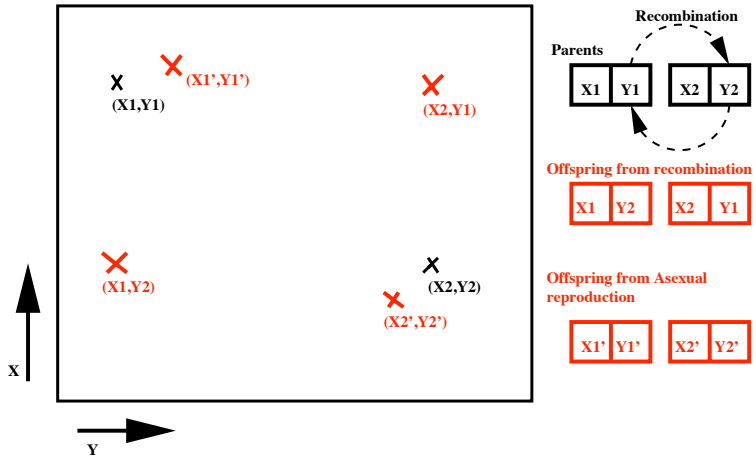


Figure 3.3: Illustration of standard genetic algorithm recombination process, and an Asexual process better suited for implementing ASHE.

A different approach of producing a new generation has to be taken. This is so because the ASHE algorithm requires that further sampling is done in the neighborhood of a current good solution as indicated by the fitness criterion. We take the approach of

making the fit population member reproduce in an asexual manner. That is, a fit member produces offspring in its current vicinity. The number of offspring, and how far their neighborhood can be from their parents, are variables that will be discussed under the performance and sensitivity analysis. A very small probability that the offspring of a fit parent may die is also introduced. All unfit parents die without offspring, but the sample from their current location is accepted. A random selection of the starting population size is made from the offspring to ensure that the population size remains the same for every generation. Algorithm 3 shows the evolutionary algorithm model implementation of ASHE.

Algorithm 3 . Evolutionary algorithm model implementation of ASHE

Initial definitions:
Objective function, e.g. function value
Variables/factors the objective function is dependent on
Range and possible values that all variables/factors can take

Sampling initialization:
Obtain initial randomly located samples/starting population
Compute/obtain the objective function values from initial sample points
Compute normalized histogram from initial sampled function values
Compute Overall Fitness Criterion OFC

while Sample points $\leq$ required no. of samples **do**
for all Members in current generation **do**
Obtain new sample point
if Location has already been sampled
Obtain alternate, close sample point
end if
Add new sample from population member to existing samples
Compute new normalized histogram after single addition, and
Compute New Fitness Criterion NFC
if $NFC < OFC$
Parent reproduces/divides (Asexual reproduction) in its vicinity
Generate random number S between 0 and 1
if S $>$ prob. of survival ($>= 0.7$)
Offspring dies
end if
else Parent dies
end if
end for
Make random selection of starting population size from offspring
Compute new overall normalized histogram
Compute OFC
end while

CHAPTER 4

PERFORMANCE AND SENSITIVITY ANALYSIS OF MODELS

Here, we establish two measures of the performance of the models in the efficient distribution of sample points. In contrast to the basis of comparison in Section 2.3, these are independent of any reconstruction algorithm. The measures here are based on the entropy measure of information [34, 82], and the Nyquist-Shannon minimum sampling rate for band-limited signals [67, 56]. These serve as an indication of the relative levels of variation in a sampled function. For both measures, a high value will signify more complexity, thus requiring a relatively higher sample density. The purpose of the ASHE algorithm is to efficiently distribute sample points. That is, adapt the sample density such that they reflect the local, and global levels of variation in the space being sampled. The defined objective measure of performance is thus based on the correlation between these indicators of variation and the sample density. A high correlation between the sample density and either of the two measures will signify good performance of the sampling scheme, and vice-versa.

Based on the established performance criteria, we carry out a performance and sensitivity analysis of the models. The sensitivity analysis seeks to investigate how such performance varies with change in factors in the different models used in the ASHE implementation. The results from these are presented, and conclusions are drawn where appropriate.

4.1 Measure Based on Frequency Content

The Fourier transform (FT) is used to decompose a function in time or space into its sine and cosine components of different frequencies. The space or time varying function can then be represented in terms of its frequency components. This is called a frequency domain or spectral representation of the function. We use 2-dimensional functions, specifically 2-dimensional digital images to illustrate the use of this transformation in establishing a measure of complexity. Since we are considering digital images, we will further restrict the discussions to the Discrete Fourier Transform (DFT), which is a sampled version of the continuous FT. Consider an image of size $M \times N$, then the DFT is given by

$$F(k,l) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m,n) e^{-2\pi i (\frac{mk}{M} + \frac{nl}{N})} , \tag{4.1}$$

where $f(m,n)$ is the image in the spatial domain, and the exponential term multiplying it is the basis function corresponding to each $F(k,l)$ in the frequency domain. The basis functions are sine and cosine waves with increasing frequencies from $F(0,0)$, which is the DC component[1], to a maximum of $F(M-1, N-1)$. The DC component represents the average brightness in the image. The resulting Fourier transform is complex, containing the real or magnitude, and imaginary or phase components. The size of each component is the same as that of the original image.

Figure 4.1 shows two images and the corresponding DFT magnitude images. For simplicity, the images contain single spatial frequencies. The original DFT magnitude images have the DC components at the edges and the highest frequency components in the middle. These have been shifted using the MATLAB fftshift function so that the DC

[1]The zero frequency component of a signal is also referred to as the Direct Current - DC component.

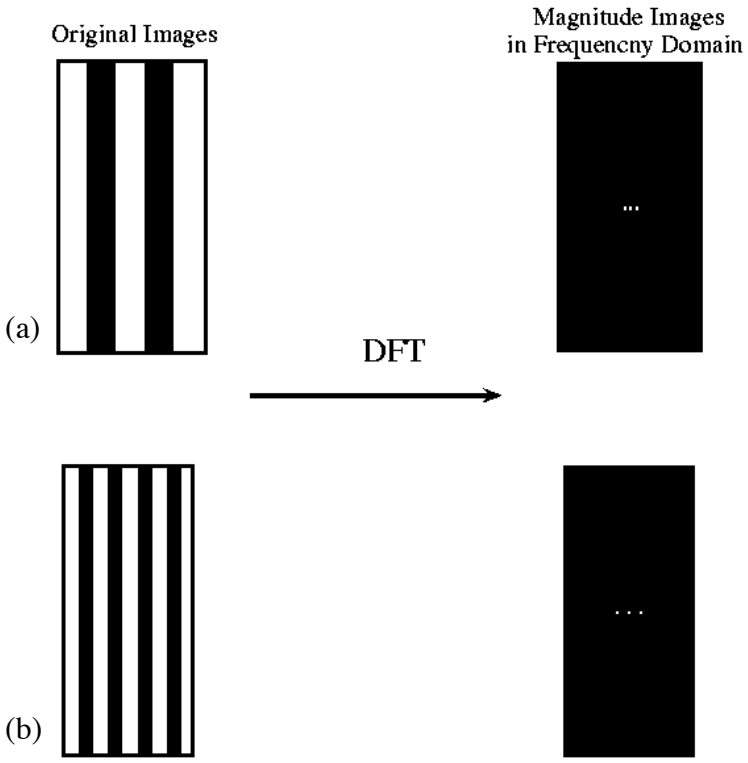


Figure 4.1: Shows two images of single spatial frequencies, and the corresponding magnitude images in the frequency domain. The image in (a) has a lower frequency than that in (b), this is reflected in the distance of the fundamental frequencies from the DC component.

components are in the middle, and the highest frequency components are at the edges. The pixels in the middle of the DFT images are the brightest, indicating that the images are dominated by their DC components. Note the two bright spots on either side of the DC component. These represent the fundamentals of the single spatial frequencies, and are mirror images of each other. Their distances from the DC component is an indication of the frequency they represent, the higher the spatial frequency, the bigger this distance. This explains why the distance is larger in the DFT from the image with the higher spatial frequency. To aid in the general interpretation of DFT images, consider the cross-sections of the power image of 2-dimensional DFTs as shown in Figure 4.2. Figure 4.2(a) gives a simplistic representation of a single frequency image in the frequency domain. Figure 4.2(b) shows multiple power plots, each representing multiple frequencies.

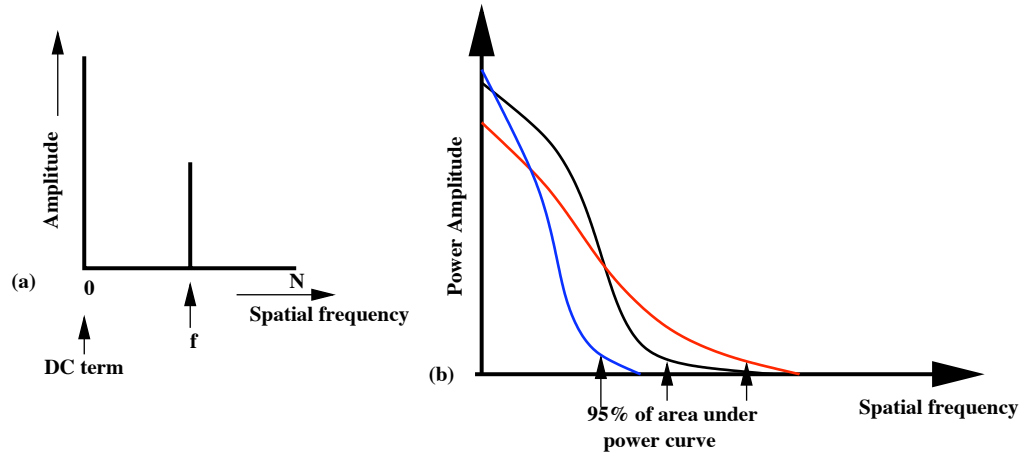


Figure 4.2: Cross-sections of images in frequency domain:(a) simplified frequency representation, (b) multiple plots showing different power spread in their spectrum

Our metric is based on the power distribution in the frequency components of the functions in question. We determine this by summing up the power in the spectrum, starting from the lowest to the higher frequencies until a value of 95% of the total power in the spectrum is obtained. We do not use 100% of the power since the function may have

support in the entire frequency domain. The DC component is also excluded to avoid a bias due to differences in function amplitudes. Generally, the higher the distance, that is spatial frequency at which this value is attained, the more the high frequency components in the function. We divide the function into 16 regions of equal sizes, and compute this value for each region. Our objective measure is the correlation coefficient (CC) between the relative sample densities in the regions and the values indicating the frequency content. The correlation coefficient takes values ($-1 \leq CC \leq 1$). A positive CC value indicates that the sample densities are higher in the regions where the function contains high frequency components. The higher the positive correlation value, the more efficient the sample distribution obtained by the employed model.

A drawback of using the DFT approach is the classic time/spatial versus frequency resolution trade-off. That is, computing the Fourier transform over a small time/spatial window will result in poor frequency resolution but good time resolution. Increasing the window size improves the frequency resolution at the expense of the time resolution. The frequency resolution is more important for the application. The window size we can use is constrained because the function has to be sub-divided into regions. We apply some zero-padding in the spatial domain before the DFT transformation. The zero-padding does not improve the frequency resolution, but it does increase the sampling rate, leading to appreciable improvements in our results.

Figure 4.3 shows two of out test functions, an image, and a 2-dimensional energy function. The functions are divided into 16 regions for the purpose of obtaining the frequency based objective measure by region. Also shown are the corresponding complexity measure images, in which the shades indicate the level of complexity. Note the general visual correlation between the frequency-content based measure, and the apparent regions of complexity in the functions.

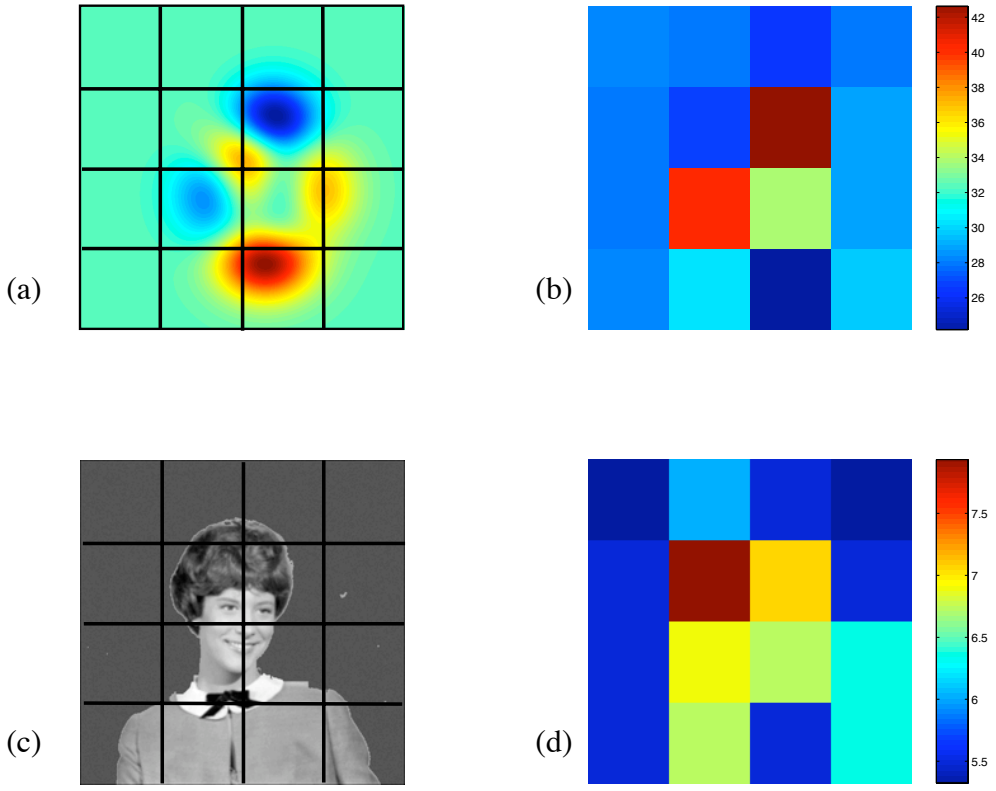


Figure 4.3: Functions divided into 16 equal regions (a) and (c), the corresponding images based on frequency content associated with the regions (b) and (d) respectively.

4.2 Measure Based on Entropy Measure of Information

Entropy as a measure of information content in a discrete system is defined as:

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i) \tag{4.2}$$

where X is a discrete random variable that can take possible values $x_1, x_2, ..., x_n$, and $p(x_i)$ is the probability that X takes the value x_i. The concept may be understood intuitively in terms of uncertainty. If the outcomes $x_1, x_2, ..., x_n$ are equally probable then uncertainty is high, and the entropy is maximal. If, however, an outcome is certain, the entropy is zero, which indicates that no additional information is obtained from the outcome.

For our purpose the function in question is divided into 16 equal sized regions as described for the frequency based measure. The entropy of each region was computed. Higher entropy values indicate greater uncertainty or complexity in the functions. An objective measure was obtained by computing the correlation coefficient between the entropy values and the relative sample densities. Again, higher positive correlation values indicate a more efficient sample distribution. Figure 4.4 shows the example functions and the corresponding entropy images. The entropies associated with each region are represented by the color shades. Note the visual correlation between the region of complexities in the functions, and their corresponding entropy shades.

4.3 Analysis of the Active Walker Model

We experiment with two functions: a smooth energy function, size 512×512 pixels, and a 2-dimensional image of size 256×256 pixels, to represent rapidly varying functions. We obtain a total of $4,096$ sample in each case, representing a ratio of $1 : 64$, and

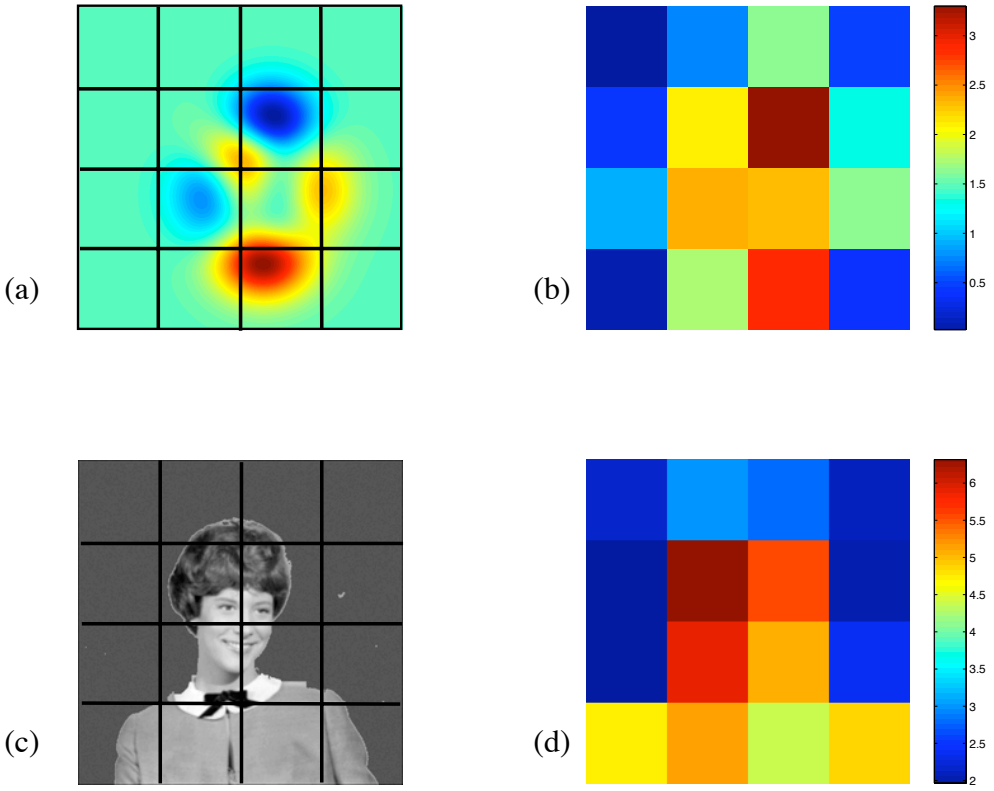


Figure 4.4: Functions divided into 16 equal regions (a) and (c), the corresponding images based on entropy associated with the regions (b) and (d) respectively.

$1:16$ respectively. The two test functions are shown in Figure 4.5. We emphasize that these test functions are used for the purpose of experimentation alone. The key elements expected in the areas of practical application are not present in them. That is, obtaining the sample values is not expensive, and we have complete prior information on these functions. They however, fully meet the requirement for these experimental purpose.

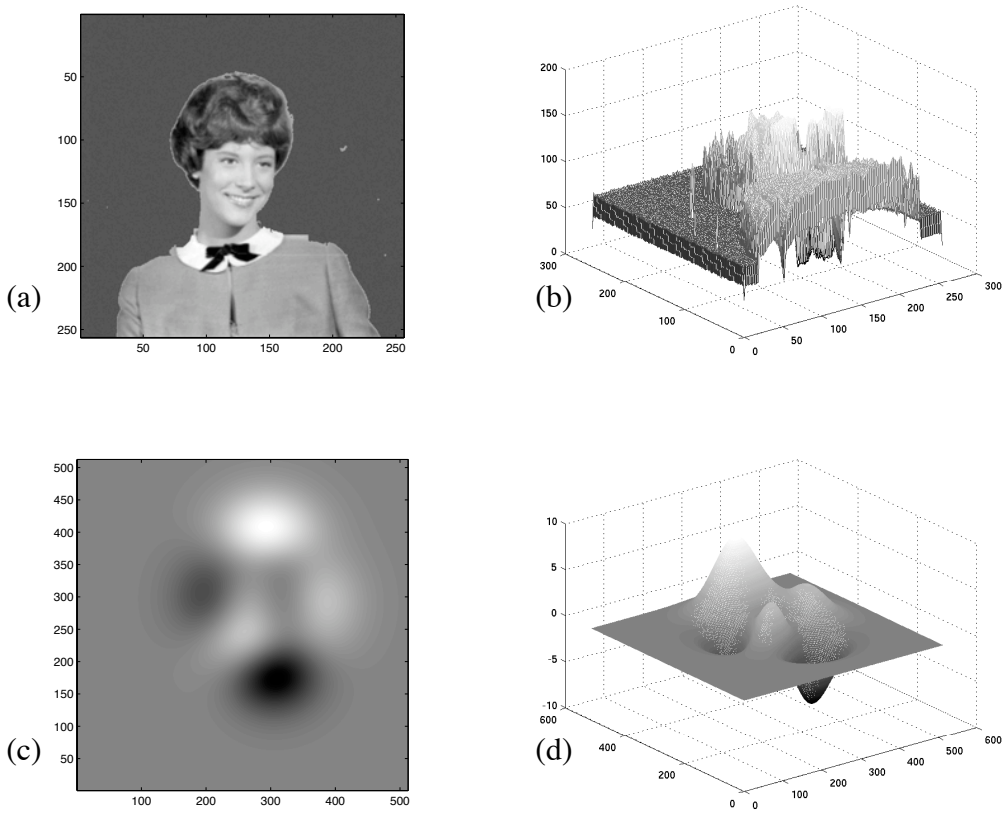


Figure 4.5: Test functions, top views (a) smooth function, (c) rapidly varying function, and their corresponding side views (b) and (d) respectively.

We identify four key factors that may affect the performance of the active walker model as described in Section 3.1.1. These are:

1. Number of active walkers (N_{aw}). This will determine the number of steps taken in the adaptive sampling process. Starting with n walkers means that $(4,096/n) - 1$ steps will be taken.

2. Number of bins (N_{ub}) in the histogram to be equalized. This is important because it constitutes a form of *resolution*. It is the number of unique groups in which the sample function values are divided.

3. Large step size (LSP). This is the step taken away from a vicinity due to an increase in the fitness criterion value. We define this as a function of the size of the space being sampled.

4. Small step size (SSP). This is the step taken in order to sample in the current vicinity of an active walker, due to a decrease in FC, signifying an improvement. This is also defined as a function of the size of the space being sampled.

We experiment with the following values of these factors:

$N_{\text{aw}} = \{4,\ 64,\ 100,\ 144,\ 1,024\}$

$N_{\text{ub}} = \{2,\ 8,\ 16,\ 32,\ 256\}$

$LSP = \{0.2,\ 0.25,\ 0.3,\ 0.35,\ 0.4\}$

$SSP = \{0.02,\ 0.04,\ 0.06,\ 0.08,\ 0.1\}$

Both LSP, and SSP are obtained by multiplying the vector containing the sizes of the dimensions of the space with these numbers. For each experiment, the function is sampled using the active walker model with a combination of these factors. One complete set of experiments thus includes 625 runs. In order to ensure adequate statistical representation, we run 100 complete sets, that is $62,500$ runs in all. Figures 4.6 and 4.7 show examples of the progression of the adaptive sampling for both test functions. In both examples, we used the following values in the active walker model: $N_{\text{aw}}= 64$, $N_{\text{ub}}= 16$, and $LSP= 0.4$. We used $SSP= 0.06$ for the smooth function, and $SSP= 0.02$ for the image. Note that nine images from intermediate stages are shown instead of the total of 64 expected in the sampling process with $N_{\text{aw}}= 64$.

In both examples, the sample density changes to reflect the complexity in the sampled function as the process progresses.

We sort the results of the tests by the measures based on frequency content, and entropy. These results are shown in Tables 4.1 - 4.4. The tables show the results for the two test functions, sorted by the two measures of performance, hence the four tables. Due to space constraint, we only show the top and bottom 20 runs based on the sorting criteria. The important information to aid in our analysis is, however, all contained in the shown portion of the results.

Note that the correlation between the sample density and the frequency based measure is lower than that between the sample density and the entropy based measure. This is probably due to the limitations of the frequency based measure for this particular purpose, as stated in Section 4.1. The correlation between both measures is generally good as the sorted table indicates. That is, both objective measures essentially give the same information.

Tables 4.1 and 4.2 show the results for the experiments using the smooth function. The results show that the active walker model requires more than three steps to achieve a good solution. This is indicated by the number of appearances of $N_{\text{aw}} = 1,024$ at the bottom of the table. $1,024$ active walkers would only take $(4,096/1,024) - 1 = 3$ steps each to complete the sampling. The sample density distribution is still essentially random after three steps. Generally, more steps improve the ability of the adaptive process in efficiently distributing samples. This has to be balanced by the need to cover the sampling space, as too few active walkers may get stuck in a locality. The results show that for this application, at least about 25 steps will yield a good sample distribution. The number of active walkers should be chosen relative to the total number of samples to be obtained.

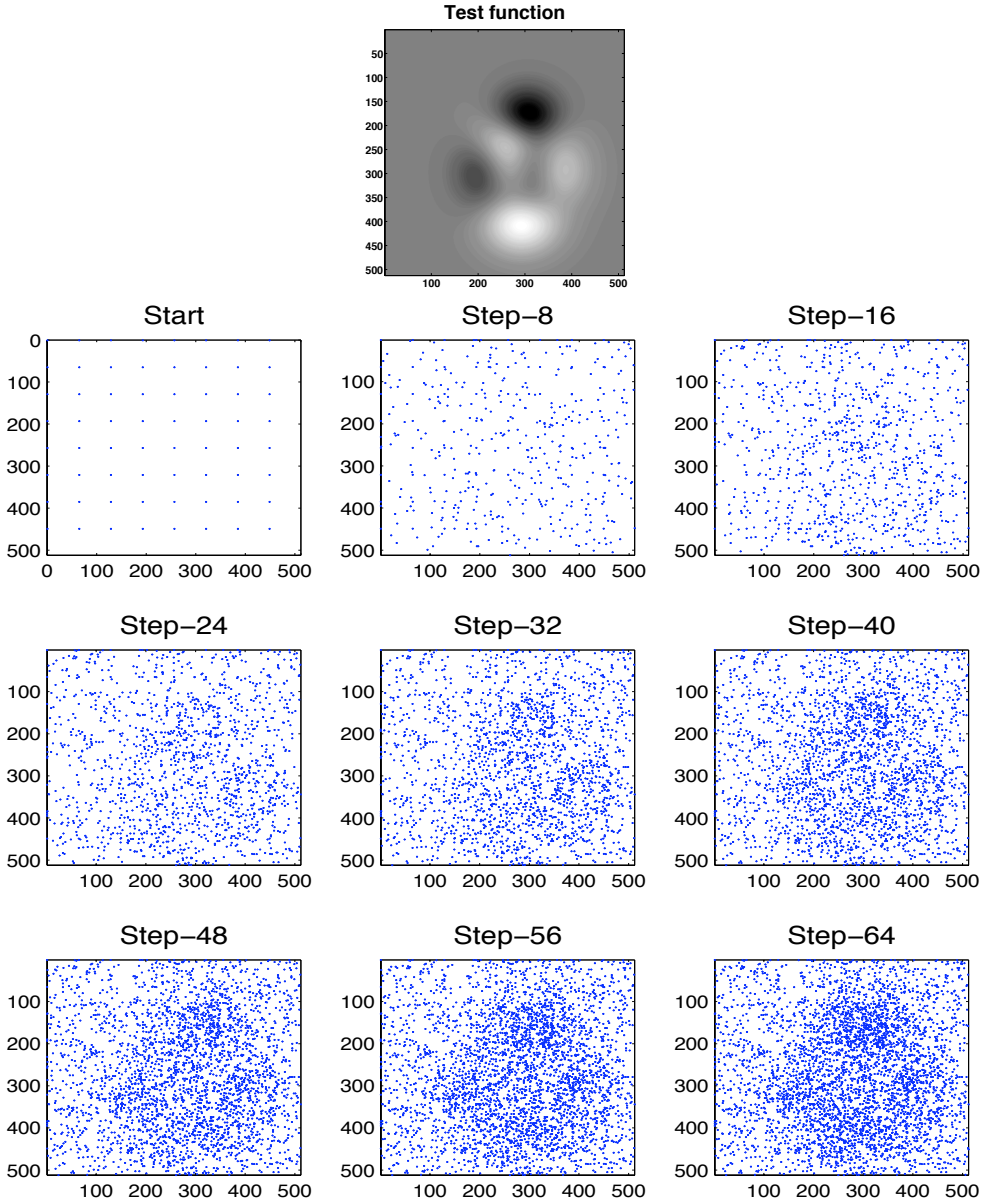


Figure 4.6: Intermediate steps from sampling a smooth test function with the active walker model using the following factors: $N_{\mathrm{aw}} = 64$, $N_{\mathrm{ub}} = 16$, $LSP = 0.4$, and $SSP = 0.06$. Total of 64 steps.

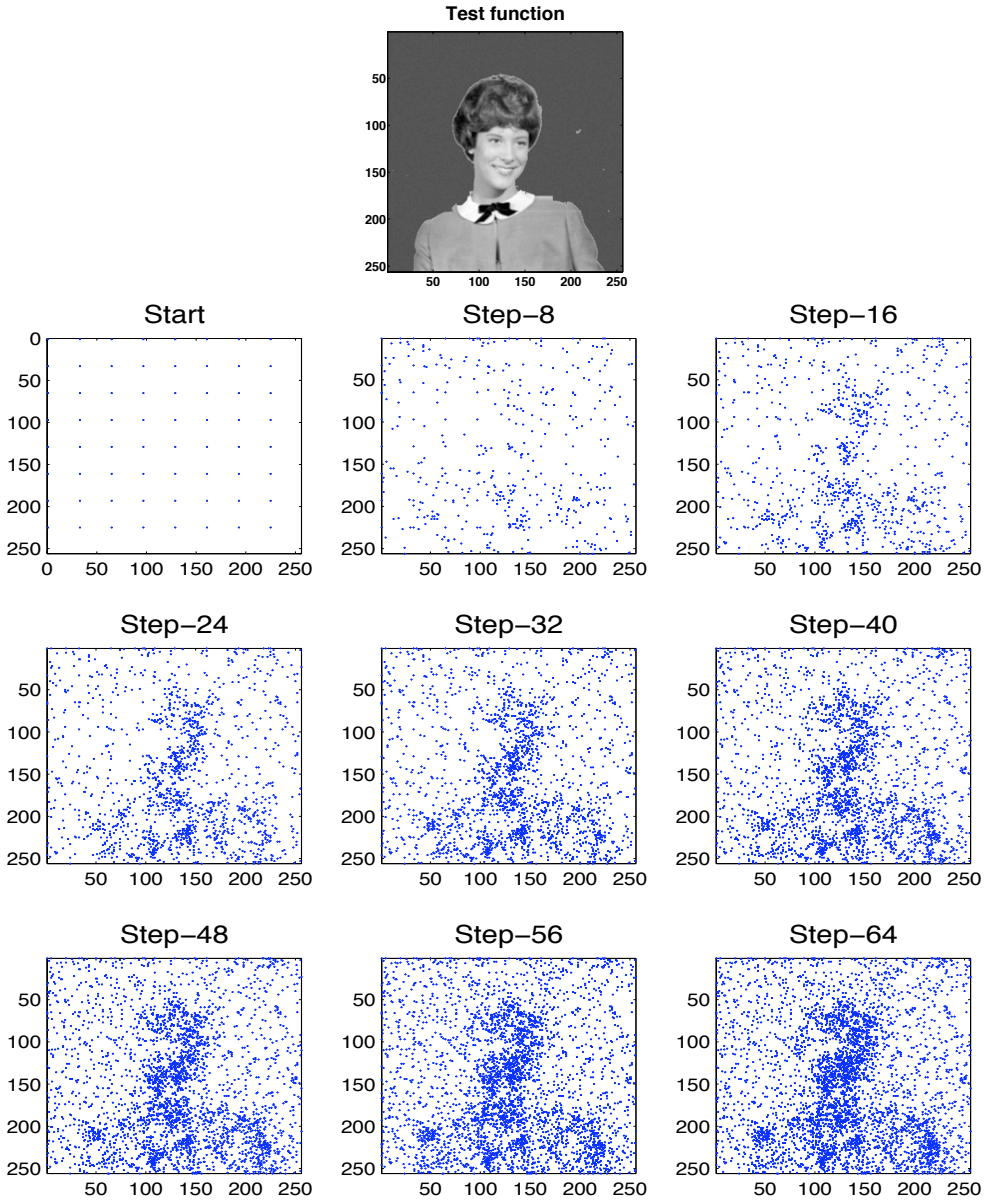


Figure 4.7: Intermediate steps from sampling a rapidly varying test function with the active walker model using the following factors: $N_{\mathrm{aw}} = 64$, $N_{\mathrm{ub}} = 16$, $LSP = 0.4$, and $SSP = 0.02$. Total of 64 steps.

Table 4.1: Performance of active walker model in sampling the smooth function, sorted by the frequency based measure.

No. of Active Walkers	No. of Bins	Long Step	Short step	CC with Freq. based Measure	CC with Ent. based Measure
100.00	16.00	0.40	0.06	0.60	0.93
144.00	16.00	0.40	0.06	0.60	0.93
144.00	16.00	0.35	0.06	0.59	0.92
144.00	16.00	0.40	0.08	0.59	0.91
4.00	16.00	0.35	0.06	0.58	0.91
4.00	16.00	0.40	0.08	0.58	0.91
144.00	16.00	0.35	0.04	0.58	0.92
64.00	16.00	0.40	0.06	0.57	0.93
100.00	16.00	0.35	0.08	0.57	0.87
64.00	16.00	0.40	0.08	0.57	0.91
100.00	16.00	0.35	0.06	0.57	0.92
100.00	16.00	0.40	0.08	0.57	0.90
64.00	16.00	0.35	0.06	0.57	0.92
144.00	16.00	0.30	0.04	0.57	0.91
4.00	16.00	0.40	0.06	0.56	0.92
64.00	16.00	0.30	0.06	0.56	0.90
4.00	16.00	0.30	0.06	0.56	0.90
144.00	16.00	0.35	0.08	0.56	0.88
144.00	32.00	0.35	0.08	0.56	0.89
144.00	16.00	0.40	0.04	0.56	0.92
:	:	:	:	:	:
:	:	:	:	:	:
:	:	:	:	:	:
1024.00	2.00	0.20	0.02	-0.26	0.14
64.00	16.00	0.25	0.10	-0.26	-0.35
1024.00	32.00	0.20	0.10	-0.30	-0.51
1024.00	16.00	0.20	0.10	-0.30	-0.48
144.00	256.00	0.20	0.10	-0.31	-0.70
64.00	256.00	0.20	0.10	-0.34	-0.74
100.00	256.00	0.20	0.10	-0.34	-0.70
4.00	256.00	0.20	0.10	-0.34	-0.73
144.00	8.00	0.20	0.10	-0.40	-0.74
64.00	8.00	0.20	0.10	-0.42	-0.76
4.00	8.00	0.20	0.10	-0.43	-0.76
100.00	8.00	0.20	0.10	-0.43	-0.73
100.00	32.00	0.20	0.10	-0.44	-0.80
4.00	32.00	0.20	0.10	-0.45	-0.83
64.00	16.00	0.20	0.10	-0.46	-0.79
64.00	32.00	0.20	0.10	-0.46	-0.81
144.00	32.00	0.20	0.10	-0.46	-0.81
4.00	16.00	0.20	0.10	-0.46	-0.80
144.00	16.00	0.20	0.10	-0.47	-0.77
100.00	16.00	0.20	0.10	-0.47	-0.80

Table 4.2: Performance of active walker model in sampling the smooth function, sorted by the entropy based measure.

No. of Active Walkers	No. of Bins	Long step	Short step	CC with Freq. based Measure	CC with Ent. based Measure
144.00	16.00	0.40	0.06	0.60	0.93
64.00	16.00	0.40	0.06	0.57	0.93
100.00	16.00	0.40	0.06	0.60	0.93
100.00	16.00	0.35	0.06	0.57	0.92
144.00	16.00	0.40	0.04	0.56	0.92
64.00	16.00	0.35	0.06	0.57	0.92
144.00	16.00	0.35	0.06	0.59	0.92
4.00	16.00	0.40	0.06	0.56	0.92
144.00	32.00	0.40	0.06	0.56	0.92
100.00	16.00	0.40	0.04	0.53	0.92
144.00	16.00	0.35	0.04	0.58	0.92
4.00	16.00	0.35	0.06	0.58	0.91
144.00	16.00	0.30	0.04	0.57	0.91
100.00	16.00	0.35	0.04	0.53	0.91
4.00	32.00	0.40	0.08	0.56	0.91
144.00	32.00	0.40	0.04	0.52	0.91
64.00	32.00	0.40	0.08	0.56	0.91
64.00	16.00	0.40	0.04	0.52	0.91
64.00	32.00	0.40	0.06	0.53	0.91
64.00	16.00	0.40	0.08	0.57	0.91
:	:	:	:	:	:
:	:	:	:	:	:
:	:	:	:	:	:
4.00	2.00	0.20	0.10	-0.10	-0.46
1024.00	16.00	0.20	0.10	-0.30	-0.48
64.00	2.00	0.20	0.10	-0.13	-0.48
1024.00	32.00	0.20	0.10	-0.30	-0.51
144.00	256.00	0.20	0.10	-0.31	-0.70
100.00	256.00	0.20	0.10	-0.34	-0.70
4.00	256.00	0.20	0.10	-0.34	-0.73
100.00	8.00	0.20	0.10	-0.43	-0.73
64.00	256.00	0.20	0.10	-0.34	-0.74
144.00	8.00	0.20	0.10	-0.40	-0.74
4.00	8.00	0.20	0.10	-0.43	-0.76
64.00	8.00	0.20	0.10	-0.42	-0.76
144.00	16.00	0.20	0.10	-0.47	-0.77
64.00	16.00	0.20	0.10	-0.46	-0.79
100.00	16.00	0.20	0.10	-0.47	-0.80
100.00	32.00	0.20	0.10	-0.44	-0.80
4.00	16.00	0.20	0.10	-0.46	-0.80
144.00	32.00	0.20	0.10	-0.46	-0.81
64.00	32.00	0.20	0.10	-0.46	-0.81
4.00	32.00	0.20	0.10	-0.45	-0.83

Table 4.3: Performance of active walker model in sampling the rapidly varying function, sorted by the frequency based measure.

No. of Active Walkers	No. of Bins	Long Step	Short step	CC with Freq. based Measure	CC with Ent. based Measure
4.00	16.00	0.40	0.02	0.72	0.88
4.00	16.00	0.35	0.02	0.69	0.86
64.00	16.00	0.40	0.02	0.68	0.87
4.00	32.00	0.40	0.02	0.68	0.86
4.00	16.00	0.30	0.02	0.67	0.85
64.00	32.00	0.40	0.02	0.66	0.89
64.00	16.00	0.35	0.02	0.66	0.86
4.00	32.00	0.35	0.02	0.65	0.84
64.00	16.00	0.30	0.02	0.64	0.85
100.00	16.00	0.40	0.02	0.63	0.86
64.00	32.00	0.35	0.02	0.62	0.86
100.00	16.00	0.35	0.02	0.62	0.85
4.00	16.00	0.25	0.02	0.61	0.82
4.00	8.00	0.40	0.02	0.61	0.80
144.00	16.00	0.40	0.02	0.61	0.84
144.00	16.00	0.35	0.02	0.59	0.84
4.00	8.00	0.35	0.02	0.59	0.79
100.00	32.00	0.40	0.02	0.59	0.85
4.00	8.00	0.30	0.02	0.59	0.79
64.00	16.00	0.25	0.02	0.59	0.82
:	:	:	:	:	:
:	:	:	:	:	:
:	:	:	:	:	:
1024.00	256.00	0.20	0.10	-0.18	-0.29
1024.00	32.00	0.20	0.10	-0.22	-0.38
1024.00	8.00	0.20	0.10	-0.23	-0.41
1024.00	16.00	0.20	0.10	-0.24	-0.43
144.00	256.00	0.20	0.10	-0.31	-0.56
4.00	256.00	0.20	0.10	-0.32	-0.55
64.00	256.00	0.20	0.10	-0.34	-0.56
4.00	8.00	0.20	0.10	-0.34	-0.62
100.00	256.00	0.20	0.10	-0.35	-0.58
64.00	8.00	0.20	0.10	-0.36	-0.64
100.00	8.00	0.20	0.10	-0.37	-0.63
144.00	8.00	0.20	0.10	-0.37	-0.65
144.00	32.00	0.20	0.10	-0.41	-0.69
144.00	16.00	0.20	0.10	-0.41	-0.69
4.00	16.00	0.20	0.10	-0.42	-0.68
100.00	16.00	0.20	0.10	-0.42	-0.67
4.00	32.00	0.20	0.10	-0.43	-0.68
100.00	32.00	0.20	0.10	-0.43	-0.68
64.00	16.00	0.20	0.10	-0.44	-0.69
64.00	32.00	0.20	0.10	-0.45	-0.70

Table 4.4: Performance of active walker model in sampling the rapidly varying function, sorted by the entropy based measure.

No. of Active Walkers	No. of Bins	Long Step	Short step	CC with Freq. based Measure	CC with Ent. based Measure
64.00	32.00	0.40	0.02	0.66	0.89
4.00	16.00	0.40	0.02	0.72	0.88
64.00	16.00	0.40	0.02	0.68	0.87
4.00	16.00	0.35	0.02	0.69	0.86
64.00	32.00	0.35	0.02	0.62	0.86
4.00	32.00	0.40	0.02	0.68	0.86
64.00	16.00	0.35	0.02	0.66	0.86
100.00	16.00	0.40	0.02	0.63	0.86
4.00	16.00	0.30	0.02	0.67	0.85
100.00	32.00	0.40	0.02	0.59	0.85
64.00	16.00	0.30	0.02	0.64	0.85
100.00	16.00	0.35	0.02	0.62	0.85
144.00	16.00	0.40	0.02	0.61	0.84
144.00	32.00	0.40	0.02	0.57	0.84
4.00	32.00	0.35	0.02	0.65	0.84
100.00	32.00	0.35	0.02	0.57	0.84
64.00	32.00	0.30	0.02	0.58	0.84
144.00	16.00	0.35	0.02	0.59	0.84
100.00	16.00	0.30	0.02	0.57	0.83
144.00	32.00	0.35	0.02	0.55	0.83
:	:	:	:	:	:
:	:	:	:	:	:
:	:	:	:	:	:
144.00	2.00	0.20	0.10	-0.11	-0.39
100.00	2.00	0.20	0.10	-0.13	-0.40
1024.00	8.00	0.20	0.10	-0.23	-0.41
1024.00	16.00	0.20	0.10	-0.24	-0.43
4.00	256.00	0.20	0.10	-0.32	-0.55
144.00	256.00	0.20	0.10	-0.31	-0.56
64.00	256.00	0.20	0.10	-0.34	-0.56
100.00	256.00	0.20	0.10	-0.35	-0.58
4.00	8.00	0.20	0.10	-0.34	-0.62
100.00	8.00	0.20	0.10	-0.37	-0.63
64.00	8.00	0.20	0.10	-0.36	-0.64
144.00	8.00	0.20	0.10	-0.37	-0.65
100.00	16.00	0.20	0.10	-0.42	-0.67
4.00	32.00	0.20	0.10	-0.43	-0.68
100.00	32.00	0.20	0.10	-0.43	-0.68
4.00	16.00	0.20	0.10	-0.42	-0.68
144.00	32.00	0.20	0.10	-0.41	-0.69
144.00	16.00	0.20	0.10	-0.41	-0.69
64.00	16.00	0.20	0.10	-0.44	-0.69
64.00	32.00	0.20	0.10	-0.45	-0.70

Also for this application, using extreme values of number of bins, such as N_{ub}= 2, or 256 produces poor results. This is because dividing the samples in to a very small number like two does not provide adequate resolution for the sample values, and therefore makes the basis of the ASHE algorithm irrelevant. Using a very high value like 256 means that it takes a longer time to form a profile in the histogram on which that ASHE is based. For this application, we obtain good results for values of N_{ub}= 16, and 32. This is to be chosen based on the expected number of unique samples in the function being sampled.

The values of LSP and SSP are the more crucial factors for the performance of this model. We always obtain poor performance when the values of LSP and SSP are comparable. This is expected since the movement of the active walkers under this condition does not reflect the effect of the fitness criterion, and is essentially random. Values of $LSP \geq 0.3$ result in good performance since it is an appreciable movement of the active walker away from its present vicinity in response to change in the fitness criterion. The SSP value is the most crucial factor. Values of SSP= 0.06 give the best result for sampling the smooth function adaptively. The stated LSP value is appropriate for all scenarios. The chosen SSP value will depend on whether a function is smooth or varying rapidly.

All the discussions for the experiments with the smooth function, are also valid for the results for the rapidly varying function shown in Tables 4.3 and 4.4. The only difference is in the crucial factor, SSP. Note here that the best performance is obtained for $SSP = 0.02$, compared to $SSP = 0.06$ for the smooth function. This is accounted for by the fact that the function in the latter case varies rapidly, and a small step results in much larger changes in sample values. Note that this is the only difference in sampling the two types of functions. This is important because algorithms like the active walker model

usually suffer the drawback of being *ad-hoc*. That is, they have to be customized for every unique purpose. Our results show that "rules of thumb" can be established for determining the other factors in the active walker model. Customization of the model for any purpose only requires minimal prior knowledge on whether a function to be sampled is smooth or varies rapidly. This prior information is available in many cases.

Figures 4.8 to 4.11 show the stages of the adaptive sampling process for the two test functions, using the active walker model. We show examples of the good (Figures 4.8 and 4.10), and poor (Figures 4.9 and 4.11) performances as indicated on the tables. In the examples with the poor performance, the sample density show a random pattern, supporting the argument given for the effect of the LSP, and SSP factors. Only intermediate steps are shown. The examples showing good performance show sample densities indicative of the regions of complexities in the function.

4.4 Analysis of the Ant Model

Here, we also experiment with the same test functions, and obtained the same number of samples as described in Section 4.3. We identify four key factors that may affect the performance of the ant model described in Section 3.2.1.

These are:

1. Number of ants (N_{as}). This will determine the number of foraging trips taken in the adaptive sampling process. Starting with n ants means that $(4,096/n) - 1$ trips will be taken.

2. Number of bins N_{ub} in the histogram to be equalized. Same explanation under the active walker model holds here.

3. The range of the effect of the deposited pheromone (RPH).

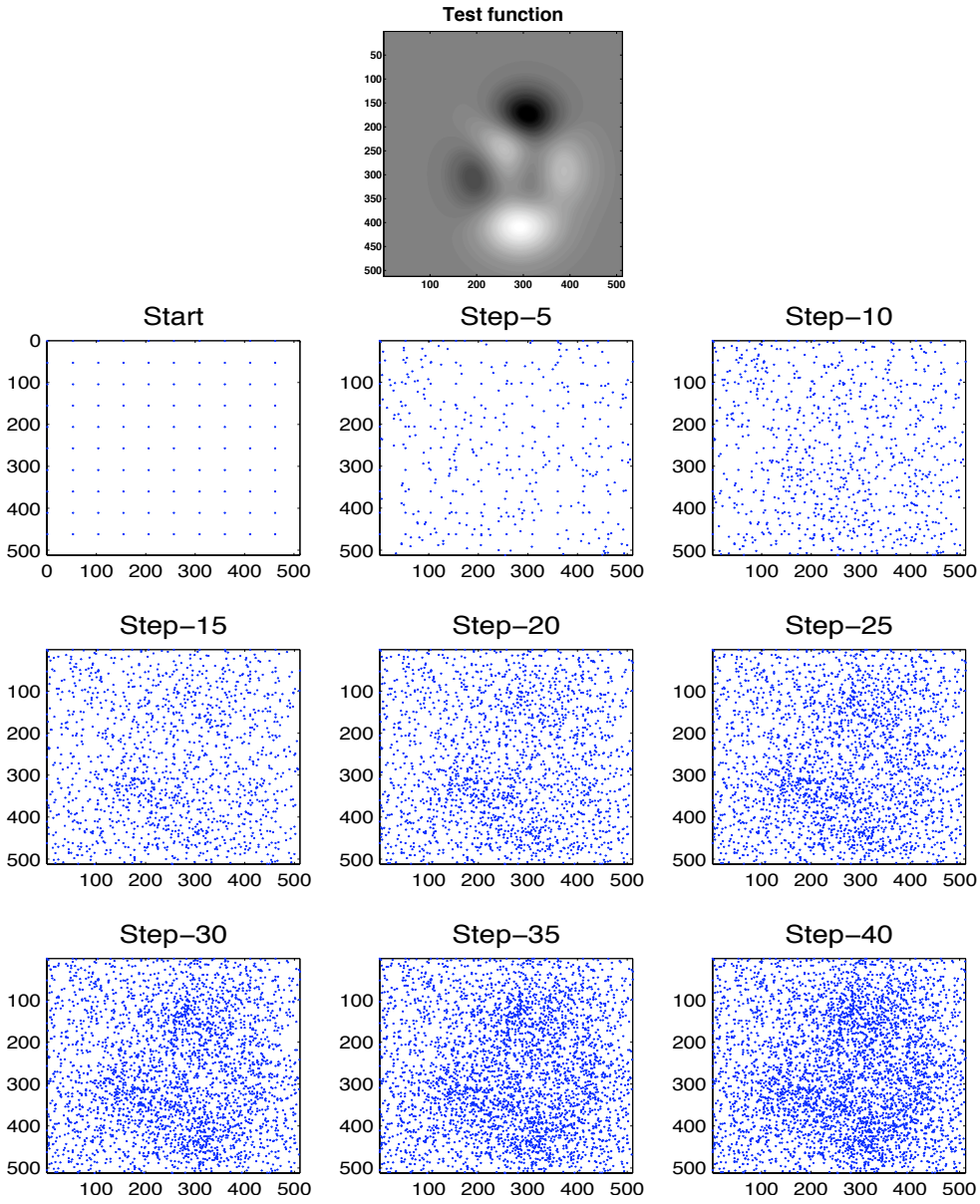


Figure 4.8: Intermediate steps from sampling a smooth test function with the active walker model. The good performance is recorded by using the following factors: $N_{\mathrm{aw}} = 100$, $N_{\mathrm{ub}} = 16$, $LSP = 0.4$, and $SSP = 0.06$. Performance measures = 0.60/0.93. Total of 40 steps.

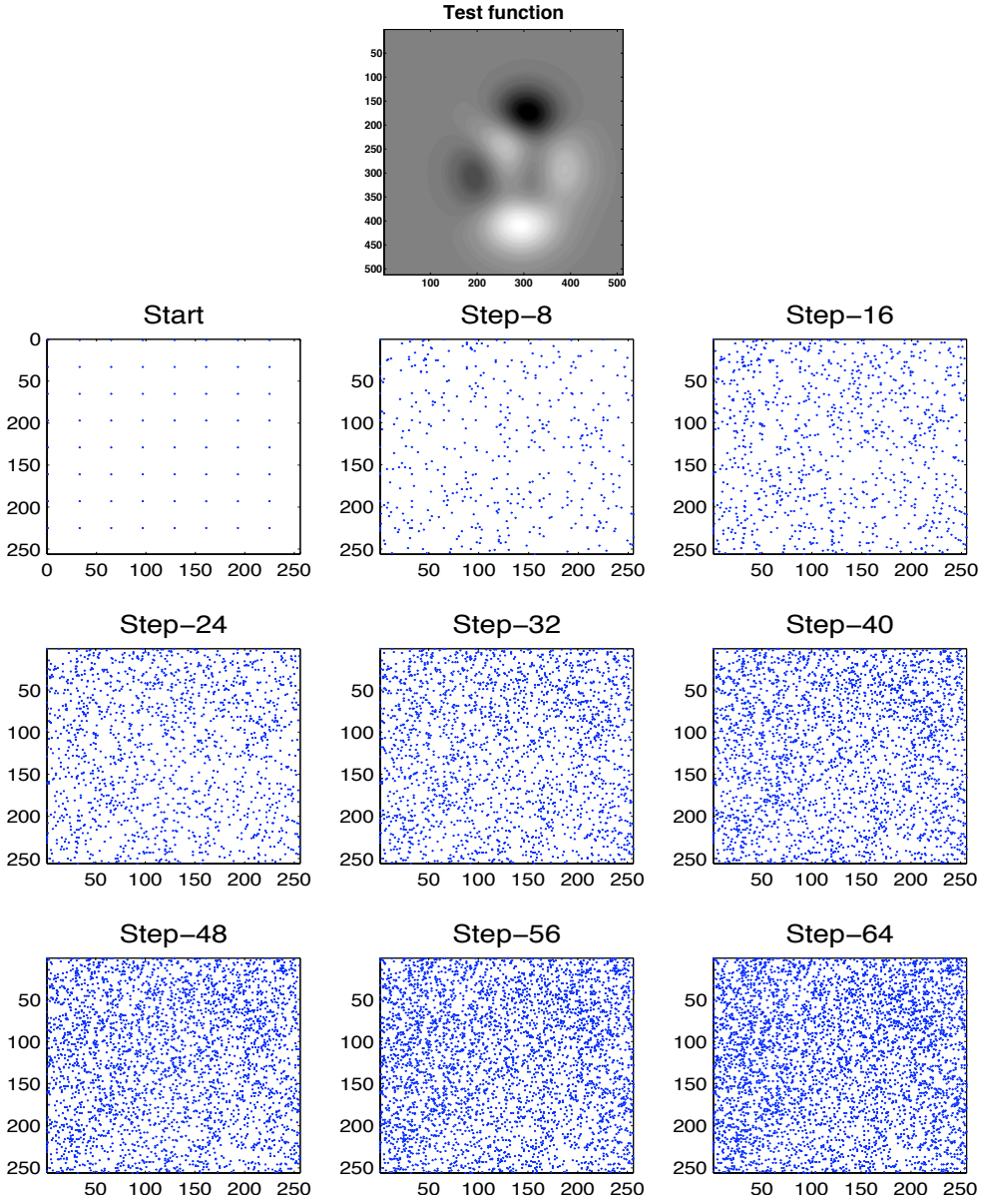


Figure 4.9: Intermediate steps from sampling a smooth test function with the active walker model. The poor performance is recorded by using the following factors: N_{aw} = 100, N_{ub} = 16, LSP = 0.2, and SSP = 0.1. Performance measures = -0.47/-0.80. Total of 40 steps.

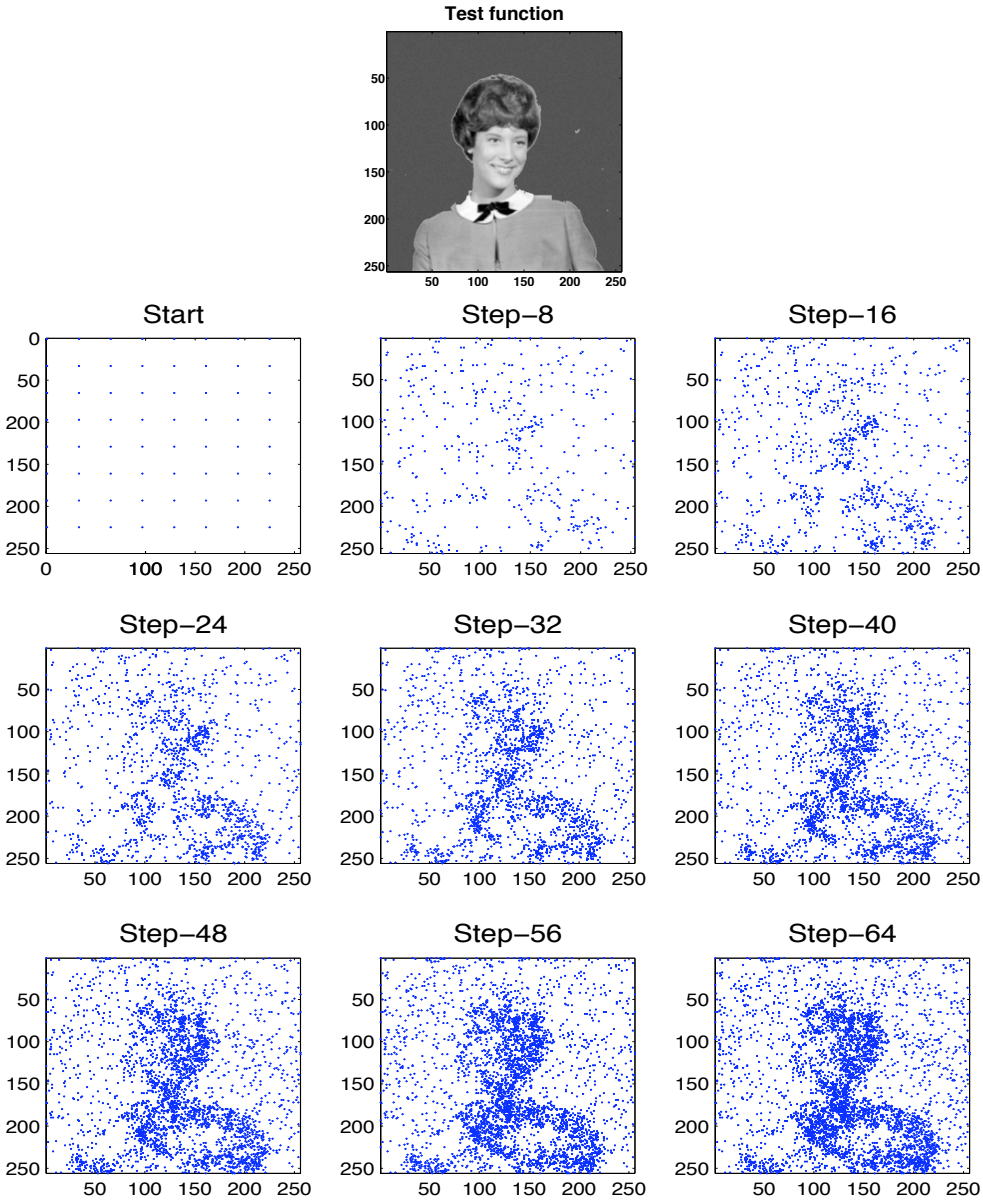


Figure 4.10: Intermediate steps from sampling a rapidly varying function with the active walker model. The good performance is recorded by using the following factors: $N_{\mathrm{aw}} = 64$, $N_{\mathrm{ub}} = 32$, $LSP = 0.4$, and $SSP = 0.02$. Performance measure = 0.66/0.89. Total of 64 steps.

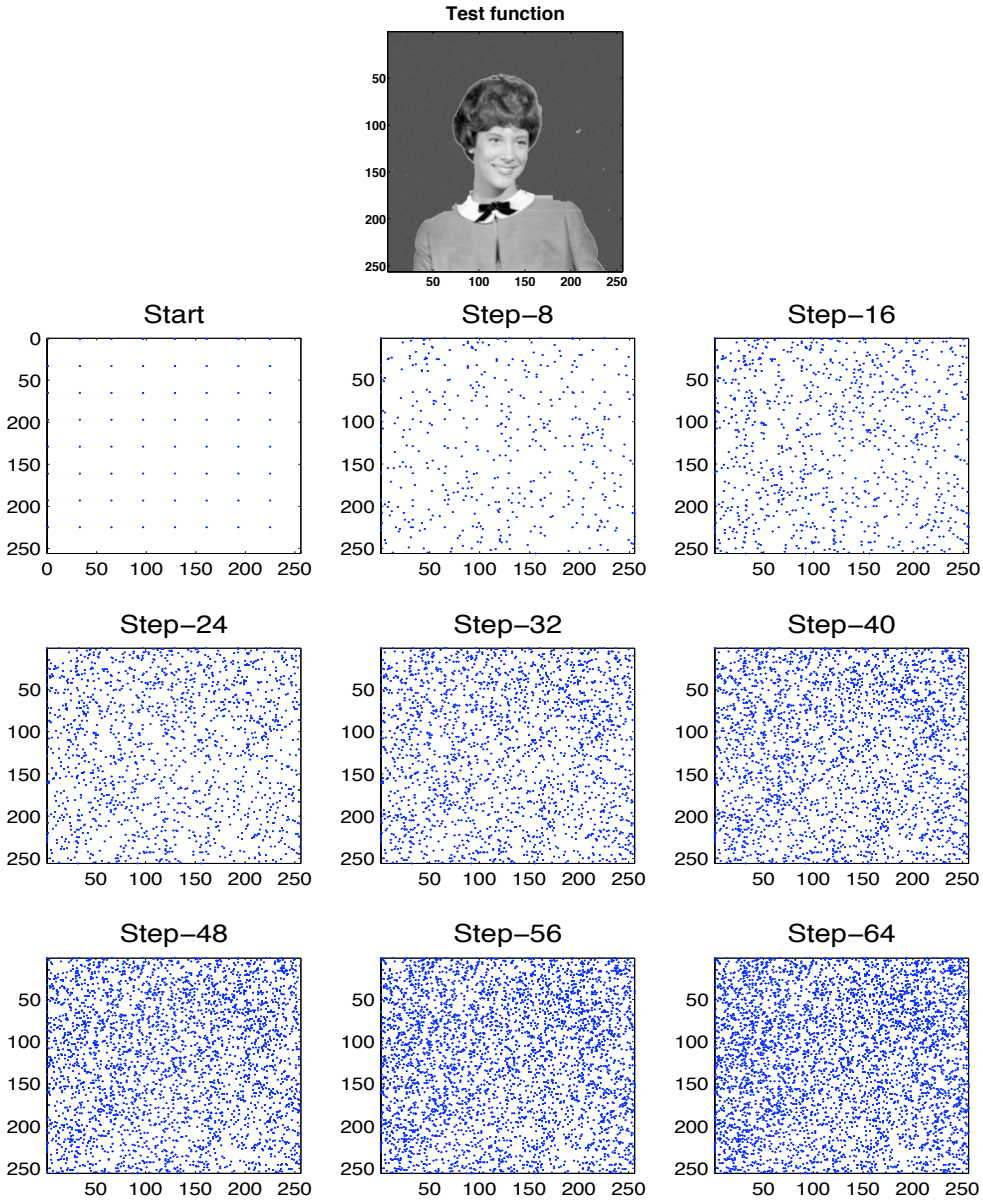


Figure 4.11: Intermediate steps from sampling a rapidly varying function with the active walker model. The poor performance is recorded by using the following factors: $N_{\mathrm{aw}} = 64$, $N_{\mathrm{ub}} = 32$, $LSP = 0.2$, and $SSP = 0.2$. Performance measure = 0.66/0.89. Total of 64 steps.

4. The range an ant can move in its neighborhood for foraging (RFO).

We experiment with the following values of these factors:

$N_{\text{as}} = \{4,\ 64,\ 100,\ 144,\ 1,024\}$

$N_{\text{ub}} = \{2,\ 8,\ 16,\ 32,\ 256\}$

$RPH = \{0.02,\ 0.05,\ 0.1,\ 0.12,\ 0.15,\ 0.2,\ 0.25\}$

$RFO = \{0.02,\ 0.05,\ 0.08,\ 0.1,\ 0.12,\ 0.15,\ 0.2\}$

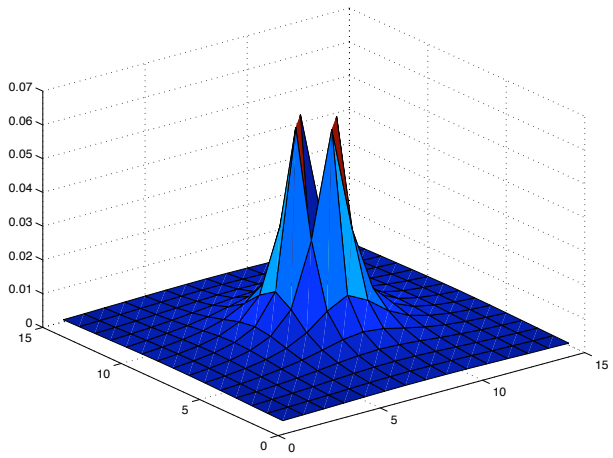


Figure 4.12: Model to simulate reduction of pheromone concentration as a function of the distance away from source. Concentration $\propto 1/d^2$.

The last two factors, RPH, and RFO are functions of the size of the sampled space. That is, they are determined by multiplying these numbers with the vector containing the spaces' dimensions. The RPH is made to fade away from the point of deposit at a rate of square of the distance d. That is the concentration $\propto 1/d^2$. The concentration, represented as probabilities, are normalized, and added to the existing value in a non-zero probability location. Figure 4.12 shows an example of how the concentration fades away with distance from the source. The central pixel is set to a probability of zero since that location has already been sampled.

Figures 4.13 and 4.14 show the change in the pheromone concentration, modeling the

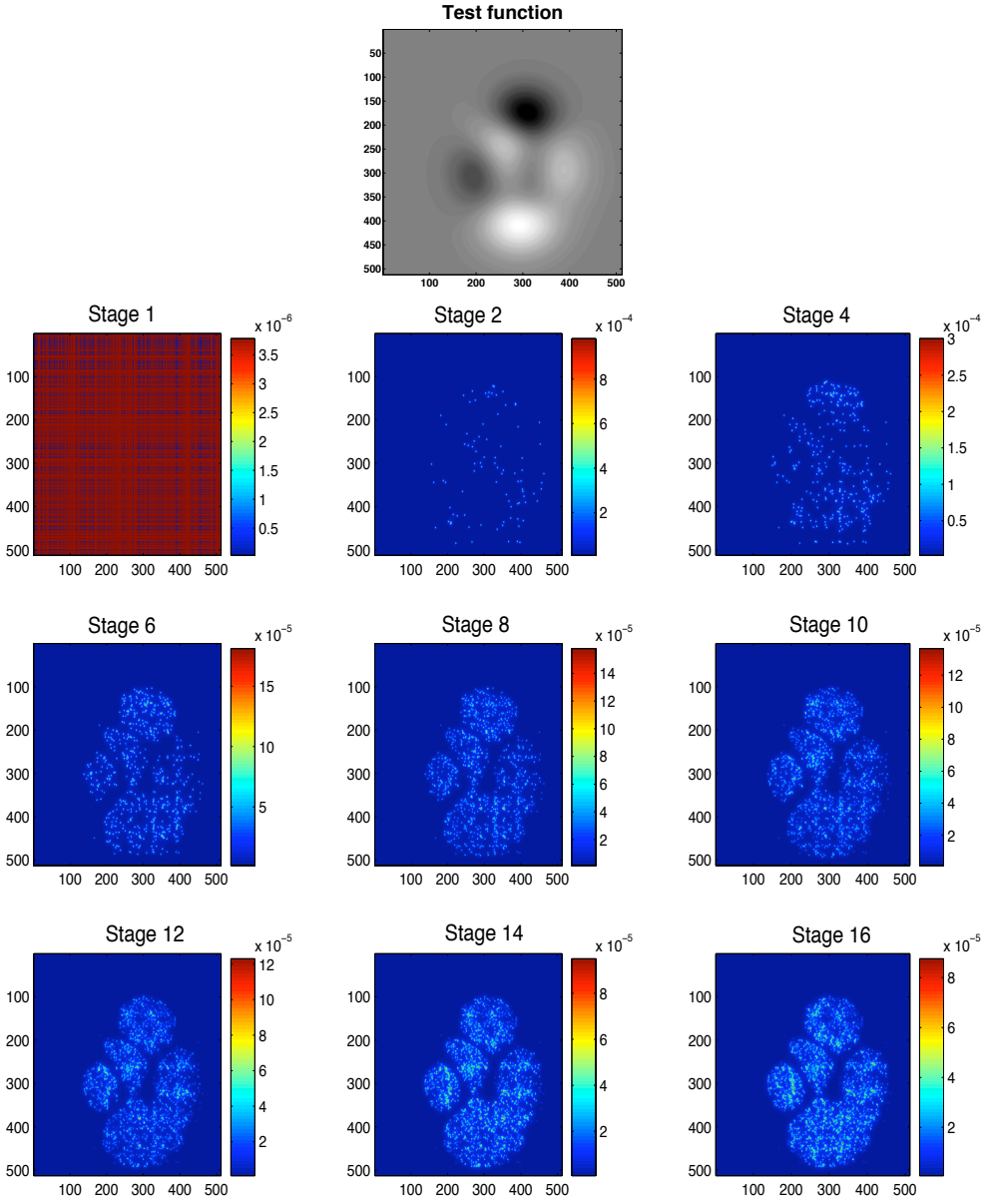


Figure 4.13: Pheromone concentration/probability change from intermediate steps in the sampling of a smooth function. Ant model used with the following factors: $N_{\text{as}} = 256$, $N_{\text{ub}} = 8$, $RPH = 0.12$, and $RFO = 0.12$. Total of 16 foraging trips by each ant.

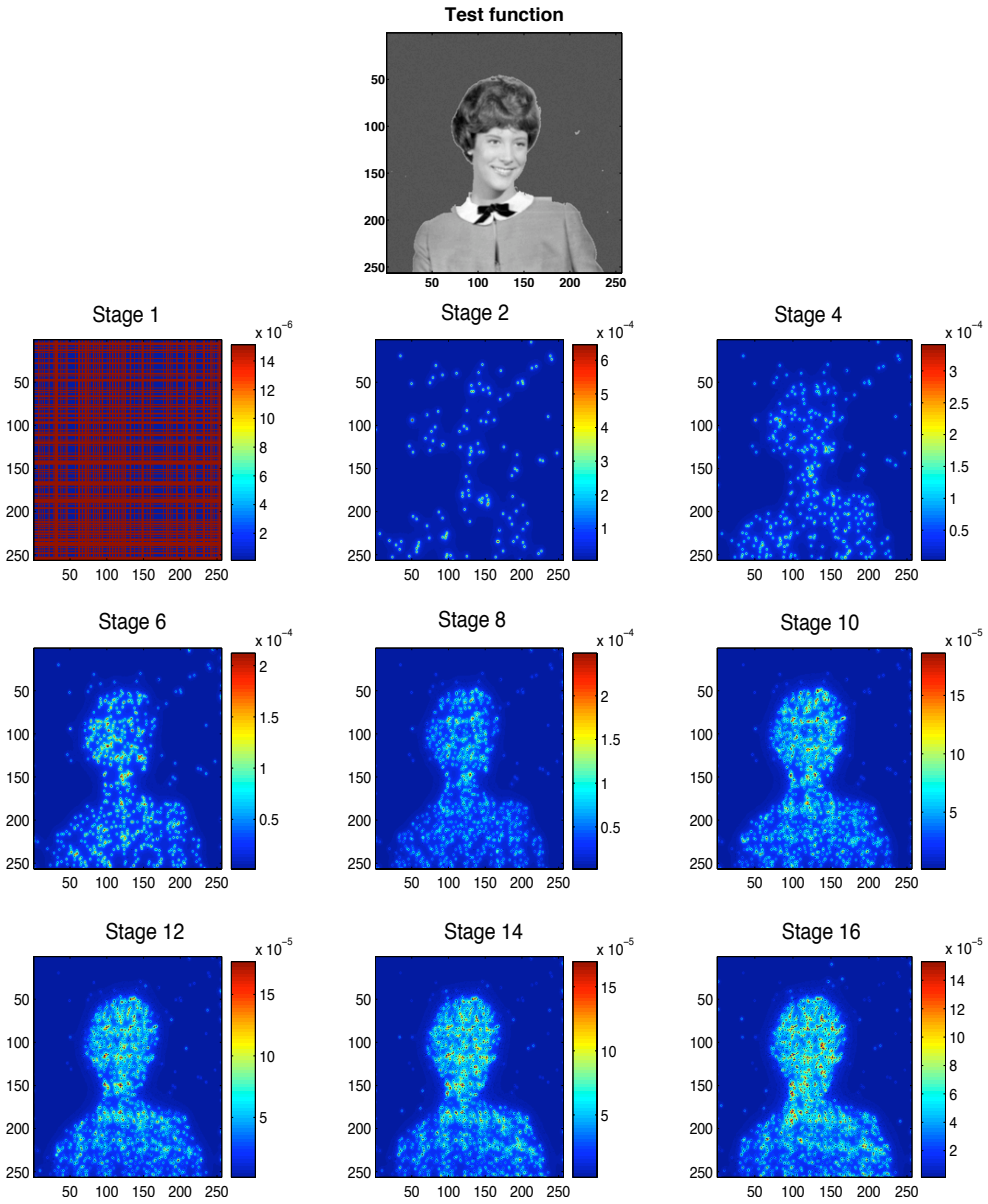


Figure 4.14: Pheromone concentration/probability change from intermediate steps in the sampling of a rapidly varying function. Ant model used with the following factors: $N_{\text{as}} = 256$, $N_{\text{ub}} = 32$, $RPH = 0.25$, and $RFO = 0.12$. Total of 16 foraging trips by each ant.

probabilities in the sampled space. Note how the probabilities change to reflect the levels of complexity in the sampled functions. Only intermediate steps are shown because of the space constraint.

Tables 4.5 to 4.8 show similarly sorted results as in Section 4.3. The higher correlation between the sample density and the entropy measure was also noted, and explained in the active walker model analysis discussion. The good correlation between both objective measures indicates that they essentially convey the same information. Tables 4.5 and 4.6 show the sorted results for the experiments with the smooth function. The results indicate that too few foraging trips yield a poor sample distribution. Generally, the poor performance results from using too many ants $N_{\text{as}} = 1,024$, each taking only three foraging trips. The reason for this is similar to that stated for the active walker model. There are too few steps to enable the feedback from the ants have an effect. The sampling under this condition is thus near random. Good performances are recorded for 15 or more foraging trips, resulting from employing $N_{\text{as}} \leq 256$. The N_{ub} also has similar effects as in the case of the active walker model. Good performance is recorded for values of at least eight bins in the histogram. Extreme N_{ub} values of 2 or 256 also result in poor performance for reasons stated earlier. No clear trend can be deciphered from the tables about the values representing the range of the effect of the deposited pheromone RPH. A clear correlation is needed between all the factors and the performance criteria, in order to establish appropriate values for the factors. Values of rfo ≥ 0.1 result in good performance.

Similar results are recorded from the experiments with the rapidly varying function shown in Tables 4.7 and 4.8. All the discussions for the experiments with the smooth function, are also valid here.

Figures 4.15 to 4.18 show the stages of the adaptive sampling process for the two test

Table 4.5: Performance of the ant model in sampling the smooth function, sorted by the frequency based measure.

No. of ANTs	No. of Bins	Range of Pheromone	Range of Foraging(1 step)	CC with Freq. based Measure	CC with Ent. based Measure
256.00	8.00	0.15	0.08	0.64	0.90
256.00	8.00	0.02	0.08	0.64	0.89
256.00	8.00	0.05	0.08	0.64	0.89
256.00	8.00	0.10	0.08	0.64	0.90
256.00	8.00	0.20	0.08	0.64	0.91
256.00	8.00	0.12	0.08	0.64	0.90
256.00	8.00	0.02	0.10	0.63	0.92
256.00	8.00	0.25	0.08	0.63	0.92
256.00	8.00	0.15	0.10	0.63	0.93
256.00	8.00	0.10	0.10	0.63	0.93
256.00	8.00	0.05	0.10	0.63	0.92
256.00	8.00	0.12	0.10	0.62	0.93
144.00	8.00	0.05	0.08	0.62	0.92
64.00	8.00	0.10	0.05	0.62	0.89
144.00	8.00	0.12	0.05	0.62	0.84
144.00	8.00	0.10	0.05	0.62	0.84
144.00	8.00	0.05	0.05	0.62	0.81
256.00	8.00	0.05	0.12	0.62	0.93
144.00	8.00	0.02	0.08	0.62	0.92
256.00	8.00	0.25	0.10	0.62	0.93
:	:	:	:	:	:
:	:	:	:	:	:
:	:	:	:	:	:
256.00	2.00	0.10	0.02	-0.08	0.10
1024.00	2.00	0.20	0.08	-0.09	0.36
256.00	2.00	0.02	0.02	-0.10	0.01
144.00	2.00	0.02	0.02	-0.10	0.05
1024.00	2.00	0.02	0.05	-0.11	0.06
256.00	2.00	0.05	0.02	-0.11	0.05
1024.00	2.00	0.05	0.10	-0.11	0.40
1024.00	2.00	0.10	0.10	-0.12	0.40
1024.00	2.00	0.12	0.05	-0.12	0.09
256.00	2.00	0.12	0.02	-0.12	0.09
1024.00	2.00	0.15	0.05	-0.12	0.10
1024.00	2.00	0.20	0.05	-0.16	0.13
1024.00	2.00	0.15	0.08	-0.16	0.28
1024.00	2.00	0.02	0.10	-0.17	0.36
1024.00	2.00	0.05	0.05	-0.18	0.02
1024.00	2.00	0.10	0.05	-0.19	0.09
1024.00	2.00	0.10	0.08	-0.21	0.23
1024.00	2.00	0.12	0.08	-0.21	0.24
1024.00	2.00	0.02	0.08	-0.24	0.17
1024.00	2.00	0.05	0.08	-0.24	0.19

Table 4.6: Performance of the ant model in sampling the smooth function, sorted by the entropy based measure.

No. of ANTs	No. of Bins	Range of Pheromone	Range of Foraging(1 step)	CC with Freq. based Measure	CC with Ent. based Measure
256.00	8.00	0.12	0.12	0.61	0.94
256.00	8.00	0.02	0.12	0.62	0.93
256.00	8.00	0.25	0.12	0.60	0.93
144.00	8.00	0.25	0.10	0.57	0.93
256.00	8.00	0.20	0.10	0.61	0.93
256.00	8.00	0.10	0.12	0.62	0.93
256.00	8.00	0.20	0.12	0.61	0.93
256.00	8.00	0.15	0.12	0.61	0.93
144.00	8.00	0.12	0.10	0.57	0.93
144.00	8.00	0.10	0.10	0.59	0.93
144.00	8.00	0.20	0.10	0.57	0.93
144.00	16.00	0.02	0.12	0.53	0.93
144.00	8.00	0.12	0.12	0.54	0.93
144.00	8.00	0.02	0.12	0.55	0.93
144.00	8.00	0.15	0.10	0.57	0.93
256.00	8.00	0.25	0.10	0.62	0.93
256.00	8.00	0.10	0.15	0.56	0.93
256.00	8.00	0.05	0.12	0.62	0.93
144.00	8.00	0.15	0.12	0.54	0.93
144.00	8.00	0.20	0.12	0.54	0.93
:	:	:	:	:	:
:	:	:	:	:	:
:	:	:	:	:	:
1024.00	8.00	0.05	0.02	0.08	0.04
1024.00	16.00	0.15	0.02	-0.00	0.04
1024.00	256.00	0.02	0.02	-0.01	0.04
1024.00	8.00	0.02	0.02	0.04	0.03
1024.00	2.00	0.10	0.02	-0.06	0.03
1024.00	32.00	0.05	0.02	-0.01	0.03
1024.00	32.00	0.10	0.02	-0.03	0.03
1024.00	2.00	0.20	0.02	-0.02	0.02
1024.00	2.00	0.05	0.05	-0.18	0.02
1024.00	256.00	0.25	0.02	-0.03	0.02
1024.00	32.00	0.20	0.02	0.00	0.02
1024.00	256.00	0.20	0.02	-0.01	0.02
256.00	2.00	0.02	0.02	-0.10	0.01
1024.00	2.00	0.25	0.02	-0.05	0.00
1024.00	16.00	0.05	0.02	-0.00	-0.00
1024.00	2.00	0.15	0.02	-0.08	-0.00
1024.00	2.00	0.02	0.02	-0.05	-0.01
1024.00	2.00	0.12	0.02	-0.06	-0.01
1024.00	32.00	0.02	0.02	-0.03	-0.02
1024.00	2.00	0.05	0.02	-0.04	-0.03

Table 4.7: Performance of the ant model in sampling the rapidly varying function, sorted by the frequency based measure.

No. of ANTs	No. of Bins	Range of Pheromone	Range of Foraging(1 step)	CC with Freq. based Measure	CC with Ent. based Measure
256.00	32.00	0.25	0.12	0.78	0.89
256.00	256.00	0.12	0.12	0.78	0.88
256.00	32.00	0.20	0.12	0.78	0.89
256.00	32.00	0.15	0.12	0.78	0.89
256.00	32.00	0.10	0.12	0.78	0.89
256.00	256.00	0.25	0.12	0.78	0.88
256.00	16.00	0.15	0.12	0.77	0.90
256.00	256.00	0.10	0.12	0.77	0.88
256.00	256.00	0.20	0.12	0.77	0.88
256.00	256.00	0.05	0.12	0.77	0.89
256.00	32.00	0.05	0.12	0.77	0.89
256.00	32.00	0.12	0.12	0.77	0.89
256.00	16.00	0.25	0.12	0.77	0.90
256.00	32.00	0.02	0.12	0.77	0.89
256.00	16.00	0.20	0.12	0.77	0.90
256.00	16.00	0.20	0.10	0.77	0.88
256.00	16.00	0.25	0.10	0.77	0.88
256.00	256.00	0.02	0.12	0.77	0.88
256.00	256.00	0.10	0.15	0.77	0.87
256.00	32.00	0.12	0.10	0.77	0.88
:	:	:	:	:	:
:	:	:	:	:	:
:	:	:	:	:	:
1024.00	256.00	0.05	0.02	0.05	0.03
144.00	2.00	0.02	0.02	0.05	0.12
1024.00	256.00	0.02	0.02	0.04	0.06
1024.00	2.00	0.10	0.05	0.03	0.16
1024.00	32.00	0.02	0.02	0.02	0.02
1024.00	16.00	0.02	0.02	0.02	0.04
256.00	2.00	0.05	0.02	0.02	0.11
256.00	2.00	0.10	0.02	0.02	0.15
1024.00	2.00	0.20	0.02	0.02	0.05
1024.00	2.00	0.15	0.02	0.02	0.05
1024.00	8.00	0.05	0.02	0.02	0.02
1024.00	32.00	0.05	0.02	0.01	0.02
1024.00	8.00	0.02	0.02	0.01	0.02
1024.00	2.00	0.02	0.05	0.01	0.10
1024.00	2.00	0.02	0.02	-0.00	0.00
1024.00	2.00	0.12	0.02	-0.00	0.02
256.00	2.00	0.02	0.02	-0.01	0.03
1024.00	2.00	0.05	0.05	-0.02	0.07
1024.00	2.00	0.10	0.02	-0.02	-0.00
1024.00	2.00	0.05	0.02	-0.03	-0.01

Table 4.8: Performance of the ant model in sampling the rapidly varying function, sorted by the entropy based measure.

No. of ANTs	No. of Bins	Range of Pheromone	Range of Foraging(1 step)	CC with Freq. based Measure	CC with Ent. based Measure
144.00	16.00	0.05	0.12	0.73	0.91
144.00	16.00	0.02	0.12	0.72	0.91
144.00	16.00	0.12	0.12	0.74	0.91
144.00	8.00	0.05	0.12	0.71	0.91
144.00	8.00	0.10	0.12	0.72	0.91
144.00	16.00	0.15	0.12	0.73	0.90
144.00	32.00	0.12	0.12	0.74	0.90
144.00	8.00	0.15	0.10	0.73	0.90
144.00	16.00	0.10	0.12	0.73	0.90
144.00	8.00	0.20	0.12	0.72	0.90
144.00	16.00	0.10	0.10	0.74	0.90
256.00	16.00	0.05	0.12	0.77	0.90
144.00	16.00	0.25	0.12	0.73	0.90
144.00	8.00	0.02	0.12	0.70	0.90
144.00	32.00	0.20	0.12	0.73	0.90
144.00	32.00	0.15	0.12	0.74	0.90
144.00	8.00	0.15	0.12	0.71	0.90
144.00	8.00	0.12	0.12	0.71	0.90
144.00	32.00	0.05	0.12	0.74	0.90
144.00	16.00	0.20	0.12	0.73	0.90
:	:	:	:	:	:
:	:	:	:	:	:
:	:	:	:	:	:
1024.00	256.00	0.15	0.02	0.08	0.07
1024.00	2.00	0.05	0.05	-0.02	0.07
1024.00	256.00	0.12	0.02	0.06	0.06
1024.00	8.00	0.10	0.02	0.06	0.06
1024.00	256.00	0.10	0.02	0.06	0.06
1024.00	256.00	0.02	0.02	0.04	0.06
1024.00	16.00	0.05	0.02	0.09	0.06
1024.00	2.00	0.15	0.02	0.02	0.05
1024.00	2.00	0.20	0.02	0.02	0.05
1024.00	16.00	0.02	0.02	0.02	0.04
1024.00	256.00	0.05	0.02	0.05	0.03
256.00	2.00	0.02	0.02	-0.01	0.03
1024.00	2.00	0.12	0.02	-0.00	0.02
1024.00	8.00	0.05	0.02	0.02	0.02
1024.00	32.00	0.02	0.02	0.02	0.02
1024.00	8.00	0.02	0.02	0.01	0.02
1024.00	32.00	0.05	0.02	0.01	0.02
1024.00	2.00	0.02	0.02	-0.00	0.00
1024.00	2.00	0.10	0.02	-0.02	-0.00
1024.00	2.00	0.05	0.02	-0.03	-0.01

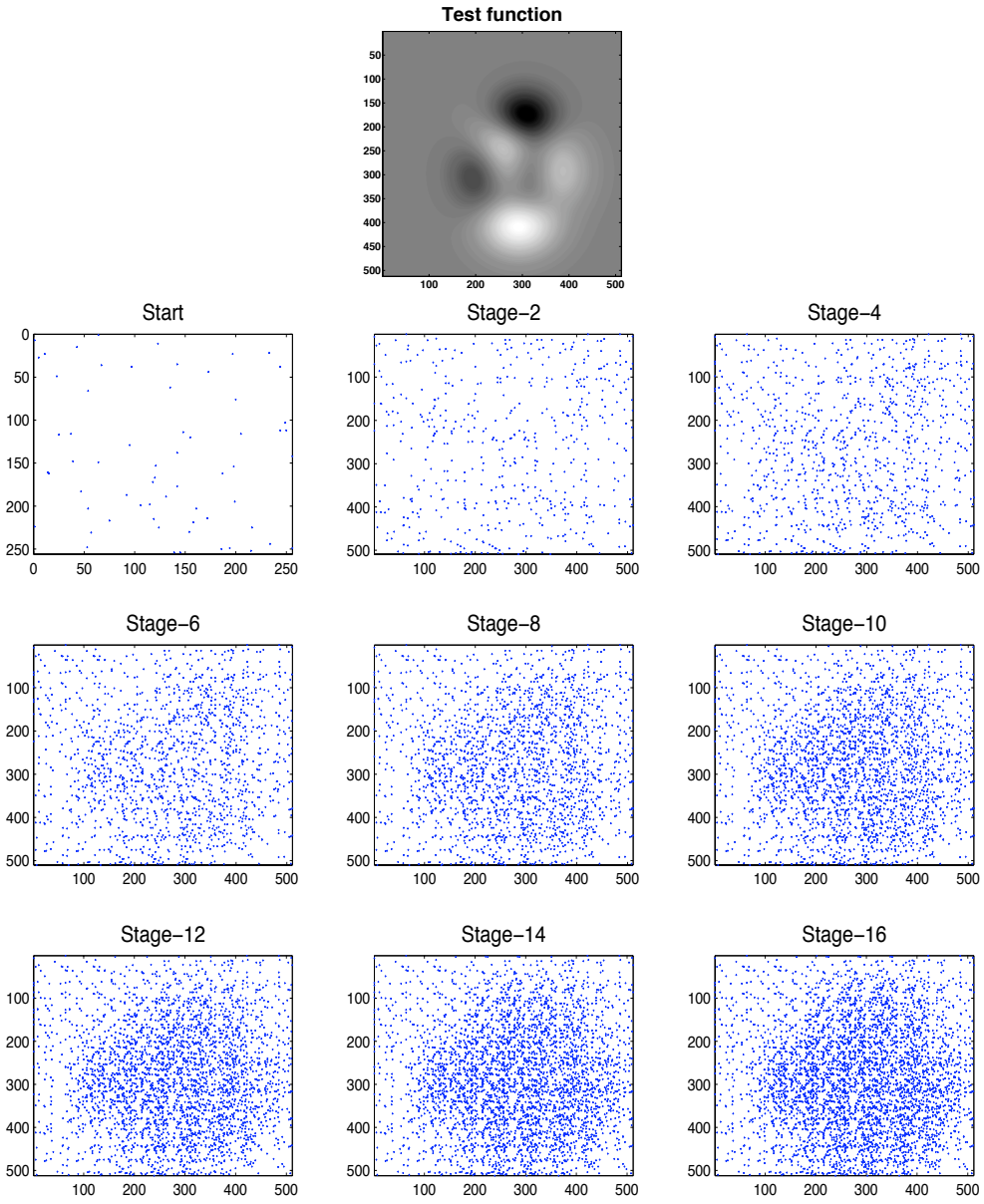


Figure 4.15: Intermediate stages from sampling a smooth test function with the ant model. The good performance is recorded by using the following factors: $N_{\text{as}} = 256$, $N_{\text{ub}} = 8$, $RPH = 0.12$, and $RFO = 0.12$. Performance measures = 0.60/0.91. Total of 16 foraging trips.

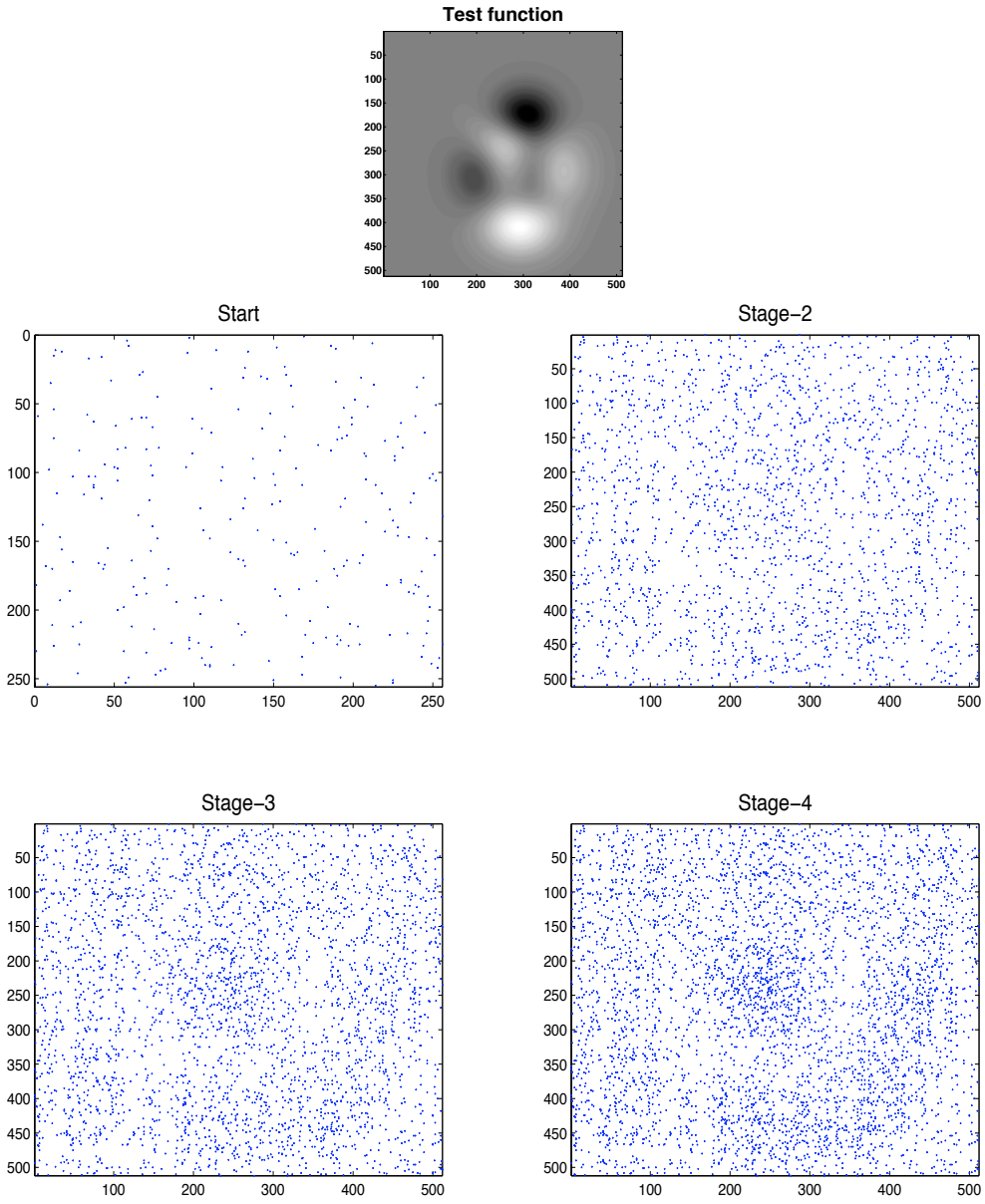


Figure 4.16: All the stages from sampling a smooth test function with the ant model. The poor performance is recorded by using the following factors: N_{as} = 1,024, N_{ub} = 2, RPH = 0.05, and RFO = 0.08. Performance measures = -0.24/0.19. Total of four foraging trips.

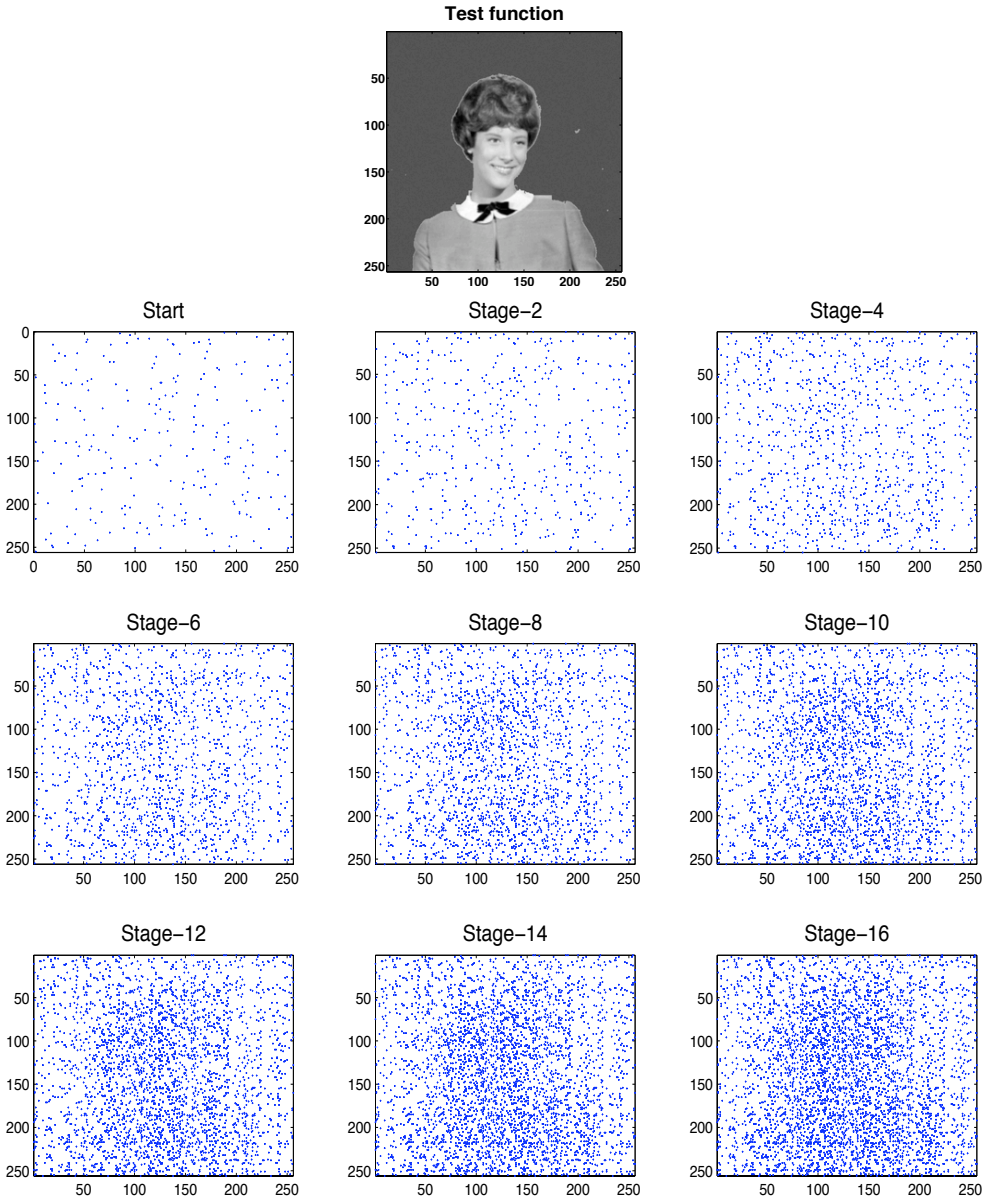


Figure 4.17: Intermediate stages from sampling a rapidly varying test function with the ant model. The good performance is recorded by using the following factors: $N_{\text{as}} = 256$, $N_{\text{ub}} = 32$, $RPH = 0.25$, and $RFO = 0.12$. Performance measures = 0.78/0.89. Total of 16 foraging trips.

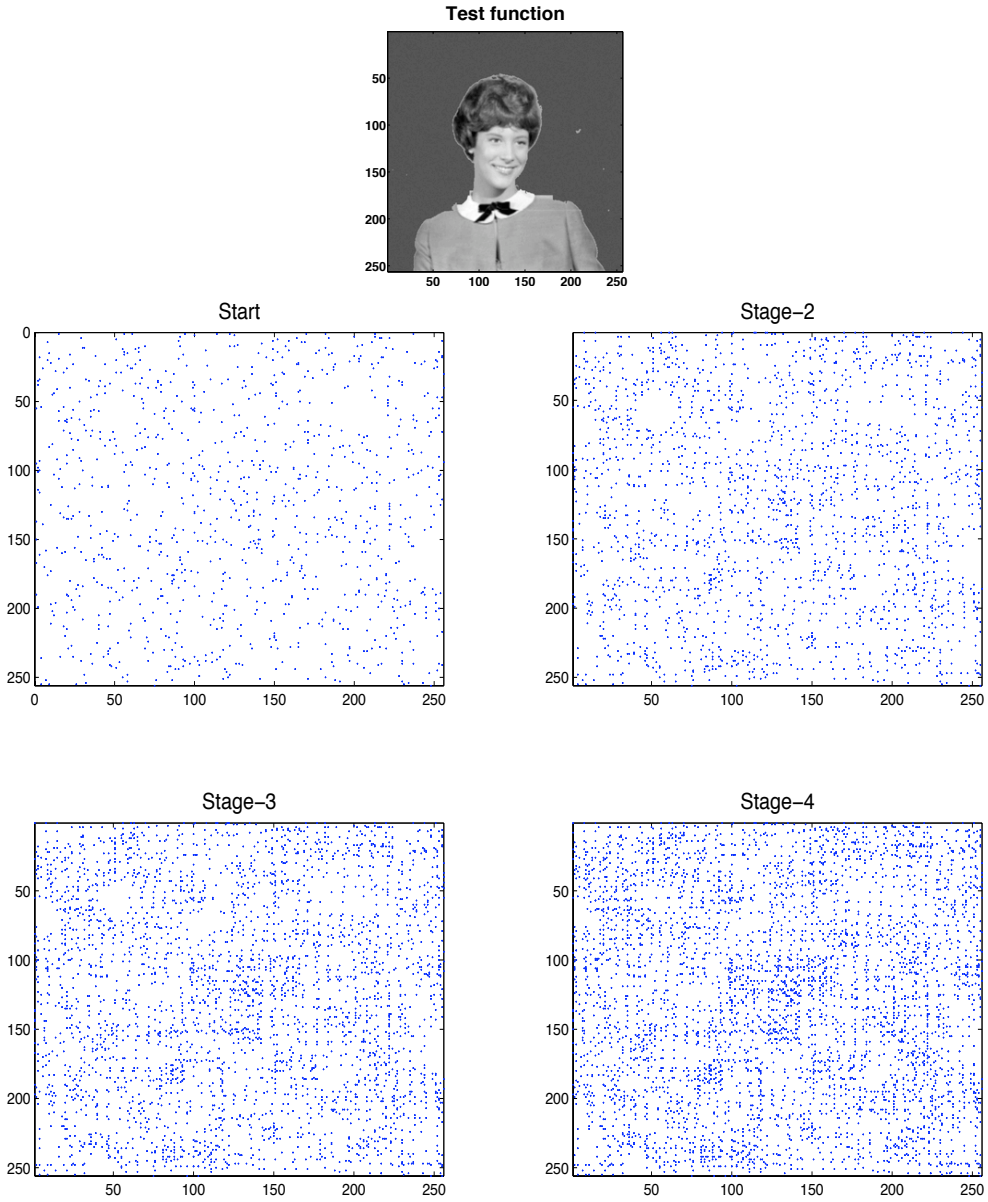


Figure 4.18: All the stages from sampling a rapidly varying test function with the ant model. The poor performance is recorded by using the following factors: N_{as} = 1,024, N_{ub} = 2, RPH = 0.05, and RFO = 0.02. Performance measures = -0.03/0.01. Total of four foraging trips.

functions, using the ant model. We show examples of the good, and poor performances as indicated on the tables. Only intermediate steps are shown in some of the cases.

4.5 Analysis of the Evolutionary Algorithm (EA) Model

The test functions, and the number of samples obtained using the EA model, are the same as in the two models considered earlier. We identify five key factors that may affect the performance of the EA model as described in Section 3.3.1. These are:

1. Size of starting population (SOP). This will determine the number of generations in the adaptive sampling process. Starting with a population of size n means that there will be $(4,096/n) - 1$ generations.

2. Number of offspring (N_{of}) from each parent.

3. Number of bins (N_{ub}) in the histogram to be equalized. Same explanation under the previously discussed models holds here.

4. Distance away from parent (DOP) or neighborhood where an offspring will reside.

5. The probability of survival of an offspring (POS).

We experiment with the following values of these factors:

$SOP = \{4,\ 64,\ 144,\ 1,024\}$

$N_{\text{of}} = \{1,\ 2,\ 4,\ 6\}$

$N_{\text{ub}} = \{4,\ 16,\ 32,\ 256\}$

$DOP = \{0.02,\ 0.05,\ 0.08,\ 0.1\}$

$POS = \{1,\ 0.9,\ 0.8,\ 0.7\}$

The DOP factor is a function of the size of the sampled space. That is, it is determined by multiplying the listed number with the vector containing the spaces' dimensions. A complete set of experiments included $1,024$ runs, resulting from all combinations of the factors. We ran 100 complete sets, requiring $102,400$ experiments.

The performance results from the experiments with the smooth function are sorted as in Section 4.3, and shown in Tables 4.9 to 4.12. For both test functions, the only factor that shows a general trend in indicating sampling performance is the N_{ub}. Extreme values of 4 and 256 generally result in the poor performance of the algorithm. Values of 16 and 32 result in the recorded good performance. The reason for this trend is the same as in the previously discussed models. It is important to note that this factor is important for all the employed models. Note that the EA algorithm is able to *converge* at a much faster rate than the other two models. This is indicated by the fact that a starting population of $1,024$ may still result in a good performance. As shown on the tables, this is only recorded when parents have multiple offspring, that is, $N_{\text{of}} \geq 2$. This rapid convergence is true for the smooth function in particular. The other four factors, do not show any unique correlation to the sampling performance. This is a significant drawback in applying the model in a practical application.

Figures 4.19 to 4.22 show the stages of the adaptive sampling process for the two test functions, using the EA model. We show examples of the good, and poor performances as indicated on the tables. As before, only intermediate steps are shown in some of the cases due to the space constraint.

Table 4.9: Performance of the Evolutionary Algorithm model in sampling the smooth function, sorted by the frequency based measure.

Population size	No. of Offsprings	No. of Bins	Neighborhood	Prob.of Survival	CC with Freq. based Measure	CC with Ent. based Measure
144.00	6.00	16.00	0.10	1.00	0.62	0.89
144.00	1.00	16.00	0.08	1.00	0.62	0.94
144.00	2.00	16.00	0.10	1.00	0.62	0.89
1024.00	6.00	16.00	0.05	1.00	0.62	0.94
64.00	1.00	16.00	0.10	1.00	0.61	0.92
144.00	2.00	16.00	0.10	0.90	0.61	0.89
64.00	1.00	16.00	0.05	0.90	0.61	0.94
144.00	1.00	16.00	0.10	1.00	0.61	0.92
144.00	1.00	16.00	0.05	1.00	0.61	0.94
144.00	2.00	16.00	0.08	0.90	0.61	0.92
4.00	1.00	16.00	0.10	1.00	0.61	0.92
144.00	6.00	16.00	0.10	0.90	0.61	0.88
144.00	1.00	16.00	0.08	0.90	0.61	0.93
4.00	1.00	16.00	0.08	1.00	0.61	0.93
144.00	1.00	16.00	0.08	0.80	0.61	0.92
144.00	1.00	16.00	0.05	0.90	0.61	0.94
1024.00	6.00	16.00	0.08	0.80	0.61	0.92
64.00	1.00	16.00	0.08	1.00	0.61	0.94
144.00	4.00	16.00	0.10	1.00	0.61	0.88
144.00	2.00	16.00	0.10	0.80	0.61	0.89
:	:	:	:	:	:	
:	:	:	:	:	:	
:	:	:	:	:	:	
144.00	6.00	4.00	0.05	0.80	0.15	0.73
64.00	6.00	4.00	0.08	0.80	0.15	0.71
64.00	4.00	4.00	0.05	0.70	0.14	0.69
4.00	2.00	4.00	0.08	0.90	0.14	0.70
144.00	2.00	4.00	0.08	0.80	0.14	0.69
64.00	6.00	4.00	0.05	0.70	0.14	0.67
4.00	1.00	4.00	0.05	0.90	0.14	0.71
64.00	2.00	4.00	0.08	1.00	0.14	0.72
64.00	2.00	4.00	0.08	0.80	0.13	0.69
144.00	6.00	4.00	0.08	0.80	0.13	0.71
64.00	2.00	4.00	0.08	0.90	0.13	0.70
144.00	2.00	4.00	0.05	0.80	0.12	0.74
144.00	2.00	4.00	0.08	0.90	0.12	0.69
144.00	2.00	4.00	0.08	1.00	0.11	0.70
144.00	1.00	4.00	0.05	1.00	0.10	0.69
144.00	2.00	4.00	0.05	0.70	0.09	0.71
64.00	1.00	4.00	0.05	1.00	0.08	0.67
144.00	4.00	4.00	0.05	0.70	0.08	0.70
144.00	6.00	4.00	0.05	0.70	0.07	0.68
4.00	1.00	4.00	0.05	1.00	0.05	0.67

Table 4.10: Performance of the Evolutionary Algorithm model in sampling the smooth function, sorted by the Entropy based measure.

Population size	No. of Offsprings	No. of Bins	Neighborhood	Prob.of Survival	CC with Freq. based Measure	CC with Ent. based Measure
1024.00	2.00	16.00	0.05	1.00	0.60	0.95
64.00	1.00	16.00	0.05	1.00	0.60	0.94
4.00	1.00	16.00	0.05	1.00	0.59	0.94
1024.00	4.00	16.00	0.08	1.00	0.60	0.94
1024.00	6.00	16.00	0.08	1.00	0.60	0.94
144.00	1.00	16.00	0.05	1.00	0.61	0.94
1024.00	2.00	16.00	0.08	1.00	0.60	0.94
144.00	1.00	16.00	0.05	0.90	0.61	0.94
1024.00	4.00	16.00	0.05	1.00	0.60	0.94
64.00	1.00	16.00	0.05	0.90	0.61	0.94
4.00	1.00	16.00	0.05	0.90	0.60	0.94
144.00	1.00	16.00	0.08	1.00	0.62	0.94
1024.00	4.00	16.00	0.02	1.00	0.57	0.94
1024.00	6.00	16.00	0.05	1.00	0.62	0.94
1024.00	6.00	16.00	0.08	0.90	0.59	0.94
64.00	1.00	16.00	0.08	0.90	0.60	0.94
1024.00	2.00	16.00	0.05	0.90	0.60	0.94
1024.00	4.00	32.00	0.08	1.00	0.55	0.94
4.00	1.00	16.00	0.05	0.80	0.60	0.94
64.00	1.00	16.00	0.08	1.00	0.61	0.94
:	:	:	:	:	:	
:	:	:	:	:	:	
:	:	:	:	:	:	
64.00	2.00	256.00	0.02	0.70	0.38	0.68
64.00	6.00	32.00	0.02	0.70	0.35	0.67
64.00	6.00	4.00	0.05	0.70	0.14	0.67
64.00	1.00	4.00	0.05	1.00	0.08	0.67
64.00	6.00	4.00	0.02	0.80	0.38	0.67
64.00	4.00	16.00	0.02	0.70	0.32	0.67
4.00	1.00	4.00	0.05	1.00	0.05	0.67
64.00	6.00	256.00	0.02	0.90	0.37	0.66
1024.00	1.00	4.00	0.10	0.80	0.29	0.66
64.00	4.00	256.00	0.02	0.80	0.33	0.66
64.00	4.00	4.00	0.02	0.70	0.30	0.66
64.00	6.00	16.00	0.02	0.70	0.33	0.65
64.00	6.00	256.00	0.02	0.80	0.41	0.65
1024.00	1.00	4.00	0.05	0.70	0.24	0.64
1024.00	1.00	4.00	0.08	0.70	0.27	0.64
64.00	4.00	256.00	0.02	0.70	0.41	0.63
1024.00	1.00	4.00	0.02	0.70	0.20	0.63
64.00	6.00	4.00	0.02	0.70	0.27	0.62
64.00	6.00	256.00	0.02	0.70	0.37	0.61
1024.00	1.00	4.00	0.10	0.70	0.23	0.57

Table 4.11: Performance of the Evolutionary Algorithm model in sampling the rapidly varying function, sorted by the frequency based measure.

Population size	No. of Offsprings	No. of Bins	Neighborhood	Prob.of Survival	CC with Freq. based Measure	CC with Ent. based Measure
4.00	2.00	16.00	0.05	1.00	0.87	0.83
4.00	2.00	32.00	0.05	1.00	0.87	0.83
64.00	2.00	256.00	0.05	1.00	0.87	0.82
4.00	2.00	256.00	0.05	1.00	0.87	0.84
64.00	2.00	16.00	0.05	1.00	0.87	0.79
64.00	6.00	256.00	0.05	1.00	0.87	0.81
4.00	4.00	16.00	0.05	1.00	0.87	0.82
4.00	6.00	16.00	0.05	1.00	0.87	0.82
64.00	2.00	16.00	0.05	0.80	0.87	0.81
64.00	4.00	256.00	0.05	1.00	0.87	0.82
4.00	4.00	32.00	0.05	1.00	0.86	0.83
64.00	2.00	32.00	0.05	1.00	0.86	0.79
64.00	4.00	16.00	0.05	1.00	0.86	0.78
64.00	4.00	32.00	0.05	0.90	0.86	0.80
64.00	6.00	32.00	0.05	1.00	0.86	0.80
64.00	2.00	16.00	0.05	0.90	0.86	0.80
4.00	6.00	16.00	0.05	0.90	0.86	0.85
64.00	6.00	16.00	0.05	1.00	0.86	0.78
4.00	4.00	32.00	0.05	0.90	0.86	0.86
64.00	4.00	32.00	0.05	1.00	0.86	0.79
:	:	:	:	:	:	
:	:	:	:	:	:	
:	:	:	:	:	:	
4.00	2.00	4.00	0.10	0.80	0.36	0.64
4.00	1.00	4.00	0.05	1.00	0.36	0.65
4.00	1.00	4.00	0.10	1.00	0.36	0.65
144.00	1.00	4.00	0.08	0.90	0.35	0.66
64.00	1.00	4.00	0.02	0.70	0.35	0.66
64.00	1.00	4.00	0.05	0.80	0.35	0.65
64.00	1.00	4.00	0.08	0.90	0.35	0.65
64.00	1.00	4.00	0.02	0.80	0.34	0.65
4.00	1.00	4.00	0.02	0.70	0.34	0.65
144.00	1.00	4.00	0.05	0.80	0.34	0.66
144.00	1.00	4.00	0.02	0.80	0.34	0.65
1024.00	1.00	4.00	0.08	0.70	0.34	0.57
4.00	1.00	4.00	0.02	0.80	0.34	0.65
4.00	1.00	4.00	0.05	0.80	0.33	0.64
144.00	1.00	4.00	0.08	1.00	0.33	0.63
144.00	1.00	4.00	0.05	0.90	0.32	0.63
64.00	1.00	4.00	0.05	0.90	0.32	0.64
64.00	1.00	4.00	0.08	1.00	0.32	0.63
4.00	1.00	4.00	0.08	1.00	0.32	0.63
4.00	1.00	4.00	0.05	0.90	0.32	0.63

Table 4.12: Performance of the Evolutionary Algorithm model in sampling the rapidly varying function, sorted by the entropy based measure.

Population size	No. of Offsprings	No. of Bins	Neighborhood	Prob.of Survival	CC with Freq. based Measure	CC with Ent. based Measure
4.00	1.00	32.00	0.02	1.00	0.76	0.93
64.00	1.00	32.00	0.02	1.00	0.74	0.93
4.00	1.00	256.00	0.02	1.00	0.77	0.92
64.00	1.00	256.00	0.02	1.00	0.72	0.91
144.00	1.00	32.00	0.02	1.00	0.67	0.90
4.00	2.00	32.00	0.05	0.80	0.81	0.89
64.00	1.00	16.00	0.02	1.00	0.72	0.89
4.00	1.00	16.00	0.02	1.00	0.73	0.88
4.00	2.00	256.00	0.05	0.90	0.84	0.88
4.00	6.00	32.00	0.05	0.80	0.82	0.88
4.00	4.00	32.00	0.05	0.80	0.82	0.88
4.00	2.00	256.00	0.05	0.80	0.74	0.88
4.00	6.00	32.00	0.08	1.00	0.79	0.88
144.00	2.00	256.00	0.05	1.00	0.82	0.88
144.00	1.00	16.00	0.02	1.00	0.67	0.88
4.00	6.00	256.00	0.05	0.90	0.83	0.87
4.00	4.00	32.00	0.08	1.00	0.79	0.87
4.00	4.00	32.00	0.05	0.70	0.74	0.87
4.00	2.00	16.00	0.05	0.80	0.81	0.87
4.00	4.00	256.00	0.05	0.90	0.83	0.87
:	:	:	:	:	:	
:	:	:	:	:	:	
:	:	:	:	:	:	
1024.00	1.00	16.00	0.08	0.70	0.41	0.65
1024.00	1.00	4.00	0.05	0.80	0.38	0.65
4.00	1.00	4.00	0.05	0.80	0.33	0.64
4.00	2.00	4.00	0.10	0.80	0.36	0.64
64.00	1.00	4.00	0.05	0.90	0.32	0.64
144.00	1.00	4.00	0.08	1.00	0.33	0.63
1024.00	1.00	4.00	0.02	0.70	0.38	0.63
64.00	1.00	4.00	0.08	1.00	0.32	0.63
64.00	6.00	4.00	0.02	0.80	0.61	0.63
4.00	1.00	4.00	0.08	1.00	0.32	0.63
1024.00	1.00	256.00	0.10	0.70	0.44	0.63
144.00	1.00	4.00	0.05	0.90	0.32	0.63
64.00	4.00	4.00	0.02	0.70	0.60	0.63
4.00	1.00	4.00	0.05	0.90	0.32	0.63
64.00	6.00	4.00	0.02	0.70	0.60	0.62
1024.00	1.00	4.00	0.05	0.70	0.36	0.62
1024.00	1.00	4.00	0.10	0.80	0.37	0.61
1024.00	1.00	16.00	0.10	0.70	0.40	0.61
1024.00	1.00	4.00	0.10	0.70	0.38	0.58
1024.00	1.00	4.00	0.08	0.70	0.34	0.57

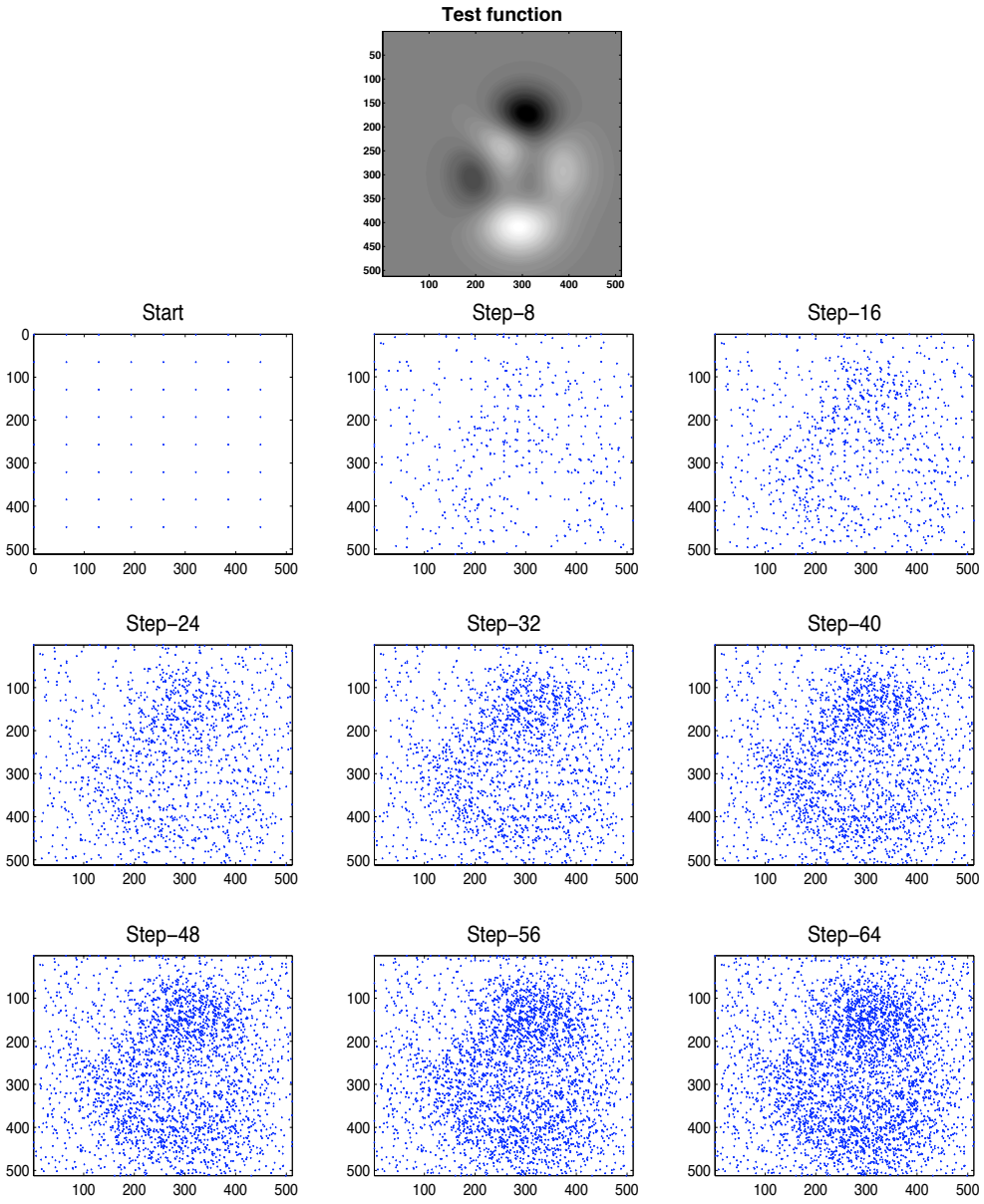


Figure 4.19: Intermediate stages from sampling a smooth test function with the Evolutionary Algorithm model. The good performance is recorded by using the following factors: SOP = 64, N_{of} = 1, N_{ub} = 16, DOP = 0.05 and POS = 1.0. Performance measures = 0.60/0.94. Total of 64 generations.

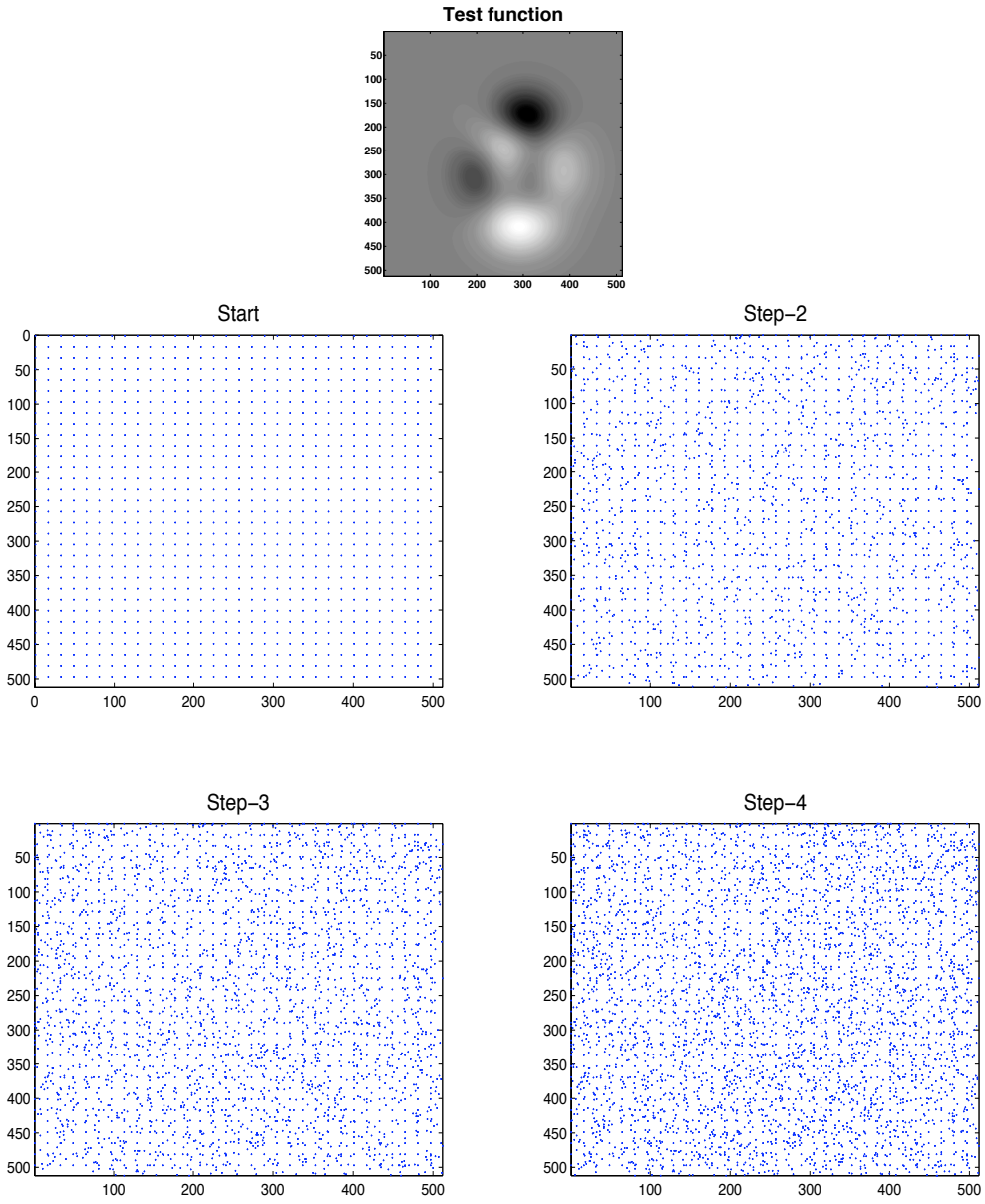


Figure 4.20: All the stages from sampling a smooth test function with the Evolutionary Algorithm model. The poor performance is recorded by using the following factors: SOP = 1,024, N_{of} = 1, N_{ub} = 4, DOP = 0.10 and POS = 0.7. Performance measures = 0.23/0.57. Total of four generations.

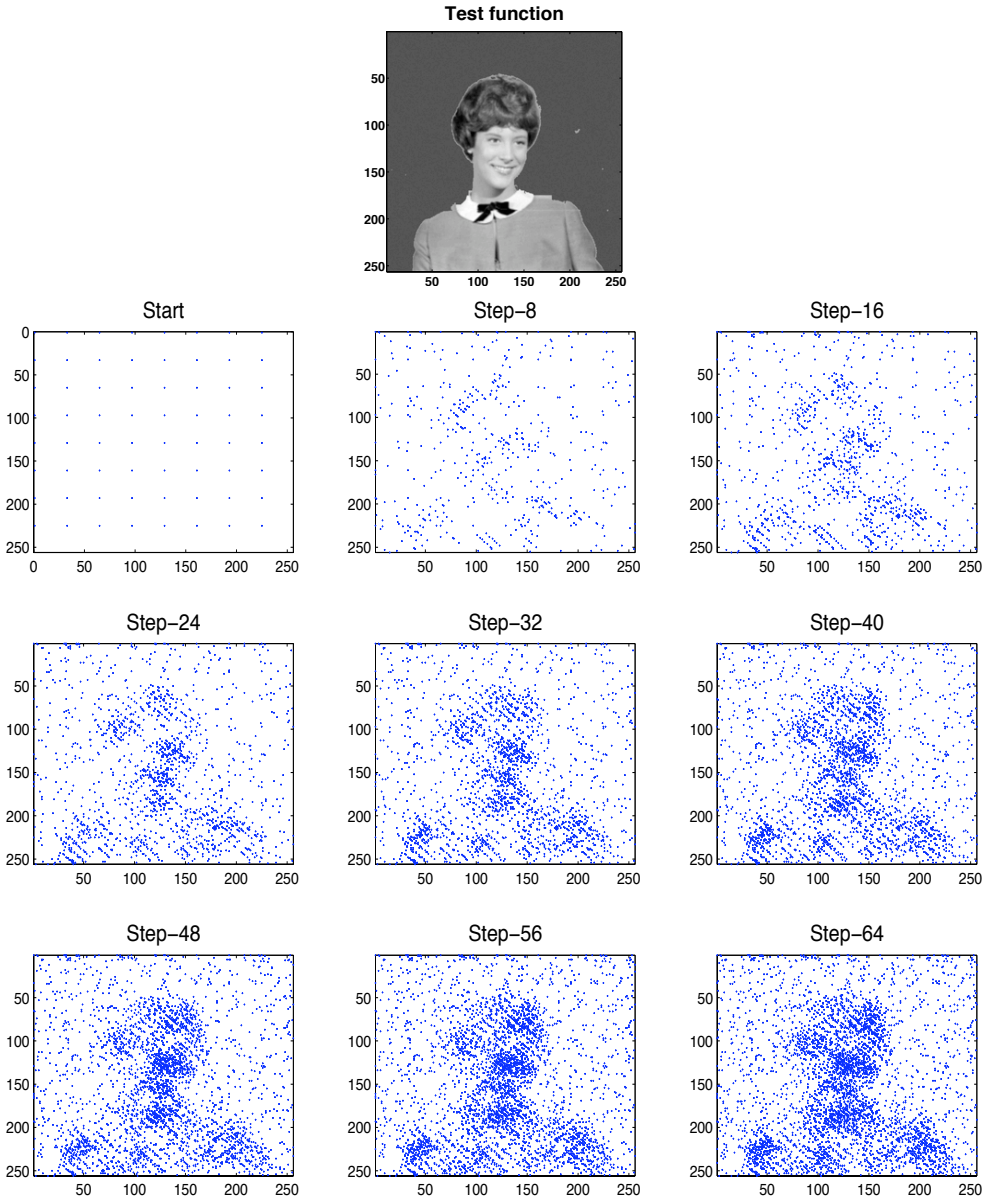


Figure 4.21: Intermediate stages from sampling a rapidly varying test function with the Evolutionary Algorithm model. The good performance is recorded by using the following factors: $SOP = 64$, $N_{\text{of}} = 1$, $N_{\text{ub}} = 32$, $DOP = 0.02$ and $POS = 1.0$. Performance measures = 0.76/0.93. Total of 64 generations.

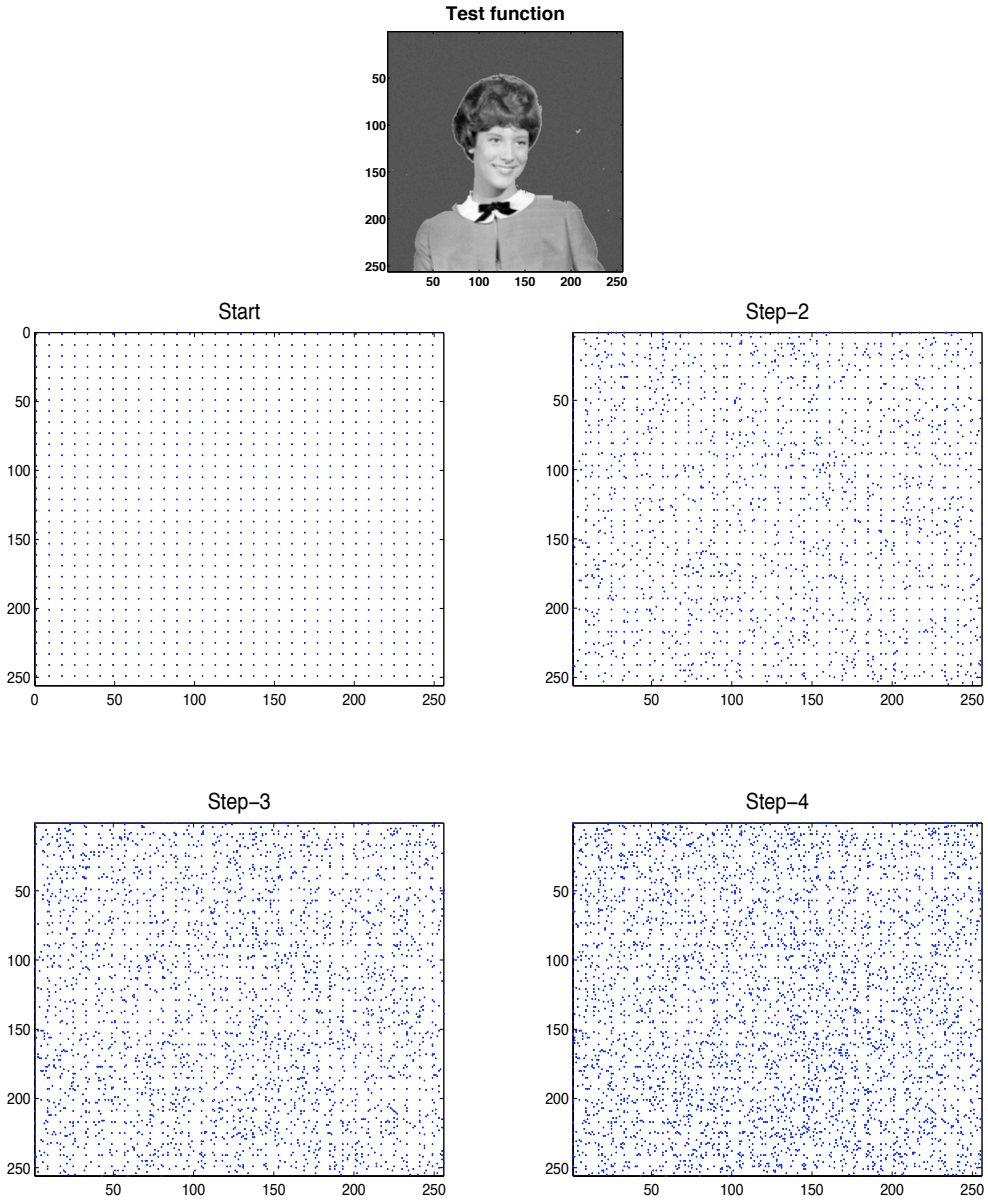


Figure 4.22: All the stages from sampling a rapidly varying test function with the Evolutionary Algorithm model. The poor performance is recorded by using the following factors: SOP = 1,024, N_{of} = 1, N_{ub} = 4, DOP = 0.08 and POS = 0.7. Performance measures = 0.23/0.57. Total of four generations.

The performance of the three models is summarized in Tables 4.13(a) - (c). They show the ranges of factors that results in the good performance recorded for each model. The tables also show corresponding averages of the entropy based performance measure listed in Tables 4.2, 4.4, 4.6, 4.8, 4.10, and 4.12.

Table 4.13: A summary of the results from the analysis of the three sampling models.

(a) Summary results for the active walker model.

	No. of active walkers	**No. of bins**	**Long step**	**Short step**	**Average CC**
Smooth	4 - 144	16 - 32	0.35 - 0.4	0.04 - 0.08	0.92
Rapidly varying	4 - 144	16 - 32	0.30 - 0.4	0.02	0.85

(b) Summary results for the ant model.

	No. of ants	**No. of bins**	**Range of pheromone**	**Range of foraging**	**Average CC**
Smooth	144 - 256	8 - 16	0.02 - 0.25	0.10 - 0.15	0.93
Rapidly varying	144 - 256	8 - 32	0.02 - 0.25	0.10 - 0.12	0.90

(c) Summary result for the evolutionary algorithm model.

	Pop. size	**No. of offspring**	**No. of bins**	**Neighb-orhood**	**Prob. of survival**	**Average CC**
Smooth	4 - 1024	1 - 6	16 - 32	0.02 - 0.08	0.8 - 1.0	0.94
Rapidly varying	4 - 144	1 - 6	16 - 256	0.02 - 0.08	0.7 - 1.0	0.89

The table shows the ranges of input factors that result in good performance, and averages of the entropy based performance values. The factors in the tables are described in Sections 4.3, 4.4, and 4.5. The shown average correlation coefficient (CC) values are the means of the entropy based performance values, which is computed as the CC between the sample density and the entropy in a region of the sample space.

4.6 Further Analysis of the Active Walker Model

Based on our discussions so far, the active walker model is the most useful of the three considered. This is because of our ability to correlate each factor in the model to its performance. The active walker model is thus given further consideration by investigating its scaling properties. We extend our analysis to three dimensions by considering a 3-dimensional test function defined as $(Sinc(x) \times Sinc(y) \times Sinc(z))$, of size $512 \times 512 \times 20$. Figure 4.23 shows a 2-dimensional slice of the function.

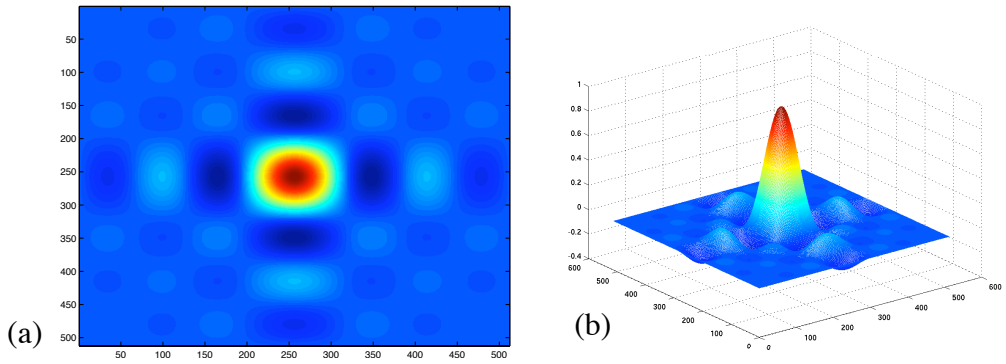


Figure 4.23: *Slice* from 3-dimensional Sinc test function. (a) top view, (b) side view.

We obtain $81,920$ samples, resulting in the same sample ratio of $1 : 64$ used in the test of the 2-dimensional functions considered earlier. Our derived measures of objective performance are extended for the 3-dimensional case. Our subsequent analysis are similar to that done for the 2-dimensional functions. Tables 4.14 and 4.15 show results of our tests, sorted in a manner similar to that done in the 2-dimensional analysis. The number of active walkers was scaled by 20, the size of the third dimension. This is to ensure proper comparison to the 2-dimensional experiment, especially on the basis of the number of steps taken by the active walkers. Note the similarities between the factors that result in good performance for the 2-, and 3-dimensional functions. This is an indi-

cation that the performance of the active walker model does not change appreciably with change in number of dimensions.

4.7 Conclusions

The findings of the analysis of the three models employed in implementing the ASHE algorithm are summarized as follows:

1. Generally, the entropy measure indicates that all three models performed better than the frequency based measure. This is indicated by the higher positive CC values. The frequency based measure is limited by the small spatial sample used in the experiments. This results in poor frequency resolution.

2. Using both measures of performance, the EA, and ant models performed marginally better than the active walker model.

3. The ant, and EA models show no apparent correlation between one or more factors, and their performance. This makes it difficult to come up with a combination of factors that are appropriate for a particular application.

4. We are able to establish correlations, separately, between the active walker model factors, and the performance of the model. This makes it possible to establish general "rules of thumb" in its application.

5. The active walker model is more robust since there is always a possibility for walkers to sample in all regions of the space all through the sampling process. In the process of sampling using the ant and EA models, some regions may be completely excluded due to good solutions obtained from other regions. This is similar to obtaining a local minimum.

Table 4.14: Performance of Active Walker model in sampling the 3-dimensional Sinc function, sorted by the frequency based measure.

No. of Walkers	No. of Bins	Long step	Short step	CC with Freq. based Measure	CC with Ent. based Measure
1280.00	8.00	0.40	0.04	0.97	0.97
2000.00	8.00	0.35	0.04	0.97	0.97
2000.00	8.00	0.40	0.04	0.97	0.97
2880.00	8.00	0.40	0.04	0.97	0.97
1280.00	8.00	0.35	0.04	0.97	0.97
2880.00	8.00	0.35	0.04	0.97	0.97
2880.00	8.00	0.30	0.04	0.97	0.96
80.00	8.00	0.40	0.04	0.96	0.96
80.00	8.00	0.35	0.04	0.96	0.96
1280.00	8.00	0.30	0.04	0.96	0.96
80.00	8.00	0.30	0.04	0.96	0.96
2000.00	8.00	0.30	0.04	0.96	0.96
80.00	8.00	0.40	0.06	0.96	0.95
2000.00	8.00	0.25	0.04	0.96	0.95
2880.00	8.00	0.25	0.04	0.96	0.95
1280.00	8.00	0.25	0.04	0.96	0.95
80.00	8.00	0.35	0.06	0.95	0.95
1280.00	8.00	0.40	0.06	0.95	0.95
1280.00	8.00	0.35	0.06	0.95	0.95
2880.00	8.00	0.40	0.06	0.95	0.95
:	:	:	:	:	:
:	:	:	:	:	:
:	:	:	:	:	:
2000.00	32.00	0.25	0.10	-0.14	-0.15
20480.00	8.00	0.20	0.10	-0.20	-0.20
2000.00	256.00	0.20	0.10	-0.21	-0.20
2880.00	256.00	0.20	0.10	-0.24	-0.23
1280.00	256.00	0.20	0.10	-0.26	-0.25
80.00	256.00	0.20	0.10	-0.28	-0.27
20480.00	16.00	0.20	0.10	-0.31	-0.31
20480.00	32.00	0.20	0.10	-0.36	-0.36
2000.00	8.00	0.20	0.10	-0.44	-0.45
2880.00	8.00	0.20	0.10	-0.46	-0.46
1280.00	8.00	0.20	0.10	-0.47	-0.47
80.00	8.00	0.20	0.10	-0.49	-0.49
2880.00	16.00	0.20	0.10	-0.60	-0.60
2000.00	16.00	0.20	0.10	-0.60	-0.60
1280.00	32.00	0.20	0.10	-0.61	-0.60
2880.00	32.00	0.20	0.10	-0.62	-0.61
1280.00	16.00	0.20	0.10	-0.63	-0.62
80.00	32.00	0.20	0.10	-0.63	-0.62
2000.00	32.00	0.20	0.10	-0.64	-0.62
80.00	16.00	0.20	0.10	-0.65	-0.64

Table 4.15: Performance of Active Walker model in sampling the 3-dimensional Sinc function, sorted by the entropy based measure.

No. of Walkers	No. of Bins	Long step	Short step	CC with Freq. based Measure	CC with Ent. based Measure
1280.00	8.00	0.40	0.04	0.97	0.97
2000.00	8.00	0.35	0.04	0.97	0.97
2000.00	8.00	0.40	0.04	0.97	0.97
2880.00	8.00	0.40	0.04	0.97	0.97
1280.00	8.00	0.35	0.04	0.97	0.97
2880.00	8.00	0.35	0.04	0.97	0.97
2880.00	8.00	0.30	0.04	0.97	0.96
80.00	8.00	0.40	0.04	0.96	0.96
80.00	8.00	0.35	0.04	0.96	0.96
1280.00	8.00	0.30	0.04	0.96	0.96
80.00	8.00	0.30	0.04	0.96	0.96
2000.00	8.00	0.30	0.04	0.96	0.96
2000.00	8.00	0.25	0.04	0.96	0.95
2880.00	8.00	0.25	0.04	0.96	0.95
1280.00	8.00	0.25	0.04	0.96	0.95
80.00	8.00	0.40	0.06	0.96	0.95
80.00	8.00	0.35	0.06	0.95	0.95
1280.00	8.00	0.40	0.06	0.95	0.95
2000.00	8.00	0.40	0.06	0.95	0.95
1280.00	8.00	0.35	0.06	0.95	0.95
:	:	:	:	:	:
:	:	:	:	:	:
:	:	:	:	:	:
2000.00	32.00	0.25	0.10	-0.14	-0.15
2000.00	256.00	0.20	0.10	-0.21	-0.20
20480.00	8.00	0.20	0.10	-0.20	-0.20
2880.00	256.00	0.20	0.10	-0.24	-0.23
1280.00	256.00	0.20	0.10	-0.26	-0.25
80.00	256.00	0.20	0.10	-0.28	-0.27
20480.00	16.00	0.20	0.10	-0.31	-0.31
20480.00	32.00	0.20	0.10	-0.36	-0.36
2000.00	8.00	0.20	0.10	-0.44	-0.45
2880.00	8.00	0.20	0.10	-0.46	-0.46
1280.00	8.00	0.20	0.10	-0.47	-0.47
80.00	8.00	0.20	0.10	-0.49	-0.49
2000.00	16.00	0.20	0.10	-0.60	-0.60
2880.00	16.00	0.20	0.10	-0.60	-0.60
1280.00	32.00	0.20	0.10	-0.61	-0.60
2880.00	32.00	0.20	0.10	-0.62	-0.61
1280.00	16.00	0.20	0.10	-0.63	-0.62
80.00	32.00	0.20	0.10	-0.63	-0.62
2000.00	32.00	0.20	0.10	-0.64	-0.62
80.00	16.00	0.20	0.10	-0.65	-0.64

6. Further experiments with the active walker model indicate that it scales well.

In addition to the foregoing, it is also straightforward to extend the active walker model into sampling in higher dimensions. The locations in an n-dimensional space are defined as vectors of length n, and distances moved by the active walkers are computed as simple vector operations. These are the reasons for the choice of the active walker model for the application discussed in Chapter 6.

CHAPTER 5

NATURE, USES AND SYNTHESIS OF HYPERSPECTRAL IMAGES

Here, we give a description of the nature, and uses of hyperspectral images. Next, we state the need for image synthesis, and describe the process in general. Finally, the process of generating a database of hyperspectral images is described, and some of the results are shown.

5.1 Nature of Hyperspectral Images

Hyperspectral images are *cubes* of data, with each value in the *cube* representing the electromagnetic energy response from an imaged scene, at a particular wavelength. Two of the dimensions in the *cube* are spatial, and the third dimension is spectral. That is, each spectral component, called a *band*, is made up of a 2-dimensional spatial image. The bands in a hyperspectral image are contiguous, and occupy a region of the electromagnetic spectrum. For example, an image with contiguous spectral bands of wavelengths in the micrometer range will be a hyperspectral infrared (HSI) image because of its location on the electromagnetic spectrum. Assuming that the bands are not completely correlated, integration of data in more than one band will result in increase in information about an image. There is usually a level of independence between the bands, and this results in a spectral signature for each spatial pixel. That is, an imaging device will record varying responses at the different wavelengths in the hyperspectral image. These responses depend on the intrinsic nature of the imaged material, thus a unique signature is recorded for each material. The information in the spectral signature is particularly useful for, but not limited to situations, in which there is a limitation on the spatial resolution that can

be obtained. Figure 5.1 shows an example of a hyperspectral image, and illustrates the foregoing about their nature.

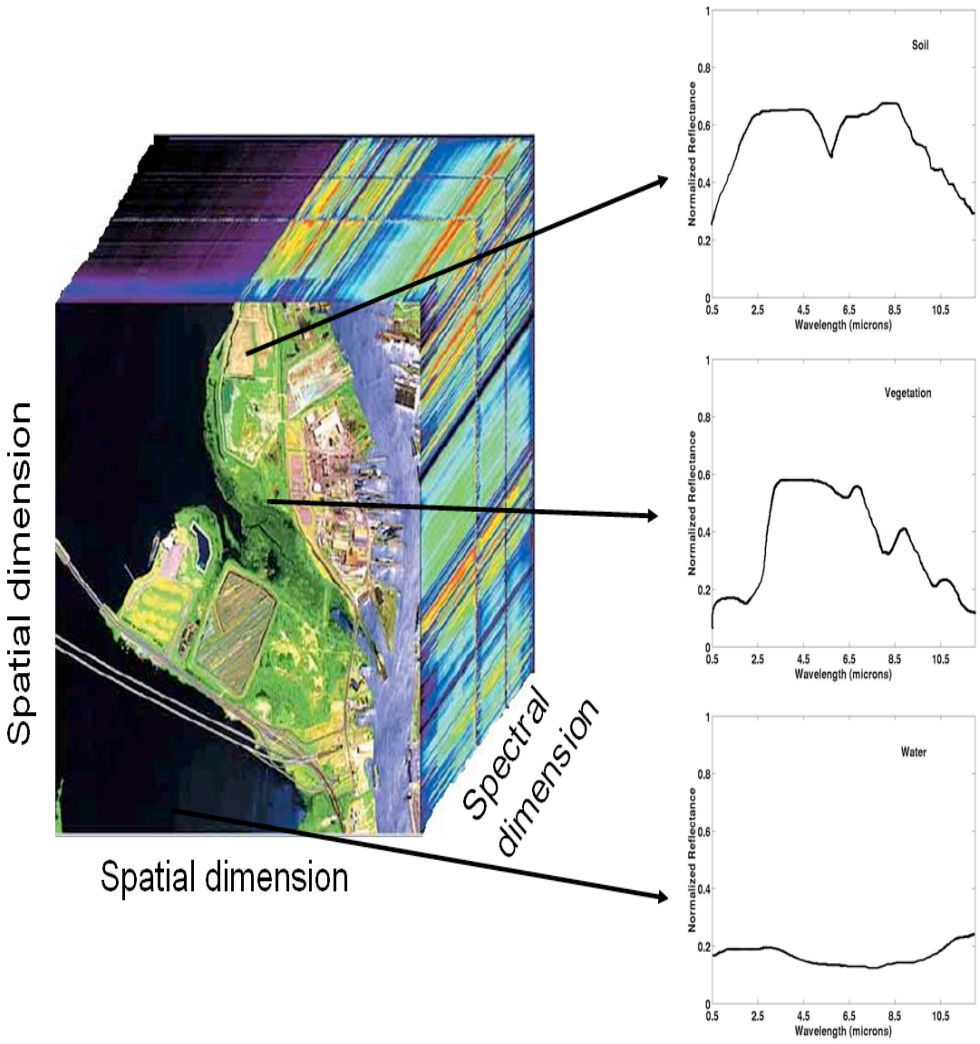


Figure 5.1: Example hyperspectral image, and material signatures. (Source: IEEE Signal Processing Magazine, Vol.19(1), 2002.)

Using multi-spectral Automatic Target Recognition (ATR) algorithms, objects in hyperspectral scenes that only span one pixel in the spatial dimensions or are even sub-pixel, can be identified from their spectral signatures. Generally, the approach of multi-spectral

ATR algorithms focuses on the spectral rather than the spatial information in the images [50]. There are numerous military [7, 44], and non-military [89, 90, 5, 93, 4] uses of multi-, and hyperspectral images.

5.2 Synthesizing Hyperspectral Images

Hyperspectral images obtained for military purposes are generally not available in the public domain because of security reasons. Even when images are available, they usually do not exist in the quantity or with the specifications required by many applications. A solution to this problem is to synthesize images with these required specifications. Synthetic images have been used as aid in the design and development stages of imaging sensors by providing an avenue to pre-evaluate the imaging products from the sensors [54, 73]. They also serve as test data for algorithm design, either because of the lack of real data [1, 80], or to augment the available real data [77]. Some examples of Synthetic Image Generation (SIG) models are the Strategic High Altitude Atmospheric Radiance Code (SHARC) [9], full spectrum scene simulator (MCScene) [70], and the Digital Imaging and Remote Sensing Image Generation (DIRSIG) [76]. All these models generate scenes by tracing rays between a simulated imaged scene, and an imaging sensor. Models of the intervening space between these two are included in the ray-tracing process.

The images we require are used primarily in a military application described in Chapter 7. The DIRSIG model has been used extensively in military applications because of the good radiometric fidelity of the images it generates. We thus decided to use this model for image synthesis.

5.2.1 The Digital Imaging and Remote Sensing Image Generation model (DIRSIG)

The model is an integrated collection of first principle based sub-models that account for scene geometry, atmospheric contributions, illuminating sources, and properties of materials in the imaged scene. After these factors are established, a ray-tracing process is employed in rendering the scene. It has be been used for generating high spatial and spectral, multi- or hyperspectral images in the 0.3 to 20 micron region [76]. The following is a brief description of some of the components of, and tools used by DIRSIG. A full description can be found in the DIRSIG manual [10].

Scene

This is a 3-dimensional space, and comprises of terrains and objects. Each of these consists of single or multiple facets. Associated with each facet in a scene are pre-defined radiometric properties obtained from experimenting with different materials. These properties determine the response from the surfaces as recorded by the imaging sensor. The shape, and number of facets on an object is fixed, but the user is allowed to associate any material with a facet. The user is also allowed to define the 3-dimensional location, size, and orientation of objects in the scene. For the imaging geometry, the relative positions of the sensors and scene can be specified in the 3-dimensional coordinate system, or in terms of distances from sensor to scene and angles relative to a reference. This allows for all the practical imaging geometries that may be needed.

Sensors

All the sensors modeled in DIRSIG are passive. This means that they register the energy that is reflected from an external source, or the energy that is emitted from the object itself. Some examples are frame cameras, and line scanners. Example parameters that

may be set for these sensors are focal length, flight paths (for sensors mounted on moving carriers), number of scan lines, and number of samples per line.

Radiometry

DIRSIG uses the MODerate spectral resolution atmospheric TRANsmittance (MODTRAN) algorithm and computer model [6] for its radiometric computations. It utilizes bidirectional reflectance data, and accounts for specular and diffuse background contributions. It also models length dependent extinction and emission properties of plumes, clouds, targets, and backgrounds. In summary, it models the intervening space between an imaged scene and a sensor. Based on this model, a database or lookup table of values is computed for each pixel in every spectral band. MODTRAN has a current limitation of 2cm^{-1} spectral resolution.

Ray Tracing

A ray tracing component utilizes the geometry information to generate a list of facets intersecting a given pixel. This is combined with information from the the radiometry model, and used in the radiance computations.

Other Software Tools

The DIRSIG comes packaged with an image viewing software called FREELOOK. This is used for previewing, and for spectral analysis of the generated hyperspectral images.

5.3 Image Synthesis with DIRSIG

The images we synthesized are used as aid in the development of Automatic Target Recognition (ATR) algorithms. Specifically, they serve as test images, used in the eval-

uation of the performance of ATR algorithms developed for military applications. The required images are Forward-Looking Infrared Images (FLIR). The database is generated according to the following specifications for each image:

1. Sensor type: single-shot images are required, thus a framing array sensor is used. This sensor's focal length is set to 50 mm.

2. Imaging geometry: the sensor is placed at a stand-off from the imaged scene in a forward-looking arrangement. The distance between the sensor and the imaged scene is 2 km. The sensor is elevated at 50 m above the imaged scene to give a larger field of view.

3. Spatial resolution: each band is of size 512×512 pixels. The spatial resolution is computed by using similar triangles. This is computed based on the DIRSIG default image length and breadth of 24748.7 microns, a framing array sensor of focal length 50 mm, and the distance between the sensor and imaged scene of 2 km. This results in a resolution of 1.93 meters.

4. Spectral span and resolution: the images range in wavelength from $8 - 13$ microns, with 40 nanometer steps between bands. This results in 126 bands per image.

Based on the stated use of the images, it is required that there is diversity in the database with respect to ATR performance. That is, images of varying degrees of difficulty should be represented in order that the ATR algorithms are adequately tested. We attempt to manually include such diversity, by varying the following factors in the imaged scene:

1. Time of day: we generate images for two different times of the day. These are 3.00 AM in the morning and 3.00 PM in the afternoon. Changes in this factor will

generally result in radiometric changes.

2. Clutter: all objects different from the target of interest are considered clutter. This includes all objects and background that can be mistaken for, or hinder in the detection of a target of interest. We introduce clutter of varying types, and in different quantities into the scenes. DIRSIG has models for both man-made clutter such as fuel drums and tents, and natural clutter such as trees and hilly terrains.

3. Target: we generate some images with a military truck as target, and others with an armored tank.

A combination of all these results in 216 hyperspectral images. Figure 5.2 shows the combination of factors that result in the database. The information containing all the image specifications are included in configuration text files required as arguments by DIRSIG for execution.

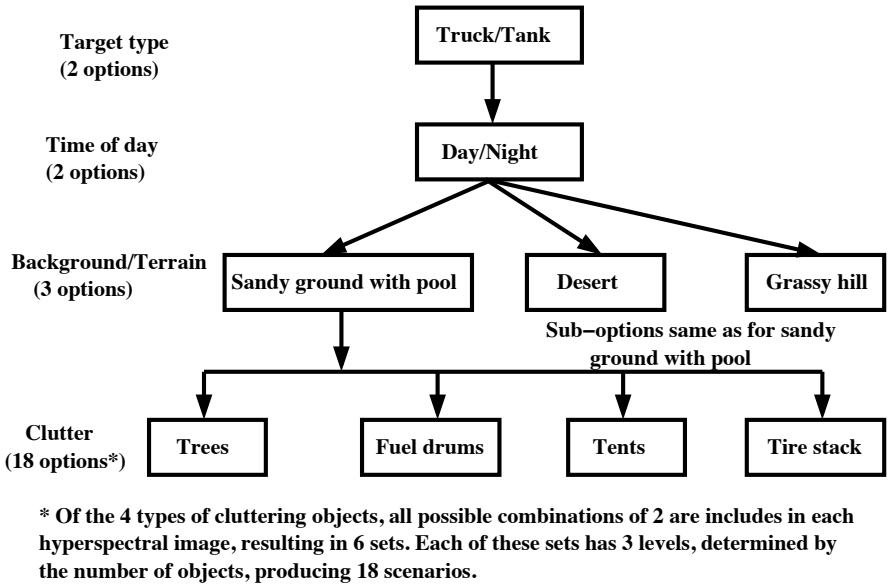


Figure 5.2: Combination of factors used to generate images in synthesized database. Total of 216 images.

Figures 5.3 - 5.5 show some example images from the synthesized database. Figure 5.6 shows example spectral signatures from some of the images.

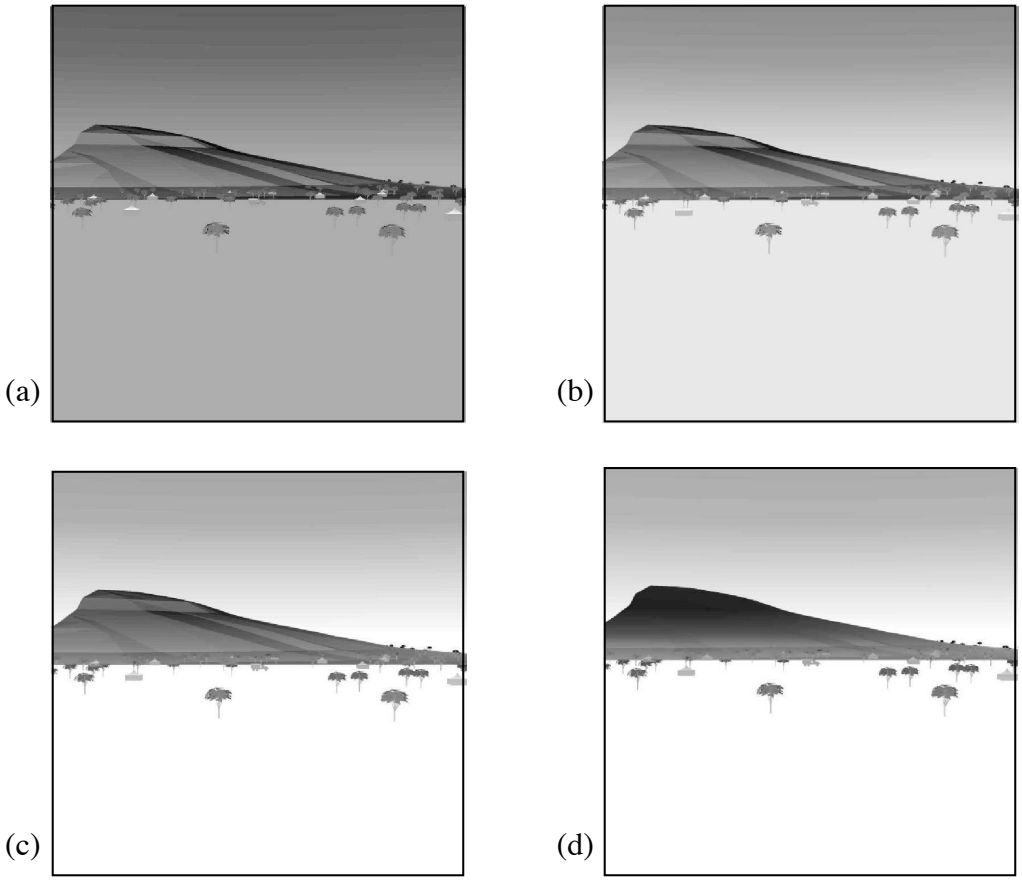


Figure 5.3: Hyperspectral image scene with a target truck on a flat surface with a hilly background. The cluttering objects are tents and trees. The wavelengths of the shown bands are (a) $\lambda = 8$ microns, (b) $\lambda = 10.6$ microns, (c) $\lambda = 11.96$ microns, (d) $\lambda = 12.76$ microns.

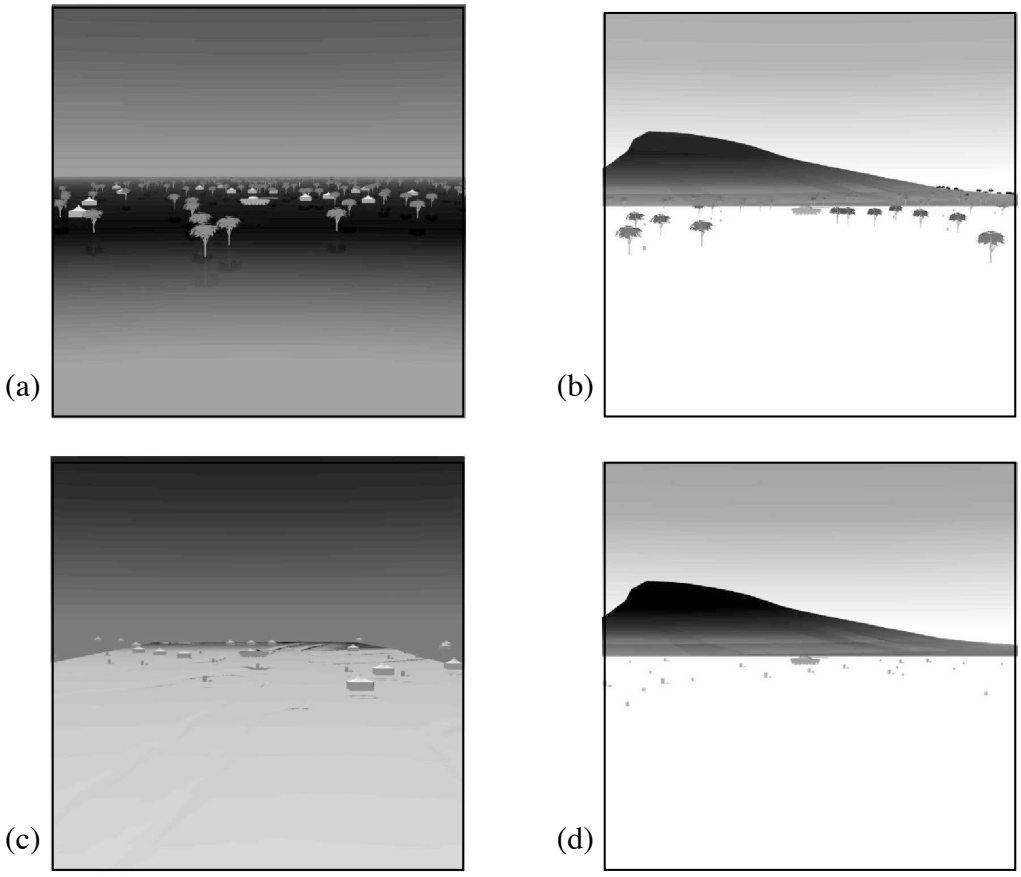


Figure 5.4: Images of different bands from different scenes, with the same target tank. (a) flat sand ground with trees and closed tents as cluttering objects, $\lambda = 8.36$ microns, (b) flat ground with hilly background, trees, and fuel drums as cluttering objects, $\lambda = 12.76$ microns, (c) desert terrain with closed tents and tire stacks as cluttering objects, $\lambda = 8.36$ microns, and (d) flat ground with hilly background, tire stacks and fuel drums as clutter, $\lambda = 12.76$ microns.

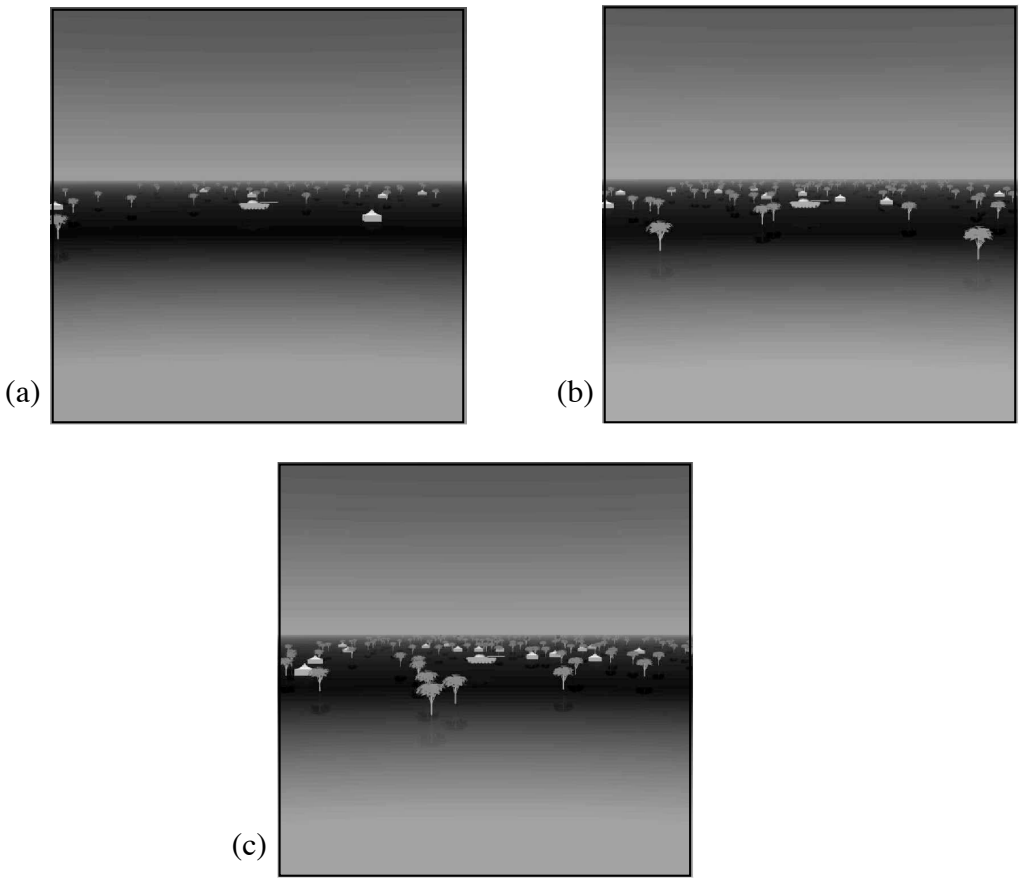


Figure 5.5: Example images of the same scene, with varying levels of clutter, determined by the quantity. (a) low clutter, (b) medium clutter, and (c) high clutter.

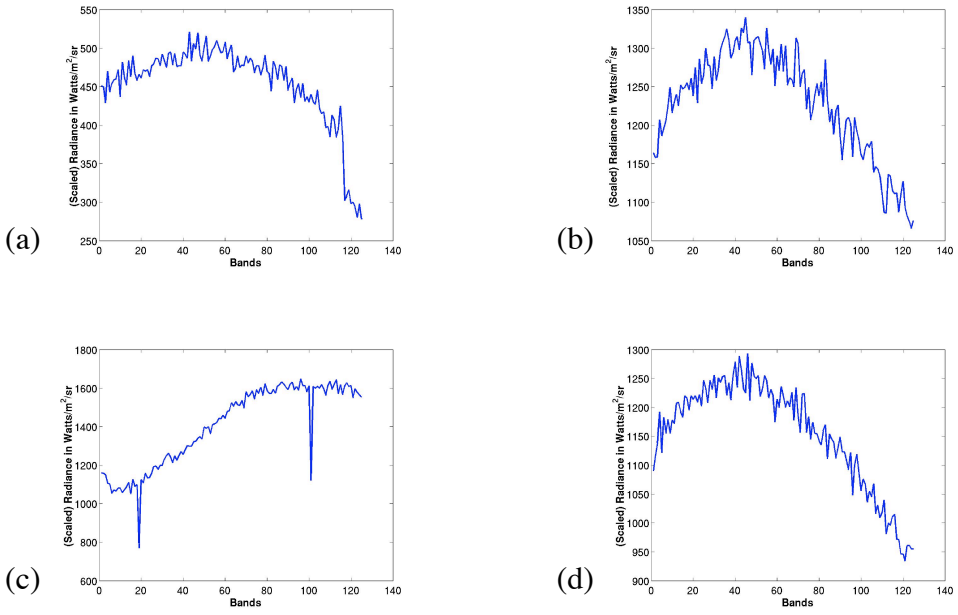


Figure 5.6: Examples of spectral signatures of different materials in the synthesized hyperspectral scenes. (a) side of hilly background, (b) side of closed tent, (c) flat sandy ground, and (d) metal front of truck.

In conclusion, it is important to note that the image synthesis process is computationally expensive. Each hyperspectral image in the database took about 150 minutes to synthesize on a 3.2 GHz Pentium IV processor machine, with 2 GB of memory. The whole database creation took about 540 hrs. DIRSIG stores each pixel as a float. Each synthesized hyperspectral image is of size 512×512 pixels $\times$ 126 bands $=$ 126MB.

CHAPTER 6

EFFICIENT HYPERSPECTRAL IMAGE SYNTHESIS USING ASHE

In this chapter, we present a more efficient approach to the hyperspectral image synthesis process described in Chapter 5. This approach is based on the Adaptive Sampling by Histogram Equalization (ASHE) algorithm. As mentioned in the hyperspectral image synthesis discussion, our aim is to generate a set of images utilized in the performance evaluation of Automatic Target Recognition (ATR) algorithms. In general, data analysis of any sort requires adequate, and a statistically representative population of the dataset in question in order to make reliable inferences. Two specific requirements of our synthesized images are:

1. Fidelity of each image. This depends on the ability of the synthetic image generation system to adequately model, and reproduce the complex interactions that exist in a real scene. There is continuous research work aimed at developing this ability [74]. It is however, beyond the scope of this work.

2. Representation in all categories of ATR difficulty in the database.

The latter requirement is the focus of this application, and it ensures that the ATRs in question are evaluated for all levels of target detection and recognition. This is an important requirement for drawing an unbiased, and conclusive inference about the performance of ATRs. The following sections describe the process of image synthesis based on the ASHE algorithm, and present some results.

6.1 ASHE based Image Synthesis

We model each generated image as a function of multiple factors, each image is thus a point in the multi-dimensional space. Some of these factors, such as time of day, are described in Section 5.3. Each synthesized image is thus a result of combining these factors as inputs to the DIRSIG model. Joining these points obtained from all possible combinations of factors yields a surface in the space. There is usually no prior knowledge of how a particular combination of conditions will affect the performance of an ATR. Without such knowledge, the typical approach is to generate images for a random combination of factors, or to generate images for combinations of factors that are evenly spaced within their possible ranges. These approaches are inefficient for situations in which there are varying slopes in the described multi-factor space. Also, due to the computational complexity of hyperspectral image synthesis described in Section 5.3, a *brute force* approach, which requires the generation of images from all combinations of factors is not feasible. Other approaches such as the gradient based search are also not feasible for the same reason.

The optimal reconstruction of such a surface from a limited number of points will result from concentrating relatively more points in regions of rapid image variation. Thus, it is desired to generate images for values, or ranges of values of these factors that are significant for change in target recognition difficulty. As shown in the description of the ASHE algorithm in Chapter 2, sampling this surface in this manner results in a distribution of ATR performance values that tends towards the uniform. Thus, sampling the surface to maximize diversity in values indicative of ATR performance, results in efficient sampling of the surface. We employ the ASHE algorithm, using the active walker model to sample the surface. The decision to use the active walker model is based on our

conclusions in Section 4.7. It is straightforward to extend the described 2-dimensional version of the model to higher dimensions. More importantly, we are able to establish appropriate input parameters

6.1.1 Imaged Scene

For this experiment, we generate images according to the urban scene from the DIRSIG manual [10]. A single band from this scene, spatial size 128×128 pixels, is shown in Figure 6.1.

Figure 6.1: A single band ($\lambda = 0.56$ nm) from the hyperspectral image of the urban scene. The spatial size is 128×128 pixels. The arrow indicates the region cropped as target.

6.1.2 Input Factors to DIRSIG

We generate images from the visible to near infrared ($0.35 - 1.0$nm) regions of the electromagnetic spectrum. We identify factors that will generally result in radiometric changes, and thus spectral signatures, for this spectral range. Some of these are time of day, day of year, visibility parameter, aerosol type parameter, wind speed, and the parameter representing the modeled atmospheric profile. The significance of each of these factors is described in detail in the DIRSIG manual. We place a further constraint on the factors utilized in image synthesis. The extra requirement is that the image synthesis based on ASHE only utilizes factors that consist of ordered sets. This ensures that a move in any single dimension generally results in an increase or decrease in the radiometric effect from that factor. This criterion excludes the parameter for the aerosol type, and atmospheric profile. These are unordered sets, and the implication is that an active walker's movement in these dimensions is random. There has to be a correlation between the step sizes of the active walker in the input parameter space, and their sample contribution to the distribution. The wind speed factor was excluded based on further experience with the image synthesis process. Our image synthesis is thus based on the following three factors:

- Time of day ($1 - 24$ hours)
- Month of year ($1 - 12$)
- Visibility parameter ($0 - 40$ km)

6.1.3 Baseline ATR Performance

In order to utilize the ASHE algorithm in the image synthesis process, we need to associate a value, indicative of baseline ATR performance with each image. The ASHE

algorithm then attempts to equalize the distribution of these values, as the sampling process progresses. We establish this through the performance of an idealized ATR. We implemented a normalized, multi-spectral matched filter ATR via the Adaptive Coherence Estimator (ACE).

$$\mathbf{ACE_{statistic}} = \frac{|\mathbf{s}^T\hat{\mathbf{R}}_b^{-1}\mathbf{x}|^2}{(\mathbf{s}^T\hat{\mathbf{R}}_b^{-1}\mathbf{s})(\mathbf{x}^T\hat{\mathbf{R}}_b^{-1}\mathbf{x})} \tag{6.1}$$

The ATR uses a spectral signature of a target in question as a template. The resulting ACE statistic is bounded between 0 and 1, and it is expressed in (6.1), where $\mathbf{s} \in \mathbb{R}^L$ and $\mathbf{x} \in \mathbb{R}^L$ are the target template and pixel under test respectively, and L is the number of bands in the hyperspectral image. The vectors $\mathbf{s}$ and $\mathbf{x}$ may also be composed of multiple pixels in the spatial dimension. In this case, 2-dimensional averages of the target and test pixels are taken in the spatial dimensions to obtain column vectors of the previously stated lengths. $\hat{\mathbf{R}}_b$, with dimensions $L \times L$ is an estimate of the covariance matrix of the background [57].

This ATR is idealized since it uses a hyperspectral image target template that is cropped from the scene. A 3×3 pixel target is cropped from the area indicated by the arrow in Figure 6.1. A 2-dimensional average of this is taken in the spatial dimension to obtain a vector of length $L =$ Number of bands, as described earlier. The false alarm rate at a particular threshold is an indication of the baseline ATR performance for a scene. The same threshold is used for all scenes to obtain this baseline performance. Note that this indicated performance is specific for the target. The use of different targets may result in a different false alarm rates at the same threshold. This is common practice, since most practical ATR algorithms are evaluated based on the detection of specific targets using the known target template. The diversity in the synthesized images is thus with respect to a particular target.

6.1.4 Image Synthesis

The arguments to DIRSIG are contained in a series of parameter files. These files contain the values of the factors that determine the nature of the synthesized images among other information. We keep other factors constant while varying the values that make up the multi-dimensional space as needed. The details of using the Adaptive Sampling by Histogram Equalization (ASHE) algorithm to achieve adaptive sampling are shown in Algorithm 4. In summary, the ASHE algorithm attempts to equalize the histogram of the baseline ATR performance values obtained from the synthesized images. The algorithm is implemented with a MATLAB script. DIRSIG and ancillary programs that are used for synthesizing the images are also called from MATLAB.

6.2 Experiments

We synthesize images by keeping all other factors that contribute to variation in the imaged scene constant while varying the three factors identified in Section 6.1.2. We synthesize a set of images using a random combination of these factors, and another set using combinations of factors that are evenly spaced within their possible ranges. These are compared to the set of images generated by the set of factors determined by the ASHE algorithm. The following are used in the active walker model in implementing ASHE: $N_{\text{aw}}= 5$, $LSP= 0.3$, and $SSP= 0.04N_{\text{aw}}= 5$. These are based on the results from our analysis in Sections 4.3 and 4.6. Each of the sets consists of 125 images, of spatial size 128×128 pixels, and 44 equally spaced spectral bands spanning $0.35 - 1.0$ nm. Each image took about 26 minutes to synthesize on a Linux workstation with a 3.2 GHz Pentium IV processor.

The baseline ATR performance values are also computed for the sets of images syn-

Algorithm 4 . Synthesizing hyperspectral images using the ASHE algorithm

Initial definitions:

Objective function - Baseline ATR performance
Factors that the Objective function is dependent as identified in Section 6.1.2
Range and possible values that these factors can take, also listed in Section 6.1.2

Sampling initialization:

Obtain initial random locations in n-dimensional space using active walkers
Synthesize images for combination of factors from these locations
Compute Baseline ATR performance from initial sample image points
Compute normalized histogram from initial sample performance values
Compute Overall Fitness Criterion OFC

While no. of synthesized images $\leq$ required number of images **do**
For all active walkers **do**
Obtain new sample point in multi-dimensional space
If location has already been sampled
Obtain alternate close sample point
end if
Synthesize new image based on active walker position
(*DIRSIG arguments are coordinates of active walker position*)
Add new image sample from active walker to existing images
Compute baseline ATR performance for new image addition
Compute new normalized histogram of performance values, and
Compute New Fitness Criterion NFC
If $NFC < OFC$
Single walker takes *short* step size in random direction
Else Single walker takes *long* step size in random direction
End if
End for
Compute new overall normalized histogram
Compute OFC
End while

thesized by a random combination of these factors, and those synthesized using combinations of factors that are evenly spaced within their possible ranges. These image sets are then compared to the adaptively synthesized images on the basis of representation across the range of performance values. This is determined as the range between the minimum, that is zero false alarm rate, to the maximum of all performance values recorded from the three methods used for image synthesis. By representation, we refer to each bin having at least one image so that an ATR algorithm test on the database would have considered all levels of difficulty. The images are also considered based on the distribution among the different levels of difficulty. That is, a measure of the uniformity in the distribution of images across the different levels of difficulty so that ATR algorithm tests are not biased by over-representation in a particular category of difficulty.

Figure 6.2 shows histograms indicating the spread of representation over the defined baseline ATR performance range, and the levels of representation for each performance value. There are 106 possible performance values in the range. As shown by the count of the number of bins with at least one image representation, the image set generated using the adaptive algorithm show representation of more ATR performance values than the other two methods. Note that none of the methods produce images that have performance values between 0 and 33. This is due to the threshold value used to determine the false alarm rate for the images. A higher value will result in lower baseline ATR performance values for all three methods.

Also, a comparison of the normalized versions of these histograms to a normalized uniform distribution with the same number of bins, shows that there is a more even distribution of the ATR performance values from the image set obtained using the ASHE algorithm. We use the fitness criterion given in (3.2) as an objective measure of this. Thus, the lower the value of the deviation, the more the distribution tends towards the

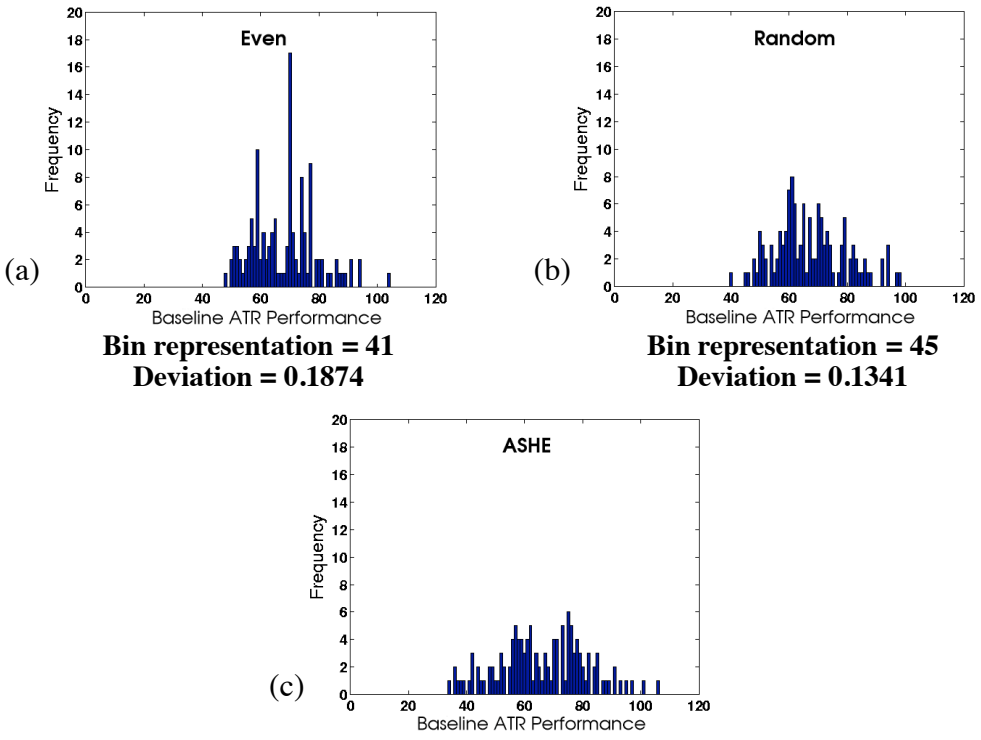


Figure 6.2: Distribution of baseline ATR performance values. Representation for images obtained from (a) combinations of evenly spaced factors, (b) random combination of factors, and (c) combination of factors obtained based on the ASHE algorithm. The bin representation is the count of bins that have at least one image, there are 106 bins in all. The deviation values are computed in a similar manner to the fitness criterion described in (3.2) earlier.

uniform. The lack of representation in the range of values between 0 and 33 diminishes the improvement recorded by using the ASHE algorithm. This is noted in the recorded deviation values. Excluding the range in the computation will make the recorded improvement more apparent.

CHAPTER 7

DEVELOPMENT OF A MEASURE OF CLUTTER FOR HYPERSPECTRAL IMAGES

In this chapter, we present our main application, which is the development of a measure of clutter for hyperspectral images. An image is said to be cluttered if some of its background and other objects may be mistaken by an Automatic Target Recognition (ATR) algorithm as the desired target. The quantity, locations and nature of these objects will determine the clutter level in the image. Motivations for characterizing and quantifying clutter in images include:

- a need to compare ATR performance on a common objective basis [83],
- the need for a measure to form the basis for a pre-processing step to discard images, or make a decision on further processing,
- the need for a measure to form the basis for a post-processing step to determine the reliability of the result of running an ATR on a scene, and
- the inverse-problem problem of creating clutter on ground scenes. e.g. camouflaging.

Such a measure of clutter will indicate the inherent difficulty for an ATR algorithm to detect targets. That is, a means to determine the degree of difficulty to detect and identify a target in a scene.

Since it is difficult to capture the multifaceted nature of image clutter in a single number, our aim is to obtain bounds on the performance of any ATR on a scene based on

a general clutter quantification scheme. That is, a high value of this quantity will indicate that any ATR will produce a high false alarm (FA) rate. A low value may, however, not result in a low FA rate. This will depend on the exact nature of the ATR.

Previous works attempting to characterize or quantify clutter in images include [94, 66, 58, 45, 51, 91, 81]. However, all these works focus on deriving clutter measures for single-band images. To the best of our knowledge, no research effort has addressed the problem of deriving a clutter measure for complete hyperspectral images.

Next, we describe our approach for developing this measure in its general form. Then, we present results from obtaining the measure for single band images, and for multi-band hyperspectral images. We also present specific applications of the derived measure in both cases.

7.1 Clutter Complexity Measure

In its general form, our approach is to obtain an aggregation of statistical image features or metrics that correlates best with baseline ATR performance. We use the terms 'features' and 'metrics' interchangeably. We compute metrics that fulfill the following criteria from the images:

1. Descriptive of scene parametric variation and significant for ATR performance.
2. Computing them only requires *a priori* information on the order of spatial extent of the target in the scene at the most.
3. Algorithmically uncomplicated, and easy to implement.

These are similar to the requirements listed in [66]. We then obtain a value indicative of baseline ATR performance from the images, and obtain the measure as an aggregation

of these metrics that correlates best with this performance. We call the derived value the Clutter Complexity Measure (CCM).

The process of combining these metrics to yield the required result is obtained through a training process on a subset of available image data. Once established, this is generalized over the complete dataset. This training process requires image data in numbers that are statistically significant. As stated earlier in Section 5.2, the availability of these images is limited. Thus, we synthesize test images as described in Chapter 5. Generalization of the derived measure from a random subset of images requires that there is a good representation of the values indicative of the ATR performance in each subset. It also requires that the range of these values is represented in the test images. Fulfillment of these requirements is improved by synthesizing images based on the ASHE algorithm. This is described in Chapter 6.

7.2 Clutter Complexity Measure for Single Hyperspectral Bands using Real Data

An ATR can utilize a combination of the information in the separate bands that make up the hyperspectral image of a particular scene. Intuitively, using multiple bands of the same scene for the purpose of target detection should yield fewer false alarms for the same probability of detection P_d when compared to using a single band. The computational resources needed by the ATR increases with each additional band, resulting in the need for an efficient selection of the bands utilized by the ATR. It will be beneficial to be able to select the bands that contain the required target information surrounded by clutter of low complexity. These fewer bands can then be used in the multiple band detection with results comparable to using all the available bands.

The clutter complexity measure (CCM) of a band can indicate the bands utility for detection. That is, bands are prioritized by their clutter complexity measure. Thus, an L

band detector will use the L bands with the least clutter complexity in the hyperspectral cube. In the following sections, we describe the process of obtaining a CCM for the bands in a hyperspectral image, and present results from experimenting with the derived CCM. The test images for the single band analysis are real, forward looking infrared (FLIR) images.

7.2.1 Baseline ATR Performance

We establish the required baseline ATR performance described in Section 7.1 by using the RX algorithm [68]. This is an anomaly detector that is capable of integrating data for multiple bands. In summary, the RX algorithm determines how much a region is different from its surrounding region relative to the arithmetic mean and variance of the pixels in this surrounding region. For each pixel, the RX algorithm computes a statistic given by:

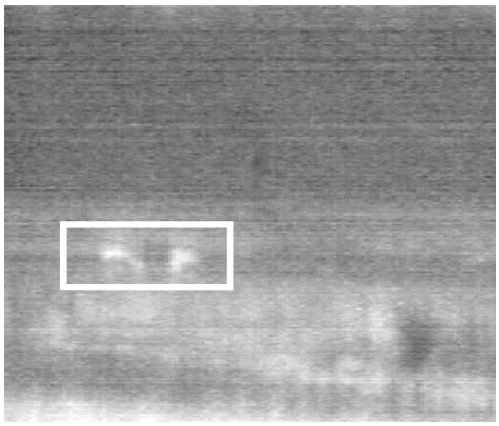
$$S = (\mathbf{x} - \mathbf{u})' \mathbf{R}^{-1} (\mathbf{x} - \mathbf{u}) \ , \tag{7.1}$$

where

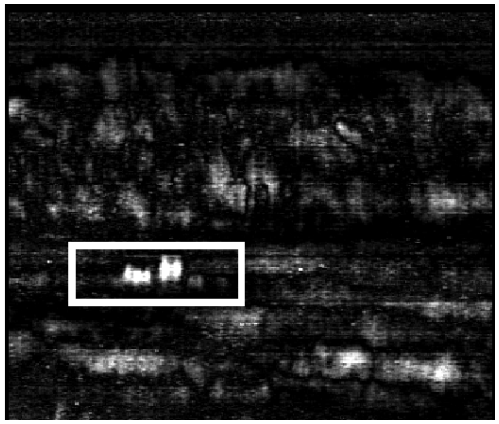
$$\mathbf{u} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i, \qquad \mathbf{R} = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{x}_i - \mathbf{u})(\mathbf{x}_i - \mathbf{u})',$$

and $\mathbf{x}_i$ is the vector of pixels from a surrounding annular ring of length N. The computed statistic S in (7.1) is then compared to a threshold to determine the presence of a possible target.

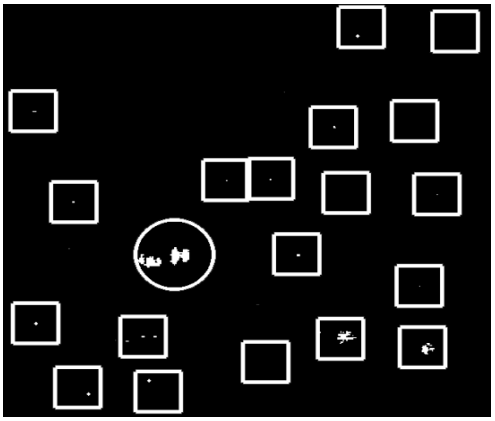
Figure 7.1(a) shows an HMMWV military vehicle at a stand-off of 1.2 km in a forward looking infrared (FLIR) scene and Figure 7.1(b) shows the result of running the RX detector on the scene. The white portions in Figure 7.1(b) indicate high values of S and black low values. These patches are clustered together for the purpose of detection and counting false alarms. The threshold of the RX algorithm is set so that $P_d = 1$ and



(a)



(b)



(c)

Figure 7.1: RX detection in FLIR images: (a) Original image (b) Image of RX statistic (c) Resulting detection image.

Table 7.1: List of some of the image statistical features used in deriving the clutter complexity measure.

Feature Name	Description
FBM Hurst Parameter	Texture roughness
Standard Deviation	Global standard deviation
Schmieder Weathersby	Average local standard deviation
Homogeneity	Average pixel variation
Energy	Average histogram energy
Entropy	Average histogram entropy
Target Interference Ratio	Average contrast
Outlier Ratio	Average percentage of outliers

the false alarm count is minimized for all the experiments. The vehicle was detected in the region marked with a circle in Figure 7.1(c). The other clustered regions that contain S values greater than or equal to the threshold are marked with squares in the detection image, these constitute false alarms.

7.2.2 Multiple-feature CCM

We obtained a clutter complexity measure as a weighted sum of statistical image features. In [45], the measure was formed by the eight features listed in Table 7.1. In addition to these, we also used five variations of the Gaussian based decomposition of images obtained by analysis-by-synthesis [8] resulting in the use of 13 statistical image features in all. We attempt to obtain the weighted sum of these 13 image features that correlates best with the performance of the RX algorithm. We computed the RX false alarm counts over a set of training images containing a given target. The false alarm count resulting from these were trained for a partition of FLIR images. Then, a set of weights were obtained that resulted in the best correlation between these false alarm counts and weighted sums of the image features. This approach is similar to that in [45] with the exception of

using the single-band RX algorithm instead of template matching to determine the ATR performance bounds.

In order to evaluate the generalization of the derived weights, we would need to obtain comparable results from them on a different partition. The work in [45] showed that such weights are not independent of the target in an image so the target in the partitions have to be the same. Given the same target, if other objects in the scene are altered, the clutter complexity should still be able to predict the ATR performance. As a result, the clutter complexity measure should yield good correlation with ATR performance bounds for a disparate set of test images that include the same target object. Our experiments only obtained such weights resulting in a good correlation on a per partition basis. That is, our training and test data set were the same. This is due to the limited number of real image data that was available for the training process described earlier. Hence, the use of synthesized hyperspectral images, described in the subsequent experiments.

7.2.3 Single-feature CCM

To avoid the questions raised by the inadequate training to obtain required weights, we tested each of the statistical image features that made up the weighted sums to see if any of them had good correlation to the false alarm count rate for all the images. Such correlation in all the image sets suggests that the statistical image feature is a good indicator of complexity. The important distinction between this and the multiple-feature clutter complexity measure is that there is no need for training. We chose the feature with the best average correlation to the false alarm count rates for all the images.

7.2.4 Single-Band CCM Experiments

Our test data set comprises of 5, 28-band (*cubes*) of forward looking hyperspectral images taken with the same polarization of 90^0 and wavelength ranging from $460 - 1,000$ nm in steps of 20 nm. Each of these 28 bands represents the same target at the same pose and stand-off distance. These criteria also formed the basis of the partitioning in [45]. Due to a computer memory constraint in running the multiple band RX algorithm, alternate bands are chosen resulting in 14 bands for the experiments. The choice of alternate bands ensures the availability of a good spread of information across the hyperspectral bands.

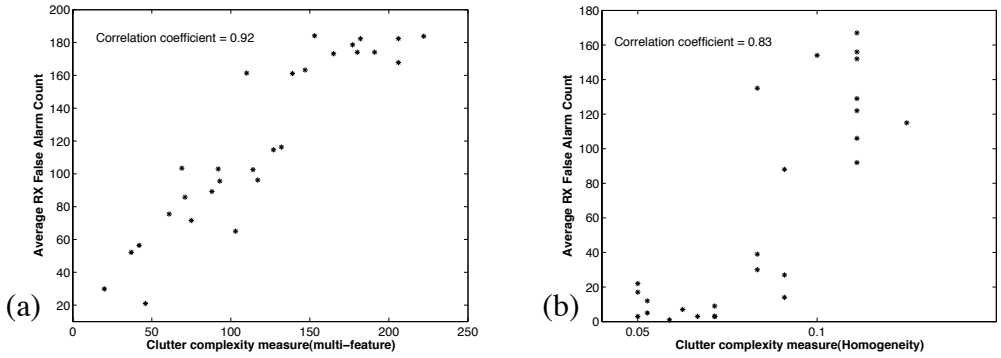
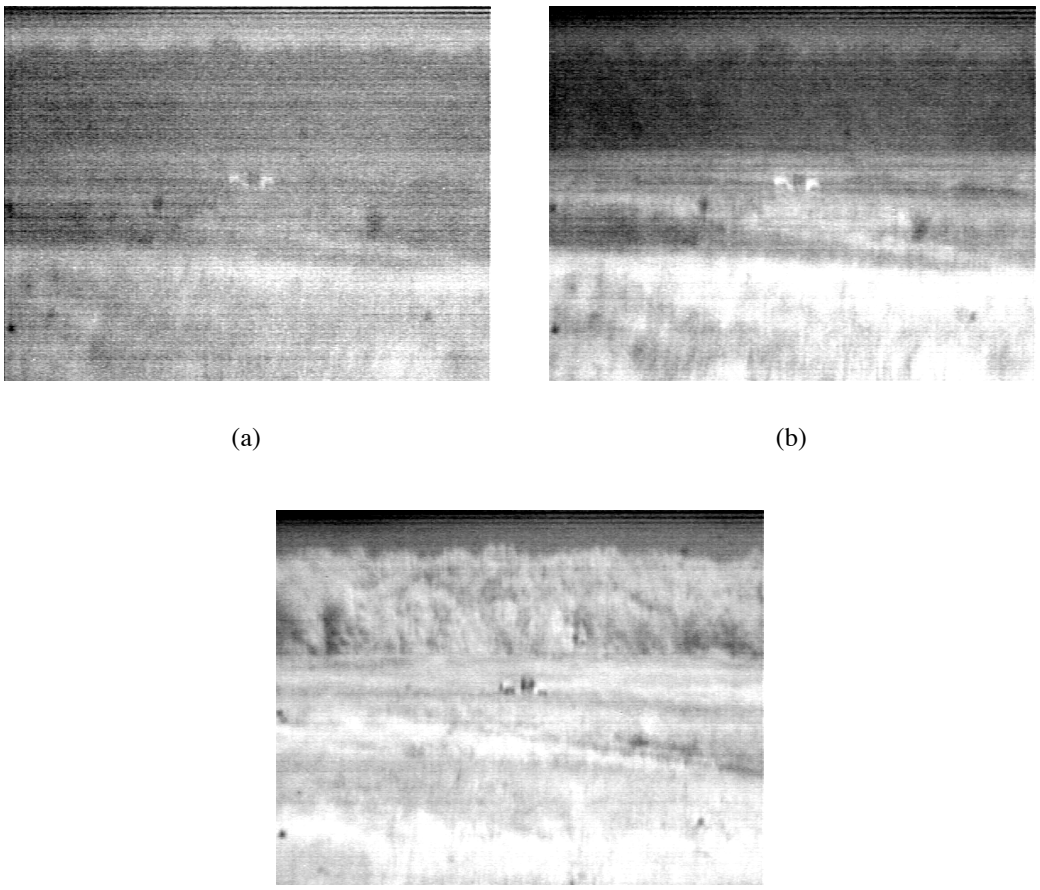


Figure 7.2: Scatter plot of clutter complexity measure and false alarm count: (a) Weighted sum of multiple features (b) Single feature.

Figure 7.2(a) shows the good correlation obtained between a weighted sum of features, i.e. the clutter complexity measure, and the false alarm count for a partition of images. Figure 7.2(b) also shows good correlation between the false alarm count and the chosen single image statistical feature (homogeneity). These results are typical for all the test data and suggest that these measures are good indicators of complexity in our test images. Figure 7.3 shows examples of images with low, medium and high complexity as classified using the weighted sum clutter complexity measure.



(a) (b)

(c)

Figure 7.3: FLIR band classification by clutter complexity measure: (a) Low clutter complexity number $= 38.84$ (b) Medium clutter complexity number $= 73.56$ (c) High clutter complexity number $= 112.91$

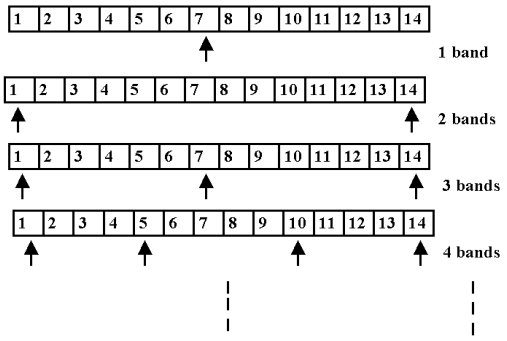


Figure 7.4: Example of band selection based on wavelength to ensure a uniform distribution of the choice of hyperspectral bands.

To test the utility of the clutter complexity measure for band selection, we start running the RX detector using a single band. More bands are then added with the choice of each extra band based on one of the following: clutter complexity measure, derived from (1) single, (2) multiple features, and (3) wavelength. The ordering by wavelength is done in order to ensure a uniform spread of the choice of bands over all available wavelengths as shown in Figure 7.4.

Figure 7.5 shows a plot of the average false alarm count over the five *cubes* of hyperspectral data against the number of bands used by the ATR for three scenarios over all 14 bands. To obtain an optimal subset of k bands, all the possible combinations of the 14 bands are considered. This was done for $k = 1$ to 5 resulting in $3,472$ ATR experiments for each set. A plot of the average false alarm count for the optimal choice of 1 to 5 bands is also shown in Figure 7.5. The probability of detection (P_d) was set to 1 for all experiments. The false alarm counts shown are obtained by averaging over the five hyperspectral images used for the experiment.

As expected, the false alarm count reduces as more bands are added for all experiments. This shows that the information in multiple bands of the hyperspectral data are complementary. The target information adds up more rapidly than the information in the

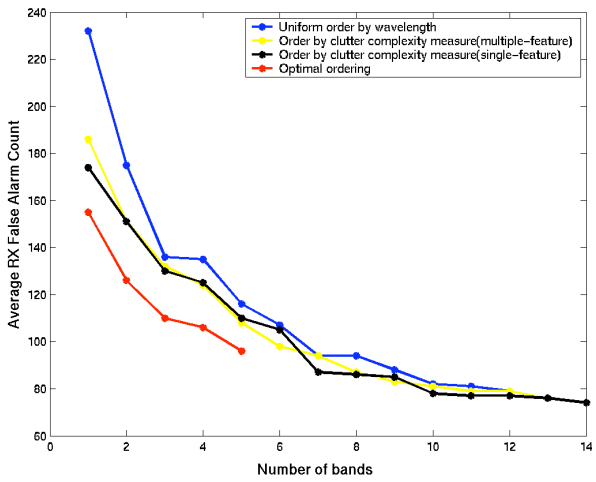


Figure 7.5: Performance of clutter complexity measures represented as average false alarm count versus number of bands integrated into the RX detector.

surrounding clutter, resulting in fewer false alarm counts for the same P_d.

The average false alarm count is less for the bands ordered using our derived clutter complexity measure compared to when the bands are ordered by wavelength. This indicates that the clutter complexity measure criteria for band selection results in an improvement in the performance of the ATR. How much of an improvement is seen by comparing the result of ordering on the basis of the clutter complexity measures to the optimal ordering. The false alarm count for the optimal ordering is about 30% less than ordering by the clutter complexity measures for 1 to 5 bands. This is about the same improvement of the ordering by clutter complexity over the uniform ordering by wavelength which does not take any clutter information into account.

There is a rapid drop in false alarm count for all the experiments from 1 to 3 bands. A *knee* is seen when between 4 to 7 bands are utilized by the ATR and there is little improvement after the use of 7 bands. The basis for ordering the bands becomes less important as more bands are added for the detection process beyond 8 bands. The curves

in Figure 7.5 merge as expected when all 14 bands are used.

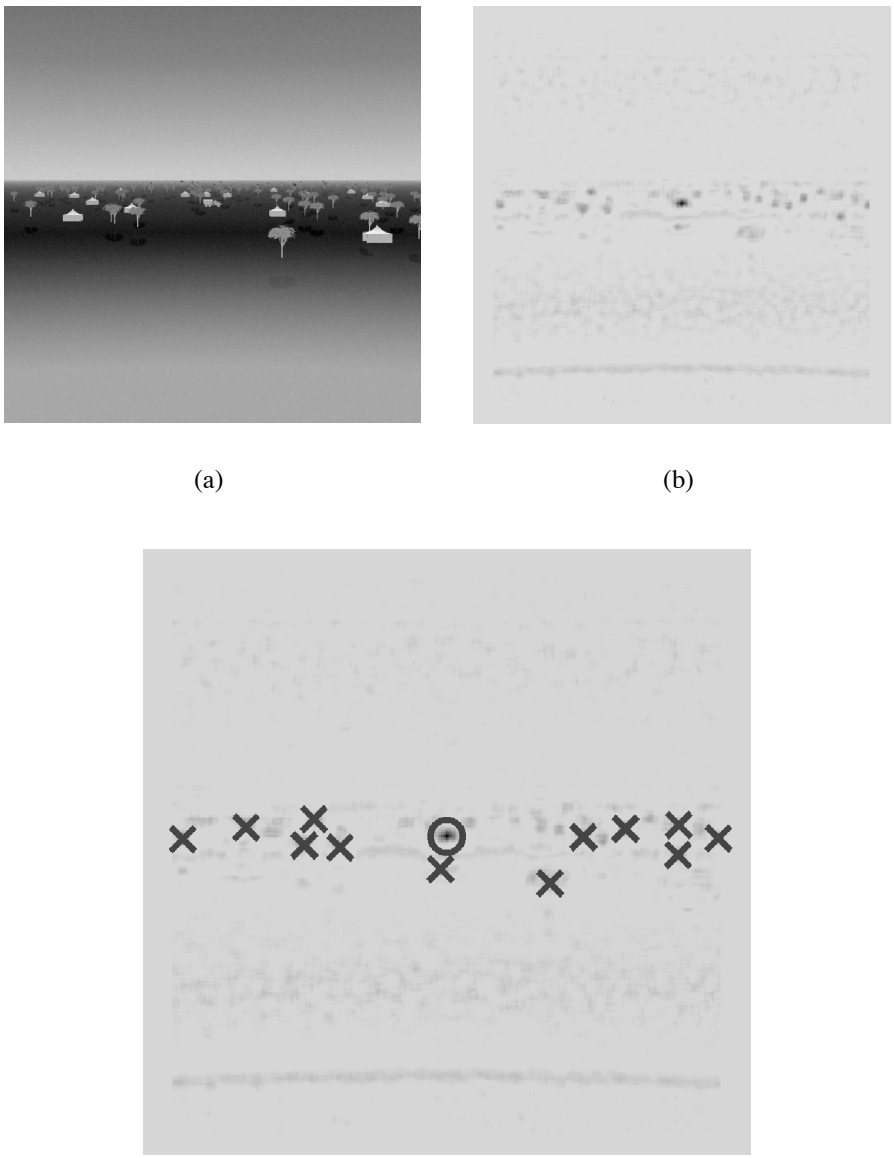
The derived CCM for single bands is shown to be a useful criterion for choosing bands in a multi-band ATR detection. It is noted that the performance of the single-feature and multi-feature clutter complexity measures are comparable. Clutter complexity measures derived from relevant multiple features should generally be more reliable than that from a single feature because of the usual multi-faceted nature of clutter. The next set of experiments reports work on deriving such multiple-feature clutter complexity measures for complete hyperspectral image *cubes*.

7.3 Clutter Complexity Measure for Hyperspectral Images

The previous experiments established the feasibility of our approach to obtain a CCM. Our goal is to establish such a measure for complete hyperspectral images. We follow the same approach outlined in Section 7.1. Here, we use synthesized hyperspectral infrared (HSI) images as our test data. We are able to follow the described training process because we have synthesized images in statistically significant numbers. We describe the process for deriving a CCM for hyperspectral images, and present subsequent experiments and results.

7.3.1 Baseline ATR Performance

We establish ATR baseline performance by utilizing an idealized implementation of a normalized, multispectral matched filter ATR via the Adaptive Coherence Estimator (ACE). This was described fully in Section 6.1.3. Its application in deriving a baseline ATR performance is similar. Figure 7.6 shows one of the bands from an example synthetic hyperspectral image *cube*, the resulting ACE statistic image, and the final detection image after thresholding. The statistic image is on a gray scale, with black (0) - detection



(a) (b)

(c)

Figure 7.6: Target detection in HSI using ACE: (a) Band ($\lambda = 8.40$ microns) from HSI image (b) Image of ACE statistic (c) Detection image, with 'O' representing the target and 'X' the false alarms.

with certainty, and white (1) - no detection at the two extremes. The target was detected in the location marked by the 'O' and the 'X's indicate false alarms in Figure 7.6(c).

7.3.2 Image Clutter Metrics

The utilized image features or metrics also fulfill the requirements outlined earlier in Section 7.1. They fall into two broad groups of features derived from hyperspectral single bands, and those derived from the complete hyperspectral cube.

Metrics Derived from Single Bands

The image clutter metrics that were used in [62] and [23] were mostly based on statistical features of the images. We implemented these and computed them for each band of the hyperspectral images. In addition to these, we also computed a metric based on parameters derived from Gabor filtering of the hyperspectral image bands. The Gabor filter extracts edges from an image at different orientations [85]. Two parameters are derived from these filtered images: the first parameter, p, is an indication of the distinctness and frequency of edges in the filtered image, while the second, c, is related to the range of pixel values in the image. All these fall under the category of single-band clutter metrics. To extend these for hyperspectral images, we compute distribution representative values like maximum, minimum, mean, median, and range for each metric resulting in five hyperspectral clutter metrics derived from each single-band metric.

Metrics Derived from Hyperspectral Image

Image clutter metrics were also computed directly from the hyperspectral image *cube*. A metric was derived from the correlation between the hyperspectral bands in an image. Generally, lower correlation between the bands signifies more unique information in each

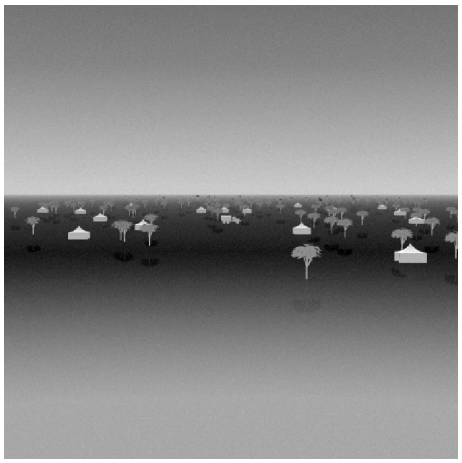
band, resulting in the better performance of multispectral ATR algorithm.

Two metrics were also computed using the vector of the pixels in the spatial dimensions along the spectral dimension. The length of the vector is equal to number of bands. One of the metrics we computed was based on the dot product between a pixel vector and the surrounding pixels. A high value indicated that the pixel vectors are from a homogeneous region, and dissimilar otherwise. The other one was based on the Kullback-Leibler distances which is the relative entropy between a pixel and its surrounding pixels [82]. Hereby, each pixel vector is modeled as a distribution, and the distance is a measure of the difference between a pixel and another. Thus, pixels in homogeneous regions will result in lower values for this metric. Finally, we derive a set of image clutter metrics from Gray Level Co-occurrence Matrices (GLCM) as proposed in [35]. This method has been used for texture characterization in images [36]. We extend the spatial-spatial offsets implemented in single-band image processing into the spectral dimension. We also experiment with a variant of the GLCM in which, the pixel locations are randomly chosen over the whole hyperspectral image cube. The five metrics derived from each variant of the GLCM are maximum value, energy, entropy, contrast and homogeneity. A more detailed description of the clutter metrics is contained in Appendix A.

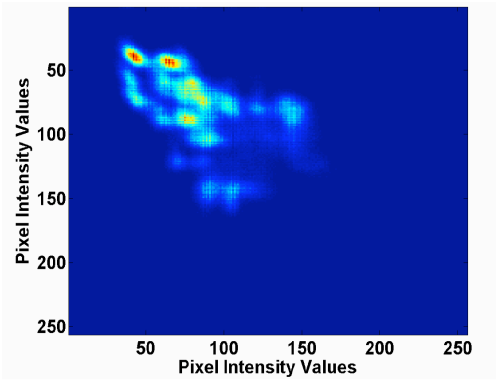
A summary of the image clutter metric categories and brief descriptions are shown in Table 7.2. We implemented a total of 129 clutter metrics. We show some examples of these metrics in Figures 7.7 and 7.8. The hyperspectral image clutter metrics derived from these were described earlier.

7.3.3 Determining Significant Metrics

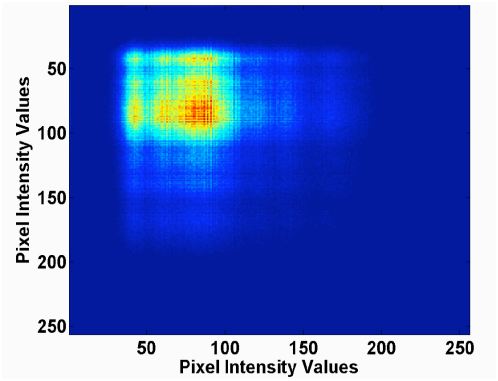
A factor analysis scheme is implemented to remove clutter metrics that are not significant for ATR performance, and to reduce redundancy among the remaining. This will result in



(a)

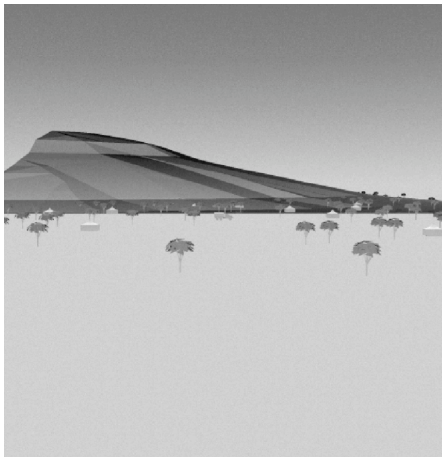


(b)



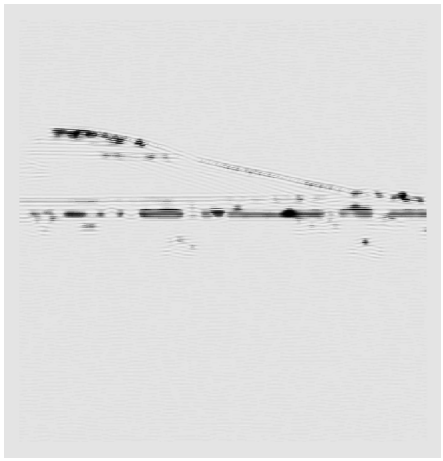
(c)

Figure 7.7: Gray Level Co-occurrence Matrices from hyperspectral images: (a) Single band from test image, (b) GLCM using offset based on target size and (c) GLCM using random pixel locations.



(a)

(b)



(c)

Figure 7.8: Gabor filtered band from hyperspectral image: (a) Hyperspectral image band, (b) Gabor filtered image, filter at 15° orientation, extracts near vertical edges, and (c) Gabor filtered image, filter at 90° orientation, extracts near horizontal edges.

Table 7.2: Summarized list of clutter metrics used in deriving the clutter complexity measure for hyperspectral images.

Metric Name	Description	No. of Metrics
Single-band clutter metrics[1]		
Standard deviation	Global standard deviation	5
Schmieder Weathersby	Average local standard deviation	5
Homogeneity	Average pixel variation	5
Energy	Average histogram energy	5
Entropy	Average histogram entropy	5
Target Interference Ratio	Average contrast	5
Outlier Ratio	Average percentage of outliers/edges	5
FBM Hurst Parameter	Texture roughness	5
GGABS(5 variations, $I - V$)	Generalized Gaussian Analysis-By-Synthesis	25
Gabor filter(5 orientations)	Parameters p (edge content), c (pixel intensity range)	2×5×5=50
Derived from band information content		
Band correlation	Mean/Median correlation in HSI bands	2
Anomaly detectors		
DotProduct	Average dot product of pixel vectors	1
Kullback-Leibler	Average relative entropy of pixel vectors	1
Derived from GLCM[2]		
GLCM Imax.	Inverse of maximum value from matrix	2×1=2
GLCM Energy	Energy computed from matrix	2×1=2
GLCM Entropy	Entropy computed from matrix	2×1=2
GLCM Contrast	Contrast computed from matrix	2×1=2
GLCM Homogeneity	Homogeneity computed from matrix	2×1=2
	Total	129

[1] 5 metrics - Min., Max., Mean, Median and Range are computed from the distribution obtained from computing these from the HSI image single bands
[2] Same values computed for both implemented variants of GLCM described

a reduction in the dimensions of the clutter metrics space, and a reduction in the required number of operations to compute them. The aim is to reduce the dimensionality yet retain significant information about clutter in the images in the clutter metrics space. In contrast to Principal Component Analysis (PCA) [50], in which the resulting dimensions in a reduction process do not map directly into the original space, our factor analysis algorithm allows the identification of the retained dimensions from the original space. This is shown in Algorithm 5.

Algorithm 5 . Factor Analysis to Determine Significant Metrics

Randomly select images of required number from database
to form a training set

for all $\alpha \in$ set of clutter metrics **do**
    compute $|\text{CC}(\alpha, \text{false alarm rate (FA)})|$
    discard α from the set if CC is
    'insignificant' i.e. ≤ 0.5
end for

compute correlation matrix of the remaining metrics

for all combinations of α and β of the remaining metrics, **do**
    if $|\text{CC}(\alpha, \beta)|$ is 'significant' i.e. ≥ 0.8 **then**
      if $|\text{CC}(\alpha, \text{FA})| > |\text{CC}(\beta, \text{FA})|$
        discard β
      else
        discard α
      end if
    end if
end for

where CC(x,y) is the correlation coefficient between variables x, and y

7.3.4 Hyperspectral Images CCM Experiments

The clutter metrics computed for each hyperspectral image are normalized for all images to avoid a bias in further processing results due to large ranges of absolute values from metric to metric. We employ linear regression to obtain a weighted combination of the subset of image clutter metrics that correlate best with the baseline clutter levels represented by false alarm rates. This weighted sum is the clutter complexity measure (CCM). A high correlation coefficient (CC) will indicate that the CCM is a good indicator of the baseline clutter levels, that is, monotonically related to ATR task difficulty. The correlation coefficient is the normalized measure of covariance between false alarm rate and the computed clutter complexity measure, and serves as our performance measure.

Data Description

We experimented with two sets of images. The first set consisted of 216 synthesized hyperspectral infrared images.he process of image synthesis, and the image specifications are described in Chapter 5. The target template under test are of size $9 \times 9 \times 126$ pixels. These are averaged as described in Section 6.1.3 to obtain column vectors of length 126 used as arguments by the ACE filter ATR. Each image had either a truck or tank as target, and contained varied clutter at varying levels. Chapter 5 also shows example images, and targets of interest.

The second set consisted of 125 images, synthesized based on the ASHE algorithm as described in Section 6.1. Each image in the set is of spatial size 128×128 pixels, and 44 equally spaced spectral bands spanning $0.35 - 1.0$ nm. The target template under test is of size $3 \times 3 \times 44$ pixels, and the location is indicated by the arrow in the scene template shown in Figure 6.1.

Procedures and Results

We conduct similar experiments with both image sets. Our first sets of experiments sought to obtain the subset of metrics that result from the described factor analysis algorithm. To achieve this, we made random selections of images, and record the resulting clutter metrics subset. This subset results from the factor analysis of all the computed clutter metrics from the images. This process is repeated $1,000$ times, each using a unique combination of images of the same number. By applying linear regression, a weighted combination of these clutter metrics which correlates best with the false alarm rates in the selected images is obtained. The selected images thus serve as a training sample set and the obtained weight from the training process is applied to the remaining images, which serve as the test set. We experimented with different training image sample sizes - from 5% to 40% of the total database size, that is, sets of 11 to 86 images. In each case, all the unselected images serve as the test set.

Figures 7.9(a-h) show histograms of the occurrences of the clutter metrics in the selection process using the first image set. The smaller sets do not show a clear dominance in terms of occurrence of any particular metric. As the training sample set size increases, for example, at 20% of the database size, there is a clear increase in the frequency of a few of the metrics while many others do not occur at all. This trend continues as the training sample set size is increased.

Table 7.3(a) shows the average values of the correlation coefficient for different training sample set sizes. Results from training with smaller set sizes show a perfect correlation between the computed CCM with false alarm rate for the test image sets, but relatively poor generalization to the whole database. This signifies an over-training. Increasing the train dataset alleviates the over-training problem and improves on the gen-

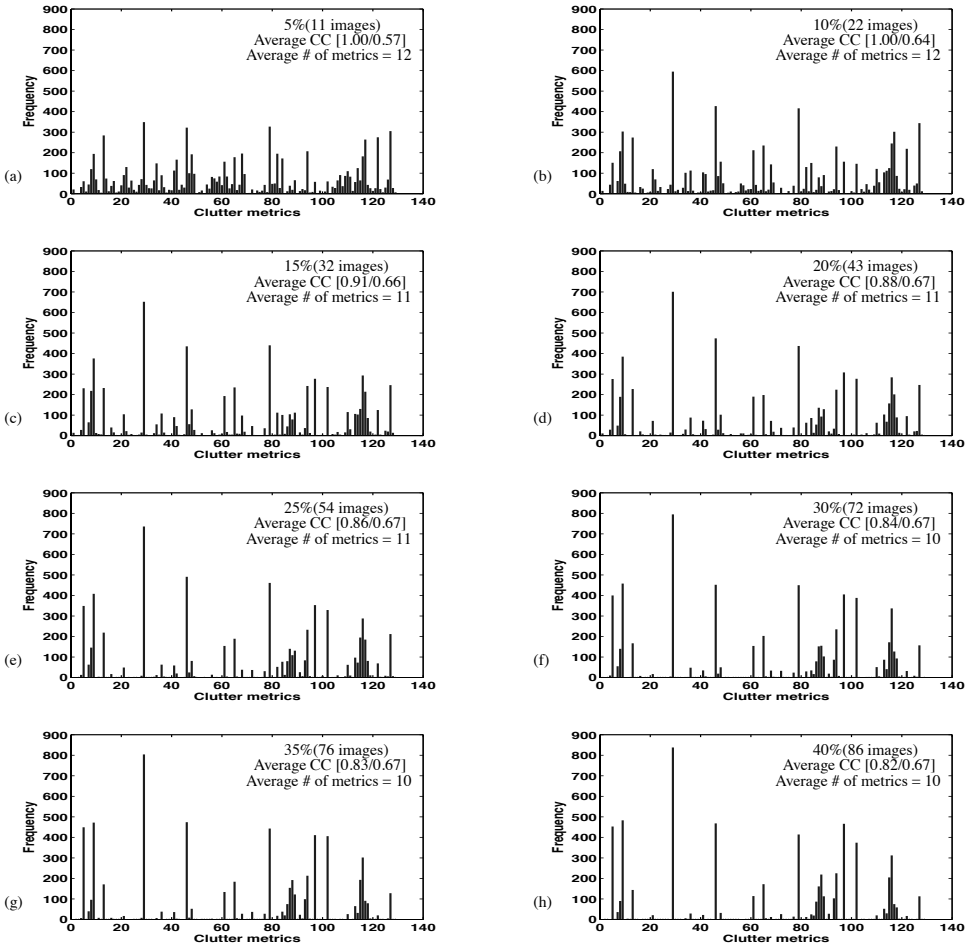


Figure 7.9: Frequency of selection of clutter metrics for training image sets ranging in sizes from 5% to 40% of entire database (11 to 86 images). Noted on the plots are the average number of selected metrics (rounded to the nearest integer), and the average CC values e.g. [0.82/0.67], which are the average CC for the training set and the test sample set respectively. The numbers in the abscissa represent an arbitrary but consistent indexing of the clutter metrics.

Table 7.3: Averaged correlation coefficients obtained between the clutter measure and false alarm rates using different training sample sizes.
(a) Using the subset of clutter metrics selected by the factor analysis process

		Test partitions	
Training partitions	Sample size	Same as train sample	Test sample
	11 images (5%)	1.00	0.57
	22 images (10%)	1.00	0.64
	32 images (15%)	0.91	0.66
	43 images (20%)	0.88	0.67
	54 images (25%)	0.86	0.67
	65 images (30%)	0.84	0.67
	76 images (35%)	0.83	0.67
	86 images (40%)	0.82	0.67

(b) Using a further subset of the metrics used to generate the results in Table 7.3(a) - only the eight metrics with the highest overall frequencies

		Test partitions	
Training partitions	Sample size	Same as train sample	Test samples
	11 images 5%	0.84	0.40
	22 images (10%)	0.77	0.57
	32 images (15%)	0.76	0.62
	43 images (20%)	0.74	0.64
	54 images (25%)	0.73	0.65
	65 images (30%)	0.72	0.66
	76 images (35%)	0.73	0.66
	86 images (40%)	0.72	0.67

Total of $1,000$ experiments with first image set. Sizes are listed as percentages of the total database.

eralization. This improvement saturates with the use of about 20% of the entire database as training samples, which is 43 images in this case.

For training set samples $\geq 20\%$, only 8 clutter metrics were consistently chosen at least 30% of the times during the selection process. The ratio of the frequency of selection of these clutter metrics to all others is also generally large. The indices (arbitrarily assigned) and brief description of these 8 metrics are: $\#5$ - homogeneity derived from the GLCM with known offset, $\#9$ - contrast derived from the GLCM with random offset, $\#29$ - range of the p values from the Gabor filtered images at 90° orientation, $\#46$ - median of the c values from the Gabor filtered images at 60° orientation, $\#79$ - range of the FBM Hurst parameter obtained from images' single bands, $\#97$ - minimum of the homogeneity obtained from the images' single bands, $\#102$ - minimum of the outlier/edge parameters obtained from the images' single bands, and $\#116$ - median of the third parameter of the Gaussian decomposition of the images' single bands.

We performed further experiments with these metrics and show the results in Table 7.3(b). It shows the result of using only the combination of these dominant image metrics for obtaining the CCM for different train sample set sizes. The same trends noted and discussed in the previous experiment, in which the complete subset of clutter metrics resulting from the factor analysis algorithm are employed, is also noticed here. The correlation coefficient values are lower in some cases, this is due to a further reduction in the clutter metric space used to determine the complexity measure.

Empirical timing tests show that it takes about 8.4 minutes to compute these 8 clutter metrics from an image, compared to 75.4 minutes taken for running the ATR for the same image. Both the ATR and clutter measures were implemented in Matlab 6.0 and the tests were carried out on a workstation with a 3.2 GHz Pentium IV processor.

Correlation coefficient results obtained using the second image set are shown in Ta-

Table 7.4: Averaged correlation coefficients obtained between the clutter measure and false alarm rates using different training sample sizes.
(a) Using the subset of clutter metrics selected by the factor analysis process

		Test partitions	
Training partitions	Sample size	Same as train sample	Test sample
	7 images (5%)	1.00	0.29
	13 images (10%)	0.94	0.42
	19 images (15%)	0.89	0.62
	25 images (20%)	0.88	0.67
	32 images (25%)	0.86	0.69
	38 images (30%)	0.85	0.72
	44 images (35%)	0.84	0.72
	50 images (40%)	0.84	0.74

(b) Using a further subset of the metrics used to generate the results in Table 7.4(a) - only the eight metrics with the highest overall frequencies

		Test partitions	
Training partitions	Sample size	Same as train sample	Test samples
	7 images (5%)	0.93	0.16
	13 images (10%)	0.91	0.31
	19 images (15%)	0.87	0.48
	25 images (20%)	0.86	0.59
	32 images (25%)	0.85	0.62
	38 images (30%)	0.84	0.67
	44 images (35%)	0.83	0.69
	50 images (40%)	0.83	0.71

Total of $1,000$ experiments with second image set. Sizes are listed as percentages of the total database.

bles 7.4(a) and 7.4(b). Note the similar trends to the previous experiments in terms of generalization of the derived CCM. Eight dominant clutter metrics were also recorded when the performance, indicated by the average CC values saturates. These are median of p values from Gabor filtered images at 120° orientation, minimum of FBM Hurst values from images' single bands, minimum, median and range of the target interference ratio from the images' single bands, maximum of the first parameter of the Gaussian decomposition of the images' single bands, and range of the second parameter of the Gaussian decomposition of the images' single bands. These metrics are different from those obtained from the initial experiments, indicating that the derived CCM is image set specific. Also, the generalization performance saturates with the use of 30% of the entire image set for training in the second experiment, compared to 20% in the first. Both fractions of the image sets result in approximately 38 images. Using this training image size, the derived clutter measure is dominated by eight clutter metrics in both cases. This indicates the the required number of training images is function of the number of dominant clutter metrics used in the CCM derivation, and not the total number of images in the experimental set.

We also show the distribution of the CC values resulting in the averages shown in Tables 7.4(a). Note that when a random selection of 38 or more training images are used, 90% or more of the CC values are ≥ 0.6. This is important because it shows that the CCM for an image set can be obtained using any random selection of training images from the complete set.

In summary, our results show a more frequent selection of a further subset of metrics used to determine our clutter measure. We refer to these as the dominant metrics. These metrics are unique for each experiment, indicating that the derived CCM is image set specific. A random set of about 38 images is shown to be sufficient to define the CCM

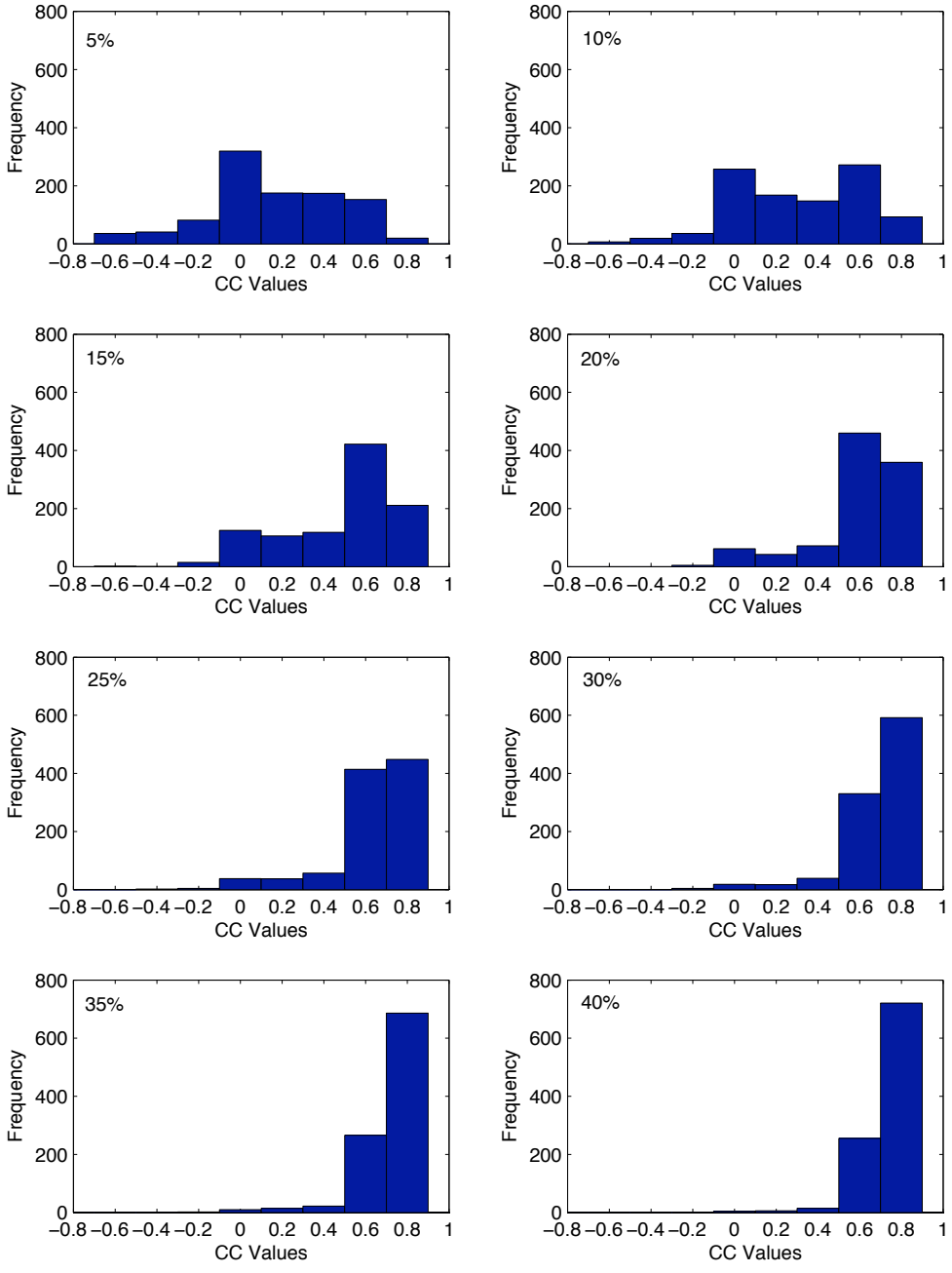


Figure 7.10: Distribution of CC values resulting in the average values shown in Table7.4(a). The indicated percentages, and the actual number of images that they represent are also shown in the same table.

using eight dominant metrics. The derived clutter measure from these training images, generalizes well for the entire database by predicting the amount of clutter in them. Further experiments to determine the clutter measure using only the dominant clutter metrics yielded similar results. Comparison of the time taken to compute the CCM from these dominant clutter metrics from and running the ATR on an image shows a ratio of about $1:9$ in the first set of experiments.

CHAPTER 8

CONCLUSIONS

We presented a novel, progressive adaptive sampling algorithm, called Adaptive Sampling by Histogram Equalization (ASHE). The algorithm adapts the local sampling density on a function, based on the distribution of already obtained samples. The aim of adaptive sampling is the efficient distribution of discrete samples used in representing a continuum. Efficient sample distribution reduces the inherent error that results from a sampling process. For nonstationary functions, adaptive schemes produce higher sampling densities in regions of higher complexities, that is, where the rate of change in the sampled function is higher. In numerous scientific applications, there is no prior knowledge of the local complexities in the sampled function, and the cost of obtaining each sample is prohibitive. Examples of such costs, which limit the number of samples that can be obtained, are time and computational resources. Thus, extra constraints are placed on adaptive sampling schemes. For efficient adaptive sampling, existing algorithms either require prior knowledge of the local complexities in the function, a high computational overhead, such as an acceptance or rejection step, or they require a large number of samples to converge. The ASHE algorithm requires no prior knowledge of the local variations in the sampled function. Also, it only adds a minimal overhead of computing a histogram of sample values at each step of the sampling process.

In the following sections, we summarize the main contributions of this dissertation, and discuss the findings from an application in which our developed algorithm was utilized. Finally, we make suggestions on further work.

8.1 Summary of contributions

In Chapter 2, we presented the basis of the ASHE algorithm as progressive sampling based on the distribution of already obtained samples. Typical sampling algorithms focus on the domain of the independent variables. Our focus is on the co-domain of the sampled quantity. We showed that, for a nonstationary function, even spaced sampling on the co-domain results in a sampling density that is proportional to the rates of change in the sampled function. We thus sampled in order to equalize the distribution of sample values. This results in sample densities in the domain that are proportional to the rate of change in the function, hence the adaptive sampling. To the best of our knowledge, this is a novel approach to adaptive sampling. Since the sampling scheme attempts to equalize the distribution of samples, we called it Adaptive Sampling by Histogram Equalization - ASHE algorithm. We illustrated the improved performance by the ASHE algorithm by comparing it to even spaced and random sampling. Even spaced or random sampling are the obvious options for obtaining expensive samples when there is no prior knowledge on the local complexity in a function. We identified the reasons precluding a rigorous mathematical proof of the improvement recorded by the ASHE algorithm. The most basic of these being the assumption that there is no prior knowledge on the nature of the sampled function. We however studied the algorithm further, by conducting performance and sensitivity analysis in a manner similar to those in other heuristic algorithm studies. Finally, we discussed broad areas of possible applications of the ASHE algorithm.

We introduced three stochastic optimization models in Chapter 3. These are: (1) an active walker model, based on elements of the random walk and Brownian motion, (2) an ant model, based on the simulation of foraging habits of insects, and (3) an evolutionary algorithm model, based on the simulation of natural dynamics in a population of organ-

isms. The basic forms of the models are described. We then developed three algorithms by adapting each model to implement the ASHE algorithm.

We conducted a performance and sensitivity analysis of the three models in Chapter 4. First, we established two objective measures for comparison. These are based on the frequency content, and the entropy measure of information in a sampled function. Both measures are designed to have a positive correlation with increase in function complexity. Our measure of performance is defined as the correlation coefficient between these measures, and the sample density obtained from each model. A high positive value (maximum $= 1$), indicates a good performance. We identified factors that could affect the performance of each model, and recorded their performances for varying values of these factors. Comparing the best performance of the three, the ant and evolutionary algorithm models performed marginally better than the active walker model. More importantly, the active walker model showed a correlation between the individual factors, and the performance. This is a crucial requirement for heuristic algorithms. If this is not met, the algorithms are ad hoc, requiring customization for each application. The other two models contained one or more factors that showed no individual correlation with the sampling performance. This limits their practical use. Based on our findings, we studied the active walker model further, by considering its scaling properties. Our results indicated that the model performance does not change appreciably with change in the dimensions of the sampled space.

We utilized the ASHE algorithm in the synthesis of hyperspectral images. The availability of real images of these types is limited, and synthesized images are used in their place. For our purpose, we required images that are diverse with respect to Automatic Target Recognition ATR performance. In general, image synthesis is computationally expensive. Also, there is no prior knowledge of how the factors in the image synthesis

process affect the ATR performance. We described the nature, uses, and synthesis of hyperspectral images in Chapter 5. We then synthesized images using a combination of even spaced factors. In Chapter 6, we employed the ASHE algorithm in the image synthesis process, and compared the images to those synthesized using even spaced, and random placed factors. This comparison is based on the requirement of image diversity with respect to ATR performance. Our results showed a marked improvement over the other methods. The worst performance was recorded for images synthesized using a combination of even spaced factors.

In Chapter 7, we developed a framework for quantifying clutter in hyperspectral images. By clutter, we mean any object or structure in an image that inhibits the detection of a target of interest. We derived this measure as an aggregation of image features that correlates best with ATR performance bounds. We called this the Clutter Complexity Measure CCM. This is an indication of the inherent difficulty for an ATR to identify a target in a scene. It is however, not based on any particular ATR, thus making it a good objective basis for comparing the performance of disparate ATRs. Our initial experiment to investigate the feasibility of this approach used single bands from real hyperspectral images. Our results showed that CCM derived for this images was useful in the efficient ordering of hyperspectral bands in a multi-band detection scheme. Using the band selection based on the CCM for the multi-band detection, we recorded an average of 30% improvement over the even spread band selection. We also successfully derived a clutter complexity measure for complete, synthesized hyperspectral images. In computing this, we developed 129 image features, and computed the CCM as an aggregation of a subset of these features. We obtain the subset of features through a factor analysis process. We were able to derive a CCM using any random selection of images from the complete set. We determined that the required size of the selection is dependent on the number of im-

age features aggregated to compute the CCM. In our experiments, the CCM consisted of 8 image features, and this required about 40 images. We also determined that the derived CCM is specific to an image set. The CCM derived for complete hyperspectral images was computed in 11% of the time it took to compute a baseline ATR performance. The CCM was shown to accurately predict the baseline ATR performance bounds in at least 64% of the cases.

8.2 Suggestions for further work

In the three models employed in implementing ASHE, the input from the fitness criterion is modeled as a step function in their outputs. In the active walker model for example, the fitness criterion input will result in either a long step or a short step, with nothing in between. Further work needs to be done to investigate the effect of using a different output model. That is, one in which the modeled output is a function of the amount of change in the input. A linear, exponential, or other non-linear models are examples that could be explored.

In the ASHE based image synthesis process, further work needs to be done to identify the effect that individual, and combination of factors have in the synthesized images. Factors that result in rapid image variation with respect to ATR performance can be identified using pattern analysis methods. Also, the use of a multi-dimensional objective function in the ASHE based image synthesis process needs to be investigated. This is in contrast to our use of only the baseline ATR performance. Other computationally less expensive indicators of image variability may be used to form a multi-dimensional histogram to be equalized.

Most of our test images for the experiments with the clutter complexity measure have been synthesized. The scheme to derive this measure needs to be tested using real

hyperspectral images.

APPENDIX A

CLUTTER COMPLEXITY METRICS

A.1 Single-band clutter metrics

A.1.1 Global standard deviation

$$\sigma_{\text{metric}} = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(I_i - \bar{I})^2} \tag{A.1}$$

where I are the intensity values in a hyperspectral band with mean $\bar{I}$, and T is the total number of pixels in this band.

The metrics described in appendices A.1.2 to A.1.7 are computed locally. That is, each hyperspectral image band is divided into N windows, with each window containing W pixels. The size of the window is chosen such that it is about twice the length of the largest target in spatial dimensions and, $\mathcal{W}_i$ represents the support for the i th window. The overall metric is then obtained by averaging the computed metric values for each window over all N windows.

A.1.2 Schmieder Weathersby

$$SW_{\text{metric}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2} \tag{A.2}$$

where σ_i^2 is the variance of pixels within the ith window.

A.1.3 Homogeneity

$$H_i = \sqrt{\frac{1}{W}\sum_{j\in\mathcal{W}_i}(I_j - \bar{I}_i)^2}\,, \tag{A.3}$$

where

$$\bar{I}_i = \frac{1}{W}\sum_{j\in\mathcal{W}_i} I_i \tag{A.4}$$

$$Homogeneity_{\text{metric}} = \frac{1}{N}\sum_{i=1}^{N} H_i \tag{A.5}$$

A.1.4 Energy

$$Energy_i = \sum_{j=0}^{GL-1}(P_i[j])^2 \tag{A.6}$$

$$Energy_{\text{metric}} = \frac{1}{N}\sum_{i=1}^{N} Energy_i \tag{A.7}$$

A.1.5 Entropy

$$Entropy_i = -\sum_{j=0}^{GL-1}(P_i[j])\log_2(P_i[j]) \tag{A.8}$$

$$Entropy_{\text{metric}} = \frac{1}{N}\sum_{i=1}^{N} Entropy_i \tag{A.9}$$

In both definitions in A.1.4 and A.1.5, GL is the defined number of gray-level intensity values (typically 256) in the image and, P_i is the histogram of the intensities of the pixels in the ith window.

A.1.6 Target Interference Ratio

$$TIR_i = |\mu_{\text{target}} - \mu_{\text{background}}|/\sigma_{\text{background}} \tag{A.10}$$

$$TIR_{\text{metric}} = \frac{1}{N}\sum_{i=1}^{N} TIR_i \tag{A.11}$$

where μ_{target} is the mean of the intensity values in a window of about the same size as the target, and $\mu_{background}$ and $\sigma_{background}$ are the mean and standard deviation of the target background. The target background is defined as the window centered around the target but twice the dimensions of the target. In this case, the size of the target background determines the value of N.

A.1.7 Outlier/Edge

$$Edge_i = \text{Cardinality of } \{j : |I_j - \bar{I}_i| > \bar{I}_i/4\} \tag{A.12}$$

where $j \in \mathcal{W}_i$ and $\bar{I}_i = \frac{1}{W}\sum_{j\in\mathcal{W}_i} I_i$

$$Edge_{\text{metric}} = \frac{1}{N}\sum_{i=1}^{N} Edge_i \tag{A.13}$$

A.1.8 FBM Hurst Parameter

$$f_s^x = \frac{1}{D_1 D_2}\sum_{i=1}^{D_1}\sum_{j=1}^{D_2} |I(i+2^s, j) - I(i,j)|^2 \tag{A.14}$$

$$f_s^y = \frac{1}{D_1 D_2}\sum_{i=1}^{D_1}\sum_{j=1}^{D_2} |I(i, j+2^s) - I(i,j)|^2 \tag{A.15}$$

$$f_s = f_s^x + f_s^y \tag{A.16}$$

$$FBM_{\text{metric}} = \text{slope } (\frac{1}{2}\log_2 f_s) \tag{A.17}$$

where D_1 and D_2 are the spatial dimensions of a hyperspectral band. f is computed for

$1:s$, where s is determined as

$$s = \text{floor}((\log(N_s)/\log(2)) - 2)\ , \tag{A.18}$$

where the floor operator rounds the expression to the nearest integer towards minus infinity, and $N_s = \text{minimum}(D_1, D_2)$.

A.1.9 Metrics c and p derived from Gabor filtered images [9]

The Gabor filter we used is bandpass filter with a Gaussian kernel. It is defined as

$$F_{\sigma,\theta}(z) = \exp(-\frac{1}{2\sigma^2}(z_\theta(1)^2 + z_\theta(2)^2)) \exp(-j\frac{2\pi z_\theta(1)}{\sigma}) \tag{A.19}$$

where $\sigma = 4$ denotes the resolution associated with the filter and

$$z_\theta = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \in \mathbb{R}^2 \tag{A.20}$$

and $\theta \in [0, 2\pi)$ is the filter rotation angle. For a bank of K filters, we obtain $F^{(j)}, j = 1, 2, ..., K$. For a particular rotation angle, the filtered images is obtained by the 2D convolution of the image with the filter

$$I^{(j)} = I * F^{(j)} \tag{A.21}$$

p and c are obtained as:

$$p_{\text{metric}} = \frac{3}{SK(I^{(j)}) - 3} \tag{A.22}$$

$$c_{\text{metric}} = \frac{SV(I^{(j)})}{p} \tag{A.23}$$

where SK and SV are the sample kurtosis and the sample variance of the Gabor filtered image respectively.

A.2 Metric derived from band information content

A.2.1 Band correlation

$$BC_{\text{metric}} = \sum_{i=1}^{L} \sum_{j=i+1}^{L} CC(b_i, b_j) \tag{A.24}$$

where L is the number of hyperspectral bands, and CC denotes the correlation coefficient between bands b_i and b_j.

A.3 Anomaly detectors

A.3.1 Dot product

$$DP_{\text{metric}} = \frac{1}{T} \sum_{i=1}^{T} (1 - (\frac{1}{n} \sum_{j=1}^{n} (x_i/|x_i|) \cdot (y_{ij}/|y_{ij}|))) \tag{A.25}$$

where x_i is the pixel vector under test and y_{ij} are the pixel vectors surrounding the vector under test, all of length L. Typically, $n = 4$, and the test pixels are located at the vertices of a square with the test pixel as center and sides of length typically equal to 3 pixels. T is the total number of pixels in the spatial dimensions minus the pixels at the edges.

A.3.2 Kullback-Leibler

$$KL_{ij} = \sum_{k=1}^{GL-1} P_k(x_i) \times \log(\frac{P_k(x_i)}{P_k(y_{ij})}) \tag{A.26}$$

where $P(x_i)$ is the histogram of the vector under test, $P(y_{ij})$ is the histogram of one of the surrounding pixels and $GL = 256$ is the number of gray-levels for the histogram definition. The above is thus the Kullback-Leibler distance between these two pixel

vectors. The arrangement of the surrounding pixels is the same as in A.3.1. and the metric value for a particular test pixel is obtained by averaging this distance over the $n = 4$ surrounding pixels. The overall Kullback-Leibler metric is obtained by averaging each pixel metric value over all T tested pixels, where T is also as described in A.3.1.

$$KL_{\text{metric}} = \frac{1}{T}\sum_{i=1}^{T}(\frac{1}{n}\sum_{j=1}^{n} KL_{ij}) \tag{A.27}$$

A.4 Metrics derived from the Gray Level Co-occurrence Matrix

Given intensity values $I(i, j, l)$, where (i, j) is the spatial location and l is the band location, and the number of gray-levels is GL (typically 256) the GLC matrix G is obtained thus:

for $t = 1 : T$

$m = I(i_t, j_t, b_t);\ n = I(i'_t, j'_t, b'_t)$

$G(m, n) = G(m, n) + 1$

end

T is the total number of samples used. The offsets in the 3 dimensions are $(i_t - i'_t, j_t - j'_t, l_t - l'_t)$. $G \in \mathbb{R}^2$ has size 256×256.

The derived metrics are:

$$GLCM_Imax. = 1/max(max(G)) \tag{A.28}$$

$$GLCM_Energy = \sum_m \sum_n G_{mn}^2 \tag{A.29}$$

$$GLCM_Entropy = -\sum_m \sum_n G_{mn} \log(G_{mn}) \tag{A.30}$$

$$GLCM_Contrast = \sum_m \sum_n G_{mn}(m - n)^2 \tag{A.31}$$

$$GLCM_Homogeneity = \sum_{m} \sum_{n} G_{mn}/(m-n) \tag{A.32}$$

REFERENCES

[1] ARNOLD, P. S., BROWN, S. D., AND SCHOTT, J. R. Hyperspectral simulation of chemical weapon dispersal patterns using dirsig. In *SPIE Proceedings* (2000), vol. 4029, pp. 288–299.

[2] ASMAR, D. C., ELSHAMLI, A., AND AREIBI, S. A comparative assessment of aco algorithms within a tsp environment. *Dynamics of Continuous, Discrete, and Impulsive Systems - Series B - Applications and Algorithms 1* (2005), 462–467.

[3] BACK, T. *Evolutionary algorithms in theory and practice - evolution strategies, evolutionary programming, genetic algorithms*. Oxford University Press, 1996.

[4] BALLINGER, D. Space-based hyperspectral imaging and its role in the commercial world. In *Proceedings of IEEE Aerospace Conference* (2001), vol. 2, pp. 915–923.

[5] BAUTISTA, P., ABE, T., YAMAGUCHI, M., YAGI, Y., AND OHYAMA, N. Digital staining for multispectral images of pathological tissue specimens based on combined classification of spectral transmittance. *Journal of Computerized Medical Imaging and Graphics 29(8)* (2003), 649–657.

[6] BERK, A. Moderate spectral resolution atmospheric transsmittance algorithm and computer model documentation. `www.vs.afrl.af.mil/ProductLines/IR-Clutter/modtran4`.

[7] BONGIOVI, R. P., HACKWELL, J. A., AND HAYBURST, T. L. Airborne lwir hyperspectral measurements of military vehicles. In *Proceedings of IEEE Aerospace Application Conference* (1996), vol. 3, pp. 121–135.

[8] BRITTON, D., SMITH, M., AND MERSEREAU, R. Generalized gaussian decompositions for image analysis and synthesis. In *Proceedings of the Systematics, Cybernetics, and Informatics (SCI)* (Orlando, FL, July 2000).

[9] BROWN, J. H. Strategic high altitude atmospheric radiance code documention. `www.vs.afrl.af.mil/ProductLines/IR-Clutter/sig.aspx`.

[10] BROWN, S. D. Digital imaging and remote sensing image generation model documention. `http://dirsig.cis.rit.edu/docs`.

[11] CARO, G. D., DUCATELLE, F., AND GAMBARDELLA, L. M. Anthocnet: an adaptive nature inspired algorithm for routing in mobile ad hoc networks. *European Transactions on Telecommunications 16(5)* (2005), 443–455.

[12] CARPIN, S., AND PILLONETTO, G. Motion planning using adaptive random walks. *IEEE Transactions on Robotics 21(1)* (2005), 129–136.

[13] COOK, R. L. Stochastic sampling in computer graphics. *ACM Transactions on Graphics 5(1)* (1986), 51–72.

[14] DENEUBORG, J. L., AND GOSS, S. Collective patterns and decision making. *Ethology, Ecology and Evolution 76* (1989), 579–581.

[15] DIPPE, M. A. Z., AND WOLD, E. H. Antialiasing through stochastic sampling. In *Proceedings of SIGGRAPH* (San Francisco, 1985), vol. 19(3), pp. 69–78.

[16] DJURIC, P., HUANG, Y., AND GHIRMAI, T. Perfect sampling: a review and applications to signal processing. *IEEE Transactions on Signal Processing 50(2)* (2002), 345–356.

[17] DORIGO, M., BIRATTARI, M., AND STIITZLE, T. Ant colony optimization: artificial ants as a computational intelligence technique. *IEEE Computational Intelligence Magazine* (2006), 28–39.

[18] DORIGO, M., AND GANBARDELLA, L. Ant colony system: a cooperative learning approach to the travelling salesman problem. *IEEE Transactions on Evolutionary Computing 1(1)* (1997), 53–66.

[19] DORIGO, M., MANIEZZO, V., AND COLORNI, A. The ant system: optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man and Cybernetics-Part B 26(1)* (1996), 1–13.

[20] DORIGO, M., MANIEZZO, V., AND COLORNI, A. Ant system: optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man and Cybernetics - Part B 26(1)* (1996), 29–41.

[21] ELDAR, Y. Irregular image sampling using the voronoi diagram. Master's thesis, Technicon-IIT, Israel, 1992.

[22] ELDAR, Y., LINDENBAUM, M., PORAT, M., AND ZEEVI, Y. Y. The farthest point strategy for progressive image sampling. *IEEE Transactions on Image Processing 6(9)* (1997), 1305–1315.

[23] FADIRAN, O. O., AND KAPLAN, L. M. Clutter complexity analysis of hyperspectral bands. In *Proceedings of the Thirty-Sixth I.E.E.E. Southeastern Symposium* (Atlanta, 2004), pp. 531–535.

[24] FADIRAN, O. O., AND MOLNAR, P. Adaptive sampling based on frequency distribution of function values. In *Proceedings of the SIAM Conference on Imaging Science* (May 2006).

[25] FADIRAN, O. O., MOLNAR, P., AND KAPLAN, L. M. Towards quantifying clutter in hyperspectral infrared images. In *Proceedings of the IASTED Conference on Circuits, Signals and Systems* (San Francisco, 2005).

[26] FADIRAN, O. O., MOLNAR, P., AND KAPLAN, L. M. Adaptive sampling via histogram equalization using an active-walker model. In *Proceedings of the IEEE International Conference on Information Science* (July 2006), pp. 424–432.

[27] FADIRAN, O. O., MOLNAR, P., AND KAPLAN, L. M. A statistical approach to quantifying clutter in hyperspectral infrared images. In *Proceedings of the IEEE Aerospace Conference* (Big Sky, March 2006), pp. CD–ROM.

[28] FOGEL, D. B. Evolutionary programming. In *Proceedings of the 5th annual conference on evolutionary programming* (1996).

[29] FREIMUTH, R. D., AND LAM, L. Active walker models for filamentarty growth patterns. In *Modeling Complex Phenomena* (1992), L. Lam and V. Naroditsky, Eds., New Tork: Springer, pp. 302–313.

[30] GORGES-SCHLEUTER, M. A comparative study of local and global selection in evolution strategies. In *Proceedings of the PPSN Conference* (1998), pp. 367–377.

[31] GOSS, S., ARON, S., DENEUBORG, J. L., AND PASTEELS, J. M. Self-organized shortcuts in the argentine ant. *Naturwissenschaften 1* (1989), 295–311.

[32] GOSS, S., BECKERS, R., DENEUBORG, J. L., ARON, S., AND PASTEELS, J. M. How trail laying and trail following can solve foraging problems for ant colonies. *Behavioural Mechanisms of Food Selection G20* (1990).

[33] GRAVEL, M., PRICE, W. L., AND GAGNE, C. Scheduling continuous casting of aluminium using a multiple objective ant colony optimization metaheuristic. *European Journal on Operations Research 143* (2002), 218–229.

[34] GRAY, R. M. *Entropy and information theory*. Springer-Verlag, 1990.

[35] HARALICK, R. M. Statistical and structural approaches to texture. In *Proceedings of the 4th Joint Conference on Pattern Recognition* (1978), pp. 45–69.

[36] HAUTA-KASARI, M., PARKKINEN, J., JAAKELAINEN, T., AND LENZ, R. Generalized co-occurrence matrix for multispectral texture analysis. In *Proceedings of the 13th International Conference on Pattern Recognition* (1996), vol. 2, pp. 782–789.

[37] HELBING, D., KELTSCH, J., AND MOLNR, P. Modelling the evolution of human trail systems. *Nature 388* (1997), 47–50.

[38] HELBING, D., SCHWEITZER, F., KELTSCH, J., AND MOLNR, P. Active walker model for the formation of human and animal trail systems. *Physical Review E 56* (1997), 2527–2533.

[39] HOLLAND, O., AND MELHUISH, C. Stigmergy, self-organization, and sorting in collective robotics. *MIT Press Journal 5(2)* (1999), 173–202.

[40] HOLLDOBLER, B., AND MOGLICH, M. *The ants*. Belknap, 1990.

[41] HUANG, Y., AND DJURIC, P. M. Variable selection by perfect sampling. *EURASIP Journal on Applied Signal Processing 2002(1)* (2002), 38–45.

[42] HUNG, K. L., AND CHANG, C. C. New irregular sampling coding method for transmitting images progressively. In *Proceedings of the IEEE Conference on Vision, Images and Signal Processing* (February 2003), pp. 44–50.

[43] JAIN, A. K. *Fundamentals of digital image processing*. Prentice Hall, 1989.

[44] JOHNSON, A., WINDESHEIM, E., AND BROCKHAUS, J. Hyperspectral imagery for trafficability analysis. In *Proceedings of IEEE Aerospace Conference* (1998), vol. 2, pp. 21–35.

[45] KAPLAN, L., NAMUDURI, K., DAVIES, M., NASRABADI, N., CHAN, L., BRITTON, D., SMITH, M., AND MERSEREAU, R. Development and analysis of a clutter complexity measure for ATR characterization. In *Proc. of the Fifth Annual ARL Federated Laboratory Symposium* (College Park, MD, Mar. 2001), pp. 195–199.

[46] KAYSER, D. R., ABERLE, L. K., POCHY, R. D., AND LAM, L. Active walker models: tracks and landscapes. *Physica A 191* (1992), 17–24.

[47] KOZA, J. R. *Genetic programming: on the programming of computers by means of natural selection*. MIT Press, 1992.

[48] LAM, L. Active walker models for complex systems. *Journal of Chaos, Solitons and Fractals 6* (1995), 267–285.

[49] LAM, L., AND POCHY, R. Active-walker models: growth and form in nonequilibrum systems. *Computers in Physics 7* (1993), 534–541.

[50] LANDGREBE, D. A. *Signal Theory Methods in Multispectral Remote Sensing*. Wiley-Interscience, 2003.

[51] LANTERMAN, A., O'SULLIVAN, J., AND MILLER, M. Kullback-Leibler distances for quantifying clutter and models. *Optical Engineering 38* (Dec. 1999), 2134–2146.

[52] LEE, M. E., REDNER, R. A., AND USELTON, S. P. Statistically optimized sampling for distributed ray tracing. In *Proceedings of SIGGRAPH* (San Francisco, 1985), vol. 19(3), pp. 61–67.

[53] LEE, M. E., REDNER, R. A., AND USELTON, S. P. A statistical method for adaptive stochastic sampling. In *Proceedings of Eurographics Conference and Exhibition* (Lisbon, 1986), vol. 11(2), pp. 157–162.

[54] LENTILUCCI, E. J., BROWN, S. D., SCHOTT, J. R., AND RAQUENO, R. V. Multispectral simulation environment for modelling low-light-level sensor systems. In *SPIE Proceedings* (1998), vol. 3434, pp. 10–19.

[55] MANIEZZO, V. Exact and approximate nondeterministic tree-search procedures for the quadratic assignment problem. *INFORMS Journal on Computing 11(4)* (1999), 358–369.

[56] MARKS, R. J. *Introduction to Shannon sampling and interpolation theory*. Springer Verlag, 1991.

[57] MCWHORTER, L. T., AND SCHARF, L. L. Adaptive matched subspace detectors and adaptive coherence estimators. In *Proceedings of the 30th Asilomar Conference on Signals, Systems, and Computers* (1996), pp. 1114–1117.

[58] MEITZLER, T., GERHART, G., AND SINGH, H. A relative clutter metric. *IEEE Trans. on Aerospace and Electronic Systems 34* (July 1998), 968–976.

[59] MHLENBEIN, H., GORGES-SCHLEUTER, M., AND KRMER, O. Evolution algorithms in combinatorial optimization. *Parallel Computing 7* (1988), 65–85.

[60] MITCHELL, D. P. Generating antialiased images at low sampling densities. In *Proceedings of SIGGRAPH* (1987), vol. 21(4), pp. 65–72.

[61] MITCHELL, M. *An introduction to genetic algorithms*. MIT Press, 1998.

[62] NAMUDURI, K. R., BOUYOUCEF, K. K., AND KAPLAN, L. M. Image metrics for clutter characterization. In *Proceedings of 2000 IEEE International Conference on Image Processing* (2000), vol. 2, pp. 467–470.

[63] NORRIS, J. R. *Markov chains*. Cambridge University Press, 1999.

[64] PAINTER, J., AND SLOAN, K. Antialiased ray tracing by adaptive progressive refinement. In *Proceedings of SIGGRAPH* (1989), vol. 23(3), pp. 281–288.

[65] PANUSOPONE, K., CHEEVASUVIT, F., AND RAO, K. R. Adaptive subsampling for image compression. In *Proceedings of the IEEE ASILOMAR Conference* (1996), pp. 239–243.

[66] PETERS, R. A., AND STRICKLAND, R. N. Image complexity metrics for automatic target recognizers. In *Proceedings of Automatic Target Recognizer System and Technology System* (Silver Spring, 1990).

[67] PROAKIS, J. G., AND MANOLAKIS, D. G. *Digital Signal Processing: Principles, Algorithms, and Applications*. Pearson Prentice Hall, 1996 (Fourth edition).

[68] REED, I., AND YU, X. Adaptive multi-band CFAR detection of an optical pattern with unknown spectral distribution. *IEEE Trans. on Acoustics, Speech, and Signal Processing 38* (Oct. 1990), 1760–1770.

[69] REIMANN, M., DOERNER, K., AND HARTL, R. F. D-ants: savings based ants divide and conquer the vehicle routing problem. *Computers and Operations Research 31(4)* (2004), 563–591.

[70] RICHTSMEIER, S. C., SUNDBERG, R. L., BERK, A., AND ADLER-GOLDEN, S. M. Full-spectrum scene simulation. In *SPIE proceedings* (2004), vol. 5425, pp. 530–537.

[71] RIGAU, J., FEIXAS, M., AND SBERT, M. Entropy-based adaptive sampling. In *Proceedings of Graphics Interface Conference* (Halifax, June 2003).

[72] RODNICK, J., AND GASPARI, G. *Elements of the random walk*. Cambridge University Press, 2004.

[73] SANDERS, J. S., AND BROWN, S. D. Utilization of dirsig in support of real-time infrared scene generation. In *SPIE Proceedings* (2000), vol. 4029, pp. 278–285.

[74] SCANLAN, N. W., SCHOTT, J. R., AND BROWN, S. D. Performance analysis of improved methodology for incorporation of spatial/spectral variability in synthetic hyperspectral imagery. In *SPIE Proceedings* (2003), vol. 5159, pp. 319–330.

[75] SCHIMANSKY-GEIER, L., MIETH, M., ROSE, H., AND MALCHOW, H. Structure formation by active brownian particles. *Physics Letters A 207* (1995), 140.

[76] SCHOTT, J. R., BROWN, S. D., RAQUENO, R. V., GROSS, H. N., AND ROBINSON, G. An advanced synthetic image generation model and its application to

multi/hyperspectral algorithm development. *Canadian Journal of Remore Sensing 15* (1999), 99–111.

[77] SCHOTT, J. R., KUO, S. D., BROWN, S. D., AND RAQUENO, R. V. Prediction of observed image spectra using synthetic image generation models. In *SPIE Proceedings* (1997), vol. 3118, pp. 81–93.

[78] SCHWEITZER, F., AND SCHIMANSKY-GEIER, L. Clustering of active walkers in a two-component system. *Physica A 206* (1994), 359–379.

[79] SCHWEITZER, F., AND SCHIMANSKY-GEIER, L. Clustering of active walkers: phase transition from local interaction. In *Fluctuations and Order: The New Synthesis* (1996), M. Millonas, Ed., New Tork: Springer, pp. 293–305.

[80] SHI, M., AND HEALEY, G. E. Three-dimensional hyperspectral texture recognition using multiband correlation models. In *SPIE Proceedings* (2003), vol. 5093, pp. 678–685.

[81] SHIRVAIKAR, M. V., AND TRIVEDI, M. M. Developing a texture-based image clutter measure for object detection. *Optical Engineering 31* (Dec. 1992), 2628–2639.

[82] SIEGMUND, D. *Elements of information theory*. Wiley Interscience, 1991.

[83] SIMS, S. Putting ATR performance on an equal basis: The measurement of knowledge based distortion and relevant clutter. In *Proc. of SPIE* (July 1999), vol. 3810, pp. 118–128.

[84] SRINIVAS, N., AND DEB, K. Multiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary Computation 2(3)* (1994), 221–248.

[85] SRIVASTAVA, A., LIU, X., AND GRENANDER, U. Universal analytical forms for modeling image probabilities. *IEEE Transactions on Pattern Analysis and Machine Intelligence 24(9)* (2002), 1200–1214.

[86] STUTZLE, T., AND HOOS, H. H. Max-min ant system. *Future Generation Computer Systems 16(8)* (2000), 889–914.

[87] SVENSON, P., AND SIDENBLADH, H. Determining possible avenues of approach using ants. In *Proceedings of the 6th International Conference on Information Fusion* (2003), pp. 1110–1117.

[88] TAMSTORF, R., AND JENSEN, H. W. Adaptive sampling and bias estimation in path tracing. In *Proceedings of the 8th Eurographics Workshop on Rendering* (St. Etienne, 1997), pp. 285–295.

[89] TAXT, T., AND LUNDERVOLD, A. Multispectral analysis of the brain using magnetic resonance imaging. *IEEE Transactions on Medical Imaging 13(3)* (1994), 470–481.

[90] VO-DINH, T., CULLUM, B., AND KASILI, P. Development of a multi-spectral imaging system for medical applications. *Journal of Physics D: Applied Physics 36(14)* (2003), 1663–1668.

[91] WALDMAN, G., WOOTTON, J., HOBSON, G., AND LUETKEMEYER, K. A normalized clutter measure for images. *Journal of Computer Vision, Graphics, and Image Processing 42* (1998), 137–156.

[92] WANG, Z., BOVIC, A. C., SHEIKH, H. R., AND SIMONCELLI, E. E. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image processing 13(4)* (2004), 600–612.

[93] WINDHAM, R. W., WILLIAM, R., LAWRENCE, C. K., PARK, B., SMITH, P. D., AND POOLE, G. Analysis of reflectance spectra from hyperspectral images of poultry carcasses for fecal and ingesta detection. In *SPIE proceedings* (2002), vol. 4816, pp. 317–324.

[94] ZHU, S., LANTERMAN, A., AND MILLER, M. Clutter modeling and performance analysis in automatic target recognition. In *Proceedings of Workshop on Classification and Detection of Difficult Targets* (Redstone Arsenal, 1998), pp. 477–496.