TOWARDS THE MITIGATION OF CORRELATION EFFECTS IN THE ANALYSIS OF HYPERSPECTRAL IMAGERY WITH EXTENSIONS TO ROBUST PARAMETER DESIGN

DISSERTATION

Jason P. Williams, Captain, USAF

AFIT/DS/ENS/12-07

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.

AFIT/DS/ENS/12-07

TOWARDS THE MITIGATION OF CORRELATION EFFECTS IN THE ANALYSIS OF HYPERSPECTRAL IMAGERY WITH EXTENSIONS TO ROBUST PARAMETER DESIGN

DISSERTATION

Presented to the Faculty

Department of Operational Sciences

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

Jason P. Williams, B.S., M.S.

Captain, USAF

August 2012

AFIT/DS/ENS/12-07

TOWARDS THE MITIGATION OF CORRELATION EFFECTS
IN THE ANALYSIS OF HYPERSPECTRAL IMAGERY WITH
EXTENSIONS TO ROBUST PARAMETER DESIGN

Jason P. Williams, B.S., M.S.
Captain, USAF

Approved:

Dr. Kenneth W. Bauer Dissertation Advisor	3 Aug 12 Date
Lt Col Mark A. Friend, PhD Committee Member	3 Aug 2012 Date
Dr. Mark E. Oxley Committee Member	3 Aug 2012 Date

Accepted:

M. U. Thomas Dean, Graduate School of Engineering and Management	14 Aug 2012 Date

Abstract

Standard anomaly detectors and classifiers assume data to be uncorrelated and homogeneous, which is not inherent in Hyperspectral Imagery (HSI). To address the detection difficulty, a new method termed Iterative Linear RX (ILRX) uses a line of pixels which shows an advantage over RX, in that it mitigates some of the effects of correlation due to spatial proximity; while the iterative adaptation from Iterative Linear RX (IRX) simultaneously eliminates outliers.

In this research, the application of classification algorithms using anomaly detectors to remove potential anomalies from mean vector and covariance matrix estimates and addressing non-homogeneity through cluster analysis, both of which are often ignored when detecting or classifying anomalies, are shown to improve algorithm performance.

Global anomaly detectors require the user to provide various parameters to analyze an image. These user-defined settings can be thought of as control variables and certain properties of the imagery can be employed as noise variables. The presence of these separate factors suggests the use of Robust Parameter Design (RPD) to locate optimal settings for an algorithm. This research extends the standard RPD model to include three factor interactions. These new models are then applied to the Autonomous Global Anomaly Detector (AutoGAD) to demonstrate improved setting combinations.

To My Girls

Acknowledgments

I would like to express my sincere gratitude to my advisor, Dr. Kenneth Bauer. Making the decision to get a PhD was easy when I realized I would have to opportunity to work for him once again. His knowledge and experience were invaluable, and his ability to pick me up when I was down and knock me down when solved the unsolvable truly kept me on track to finish on time.

I would also like to thank the other members of my committee, specifically Lt Col Mark Friend for his assistance and encouragement in the completion of this document. At the end of the day I can always say I had a Friend on my committee.

To the ENS guys, I guess that includes Trevor, the PhD-12S guys, and the official members of the Unofficial AFIT Homebrew Club, thanks for being there when I needed a break or I had "code running," you have all been true friends.

Finally, and most importantly, I would like to thank my wife for supporting me in the decision to return to AFIT and listening to me talk about my research for three straight years this time; you always appeared interested. I could not have done any of this without you.

Jason P. Williams

Table of Contents

Page

List of Figures

Figure **Page**

List of Tables

Table **Page**

TOWARDS THE MITIGATION OF CORRELATION EFFECTS
IN THE ANALYSIS OF HYPERSPECTRAL IMAGERY WITH
EXTENSIONS TO ROBUST PARAMETER DESIGN

1 Introduction

1.1 Background

Hyperspectral Imagery (HSI) is a method used to collect contiguous data across a large swath of the electromagnetic spectrum, which is accomplished by using a specialized camera mounted on an aircraft or satellite to take a picture of the required area, thereby recording the magnitude of the bands within the collected wavelengths. Typically, HSI encompasses the visible to infrared regions of the spectrum, containing anywhere from more than 20 to 250 plus spectral bands, whereas standard digital cameras capture three bands: red, green, and blue. The electromagnetic spectrum, shown in Figure 1, is comprised of various wavelengths, measured in micrometers (μm) or nanometers (nm), commonly by the visible region, but also includes X-rays, ultraviolet, infrared, micro-waves, etc. (Landgrebe, 2003).

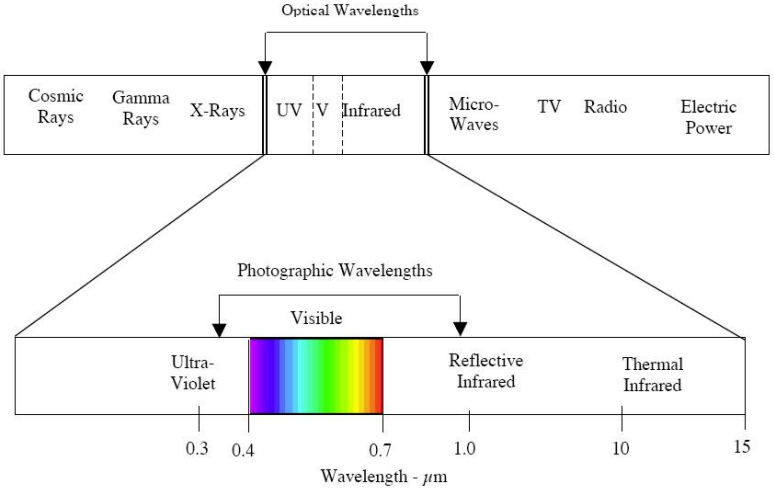

Figure 1: The Electromagnetic Spectrum (Landgrebe, 2003)

When dealing with HSI data, anomaly detection is used to find objects of interest within the image by locating pixels that are statistically different from the background. The vast amount of data contained in HSI affords a great opportunity to detect anomalies in an image using standard multivariate statistical techniques, as each material reflects individual wavelengths of the spectrum differently. Figure 2 shows a spectral space plot of water, trees, and soil. This gives a good visual representation of how various materials reflect individual wavelengths. The three plots across the entire spectrum shown are very different. However, there are regions where they overlap and become indistinguishable. This highlights the benefit of collecting a vast amount of wavelengths over the three used for a standard color image. However, the large amount of data contained within each image often requires dimensionality reduction/feature selection techniques to be employed such that analysis of the image data operates on lower dimensional, uncorrelated data (Landgrebe, 2002).

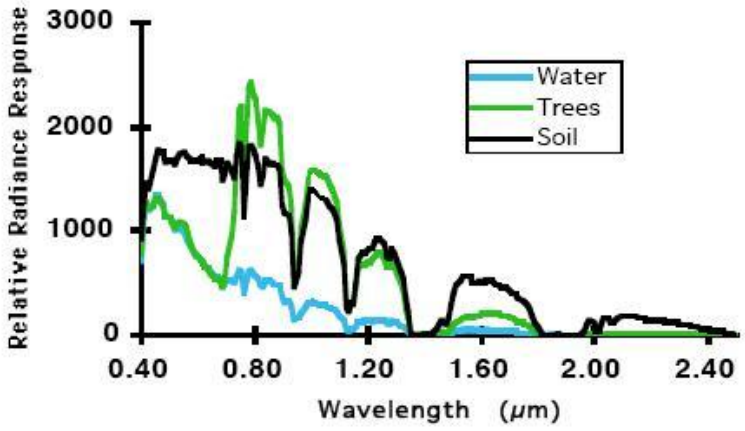

Figure 2: Spectral Space Plot (Landgrebe, 2003)

When a hyperspectral image is collected, the data is stored in a three-dimensional matrix, referred to as an image cube or data cube, displayed in Figure 3. The first two dimensions of the data cube correspond to the location of the pixel in the image, and the third dimension represents the different spectral bands that were collected (Landgrebe, 2003). Prior to processing an image, for anomaly detection or classification, it is usually transformed into a data matrix. A data matrix consists of an $n \times p$ matrix where n is the total number of pixels in the image consisting of p spectral bands, therefore a single pixel is represented by a $1 \times p$ vector.

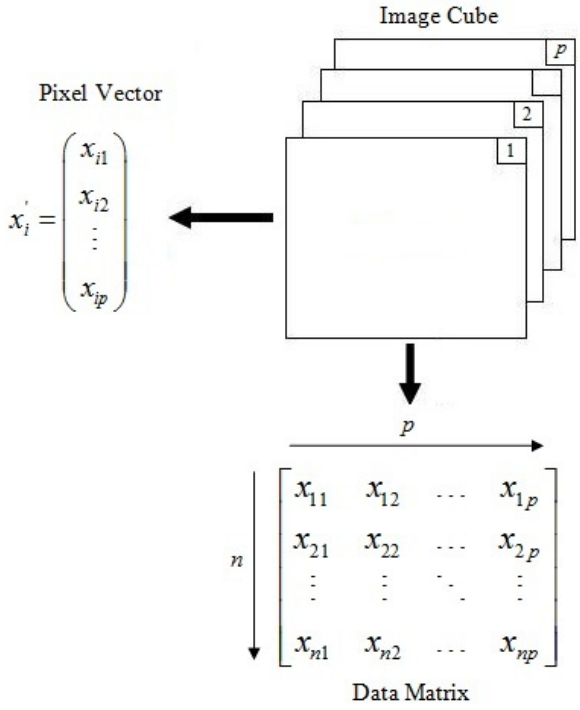

Figure 3: Image Cube

Current anomaly detectors, such as the RX anomaly detector created by Reed and Yu (1990), are likely to have a high false positive detection rates because they assume the data is modeled with a Gaussian distribution. However, it is has been shown that

hyperspectral data is not often unimodal (Banerjee et al., 2006). Further, to compound the non-Gaussian difficulty, the data, by its very nature, is correlated and heterogeneous. There are four main correlation problems inherent to HSI that if addressed properly could potentially benefit anomaly detection and classification: spatial correlation (correlation between pixels due to proximity), spectral correlation (correlation between spectral bands), the presence of outliers or anomalies, and non-homogeneity. Even though many of the current detectors, such as RX, are hindered by these correlation problems, they are still used in practice because they have a relatively fast processing time, are intuitively easy to understand, and are simple to implement.

Most anomaly detectors have numerous user-defined settings that are required to implement the algorithm. Using improper settings can have a negative effect on the overall performance of the algorithm. Additionally, a particular set of images being analyzed could benefit from one setting combination, whereas another set could be hindered by said combination. Therefore, finding a setting combination that is robust to a vast collection of images is pertinent. This leads to the idea of implementing Robust Parameter Design (RPD) to find the setting combinations which are successful across a wide range of images with little variability. To do this, the image characteristics need to be treated as noise variable and the settings are treated as control variables.

1.2 Original Contributions and Research Overview

The first goal of this research is to address correlation problems inherent to HSI that are often ignored by the research community when performing anomaly detection or classification. The four main correlation problems are: spatial correlation (correlation between pixels due to proximity), spectral correlation (correlation between spectral

bands), the presence of outliers or anomalies, and non-homogeneity. The second goal of the research is to extend the standard Noise by Noise (N × N) RPD model recently introduced by Mindrup et al. (2012) to include control by noise by noise (C × N × N) and noise by control by control (N × C × C) interactions.

Chapter 2 will introduce two new anomaly detectors: Linear RX (LRX), a variant of Reed and Yu (1990) RX detector, and Iterative Linear RX (ILRX), a variant of the Taitano et al. (2010) Iterative RX (IRX) detector. LRX addresses spatial correlation related to RX by establishing a mean vector and covariance matrix using data that is, on average, farther from each other than the standard RX window. The IRX detector allows for the exclusion of outliers in the mean vector and covariance matrix calculations, thereby promoting a more accurate assessment of the target pixel. ILRX then exploits the innovations of both LRX and IRX.

Chapter 3 continues addressing correlation in HSI, but this time with the goal of classification. The Adaptive Matched Filter (AMF) with Manolakis et al. (2009) suggested improvements, to be called the Robust AMF, is competed against the Standard AMF. The improvements suggested by Manolakis et al. (2009) are to remove the anomalies from the image prior to calculating the required mean vector and covariance matrix. Additionally, two more AMFs will be tested against the Standard AMF and Robust AMF. Clustered AMF, which clusters the image after removal of the anomalies and classifies the pixel of interest using the mean vector and covariance matrix of the cluster in which the pixels is located; and Largest Cluster AMF, which similarly clusters the image after removal of the anomalies, however, it classifies the pixel of interest using the mean vector and covariance matrix of the largest cluster in the image. Robust AMF

addresses the problem of anomalous pixels skewing the required statistics. Clustered AMF and Largest Cluster AMF exploit the idea of Robust AMF and address the concern of non-homogeneity.

Chapter 4 provides the required statistical models to extend the Mindrup et al. (2012) N × N RPD model to include higher order terms, including the C × N × N and N × C × C interaction terms. These higher order models will then be applied to the Autonomous Global Anomaly Detector (AutoGAD), a HSI anomaly detector, to locate better operating parameter settings, using properties of the hyperspectral images as system noise (Johnson et al., 2012). The benefit of the models will be demonstrated through increased R^2_{adj} and decreased Mean Squared Error (MSE), and new AutoGAD settings which provide higher mean responses and lower response variance.

2 Towards the Mitigation of Correlation Effects in Anomaly Detection for Hyperspectral Imagery

2.1 Introduction

Remote sensing involves studying a given object without initiating physical contact (Eismann, 2012; Schott, 1997); of particular interest are passive remote sensing systems which rely on natural sources of illumination. Hyperspectral Imagery (HSI) systems are passive systems which collect spectrally contiguous data across a large swath of the electromagnetic spectrum, permitting material identification through fine spectral sampling. One of the fundamental problems faced by practitioners in this area is analyzing the highly correlated data streams that are output from these models (Banks et al, 2009). Computer models, such as discrete-event simulations, are used to aid in understanding real-world processes. Simulation analysts must deal with temporal correlation. In this research, we are concerned with highly correlated data of both a spatial and spectral nature. Specifically, we will address the spatial correlation problem.

Typically, HSI encompasses the visible to infrared regions of the spectrum, containing anywhere from more than 20 to 250 plus spectral bands, whereas standard digital cameras capture three coarsely sampled bands: red, green, and blue. The vast amount of data contained in HSI affords a great opportunity to detect anomalies in an image using standard multivariate statistical techniques, as each material reflects individual wavelengths of the spectrum differently. However, the large amount of data contained within each image often requires dimensionality reduction/feature selection techniques to be employed such that analysis algorithms operate on lower dimensional, uncorrelated data as described by Landgrebe (2002).

Anomaly detection refers to the location of spectral data that does not belong within a given set. It can be used in numerous applications such as financial fraud detection, computer security, and military surveillance (Chandola et al., 2009). In HSI applications, anomaly detection is used to find objects of interest within the image by locating pixels statistically different from the non-anomaly pixels, referred to as the background. Three broad categories of anomaly detection methods exist (Chandola et al., 2009): supervised, semi-supervised, and unsupervised detection. Supervised detection requires a set of training data that includes both the background and anomaly data prior to analysis. Semi-supervised detection also requires a training set; however, it only requires background data. Differences between images, e.g., the desert and forest images in this research present a problem that effects supervised or semi-supervised methods when applied to HSI. Therefore, it is difficult to train a detector on one image and test it against another. The standard work around for semi-supervised detection is to select a random set of data. This practice is successful because the set of anomalies in the data set is assumed to be sparse; hence, the random selection should provide a representative sample of the true background. Unsupervised detection does not require a training set, and is therefore more appropriate when analyzing HSI data.

The literature on anomaly detection in HSI has increased following the publication of Reed and Yu's paper on the RX detector in 1990 (Reed and Yu, 1990), to include various articles with modifications or additions to the RX detector (Eismann, 2012; Hsueh and Chang, 2004; Yanfeng et al., 2006; Liu and Chang, 2008; Taitano et al., 2010), classification and discrimination methods (Eismann, 2012; Chang and Ren, 2000; Chang and Chiang, 2002), different fusion techniques (Acito et al., 2006; Nasrabadi,

2008), and overview articles (Manolakis and Shaw, 2002; Stein et al., 2002; Smetek and Bauer, 2008) of detection algorithms, including RX. Related work includes a number of additional detectors, such as: Support Vector Data Description (SVDD) (Tax and Duin, 1999; 2004; Banerjee et al., 2006; 2007), multiple window detectors (Yanfeng et al., 2006; Kwon et al., 2003; Liu and Chang, 2004), and various mixture models (Eismann, 2012; Smetek and Bauer, 2008; Grossman et al., 1998; Clare et al. 2003). More recently, work has been conducted using synthetically generated or simulated data to supplement the low number of hyperspectral images with available truth masks that are typically accessible to researchers (Huesh and Chang, 2004; Shi and Healey, 2005; Gaucel et al., 2005; Bellucci et al., 2010).

In practice, the RX method, when applied to hyperspectral data, is likely to have a high false positive detection percentage because the underlying statistics assumes the data being analyzed follows a Gaussian distribution. However, Banerjee et al. (2006) showed that HSI is not often unimodal. Further, to compound the non-Gaussian difficulty, an image, by its very nature, is correlated and heterogeneous. However, RX is still used in practice because it offers fast processing times, is intuitively easy to understand, and is algorithmically simple.

The purpose of this chapter is to present modifications to the standard RX algorithm. A new method, called Linear RX (LRX), has the ability to overcome some of the correlation problems hindering RX (Reed and Yu, 1990) and Iterative RX (IRX) (Taitano et al. 2010). This research contrasts the performance of LRX and, another new method, its variant Iterative Linear RX (ILRX), to RX and IRX. Additionally, to further

test the benefit of the new algorithms, both algorithms are tested against the global SVDD algorithm, a promising new supervised HSI detector (Banerjee et al., 2006; 2007).

The remainder of this chapter is organized as follows: Section 2 presents a description of the algorithms contrasted in this research, Section 3 details the methodology used to compete the five anomaly detectors, Section 4 provides the experimental results, and in Section 5, the chapter is concluded.

2.2 Algorithms

This section of the chapter describes how each of the five anomaly detection algorithms are contrasted and implemented, and explains the use of the Normalized Difference Vegetation Index (NDVI) in post-processing to realize improved results from the detectors. Due to the large amount of data contained within a given hyperspectral image, it is standard practice, prior to applying an anomaly detector, to reduce the dimensionality of the image by running Principal Component Analysis (PCA) (Farrell and Mersereau; 2005) on the whole data set, retaining the *P* largest Principal Components (PCs).

2.2.1 The RX Detector (RX)

The RX detector, introduced by Reed and Yu (1990), detects anomalies utilizing a moving window approach, where the pixel in the center is scored by comparison to the remaining pixels in the window. The window, usually square in shape, is shifted, one pixel at a time, across a row of pixels with the new center pixel being scored at each step, as displayed in Figure 4a, where the red square represents the test pixel and the box around the test pixel represents the pixels compared with the test pixel to generate an RX

score. This process is continued until all possible pixels have been analyzed. Each test pixel, x, is given a score based upon a generalized likelihood ratio test which simplifies to equation (1) if the pixels within the test window are assumed to be normally distributed with mean vector of the background pixels, μ, and covariance matrix Σ. It should also be noted that as the number of pixels in the window, N, approaches to infinity, the *RX* score becomes the squared Mahalanobis distance between the test pixel and the mean vector of the background pixels,

$$RX(x) = (x-\mu)^T \left[\left(\frac{N}{N+1} \right) \Sigma + \left(\frac{1}{N+1} \right) (x-\mu)(x-\mu)^T \right]^{-1} (x-\mu). \qquad (1)$$

Pixels with an *RX* score greater than $\chi^2_{\alpha, (N-1),}$ where α represents the corresponding significance level of the chi-squared distribution, are labeled anomalous by the RX detector.

2.2.2 The Iterative RX Detector (IRX)

The IRX detector (Taitano et al., 2010) is an extension of the standard RX detector; IRX extends the RX detector through an iterative process, where each iteration sees IRX calculating an improved estimate of the mean vector and covariance matrix of the background pixels.

The IRX algorithm is processed using the following steps:

1. Each iteration begins by running the standard RX algorithm to calculate an RX score, i.e. $RX(x)$, for each testable pixel in the image; however, to improve accuracy, pixels selected as anomalies in the previous iteration are excluded from the data used to estimate the mean vector and covariance matrix of the background. Note: At the start of the algorithm the set of anomalies is empty.

2. Using the RX scores calculated in step 1, a pixel, *x*, is considered anomalous if its RX score is greater than $\chi^2_{\alpha, (N-1)}$. This ends a given iteration, allowing for pixels to enter and exit the set of anomalies.
3. The algorithm ends if the set of anomalies determined in step 2 is identical to the set of anomalies from the previous iteration or the maximum number of iterations has been reached. Otherwise, the algorithm iterates again from step 1.

2.2.3 The Linear RX (LRX) and Iterative Linear RX (ILRX) Detectors

LRX and ILRX are similar to RX and IRX, respectively, however, instead of a window being moved through the image, they employ a vertical line of pixels above and below the test pixel. If the number of pixels above or below the test pixel exceeds the height of the image, the required pixels are taken from the bottom of the previous column or from the top of the following column, Figure 4b. The line is used to increase the average distance between the pixels used to estimate the mean vector and covariance matrix. This can be seen in Figure 5, which shows the average distance between pixels using a window and a line. Increasing the average distance between pixels mitigates the deleterious effects of correlation due to the spatial proximity of the pixels. Such a step allows for the reduction of bias and error in the estimation of the mean vector and covariance matrix. A possible concern for such an approach might be that the reduction in the contribution to the bias due to spatial correlation may be offset by the contribution to the bias due to image non-stationarity. This issue is discussed later in the chapter and as demonstrated below, is not a concern for images we tested.

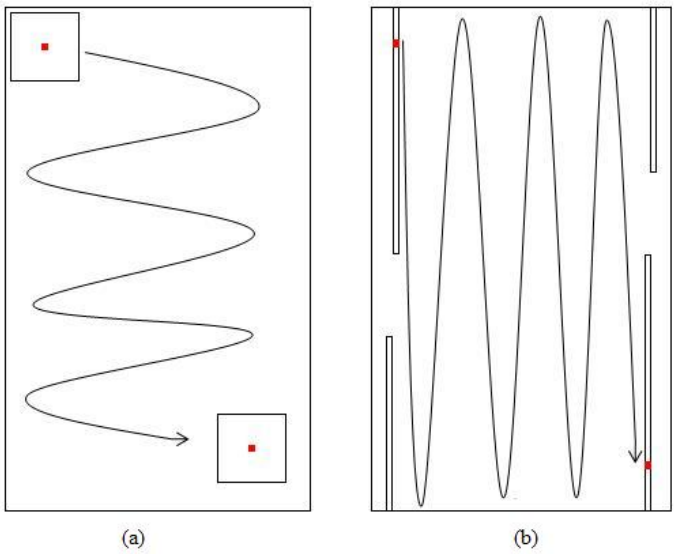

Figure 4: RX Window vs. LRX Line

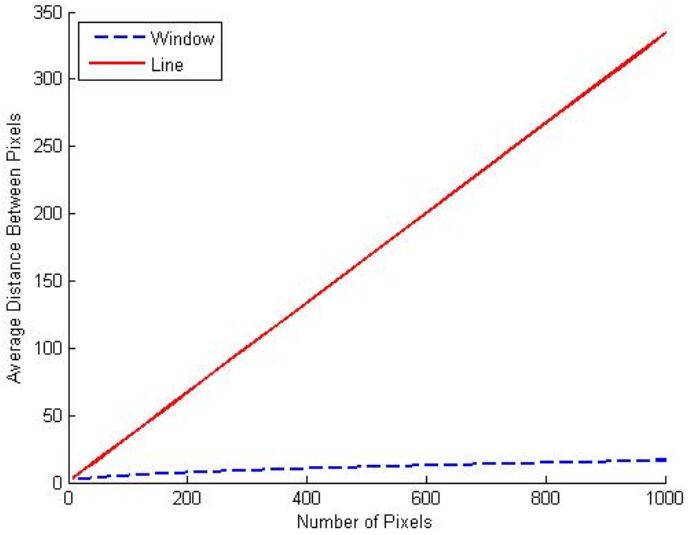

Figure 5: Average Distance Between Pixels

2.2.4 Support Vector Data Description (SVDD)

Banerjee et al. (2006; 2007) extended the SVDD algorithm by Tax and Duin (1999; 2004) into an HSI anomaly detector. SVDD is a one-class classifier, where points are considered in class or out of class, where the support of the distribution is considered

as the minimally enclosing hypersphere in the feature space. In operation, SVDD takes a training set, $T = \{x_i,\ i = 1,\ \ldots\ , M\}$, of M background pixels, x, is randomly selected from the image as the training data. SVDD then attempts to determine the minimum volume hypersphere, $S(R,a) = \{x : \|x - a\| \le R\}$, as the L^2 norm or Euclidean norm, with radius $R > 0$ and center a that contains the set of M randomly chosen pixels. This is obtained by solving the following minimization problem,

$$\min(R) \text{ subject to } x_i \in S,\ i = 1,...,M. \tag{2}$$

The radius R and center a of the hypersphere are determined by minimizing the Lagrangian, L, with respect to the weights, or support vectors, α_i,

$$L(R,a,\alpha_i) = R^2 - \sum_{i=1}^{M} \alpha_i \left\{ R^2 - \left(\langle x_i, x_i \rangle - 2\langle a, x_i \rangle + \langle a, a \rangle \right) \right\}, \tag{3}$$

where $\langle \cdot, \cdot \rangle$ represents the dot product of the operation of the two vectors.

After optimizing, the kernel technique, which transforms data to a different dimensional space for simpler computations without ever explicitly calculating the mapping, can be applied which leads to the SVDD statistic,

$$SVDD(y) = R^2 - K(y,y) + 2\sum_{i=1}^{M} \alpha_i K(y, x_i), \tag{4}$$

where $K(x,y)$ is the kernel mapping defined by

$$K(x,y) = \exp\left(-\|x - y\|^2 / \sigma^2 \right), \tag{5}$$

and variable σ^2 is a radial basis function parameter used as a scaling factor to determine the size of the hypersphere, hence adjusting how well the SVDD algorithm generalizes the incoming data.

When applied to HSI, the SVDD algorithm is processed in the following steps (Banerjee et al., 2006):

1. Randomly select *M* pixels from the image.
2. Estimate an optimal value for σ^2 by determining a value that will minimize the false positive rate or the number of background pixels classified as targets.
3. Estimate the parameters (R, a, α_i) needed to model the hypersphere.
4. Determine *SVDD(y)* for each test pixel. If $SVDD(y) \geq t$, for a user defined threshold t, the pixel is labeled an anomaly, otherwise if $SVDD(y) < t$ the pixel is labeled as background.

The SVDD algorithm is considered in this research because it is a novel and promising state of the art detector and as a semi-supervised method, it allows for an interesting performance contrast relative to the other unsupervised methods.

2.2.5 Normalized Difference Vegetation Index (NDVI)

It is fairly common to get false positives, i.e. pixels labeled as anomalies that are truly background pixels, when attempting to find anomalies, generally man-made objects, in HSI using one of the previously described methods. One relatively simple way to reduce false positives is to implement some form of pre- or post-processing. Since we are attempting to locate anomalies without prior knowledge, one applicable post-processing method is applying the Normalized Difference Vegetation Index (NDVI), as introduced by Rouse et al. (1973), to remove pixels that are likely to be vegetation.

NDVI gauges whether or not a given pixel is green vegetation by using the absorptive cutoff of chlorophyll between the visible and near infrared spectrum. It does this by comparing the intensity of the visible bands to the intensity of the near infrared

bands, since the reflectance in the near infrared bands is considerably larger for vegetation. The measure is given by

$$NDVI = \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}}, \tag{6}$$

where NIR denotes the radiance value of the near infrared spectral band and Red denotes the radiance value of the red spectral band (Eismann, 2012; Schott, 1997; Rouse et al., 1973; Landgrebe, 2003). Prior to locating the anomalies, an NDVI threshold for the image and an NDVI score for each pixel is calculated. The NDVI threshold was determined by plotting the NDVI scores for all of the images and setting a threshold. Subsequently, all pixels with an NDVI score above the threshold are classified as vegetation. Once the anomaly detector has been run, regardless of the indications, all declared vegetation pixels are classified as background. Since the desert images display NDVI scores that are, for the most part, below the selected threshold, very few pixels will be classified as vegetation; hence, very few potential false positives are deleted.

2.3 Methodology

The five anomaly detectors were compared using six images from the Forest Radiance I and Desert Radiance II collection events, from the Hyperspectral Digital Imagery Collection Equipment (HYDICE) push-broom, aircraft mounted sensor (Rickard et al., 1993). The HYDICE sensor collects spectral data in 210 bands between 397 nm and 2,500 nm, including visible, and infrared data. Due to atmospheric absorption effects, only 145 bands were used in the analysis of the images. A description of each image is shown in Table 1 and the images are displayed in Figure 6.

Table 1: ARES Image Data

Image	Size	Total Pixels	Anomalies	Anomalous Pixels
ARES 1D	291 x 199	57,909	6	235
ARES 1F	191 x 160	30,560	10	1,007
ARES 2D	215 x 104	22,360	46	523
ARES 2F	312 x 152	47,424	30	307
ARES 3F	226 x 136	30,736	20	145
ARES 4F	205 x 80	16,400	29	109

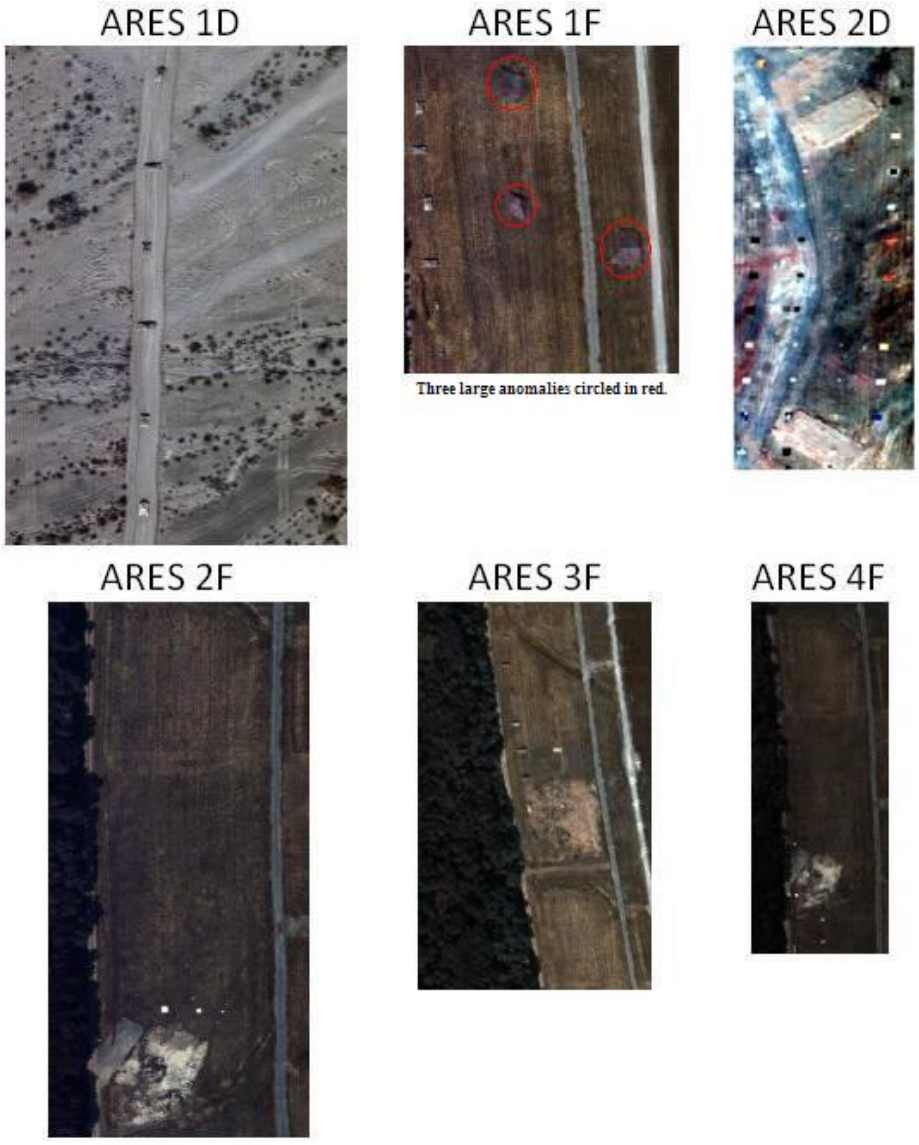

Figure 6: ARES Images

Due to the small number of images available and the need to train and validate each of the algorithms, all six images, presented in Table 1 and Figure 6, were divided in half to create a top and bottom portion of the image. Then, a top or bottom of each image, chosen randomly, was selected for the training set of images and the other half was used in the validation set. The training set included the top of ARES 2D and ARES 4F and the bottom of ARES 1D, ARES 1F, ARES 2F, and ARES 3F. It may be a stretch to try and draw too much from a comparison of five algorithms and only six images; however, we believe our experiments point to the clear potential of the new technique. Furthermore, while splitting the images in half to double our data may not be the best method for creating training and test sets, it is certainly better than using the same images for training and test. We acknowledge that the image halves are spectrally correlated due to shared weather, viewing conditions, etc. However, correlation in the spatial domain appears to be minimal.

Each of the algorithms was tested across a large combination of parameter settings in order to find the optimal settings for the algorithm. The parameters were: the number of PCs to retain, the number of pixels to use in the window/line using RX methods or the size of the training set for SVDD, the number of iterations to use for the iterative methods, whether NDVI was used in post processing, and the parameter value, σ for SVDD. This is summarized in Table 2. The last column is displayed to show how many combinations of parameters were collected for each method on a single image.

Table 2: Algorithm Parameter Settings

Algorithm	Number of PCs	Number of Pixels*	Number of Iterations	σ^2	NDVI	Total Data Points Collected (Per Image)
RX	3, 4,…, 10	17^2, 19^2,…, 25^2	1	-	Yes/No	80
IRX	3, 4,…, 10	17^2, 19^2,…, 25^2	10, 20,…, 50	-	Yes/No	400
LRX	3, 4,…, 10	0.5*H, 1*H, 1.5*H, 2*H	1	-	Yes/No	64
ILRX	3, 4,…, 10	0.5*H, 1*H, 1.5*H, 2*H	10, 20,…, 50	-	Yes/No	320
SVDD	3, 4,…, 10	17^2, 19^2,…, 25^2	1	10, 20,…, 300	Yes/No	2,400

The algorithms' anomaly detection performance on the selected test set was compared through the use of Receiver Operating Characteristic (ROC) curves (Fawcett, 2006). Since the RX algorithms' test statistics are based upon the chi-squared distribution, the significance level α was varied to serve as the threshold for the ROC curves. Similarly, the user-defined anomaly threshold t was varied in the SVDD algorithm to generate ROC curves. Due to the fact that a large number of settings for each algorithm were examined, visual inspection of the ROC curves was not feasible. Therefore, the individually-tested setting combinations for each algorithm were scored using the Neyman-Pearson technique (Kay, 1993). Specifically, the True Positive Fraction (TPF) for the anomalous pixels detected in each of the six images was averaged when the corresponding False Positive Fraction (FPF) is equal to 0.1. A FPF of 0.1 was chosen because it was deemed that if the FPF exceeded 0.1, the algorithm would no longer be of any practical use due to over-saturation of misclassified data.

After the setting combination with the highest average TPF at a FPF = 0.1 was determined for each anomaly detector, its performance was validated by taking the best settings for each individual algorithm, and running them on the six validation images.

An artifact of the RX and IRX methods, as described by Reed and Yu (1990) and Taitano et al. (2010), is an area of pixels that form a border around the image which

cannot be tested due to the requirement of the window. Methods to allow the algorithms to test the border pixels can be implemented, such as using only the part of the window that is within the image or moving the test pixel from the center of the window when it is against the border of the image. However, in this research, the RX and IRX algorithms as originally designed were competed and the border pixels that could not be tested were not considered in the performance evaluation.

2.4 Results

Relative to the training data, the results for the best settings of each algorithm by image and overall average are shown in Table 3. It can be seen that LRX achieves equivalent performance to RX in most images; ARES 1F is an exception where the spatially large objects appear to confound the RX algorithm, yet are detected by LRX. With iterations, ILRX was the best performing algorithm or tied with IRX in all cases, except ARES 2F.

Table 3: Training Data Results (TPF at FPF = 0.1)

Algorithm	ARES 1D	ARES 1F	ARES 2D	ARES 2F	ARES 3F	ARES 4F	Average
RX	0.8673	0.3410	0.9933	0.9455	0.9535	0.8649	0.8276
IRX	1.0000	0.4615	1.0000	1.0000	0.9744	0.9444	0.8967
LRX	0.9118	0.7916	0.9474	0.7632	0.8308	0.7990	0.8406
ILRX	1.0000	1.0000	1.0000	0.9912	1.0000	0.9500	0.9902
SVDD	0.9558	0.9588	0.9880	0.9386	0.9846	0.8750	0.9501

The corresponding best tested setting for each of the algorithms is displayed in Table 4. "Yes" or "No" in the NDVI column for SVDD implies the algorithm achieved the same results whether or not NDVI was used in post processing. It should be noted that ILRX was the most robust of the algorithms tested. During training, eleven different parameter settings realized the same values shown in Table 3. No other algorithm had

multiple parameter settings that obtained the optimal results. Since multiple setting combinations were found for ILRX they were all tested on the validation images.

Table 4: Best Tested Parameter Settings from Training Data

Algorithm	Number of PCs	Number of Pixels	Number of Iterations	σ^2	NDVI
RX	9	23^2	1	-	Yes
IRX	9	25^2	20	-	Yes
LRX	9	1*H	1	-	Yes
ILRX	10	2*H	30	-	Yes
SVDD	10	25^2	1	60	Yes or No

The results from the validation images are displayed in Table 5, to include the best and worst tested parameter settings of eleven training combinations validated for ILRX. It can be seen that ILRX is still the top performer overall regardless of whether the best or worst training settings were implemented. Furthermore, the ILRX algorithm received the smallest drop in average TPF when the settings were tested on the validation images, as compared to the training images.

Table 5: Validation Data Results (TPF at FPF = 0.1)

Algorithm	ARES 1D	ARES 1F	ARES 2D	ARES 2F	ARES 3F	ARES 4F	Average
RX	0.9016	0.2075	0.9920	0.8282	0.9545	0.8864	0.7950
IRX	1.0000	0.3186	1.0000	0.9495	1.0000	1.0000	0.8780
LRX	0.9645	0.4902	0.9890	0.8342	0.9104	0.7669	0.8259
ILRX (Best)	1.0000	0.9449	1.0000	0.9741	1.0000	1.0000	0.9865
ILRX (Worst)	1.0000	0.7394	1.0000	0.9646	1.0000	0.9744	0.9464
SVDD	0.9180	0.9850	0.9983	0.8641	0.9330	0.8200	0.9197

Figure 7 shows the ROC curves for each of the six validation images comparing TPF to FPF using the best tested settings for each algorithm from the training images, as

displayed in Table 4. In every case, IRX performs better than RX and ILRX performs better than LRX; hence, the comments below focus on IRX, ILRX, and SVDD.

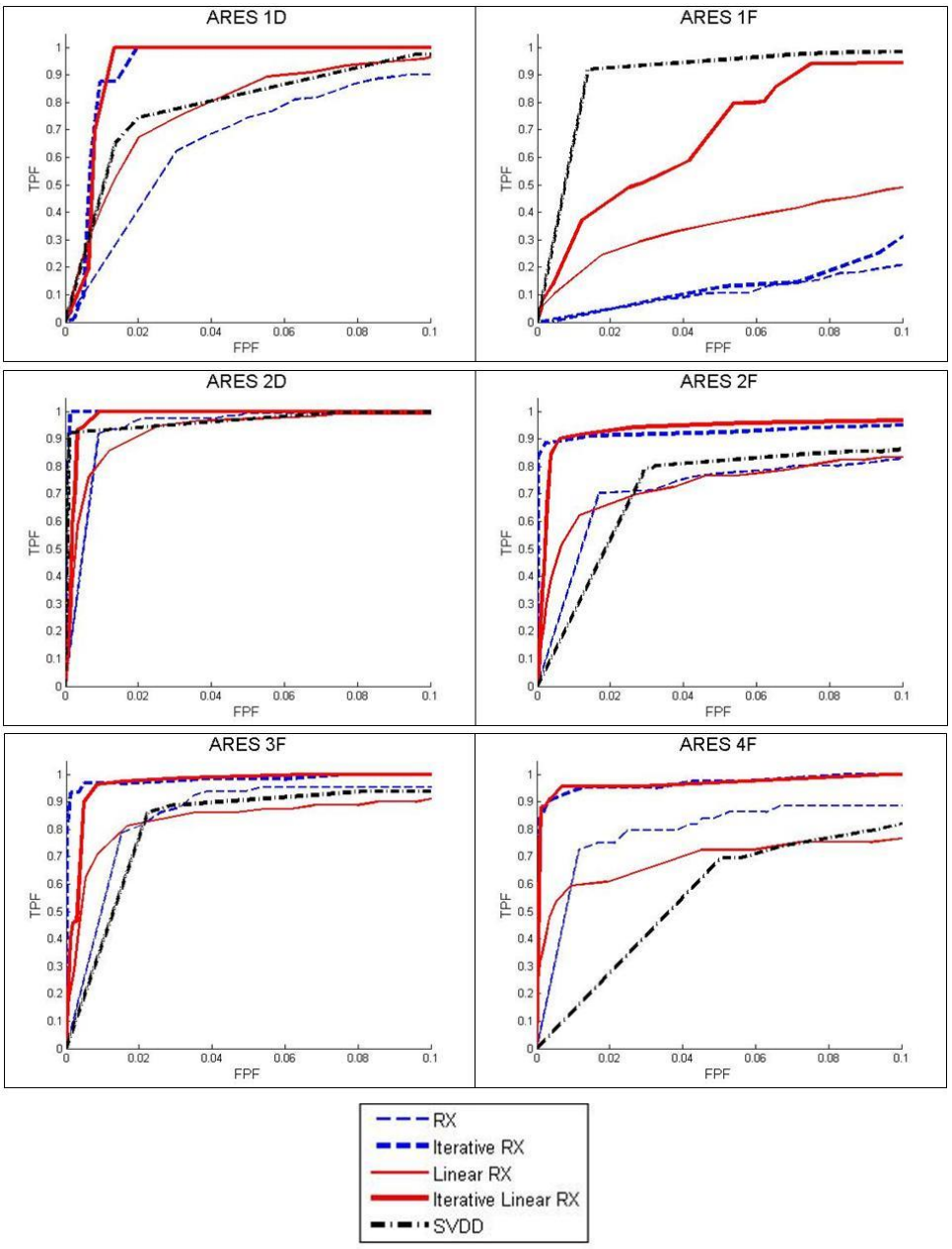

Figure 7: ROC Curves of Best Tested Parameter Settings on Validation Images

IRX did well on all of the images except when there are large anomalies, such as the ones highlighted in ARES 1F. This is because the window, as it moves through a

large anomaly, becomes dominated by the local anomalous pixels rather than the general background of the image. This defeats the purpose of the window, which is to give a good estimate of the true background of the image. As a result, the pixel being analyzed appears similar to the other pixels in the window and is not classified as an anomaly. ILRX mitigates this problem through its use of a vertical line which only contains a small portion of even a large anomaly and considerably more background pixels.

ILRX had the highest performance or was comparable with the other detectors in all of the images. It had slight problems with the rock formations in ARES 2F and 4F that IRX does not detect due to the window effect of large images; however, this is difficult to discern from the ROC curve due to ILRX detecting most of the anomalies at a relatively low threshold.

SVDD consistently performed better than both of the non-iterative methods, however, it was inconsistent with regard to its performance against the iterative methods. Also, the fact that it is a semi-supervised method that randomly selects training data can lead to less than optimal performance from the detector. The only image where SVDD outperformed the other algorithms is ARES 1F, where IRX has trouble with large anomalies and ILRX has difficulty with vertical roads.

Figure 8 shows the color representation of the image and the pixels classified as anomalies, or anomalous pixel maps, for IRX, ILRX, and SVDD on the validation images ARES 1F and 4F, note that the masks of IRX are smaller because it was not used to test the borders of the image. The anomalous pixel maps were generated at the first knee in the ROC curve so that they were not overwhelmed by false positive pixels. The corresponding TPF and FPF are displayed below each of the images. It can be easily

seen in ARES 1F that SVDD is realizing superior results, primarily because IRX is not locating the large anomalies and ILRX in addition to finding almost all of the anomalous pixels is having some difficulties with the roads. In ARES 4F both of the RX methods are giving high-quality results and the SVDD algorithm is getting inundated by the large rock formation.

The embellishments to RX follow a reasoned pattern. IRX allows for the exclusion of outliers in the local mean vector and covariance matrix calculations, thereby promoting a more accurate assessment of the target pixel (2010). LRX mitigates the correlation difficulties related to RX by establishing the mean vector and covariance matrix that is, on average, further from each other than the standard RX window. The possible concern that the reduction in the contribution to the bias due to spatial correlation may be offset by the contribution to the bias due to image non-stationarity was not realized here. We believe this is due to the following factors. If one considers the non-stationarity in the image as being characterized by distinct pixel clusters then the variation between these clusters appears to be significantly less than the variation between the background pixels, in general, and the target pixels. Further, the running covariance matrix estimate calculated across the background pixels appears to be fairly robust to the heterogeneity as evident by the algorithms performance. The notion of using separate estimates from the individual clusters is the subject of current research. Finally, ILRX exploits the innovations of both IRX and LRX. Taken together, these innovations make ILRX a very competitive algorithm.

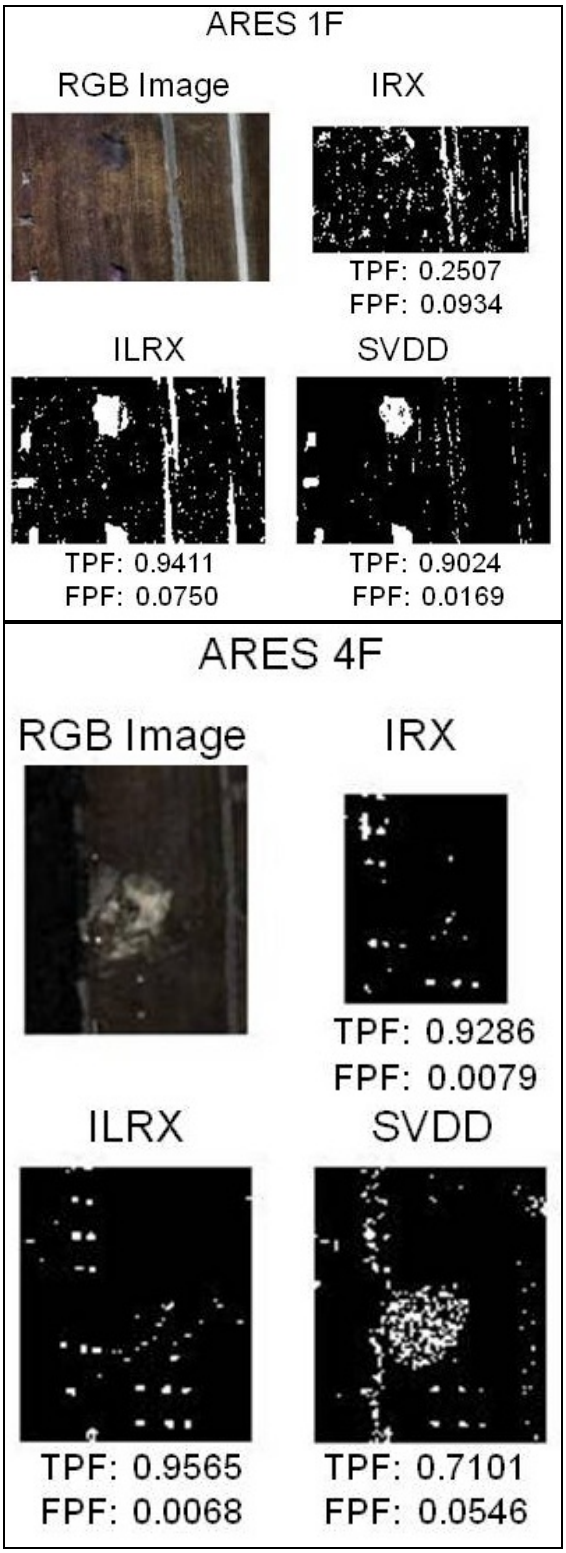

Figure 8: Anomalous Pixel Maps

2.5 Conclusions

This chapter presented LRX and ILRX, updates to the newly introduced IRX algorithm. Through experimentation, the line of pixels used by ILRX shows an advantage over RX and IRX in that it can help mitigate the deleterious effects of correlation due to the spatial proximity of the pixels while the iterative adaptation taken from IRX simultaneously eliminates outliers. Such steps allow for the reduction of bias and error in the estimation of the mean vector and covariance matrix, thus accounting for a portion of the spatial correlation inherent in HSI data. Using the HYDICE images, ILRX has been shown to be very promising unsupervised anomaly detection algorithm.

3 Clustering Hyperspectral Imagery for Robust Classification

3.1 Introduction

Hyperspectral Imagery (HSI) is a method used to collect contiguous data across a large swath of the electromagnetic spectrum. This is accomplished by using a specialized camera mounted on an aircraft or satellite to record the magnitude of the bands within the collected wavelengths of each pixel within the area of interest. The number of pixels in a hyperspectral image depends on the resolution of the camera and the size of the area being imaged. The number of bands recorded is upwards of 200 or more (Shaw and Manolakis, 2002), and typically spans the range from ultraviolet to the infrared regions of the electromagnetic spectrum. The vast amount of data contained in HSI affords an excellent opportunity to detect anomalies using multivariate statistical techniques, as each material reflects individual wavelengths of the spectrum differently (Landgrebe, 2002).

Target detection algorithms can be divided into two groups: anomaly detection algorithms and classification algorithms. Anomaly detection algorithms do not require the spectral signatures of the anomalies they are attempting to locate. A pixel is declared an anomaly if its spectral signature is statistically different than the model of the local or global background that it is being tested against. This implies these algorithms cannot distinguish between anomalies, they only make a decision on whether or not a pixel is anomalous; hence the application can be considered a two-class classification problem (Shaw and Manolakis, 2002). Classification algorithms attempt to identify targets based on their specific spectral signature, however, to accomplish this they require additional information in the form of a spectral library (Manolakis et al., 2009).

Eismann et al. (2009) claim that the mean vector and covariance matrix required for anomaly classification can be estimated globally from the entire image data under the assumption that there are a small number of anomalies in the image and this has an insignificant effect on the covariance matrix. This statement is contested in Smetek (2007), where potential ill effects of a small number of anomalies on the estimation of the covariance matrix are detailed. Similarly, Manolakis et al. (2009) state that:

> Possible presence of targets in the background estimation data lead to the corruption of background covariance matrix by target spectra. This may lead to significant performance degradation; therefore, it is extremely important that, the estimation of μ and $\sum$ should be done using a set of "target-free" pixels that accurately characterize the background. Some approaches to attain this objective include: (a) run a detection algorithm, remove a set of pixels that score high, recompute the covariance with the remaining pixels, and "re-run" the detection algorithm, and (b) before computing the covariance, remove the pixels with high projections onto the target subspace. (Manolakis et al., 2009)

This research demonstrates that classification algorithms, such as the Adaptive Matched Filter (AMF), may be improved by addressing correlation and homogeneity problems inherent to HSI that are often ignored in practice. We begin by showing the benefit of using an anomaly detector to remove potential anomalies from the mean vector and covariance matrix statistics, as suggested by Manolakis et al. (2009). In addition, we show further benefits by addressing the non-homogeneity of HSI through the use of cluster analysis prior to classification.

The remainder of this chapter is organized as follows. Section 2 reviews the basics of classification, describes the Adaptive Matched Filter (AMF) as well as AMF variants used in this research, and discusses atmospheric compensation. Section 3 briefly outlines the seven anomaly detectors implemented. Section 4 discusses clustering, more specifically, the X-means algorithm. Section 5 details the methodology implemented.

Section 6 presents the results of the experiments. Finally, Section 7 concludes the chapter.

3.2 Classification

This section describes the AMF classification algorithms used in this research. Three new variants to the AMF are introduced that have the ability to classify with improved accuracy by addressing correlation and homogeneity problems inherent to HSI. Elementary atmospheric compensation is also discussed, detailing a method to transform a spectral library into the image space to allow for proper classification.

3.2.1 Classification Algorithms

The goal of a HSI statistical classification algorithm is to determine whether or not a test pixel is likely made of the same material as a target pixel. Define the conditional probability density of the test pixel, x, as realized under the alternative hypothesis, H_a (same as the target pixel), as $f_a\left(x \mid H_a\right)$, and the conditional probability density of the test pixel, x, as realized under the null hypothesis, H_0 (not the same as the target pixel), as $f_0\left(x \mid H_0\right)$. The corresponding likelihood ratio is

$$F\left(x\right) = \frac{f_a\left(x \mid H_a\right)}{f_0\left(x \mid H_0\right)}. \tag{7}$$

If $F(x)$ is greater than the user defined threshold, t, then the null hypothesis is rejected, meaning the test pixel is considered a target pixel; otherwise, the null hypothesis cannot be rejected, implying the test pixel is not considered a target pixel (Manalokis et al., 2007). That is if $F\left(x\right) > t$ the test pixel is considered a target pixel or if $F\left(x\right) \le t$ the test pixel is not considered a target pixel.

The classification algorithm utilized in this research is the full-pixel AMF as defined by Manolakis and Shaw (2002). The algorithm assumes the target spectra and background spectra have a common covariance matrix, Σ, and is defined by

$$\text{AMF} = \frac{s^T \Sigma^{-1} x}{s^T \Sigma^{-1} s}. \tag{8}$$

Additionally, it is assumed that the global mean is removed from the estimate of target spectral signature, s, and test pixel spectral signature, x. The spectral signature of the target of interest is a fixed $1 \times p$ vector determined from a spectral library or the mean of a sample of known target pixels collected under the same conditions (Manalokis et al., 2009).

3.2.2 Variants of the AMF

The standard AMF and three variants of the AMF are implemented in this research. The first method is the standard AMF as described above, where the mean vector and covariance matrix are taken from the entire image. The first variant, to be called Robust AMF, is suggested by the quote from Manolakis et al. (2009) in the introduction of the chapter. In this method, an anomaly detector first analyzes the image, then the mean vector and covariance matrix are estimated from the image without the detected anomalies. The second variant is referred to as Clustered AMF. In this method anomalies are removed as in Robust AMF, next the image is clustered without the detected anomalies. Each of the clusters yields a mean vector and covariance matrix estimate. The corresponding background statistics for the pixels to be classified are determined through the modal class of its neighbors. A similar idea has been proposed with anomaly detection (Stein et al., 2000). Due to the time-consuming nature of

determining which cluster a pixel is located in, a third variant is developed and called Largest Cluster AMF. This method removes the anomalies and clusters the resulting data as is done in Clustered AMF; however, the mean vector and covariance matrix for the pixels to be classified are estimated from the single largest cluster of data in the image.

3.2.3 Atmospheric Compensation

Spectral signature matching within HSI typically incorporates a spectral library consisting of ground measured reflectance data from objects of interest. The difficulty with spectral signature matching is that hyperspectral images are collected using a sensor that collects pupil-plane radiance, which includes reflected and radiated energy as well as atmospheric distortions. Before spectral signatures from an image can be compared to target signatures, atmospheric compensation must be performed to bring the spectral library from the reflectance space to the pupil-plane radiance space. Since radiance data is a function of atmospheric conditions, which vary greatly by collection time, the spectral library must be processed with each image separately (Eismann, 2012).

Linear and model based approaches are available to transform data from the reflectance space to the radiance space. Model based approaches, such as MODTRAN (Berk et al., 1999), require prior knowledge about the scene collection. Linear methods assume that atmospheric content is a linear addition where the pupil-plane radiance is a function of reflectance with a scaling multiplier and offset

$$L_i = a\rho_i + b\,, \tag{9}$$

where ρ_i is a reflectance signature to be transformed into the L_i radiance space with gain, *a*, and offset, *b*, as calculated by

$$a = \frac{L_2 - L_1}{\rho_2 - \rho_1}, \tag{10}$$

$$b = \frac{L_1\rho_2 - L_2\rho_1}{\rho_2 - \rho_1}, \tag{11}$$

where ρ_1 and ρ_2 are known reflectance signatures from the spectral library, and L_1 and L_2 are the corresponding radiance measurements from the scene. Linear methods are comprised of two general types, methods such as the Empirical Line Method (ELM), which require known objects of interest to be within the spectral library and located within the image; and vegetation normalization methods, which use expected radiance and reflectance of vegetation in place of specific known objects. Both permit atmospheric compensation to be conducted for the remaining objects in the spectral library (Eismann, 2012).

For situations lacking prior knowledge of scene content, methods such as vegetation normalization are appropriate, where the linear method in equation (9) is applied with radiance measurements for materials expected in the scene. The Normalized Difference Vegetation Index (NDVI) and the Bare Soil Index (BI) are two methods that allow atmospheric compensation to be performed depending on the scene landcover (Eismann, 2012). The images used in this research come from the Hyperspectral Digital Imagery Collection Equipment (HYDICE) (Rickard et al., 1993) sensor for the Forest Radiance I and Desert Radiance II collection events in 1995. The images were collected with 210 bands between 397 nm and 2,500 nm and the ground reflectance data was collected with 430 bands between 350nm and 2,500 nm. The collection names allude to images consisting of forest and desert scenes. Atmospheric compensation was performed

with NDVI for forest images, and BI was applied for the desert images due to the lack of vegetation.

3.2.3.1 Normalized Difference Vegetation Index (NDVI)

NDVI (Rouse et al., 1973) is a method that is used to determine whether a pixel within a hyperspectral image is green vegetation. It does this by comparing the radiance of the Near Infrared (NIR) spectrum to the red spectrum

$$\text{NDVI} = \frac{\text{NIR} - \text{red}}{\text{NIR} + \text{red}}, \tag{12}$$

where red corresponds to the 600 – 700 nm bands and NIR corresponds to the 700 – 1,000 nm near infrared bands (Eismann, 2012). In this research, we used bands corresponding to 660 nm for red and 860 nm for NIR.

NDVI is calculated for each pixel within an image, and pixels with the highest scores can be used as vegetation within the radiance space. Hence, the vegetation in the spectral library can be used as ρ_2 in equations (10) and (11) , and the average spectral signature of the pixels with the highest NDVI score can be used as L_2. L_1 can be determined from the shadows within an image which can be estimated by the spectral signature, which is calculated by taking the minimum value from each band in the image across all pixels. Finally, ρ_1 is set as a vector of zeros, and interpreted as the ideal minimum radiance in the image (Eismann, 2012).

3.2.3.2 Bare Soil Index (BI)

BI (Chen et al., 2004) is a method similar to NDVI, however, it is designed for bare soil within a hyperspectral image. It can be employed in the same fashion as NDVI assuming there is a soil measurement within the spectral library

$$\mathrm{BI} = \frac{(\mathrm{SWIR} + \mathrm{red}) - (\mathrm{NIR} + \mathrm{blue})}{(\mathrm{SWIR} + \mathrm{red}) + (\mathrm{NIR} + \mathrm{blue})}, \tag{13}$$

where blue corresponds to the 450 – 500 nm bands, red corresponds to the 600 – 700 nm bands, NIR corresponds to the 700 – 1,000 nm bands, and Short Wave Infrared (SWIR) corresponds to the 1,150 – 2,500 nm short-wave infrared bands (Eismann, 2012). In this research, we used the band corresponding to 470 nm for blue, 660 nm for red, 860 nm for NIR, and 2,280 nm for SWIR.

3.3 Anomaly Detection

To employ an anomaly detection algorithm to hyperspectral data, first the atmospheric absorption bands should be removed and the data cube must be reshaped into a data matrix. The removal of the absorptions bands in the images employed in this research results in the retention of 145 of the 210 original bands. HSI data is typically stored in a three-dimensional matrix, referred to as an image cube or data cube, with the first two dimensions of the matrix being the location of the pixel in the image and the third dimension being the magnitude at each of the recorded electromagnetic bands. Therefore, it can be viewed as a stack of images with each image representing the intensity of a given band. A $n \times p$ data matrix is generated by reshaping the data cube into a matrix with the first dimension containing all n pixels in the image and the second dimension containing all p bands. After the data is in the proper form, Principal Component Analysis (PCA) (Landgrebe, 2003) is employed as a data reduction tool. In all of the algorithms except AutoGAD the user is left to determine the number of Principal Components (PCs) to retain in order to reduce the dimensionality of the data.

3.3.1 RX Detector

The RX algorithm was developed by Reed and Yu (1990). It detects anomalies through the use of a moving window. The pixel in the center of the window is scored against the other pixels in the window. The window is then shifted by one pixel and the process is repeated until each pixel, *x*, has received an $RX(x)$ score based on

$$RX(x) = (x-\mu)^T \left[\left(\frac{N}{N+1} \right) \Sigma + \left(\frac{1}{N+1} \right) (x-\mu)(x-\mu)^T \right]^{-1} (x-\mu), \qquad (14)$$

where N is the number of pixels in the window and μ and $\sum$ are the estimated mean and covariance matrix of the data within the window. Pixels are considered anomalous if their RX score is greater than a chi-squared distribution with corresponding significance level, α, and $N-1$ degrees of freedom.

3.3.2 Iterative RX (IRX) Detector

The Iterative RX (IRX) detector was introduced by Taitano et al. (2010) as an extension to the RX detector in an attempt to mitigate the effects that anomalies have on mean vector and covariance matrix calculations. The algorithm runs RX in an iterative fashion, each time removing pixels flagged in the previous iteration as anomalous from the mean vector and covariance statistics used to calculated the RX scores. This process continues until the set of anomalies in the previous iteration matches the set from the current iteration or the maximum number of iterations has been completed (Taitano et al., 2010).

3.3.3 Linear RX (LRX) and Iterative Linear RX (ILRX) Detectors

The Linear RX (LRX) and Iterative Linear RX (ILRX) detectors (Williams et al., 2012) function in the same manner as the RX and IRX except a vertical line of data is used as opposed to a window. If the number of pixels selected for the line size is larger than the image, then pixels are taken from the bottom of the previous column and the top of the subsequent column. These methods are advantageous over the previously described methods because they increase the average distance between the test pixel and the pixels used to estimate the background statistics, thereby decreasing the effects of correlation due to spatial proximity (Williams et al., 2012).

3.3.4 Autonomous Global Anomaly Detector (AutoGAD)

The Autonomous Global Anomaly Detector (AutoGAD) (Johnson et al., 2012) is an Independent Component Analysis (ICA) (Hyvärinen et al., 2001; Stone, 2004) based detector that is processed in four phases. Phase I reduces the dimensionality of the data through PCA (Landgrebe, 2003), using the geometry of the eigenvalue curve to determine the number of PCs to retain. Phase II conducts ICA on the retained PCs from Phase I via the FastICA algorithm (Hyvärinen, 1999). Phase III determines the Independent Components (ICs) that potentially contain anomalies using two filters: the potential anomaly signal to noise ratio and the maximum pixel score. Phase IV smooths the background noise in the ICs selected in Phase III using an adaptive Wiener filter (Lim, 1990) in an iterative fashion then locates the potential anomalies using the Chiang et al. (2001) zero bin method.

3.3.5 Support Vector Data Description (SVDD)

Support Vector Data Description (SVDD) was originally applied to HSI data by Banerjee et al. (2006; 2007). SVDD is a semi-supervised algorithm that requires a training set of background data. Since HSI images are usually assumed to contain few anomalies, the training set is generated by randomly selecting pixels from the image. The minimum volume hypersphere about the training set, $S(R,a) = \{x : \|x - a\| \le R\}$, is then calculated with center *a* and radius *R*. The hypersphere is determined through constrained Lagrangian optimization that simplifies to

$$SVDD(y) = R^2 - K(y,y) + 2\sum_i \alpha_i K(y,x_i), \tag{15}$$

where $K(x,y)$ is the kernel mapping defined by

$$K(x,y) = \exp\left(-\|x-y\|^2 / \sigma^2\right), \tag{16}$$

and *y* is the pixel of interest, and α_i are the weights or support vectors, and σ^2 is a radial basis function parameter used to scale the size of the hypersphere. Finally, pixels that have a SVDD score larger than a user defined threshold are considered anomalies (Banerjee et al., 2006; 2007).

3.3.6 Blocked Adaptive Computationally Efficient Outlier Nominators (BACON)

Blocked Adaptive Computationally Efficient Outlier Nominators (BACON) is a statistical outlier detector created by Billor et al. (2000). It attempts to locate outliers in a data set through the use of iterative estimates of the model with a robust starting point. The algorithm is computationally efficient, regularly requiring less than five iterations to converge, so it is applicable to HSI data. The basic idea is to start with a small subset of outlier free data and iteratively add blocks of data to the data set until all data points not

considered outliers are in the data set. The final data set is then assumed to be outlier free and thus can be used to generate robust mean vector and covariance matrix estimates (Billor et al., 2000).

The BACON algorithm begins by selecting an initial basic subset of data with $m = cp$ data points with the smallest Mahalanobis distance, where in the case of HSI p is equal to the number of bands within image and $c = 4$ or 5, as suggested by Billor et al. (2000), as long as $m \geq n$ where n is the number of pixels in the image. Next, the Mahalanobis distances are calculated for each of the pixels; remembering μ and Σ are now the mean vector and covariance matrix of the basic subset. A new basic subset of all of the pixels with distances less than $c_{npr}\chi^2_{p,\alpha/2}$ is selected, where $\chi^2_{p,\alpha/n}$ is the $1 - (\alpha/n)$ significance level of a chi-squared distribution with p degrees of freedom and $c_{npr} = c_{np} + c_{hr}$ is the correction factor where

$$c_{np} = 1 + \frac{p+1}{n-p} + \frac{2}{n-1-3p}, \tag{17}$$

$$c_{hr} = \max\left\{0, \frac{(h-r)}{(h+r)}\right\}, \tag{18}$$

$$h = \frac{(n+p+1)}{2}. \tag{19}$$

Here r is the size of the current basic subset, n is the number of observations, or pixels, and p is the dimensionality of the data, or bands. If the size of the new basic subset is the same size of the basic subset from the previous iteration, the algorithm is terminated. Otherwise, a new basic subset is calculated (Billor et al., 2000).

3.4 Clustering

Cluster analysis is a multivariate analysis technique for grouping, or clustering, a dataset into smaller subsets known as clusters. The goal of cluster analysis is to maximize the between-cluster variation while minimizing intra-cluster variation (Dillon and Goldstein, 1984). K-means is a clustering technique where the data is split into *k* user-defined number of clusters. The K-means algorithm is initialized by randomly selecting *k* starting points, or cluster centers. Next, a random data point is selected and added to the nearest cluster. The corresponding cluster center is then updated with the new data, allowing for currently clustered data to move into other clusters. This process is repeated until all of the data points are in one of the *k* clusters and no data points are moved in an iteration of the algorithm (Dillon and Goldstein, 1984).

The main difficulty when using a clustering algorithm such as K-means is selecting the *k*, the number of clusters. X-means (Pelleg and Moore, 2000) is a clustering algorithm based upon K-means that has the ability to select the number of clusters. This is accomplished by running K-means multiple times, splitting each of the original clusters in two, and scoring each possible subset of full and partial clusters using the Bayesian Information Criterion (BIC) to determine the optimal clustering (Pelleg and Moore, 2000).

Rather than supplying the X-means algorithm the specific number of clusters as in K-means, the user defines a range of possible clusters, k-lower and k-upper. The algorithm begins by running K-means on the data set with *k* equal to k-lower. Next, each of the original *k* clusters are split in two using K-means with *k* equal to two. Then all 2^k possible combinations of whole clusters and split clusters are analyzed for the

corresponding BIC scores. Then *k* is incremented and the process is repeated until k-upper has been analyzed. The algorithm then returns the *k* cluster centers with the highest corresponding BIC score and K-means is run one last time using the returned cluster centers as the initial starting points (Pelleg and Moore, 2000).

3.5 Methodology

Nine HYDICE (Rickard et al., 1993) hyperspectral images were employed in this research, six forest images and three desert images, as shown in Figure 9 with image details displayed in Table 6. The first step was to analyze an image using one of the seven anomaly detectors described in Section 3: RX, IRX, LRX, ILRX, AutoGAD, SVDD, or BACON. Each algorithm has user defined settings which influence the algorithms' performance. In these experiments, the RX detectors and SVDD used the best settings as reflected in Williams et al. (2012), the settings for AutoGAD were taken from Johnson et al. (2012), and the settings for BACON were taken from Billor et al. (2000). The anomalies detected by the anomaly detector were used twice: first, the anomalies were removed from the image to calculate a robust mean vector and covariance matrix, and second, the anomalous pixels served as the test pixels to be classified using one of the four AMF variants described in Section 2.B: standard AMF, Robust AMF, Clustered AMF, and Largest Cluster AMF.

Table 6: Hyperspectral Image Data

Image	Size	Total Pixels	Anomalous Pixels	Anomalies	Unique Targets
1F	191 × 160	30,560	994	10	5
2F	312 × 152	47,424	281	30	9
3F	226 × 136	30,736	96	20	11
4F	205 × 80	16,400	75	29	12
5F	470 × 156	73,320	440	15	20
6F	355 × 150	53,250	976	45	10
1D	215 × 104	22,360	490	46	22
2D	156 × 156	24,336	417	4	3
3D	460 × 78	35,880	405	12	12

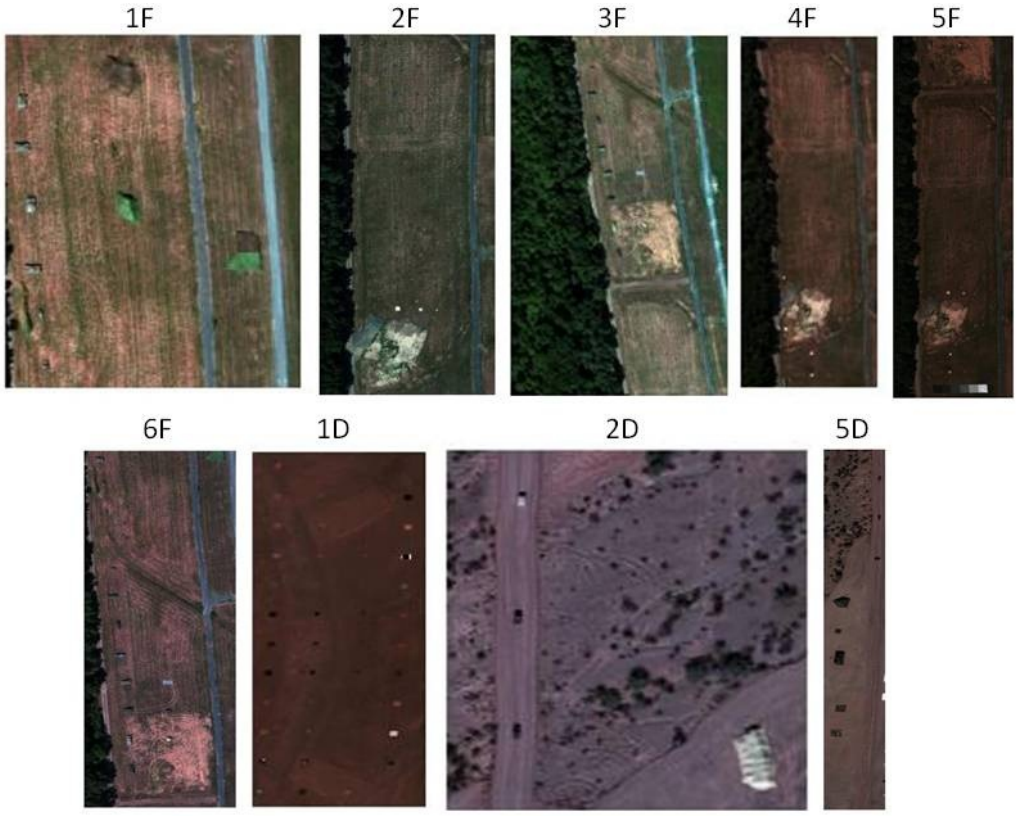

Figure 9: Hyperspectral Images

The following steps were performed to classify anomalies. First, the spectral library, consisting of 30 objects for forest images and 34 objects for desert images, was transformed from radiance space to reflectance space using NDVI or BI atmospheric compensation depending on the image scene, as depicted in the bottom Figure 10. Next,

one of the seven anomaly detectors then analyzed the image for anomalous pixels, as shown in the top Figure 10. Pixels below the anomaly detector threshold (T_1) are classified as background. Pixels above T_1 are classified as anomalies and are used as the pixels to be classified by the AMFs. The mean vector (μ) and covariance matrix ($\sum$) are estimated using the appropriate set of data for the AMF variant. The standard AMF uses μ and $\sum$ from the entire image, the robust AMF uses μ and $\sum$ from the image without the detected anomalies, and the Clustered AMF and Largest Cluster AMF use μ and $\sum$ from the appropriate cluster of data, as determined by the X-means algorithm. Finally, the data is processed through the classifier and pixels below the classifier threshold (T_2) are classified as background and pixels above T_2 are classified as appropriate target types.

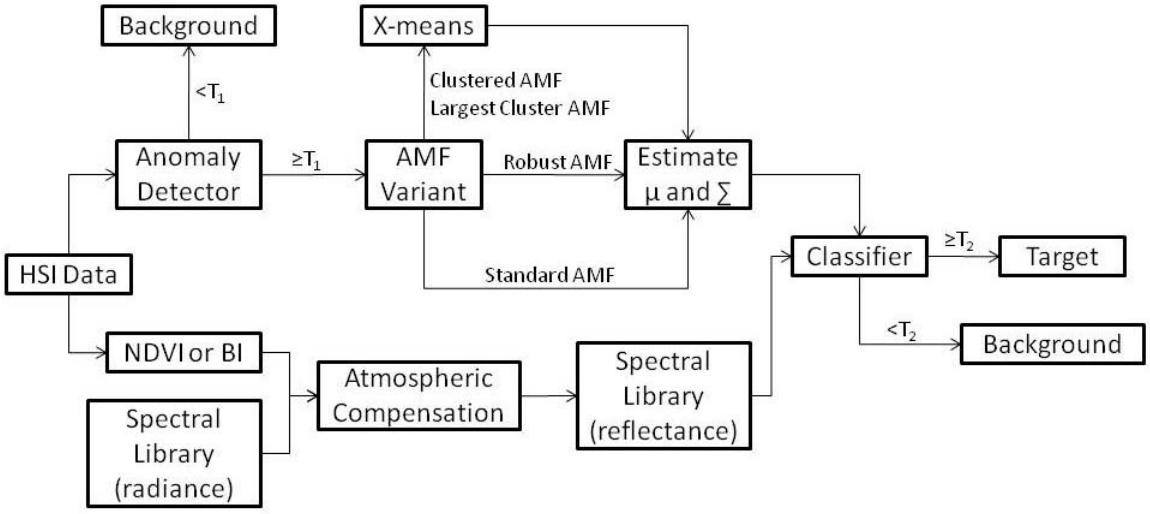

Figure 10: Classification Experimental Process Graph

In order to generate Receiver Operating Characteristic (ROC) like curves (Fawcett, 2006) for each AMF/anomaly detector pair across all nine test images, the following methodology was employed. Each anomaly detector was run across a range of

anomaly detector thresholds (T_1^i), i = 1, 2, …, 19 where $T_1^1 = 0.01$, $T_1^{i+1} = T_1^i + \Delta_1$, and $\Delta_1 = 0.005$. The pixels flagged as anomalous are then processed through the four AMFs. Each target in the spectral library is compared to a pixel of interest, and the target type with the largest resulting AMF score is declared given its score is above the AMF threshold (T_2^j), j = 1, 2, …, 100 where $T_2^1 = 0.01$, $T_2^{j+1} = T_2^j + \Delta_2$, and $\Delta_2 = 0.01$. The thresholding implemented in this research is created by finding the range of all of the AMF scores corresponding to the target type with the largest resulting AMF score and normalizing them between zero and one to allow for consistency amongst the different AMFs within an image. When a threshold is selected, any pixel with an AMF score less than the threshold is classified as background. Analyzing each of the different detector thresholds, across all four of the AMFs, while allowing T_2^j to vary across the entire range of possible values enumerates all combinations of T_1^i and T_2^j.

As the images are processed, a 2 × 3 confusion matrix, as displayed in Figure 11, was generated for each T_1^i, T_2^j, image combination, denoted $C_{T_1^i T_2^j}(k)$, where *i* and *j* represent specific threshold values and *k* is the image of interest. The confusion matrix is comprised of two sections to allow for the scoring of every pixel in the image. The anomaly detector section reflects the declaration of a test pixel as background or the passing of the pixel to be classified. Relative to the detector declaration of background, there are False Negatives (FN) and True Negatives (TN). The anomaly classifier section then reflects the ultimate classification of the pixels classified as anomalous by the anomaly detector.

	Anomaly Detector	Anomaly Classifier	
	"Background"	"Target"	"Background"
Target	False Negative (Detector)	True Positive (Classifier)	False Negative (Classifier)
Background	True Negative (Detector)	False Positive (Classifier)	True Negative (Classifier)

Figure 11: 2 × 3 Confusion Matrix

After each of the nine images are processed, new 2 × 3 confusion matrices are generated which consist of the sum across all nine images for each $C^{k}_{T_1^i T_2^j}$ combination defined by

$$C_{T_1^i T_2^j} = \sum_{k=1}^{I} C_{T_1^i T_2^j}(k), \tag{20}$$

where I is the number of images analyzed, here $I = 9$. Below, the true positive fraction (classifier) versus false positive fraction (classifier) data point for each $C_{T_1^i T_2^j}$ was plotted for a given anomaly detector, anomaly classifier combination and the frontier of the resulting data is interpreted as a rough ROC curve, as is shown, notionally, in Figure 12 where the circles are data points and the line represents the ROC curve.

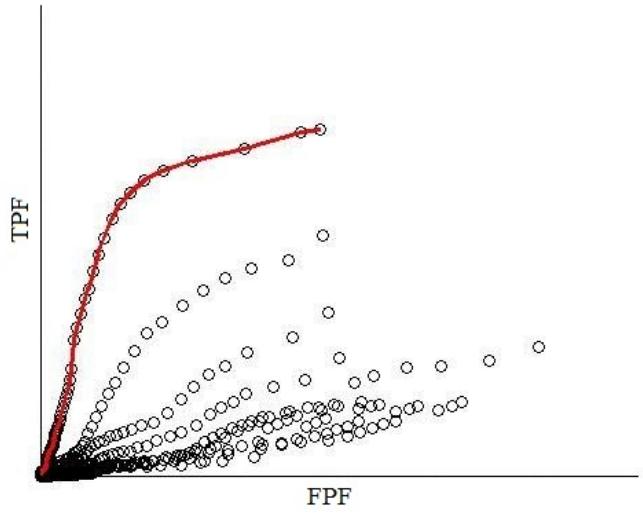

Figure 12: ROC Curve Generated from the Frontier of the Data

3.6 Results

The four AMFs and the seven anomaly detectors were scored to produce ROC curves as described in Section 5, with the results shown in Figure 13 and Figure 14. Each figure contains separate graphs of classification performance generated from each anomaly detector. Figure 13 shows the ROC curves for the full process (detection plus classification). This means taking into account the full 2 × 3 confusion matrix, implying

$$\mathrm{TPF} = \frac{\mathrm{TP_C}}{\mathrm{TP_C} + \mathrm{FN_C} + \mathrm{FN_D}}, \tag{21}$$

$$\mathrm{FPF} = \frac{\mathrm{FP_C}}{\mathrm{FP_C} + \mathrm{TN_C} + \mathrm{TN_D}}, \tag{22}$$

where the subscripts C and D denote classifier and detector, respectively. Note the low TPF and extremely small FPF come from the fact that nine images with a total of 334,226 pixels were analyzed in the process and these statistics are biased downward by a large number of $\mathrm{FN_D}$ and $\mathrm{TN_D}$. The key insight to be gained from these ROC curves is that in

all cases the variants outperform the standard AMF. Furthermore, the clustering methods enhance the Robust AMF.

To get a better visualization of the data, a set of conditional ROC curves were created. Figure 14 displays the ROC curves which account for the performance of the AMF given detections, implying

$$\mathrm{TPF} = \frac{\mathrm{TP_C}}{\mathrm{TP_C} + \mathrm{FN_C}}, \tag{23}$$

$$\mathrm{FPF} = \frac{\mathrm{FP_C}}{\mathrm{FP_C} + \mathrm{TN_C}}. \tag{24}$$

Again, we see the variants outperform the standard AMF and in most cases the clustering methods enhance the Robust AMF.

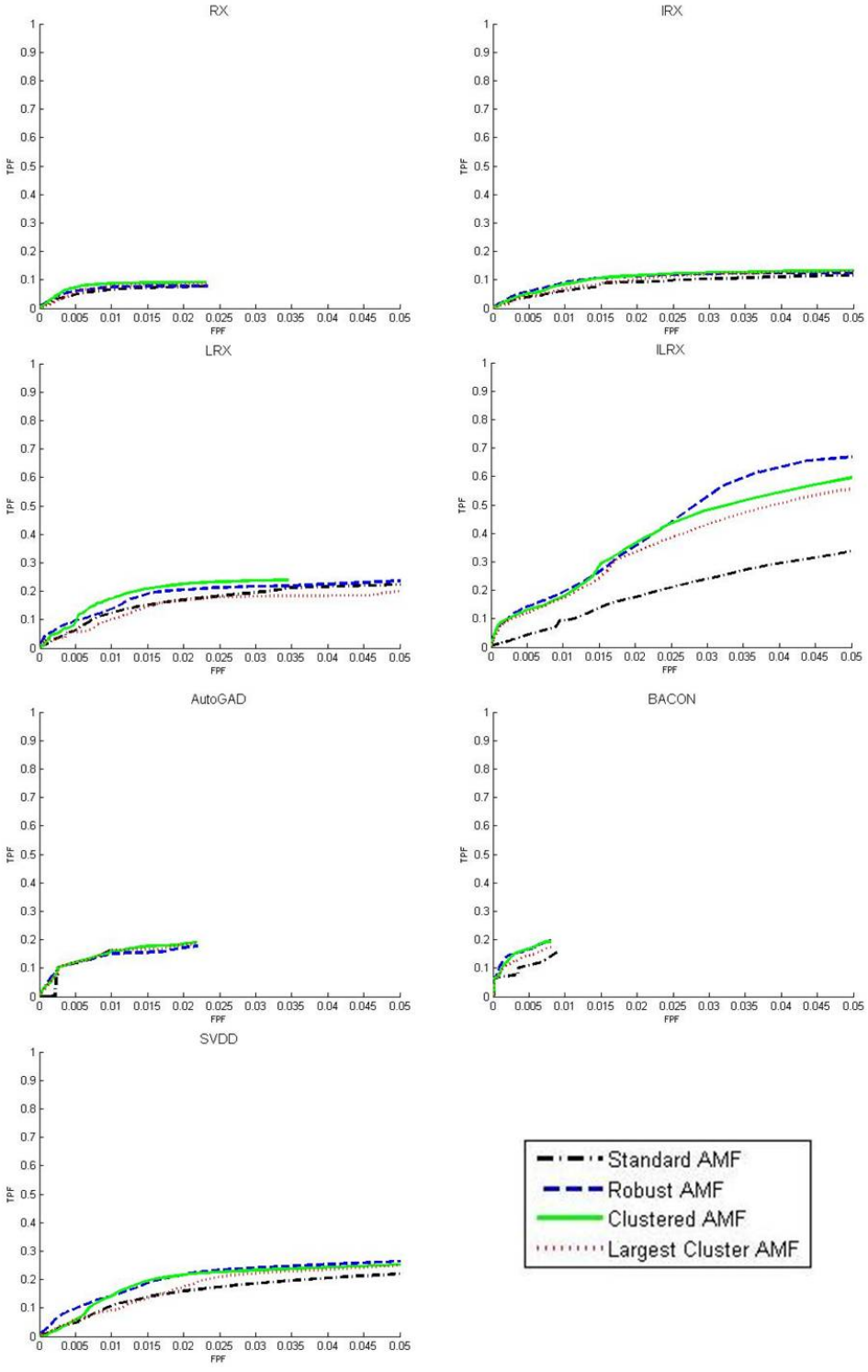

Figure 13: ROC Curves for Full Process

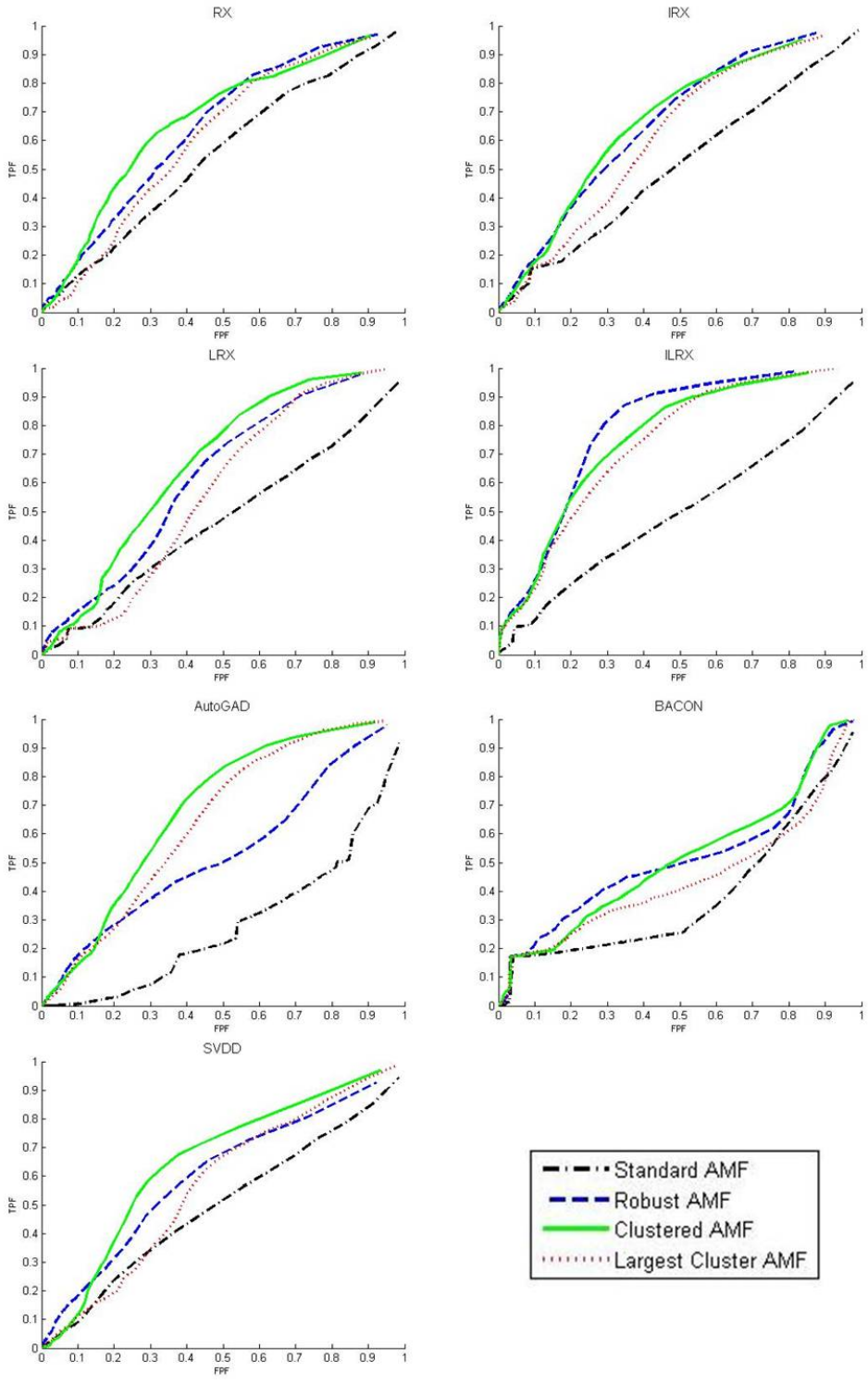

Figure 14: ROC Curves for AMF Results

3.7 Conclusions

This research demonstrates improvements to Adaptive Matched Filter (AMF) performance by addressing correlation and non-homogeneity problems inherent to HSI data, which are often ignored when classifying anomalies. The standard AMF and three variants were employed along with seven different anomaly detectors utilized prior to classification. Manolakis et al. (2009) state the estimation of the mean vector and covariance matrix should be calculated using "target-free" data, generating a Robust AMF. Through the use of prior anomaly detection the Robust AMF showed improved performance over the standard AMF. Additionally, classification was further enhanced by simultaneously addressing the non-homogeneity of HSI data by selecting the required statistics from the appropriate cluster of "target-free" data.

4 Further Extensions to Robust Parameter Design: Three Factor Interactions with an Application to Hyperspectral Imagery

4.1 Introduction

Hyperspectral Imagery (HSI) is a method of collecting vast amounts of information from the electromagnetic spectrum. In principle, a HSI sensor is similar to a standard digital camera. However, instead of the three wavelengths, or bands, collected by a digital camera, a HSI sensor collects upwards of 250 different bands. These bands typically span the visible spectrum up through parts of the near-infrared spectrum (Shaw and Manolakis, 2002).

When an object is covered with camouflage netting it can be difficult to distinguish in a photograph. However, since distinct objects reflect electromagnetic energy differently, the same hidden object could be visible within specific bands beyond those included in a color image. This potential provides opportunities for locating unusual objects within HSI (Landgrebe, 2003). This is typically accomplished by one of two methods: local or global anomaly detection. A local anomaly detector uses a window that moves through the image. The pixel at the center of the window is tested against the other pixels in the window to determine if it is anomalous (Stein et al., 2002). The classic example of a local anomaly detector is the RX detector (Reed and Yu, 1990); others include the Adaptive Causal Anomaly Detection (Hsueh and Chang, 2004), the Multi-Window Anomaly detector (Liu and Chang, 2008), the Iterative RX detector (Taitano et al., 2010) and Iterative Linear RX detector (Williams et al., 2012; 2010). A global anomaly detector uses all of the data in the image to determine a model of the background, assuming sparse anomalies. The individual pixels in the image are then

tested against this model to determine if they are anomalous (Stein et al., 2002). Examples of a global anomaly detector include Joint Subspace Detection (JSD) (Schaum, 2004), Support Vector Data Description (SVDD) (Banerjee et al., 2006), and the Autonomous Global Anomaly Detector (AutoGAD) (Johnson et al., 2012).

An anomaly detector such as AutoGAD requires the user to provide various parameters and thresholds in order to analyze an image. These user-defined settings can be considered controllable factors, or control variables, and potentially have a large effect on algorithm performance. Mindrup et al. (2010) proposed that certain attributes of hyperspectral images could be thought of as uncontrollable factors, or noise variables. The presence of controllable and uncontrollable factors in HSI anomaly detection algorithms suggests the use of Robust Parameter Design (RPD) to locate the best settings for the algorithm. Mindrup et al. (2012) showed that the standard RPD model defined by Myers et al. (2009) might not be sufficient for use with more complex data, such as HSI, and extended the model to include noise by noise (N × N) interactions. Subsequent research indicates that further extensions to the RPD framework might prove efficacious relative to the HSI application. This research extends the work by Mindrup et al. (2012) to include control by noise by noise (C × N × N) and noise by control by control (N × C × C) interactions. Similarly, these new models are applied to the AutoGAD algorithm to locate improved settings.

This chapter is organized in the following fashion. In Section 2, the statistical framework of RPD is reviewed, followed by a summary of the recent N × N embellishment of Mindrup et al. (2012). The section is closed with the addition of C × N × N and N × C × C interactive terms to the model. Section 3 describes AutoGAD,

the algorithm used to demonstrate a real world implementation of the RPD extensions. Section 4 presents the experiment conducted with AutoGAD. Finally, Section 5 concludes the chapter.

4.2 Robust Parameter Design

RPD was formally introduced to the United States in the 1980s by Genichi Taguchi (Taguchi, 1986; 1987) a Japanese quality consultant. RPD is a method for selecting the best levels of controllable factors, or control variables, within a system with a focus on the system variability. This method assumes that the majority of system variance comes from uncontrollable factors, or noise variables, and that these uncontrollable factors may be controlled in the experimental designs used for the RPD. The underlying RPD model is a function of the control variables, *x*, and the noise variables, *z*. The goal of RPD is to select the levels of control variables such that the system is robust to the variance from the noise variables (Robinson et al., 2004).

Lin and Tu (1995) suggest a Mean Squared Error (MSE) approach to determine the optimal control settings by minimizing a function of process mean, $\hat{\mu}$, and variance, $\hat{\sigma}^2$. The function to be minimized, within the experimental design space, varies according to whether a minimum mean, maximum mean, or target mean (*T*) is desired, as shown by the following functions:

$$\mathrm{LT}_{(\text{min mean})} = \hat{\mu}^2 + \hat{\sigma}^2, \tag{25}$$

$$\mathrm{LT}_{(\text{max mean})} = -\hat{\mu}^2 + \hat{\sigma}^2, \tag{26}$$

$$\mathrm{LT}_{(\text{target mean})} = \left(\hat{\mu} - T\right)^2 + \hat{\sigma}^2. \tag{27}$$

The general form of the standard RPD model (Std) assumes that there are no noise by noise and quadratic noise terms within the model resulting in

$$y_{(\text{Std})} = \beta_0 + x'\beta + x'\text{B}x + z'\gamma + x'\Delta z + \varepsilon, \tag{28}$$

which can be rewritten as

$$y_{(\text{Std})} = \beta_0 + x'\beta + x'\text{B}x + \left(\gamma' + x'\Delta\right)z + \varepsilon, \tag{29}$$

where x is an $m \times 1$ vector of control variables, z is an $n \times 1$ random vector of noise variables, β_0 is the intercept, β is the $m \times 1$ vector of control variable coefficients, B is the $m \times m$ matrix of quadratic control variable coefficients, γ is the $n \times 1$ vector of noise variable coefficients, Δ is the $m \times n$ matrix of control by noise variable interaction coefficients, and the random error associated with the model is $\varepsilon \sim \text{N}\left(0,\ \sigma^2\right)$. The random noise variables, z, are assumed such that $E\left(z\right) = 0$ and $\text{var}\left(z\right) = \Sigma_z$. The expected value and variance models, see the appendix, are then

$$E\left[\, y_{(\text{Std})} \,\right] = \beta_0 + x'\beta + x'\text{B}x \tag{30}$$

and

$$\text{var}\left[\, y_{(\text{Std})} \,\right] = \left(\gamma' + x'\Delta\right)\Sigma_z\left(\gamma' + x'\Delta\right)' + \sigma^2, \tag{31}$$

where σ^2 is estimated by the MSE from the fitted model (Myers et al., 2009). Substituting equation (30) for $\hat{\mu}$ and equation (31) for $\hat{\sigma}^2$ in any of the LT functions described in equations (25) – (27) reveals that the LT function is only a function of the control variables, x. This implies that the optimal control variables can be determined through constrained optimization of the LT function without explicit noise variable consideration (Köksoy, 2006).

Mindrup et al. (2012) derived the mean and variance model necessary to implement the noise by noise RPD model herein denoted as N × N.

$$y_{(\mathrm{N}\times\mathrm{N})} = \beta_0 + x'\beta + x'\mathrm{B}x + \left(\gamma' + x'\Delta\right)z + z'\Phi z + \varepsilon\,, \tag{32}$$

$$E\left[y_{(\mathrm{N}\times\mathrm{N})}\right] = \beta_0 + x'\beta + x'\mathrm{B}x + tr\left(\Phi\Sigma_z\right), \tag{33}$$

$$\mathrm{var}\left[y_{(\mathrm{N}\times\mathrm{N})}\right] = \left(\gamma' + x'\Delta\right)\Sigma_z\left(\gamma' + x'\Delta\right)' + 2tr\left(\Phi\Sigma_z\right)^2 + \sigma^2\,, \tag{34}$$

where Φ is the $n \times n$ matrix of quadratic noise variable coefficients and *tr* represents the trace of a matrix. The derivations of the expectation and variance models are given in the appendix.

Here the Mindrup et al. (2012) model is extended two steps forward. The first model includes control by noise by noise (C × N × N) interactions. Its expansion is given in equation (35) with mean and variance expressions given in equations (37) and (38). The second model adds noise by control by control (N × C × C) interactions to the previous model. Its expression is given in equation (39) with mean and variance expressions given in equations (41) and (42). The derivations of the expectation and variance expressions for both models are provided in the appendix.

Here $\Psi_x = \sum_{i=1}^{m}\Psi_i x_i$ where Ψ_i is an n × n matrix of control by noise by noise terms corresponding to x_i, and $\tilde{x} = \left[x'\Omega_1 x,\ x'\Omega_2 x,\ \ldots\ , x'\Omega_n x\right]$, where Ω_j is an m × m matrix of noise by control by control terms corresponding to z_j. The C × N × N model is

$$y_{(\mathrm{C}\times\mathrm{N}\times\mathrm{N})} = \beta_0 + x'\beta + x'\mathrm{B}x + \left(\gamma' + x'\Delta\right)z + z'\Phi z + z'\Psi_x z + \varepsilon\,, \tag{35}$$

which can be rewritten as

$$y_{(\mathrm{C}\times\mathrm{N}\times\mathrm{N})} = \beta_0 + x'\beta + x'\mathrm{B}x + \left(\gamma' + x'\Delta\right)z + z'\left(\Phi + \Psi_x\right)z + \varepsilon\,, \tag{36}$$

$$E\left[y_{(\mathrm{C}\times\mathrm{N}\times\mathrm{N})} \right] = \beta_0 + x'\beta + x'\mathrm{B}x + tr\left(\left(\Phi + \Psi_x \right) \Sigma_z \right), \tag{37}$$

$$\mathrm{var}\left[y_{(\mathrm{C}\times\mathrm{N}\times\mathrm{N})} \right] = (\gamma' + x'\Delta)\Sigma_z(\gamma' + x'\Delta)' + 2tr\left(\left(\Phi + \Psi_x \right) \Sigma_z \right)^2 + \sigma^2 . \tag{38}$$

The N × C × C model, which is an extension to the C × N × N model, is

$$y_{(\mathrm{N}\times\mathrm{C}\times\mathrm{C})} = \beta_0 + x'\beta + x'\mathrm{B}x + \left(\gamma' + x'\Delta \right) z + z'\left(\Phi + \Psi_x \right) z + \tilde{x}z + \varepsilon , \tag{39}$$

which can be rewritten as

$$y_{(\mathrm{N}\times\mathrm{C}\times\mathrm{C})} = \beta_0 + x'\beta + x'\mathrm{B}x + \left(\gamma' + x'\Delta + \tilde{x} \right) z + z'\left(\Phi + \Psi_x \right) z + \varepsilon , \tag{40}$$

$$E\left[y_{(\mathrm{N}\times\mathrm{C}\times\mathrm{C})} \right] = \beta_0 + x'\beta + x'\mathrm{B}x + tr\left(\left(\Phi + \Psi_x \right) \Sigma_z \right), \tag{41}$$

$$\mathrm{var}\left[y_{(\mathrm{N}\times\mathrm{C}\times\mathrm{C})} \right] = \left(\gamma' + x'\Delta + \tilde{x} \right) \Sigma_z \left(\gamma' + x'\Delta + \tilde{x} \right)' + 2tr\left(\left(\Phi + \Psi_x \right) \Sigma_z \right)^2 + \sigma^2 . \tag{42}$$

4.3 Autonomous Global Anomaly Detector

AutoGAD (Johnson et al., 2012) is a fully autonomous HSI anomaly detection algorithm that is comprised of four phases where the only required information from the user is the data in the proper form and the required input parameters, described in Section 4.4.1. This suggests that AutoGAD would be an excellent candidate for an RPD experiment using the input parameters as control variables and image properties processed as noise variables. The objective of the experiment being the determination of robust settings for the input parameters. The experiment described in Section 4.4 parallels the work of Mindrup et al. (2012).

4.3.1 Image Preprocessing

The data from a hyperspectral image is stored in a three dimensional data matrix referred to as a data cube where the vertical and horizontal dimensions refer to the height

and width of the image and the third dimension refers to the number of spectral bands collected. The data cube can be thought of as a collection different images of the scene, one at each collected spectral resolution (Landgrebe, 2003). For AutoGAD to detect anomalies within an image it must be converted into the proper form. First, the data needs to be reshaped from a three-dimensional data cube to a two-dimensional data matrix, consisting of the total number of pixel by the total number of spectral bands. The next step is to remove the atmospheric absorption bands, those bands where the energy is almost completely absorbed by the atmosphere (Eismann, 2012). This reduces the number of bands in the images used in this research from 210 to 145.

4.3.2 Feature Extraction I (Phase I)

After the image has been preprocessed into the proper form the dimensionality of the data is further reduced using Principal Components Analysis (PCA). PCA is a linear transformation that maps the original data to a new space where all of the columns of the data are uncorrelated. The resulting Principal Components (PCs) are sorted such that the first PC contains the most variance from the original data and the last PC contains the least (Anderson, 2003; Landgrebe, 2003). Therefore, the original data can be reduced by selecting only the most significant PCs.

One of the more popular methods to select the significant PCs involves using the percentage of total variance explained (Dillon and Goldstein, 1984). Farrell and Mersereau (2005) showed that this method was extremely erratic in HSI applications. Johnson et al. (2012) proposed a method they call the Maximum Distance Secant Line (MDSL). This process finds the "knee" in the eigenvalue curve to determine how many PCs to select. It has been shown to work well in practice and is employed here.

Once the significant PCs are selected using the MDSL method, the resulting reduced data set is whitened. The new whitened reduced data set is then processed in the next phase of the algorithm.

4.3.3 Feature Extraction II (Phase II)

Independent Component Analysis (ICA) (Hyvärinen et al., 2001; Stone, 2004), via the FastICA algorithm (Hyvärinen, 1999), is performed on the data with the aim of producing statistically independent data. Each of the Independent Components (ICs) are then reshaped back into an matrix the size of the original image referred to as an abundance map (Johnson et al., 2012).

4.3.4 Feature Selection (Phase III)

In this phase of the algorithm the abundance maps are analyzed to determine those which potentially contain anomalies. This is accomplished through the use of two filters: the Potential Anomaly Signal to Noise Ratio (PA SNR) and the maximum pixel score. The PA SNR is defined as:

$$\text{PA SNR} = 10 \cdot \log_{10}\left(\frac{\text{variance}(\text{potential anomalies})}{\text{variance}(\text{background})}\right). \tag{43}$$

The difficulty in applying this statistic lies in determining the threshold between the background and potential anomalies. Johnson et al. (2012) employ a method described by Chiang et al. (2001) that suggests potentially anomalous pixels occur after the first zero bin in the histogram of the data.

Each pixel score is standardized within an abundance map. This situation allows for the selection of a threshold to flag the potential for anomalies. The maximum pixel

score in each abundance map is compared to the threshold. If the threshold is exceeded then the abundance map associated with that scene is a map that potentially flags anomalies (Johnson et al., 2012).

The abundance maps with corresponding values of PA SNR and maximum pixel score found to be above both user-defined thresholds are analyzed in phase four for the presence of anomalies (Johnson et al., 2012).

4.3.5 Identification (Phase IV)

The first step in the identification phase is to smooth out background noise of the data using an adaptive Wiener filter (Lim, 1990). Johnson et al. (2012) implement this process iteratively to create a smoother background/noise separation and call this method an Iterative Adaptive Noise (IAN) filter. Subsequently, to determine which pixels in the selected abundance maps are anomalous, the Chiang et al. (2001) method is implemented a second time and pixels after the first zero bin are flagged as anomalies.

4.4 RPD Experiments with AutoGAD

In this experiment the input parameters in AutoGAD are considered to be control variables and following the work of Mindrup et al. (2012), certain properties of the images are treated as noise variables. The main difference between the experimental design used in this research and the work of Mindrup et al. (2012) concerns our use of 15 images for training and 5 for test as opposed to 10 for training and 10 for test in the previous work.

4.4.1 AutoGAD Input Parameters

AutoGAD has nine user-defined input parameters:

- Dimension Adjust: number of PCs above or below the knee in the eignvalue curve to retain.
- Bin Width SNR: size of the histogram bins used to determine a separation between the background and target classes for PA SNR.
- PA SNR Threshold: SNR between the variance of the potential anomalies and the variance of the background used to nominate an abundance maps as potentially containing anomalies.
- Max Score Threshold: maximum pixel score used to nominate abundance maps as potentially containing anomalies.
- Low PA SNR Threshold: determination of whether a high or low number of iterations are used in IAN filtering.
- IAN Filtering Iterations (High SNR): number of iterations for IAN filtering when the PA SNR is determined to be high.
- IAN Filtering Iterations (Low SNR): number of iterations for IAN filtering when the PA SNR is determined to be low.
- Window Size: size of the window used in IAN filtering.
- Bin Width Identify: size of the histogram bins used to determine a separation between the background and target classes for identifying pixels as anomalies.

Johnson et al. (2012) suggests that the values displayed in Table 7 be used for favorable performance of the algorithm. They observe that if the settings for the IAN filtering are altered away from the suggested settings, specifically window size, the performance of the algorithm is negatively affected. Additionally, the original settings and the improved settings, from the Johnson et al. (2012) paper, both suggest a value of -1 be used for the dimension adjust parameter. These observations led to the decision to fix the three IAN filter parameters and the dimension adjust parameter for the purpose of

the RPD experiment. The remaining five input parameters were selected to be control variables, as displayed in Table 7, with the corresponding test range.

Table 7: AutoGAD Suggested Parameters and Test Range

Parameter	Suggested	Test Range
Dimension Adjust	-1	-1
Bin Width SNR (x_1)	0.05	[0.01 - 0.09]
PA SNR Threshold (x_2)	2	[1 - 5]
Max Score Threshold (x_3)	10	[6 - 14]
Low PA SNR (x_4)	10	[6 - 14]
IAN Filtering Iterations (High SNR)	100	100
IAN Filtering Iterations (Low SNR)	20	20
Window Size	3	3
Bin Width Identify (x_5)	0.05	[0.01 - 0.09]

4.4.2 Training and Test Images

The ten images used in this experiment come from the Hyperspectral Digital Imagery Collection Equipment (HYDICE) sensor during Forrest Radiance I and Desert Radiance II collection events. To provide sufficient imagery for training and testing the original images were each split into two images, resulting in a total of 20 images, ten top halves and ten bottom halves. Fifteen image halves were randomly selected for training set and the remaining five image halves were used in the test set.

To employ the 20 images as noise variables three noise characteristics defined by Mindrup et al. (2010) were calculated for each image: Fisher ratio, ratio of anomalous to background pixels, and number of clusters. Fisher ratio, z_1, provides a good class separablility measure (Duda et al., 2001; Mao, 2002). Mindrup et al. (2010) defined the Fisher ratio for a hyperspectral image as the average of the Fisher ratios of the K bands in the given image resulting in

$$z_{1i} = \frac{\sum_{k=1}^{\mathrm{K}} \left(\frac{\left(\mu_{a_{ik}} - \mu_{b_{ik}} \right)^2}{\sigma_{a_{ik}}^2 + \sigma_{b_{ik}}^2} \right)}{\mathrm{K}}, \tag{44}$$

where $\mu_{a_{ik}}$ and $\sigma_{a_{ik}}^2$ are the mean and variance of the anomalous pixels of image *i* and band *k*, and similarly $\mu_{b_{ik}}$ and $\sigma_{b_{ik}}^2$ are the mean and variance of the background pixels of image *i* and band *k*, with respect to the image's truth mask.

The ratio of anomalous to background pixels, z_2, is calculated using a truth mask by

$$z_{2i} = \frac{a_i}{b_i}, \tag{45}$$

where a_i is the number of anomalous pixels in image *i* and b_i is the number of background pixels in image *i*, with respect to the image's truth mask (Mindrup et al., 2010).

The number of clusters in the image, z_3, is the number of homogeneous groups within the image as calculated by the X-means algorithm (Pelleg and Moore, 2000). Table 8 displays the training and test images along with their calculated noise values.

Table 8: Training and Test Images with Calculated Noise Values

	Image	Image Half	Fisher Ratio (z_1)	Percent Target (z_2)	Number of Clusters (z_3)
Training Set	1D	Top	1.7797	0.0043	3
	1D	Bottom	1.6265	0.0028	3
	1F	Bottom	0.3148	0.0225	5
	2D	Top	0.0957	0.0247	4
	2D	Bottom	0.1762	0.0288	3
	2F	Top	0.9633	0.0084	7
	2F	Bottom	0.9311	0.0085	7
	3D	Top	0.1695	0.0034	3
	3F	Top	0.2650	0.0053	8
	3F	Bottom	0.2153	0.0078	5
	4D	Top	1.4093	0.0156	6
	4D	Bottom	2.6382	0.0275	4
	4F	Bottom	0.0779	0.0063	8
	5F	Bottom	0.7412	0.0094	7
	6F	Top	0.2658	0.0109	6
Test Set	1F	Top	0.4335	0.0392	5
	3D	Bottom	1.4299	0.0033	3
	4F	Top	0.0826	0.0046	7
	5F	Top	0.1991	0.0078	10
	6F	Bottom	1.8451	0.0052	4

4.4.3 AutoGAD Responses

With each instance of the AutoGAD algorithm the True Positive Fraction (TPF) and Label Accuracy (LA) are collected. TPF is the ratio of the number of anomalous pixels correctly classified to the total number of anomalous pixels in the image. LA is the ratio of the number of anomalous pixels correctly classified to the total number of pixels classified as anomalous. The response used for the experimental design was

$$y = \text{TPF} + \text{LA}. \tag{46}$$

This response is used for the experiment because it captures the viewpoints of both the designer of the algorithm and a potential user of the algorithm. Since both TPF and LA have a range of [0, 1] the response has the range of [0, 2].

4.4.4 Experimental Design

To conduct the training portion of the experiment a Face Centered Cube (FCC) design (Montgomery, 2008) was selected for the five control variables, to include 16 center runs and five repetitions. The resulting FCC was then crossed with the 15 images randomly selected for the training set for a total of 4,350 runs. The replicates were used to account for the variability of the FastICA algorithm implemented within AutoGAD; all other aspects of AutoGAD are deterministic.

Before performing regression on the data to create the four RPD models, the five control variables, $x_1 - x_5$, were coded [-1, 1] and the three noise variables, $z_1 - z_3$, were standardized individually. The standardization of the noise variables results in the correlation matrix shown below.

$$\Sigma_z = \begin{bmatrix} 1 & -0.0320 & -0.2330 \\ -0.0320 & 1 & -0.2460 \\ -0.2330 & -0.2460 & 1 \end{bmatrix}$$

4.4.5 Results

Table 9 displays pertinent data from the four regression models. Notice that as the model becomes more complex, i.e. more terms are added; the R^2 and R^2 adjusted (R^2_{adj}) increase and the MSE decreases.

Table 9: Regression Data

	Std	N × N	C × N × N	N × C × C
R^2	0.6269	0.7637	0.8031	0.8291
R^2_{adj}	0.6253	0.7624	0.8010	0.8264
MSE	0.1157	0.0734	0.0615	0.0536

The optimal settings were calculated using the Lin and Tu MSE method (Lin and Tu, 1995) with the focus on maximizing the mean and minimizing the variance of TPF + LA. Table 10 shows the optimal RPD settings for AutoGAD for each of the defined models along with the suggested setting from Johnson et al. (2012) and Table 11 shows the mean and variance of the responses from the five test images. Notice that as the model becomes more complex the mean increases. Likewise, the variance decreases as the model becomes more complex. The goal of RPD is to locate settings that provide a favorable response while reducing process variance. This RPD example using AutoGAD shows a real world implementation of a system that would benefit from a more complex model.

Table 10: Optimal Settings for AutoGAD by Model

Parameter	Johnson	Std	N × N	C × N × N	N × C × C
Dimension Adjust	-1	-1	-1	-1	-1
Bin Width SNR	0.05	0.009	0.09	0.01	0.0198
PA SNR Threshold	2	1.0002	3.8849	5	1
Max Score Threshold	10	6	6	6	6
Low PA SNR	10	7.0936	14	14	14
IAN Filtering Iterations (High SNR)	100	100	100	100	100
IAN Filtering Iterations (Low SNR)	20	20	20	20	20
Window Size	3	3	3	3	3
Bin Width Identify	0.05	0.0764	0.0618	0.09	0.09

Table 11: Result from Optimal AutoGAD Settings by Model

	Johnson	Std	N × N	C × N × N	N × C × C
mean(TPF + LA)	0.9459	0.9016	1.1675	1.2161	1.2263
var(TPF + LA)	0.6471	0.3198	0.2433	0.2155	0.1854

4.5 Conclusions

This research provides the required statistical models to extend the Mindrup et al. (2012) N × N RPD model to include higher order terms, including the C × N × N and N × C × C interaction terms. These higher order models were applied to AutoGAD to locate better operating parameter settings, using properties of the hyperspectral images as system noise. The expanded models provided better fits to the data demonstrated through increased R^2_{adj} and decreased MSE. Finally, the parameter settings located using these new models provided higher mean responses and lower response variance, the overall goal of RPD.

5 Conclusion

This research had three goals: address correlation problems inherent to Hyperspectral Imagery (HSI) that are often ignored by the research community when performing anomaly detection, address similar correlation problems with anomaly classification, and extend the Robust Parameter Design (RPD) model to include three factor interactions. Chapters 2 – 4 outlined each of these goals individually and gave results to show the newly implemented methods achieved the stated goals.

5.1 Original Contributions

Chapter 2 introduced two new anomaly detectors: Linear RX (LRX), a variant of Reed and Yu (1990) RX detector, and Iterative Linear RX (ILRX), a variant of the Taitano et al. (2010) Iterative RX (IRX) detector. LRX addresses spatial correlation related to RX by establishing a mean vector and covariance matrix using data that is, on average, further from each other than the standard RX window. The IRX detector allows for the exclusion of outliers in the mean vector and covariance matrix calculations, thereby promoting a more accurate assessment of the target pixel. ILRX then exploits the innovations of both LRX and IRX.

Chapter 3 continued addressing correlation in HSI, but this time with the goal being classification. The Adaptive Matched Filter (AMF) with the Manolakis et al. (2009) suggested improvements, referred to as Robust AMF, and was shown to be superior to the Standard AMF. Additionally, two more AMFs were created, Clustered AMF and Largest Cluster AMF, and were tested against the Standard AMF and Robust AMF. Robust AMF addresses the problem of anomalous pixels skewing the required

statistics. Clustered AMF and Largest Cluster AMF exploit the idea of Robust AMF and address the concern of non-homogeneity.

Chapter 4 provided the required statistical models to extend the RPD model to include higher order terms, including the Control by Noise by Noise (C × N × N) and Noise by Control by Control (N × C × C) interaction terms. Applications of the models demonstrated increased R^2_{adj} and decreased Mean Squared Error (MSE), and new parameter settings which provided higher mean responses and lower response variance.

5.2 Suggested Future Work

In the process of researching correlation and RPD, some potential extensions to this work became evident:

- A comparison of the ILRX algorithm with the wide range of additional RX based methods to address RX shortcomings.
- The new classification algorithms take into account how anomalies and non-homogeneity affect the mean vector and covariance matrix estimates. Addressing spatial and spectral correlation directly, rather than just within the anomaly detectors, could further benefit anomaly classification.
- Determine how much improvement can be gained when employing the extended RPD model to generate new parameters for the standard anomaly detection algorithms, such as: RX, IRX, ILRX, SVDD, etc.

Appendix

This appendix details the derivations of the expectation and variance of the four separate models discussed in this paper: Standard Model (Std), Noise by Noise Model (N × N), Control by Noise by Noise Model (C × N × N), and Noise by Control by Control Model (N × C × C). Within each model x is an $m \times 1$ vector of control variables, z is an $n \times 1$ random vector of noise variables distributed $\mathrm{N}(0, \Sigma_z)$, β_0 is the intercept, β is the $m \times 1$ vector of control variable coefficients, B is the $m \times m$ matrix of quadratic control variable coefficients, γ is the $n \times 1$ vector of noise variable coefficients, Δ is the $m \times n$ matrix of control by noise variable interaction coefficients, Φ is the $n \times n$ matrix of quadratic noise variable coefficients, $\Psi_x = \sum_{i=1}^{m} \Psi_i x_i$ where Ψ_i is an n × n matrix of control by noise by noise coefficients corresponding to x_i, $\tilde{x} = \left[x'\Omega_1 x,\ x'\Omega_2 x,\ \ldots,\ x'\Omega_n x\right]$ where Ω_j is an m × m matrix of noise by control by control coefficients corresponding to z_j, and the random error, ε, associated with the model is assumed to be distributed $\mathrm{N}(0, \sigma^2)$.

A.1 Standard Model (Myers et al., 2009)

$$y_{(\mathrm{Std})} = \beta_0 + x'\beta + x'\mathrm{B}x + (\gamma' + x'\Delta)z + \varepsilon$$

Since the $E(z) = 0$ and the $E(\varepsilon) = 0$

$$E\left(y_{(\mathrm{Std})}\right) = \beta_0 + x'\beta + x'\mathrm{B}x$$

Since β_0 and x are constants

$$\begin{aligned}\operatorname{var}\left(y_{(\mathrm{Std})}\right) &= \operatorname{var}((\gamma'+x'\Delta)z+\varepsilon)\\ &= (\gamma'+x'\Delta)\operatorname{var}(z)(\gamma'+x'\Delta)'+\sigma^2\\ &= (\gamma'+x'\Delta)\Sigma_z(\gamma'+x'\Delta)'+\sigma^2\end{aligned}$$

A.2 Noise by Noise Model (Mindrup et al., 2012)

$$y_{(\mathrm{N}\times\mathrm{N})} = \beta_0 + x'\beta + x'\mathrm{B}x + (\gamma'+x'\Delta)z + z'\Phi z + \varepsilon$$

$$\begin{aligned}E\left(y_{(\mathrm{N}\times\mathrm{N})}\right) &= E\left(\beta_0 + x'\beta + x'\mathrm{B}x + (\gamma'+x'\Delta)z + z'\Phi z + \varepsilon\right)\\ &= \beta_0 + x'\beta + x'\mathrm{B}x + E\left(z'\Phi z\right)\end{aligned}$$

If x is distributed $\mathrm{N}(0,\ V)$, then the $E[x'Ax] = tr(AV)$ where *tr* is the trace of a matrix (Searle, 1971). Therefore,

$$E\left(y_{(\mathrm{N}\times\mathrm{N})}\right) = \beta_0 + x'\beta + x'\mathrm{B}x + tr\left(\Phi\Sigma_z\right)$$

$$\begin{aligned}\operatorname{var}\left(y_{(\mathrm{N}\times\mathrm{N})}\right) &= \operatorname{var}\left(\beta_0 + x'\beta + x'\mathrm{B}x + (\gamma+x'\Delta)z + z'\Phi z + \varepsilon\right)\\ &= (\gamma'+x'\Delta)\Sigma_z(\gamma'+x'\Delta)'+\sigma^2 + \operatorname{var}\left(z'\Phi z\right) + 2\operatorname{cov}\left((\gamma'+x'\Delta)z, z'\Phi z\right)\end{aligned}$$

If x is distributed $\mathrm{N}(0,\ V)$, $\operatorname{var}[x'Ax] = 2tr(AV)^2$ where *tr* is the trace of a matrix (Searle, 1971). Therefore,

$$\operatorname{var}\left(z'\Phi z\right) = 2tr\left(\Phi\Sigma_z\right)^2$$

Theorem 1. Let $\tilde{z} = z'\Gamma z$, where z is an $n \times 1$ random vector such that $E(z) = 0$ and $\operatorname{var}(z) = \Sigma_z$ and Γ is a constant matrix of size $n \times n$. If α is a constant vector of size $n \times 1$, then $\operatorname{cov}(\alpha' z, \tilde{z}) = 0$.

Proof.

$$\begin{aligned}\operatorname{cov}(\alpha' z, \tilde{z}) &= \operatorname{cov}(\alpha' z, z'\Gamma z) \\ &= E(\alpha' z z'\Gamma z) - E(\alpha' z)E(z'\Gamma z) \\ &= E(\alpha' z z'\Gamma z) \\ &= E\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\alpha_k z_k z_j \gamma_{ji} z_i\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\alpha_k \gamma_{ji} E(z_k z_j z_i)\end{aligned}$$

The resulting summation consists of three different types of expectation terms: $E\left[z_a^3\right]$, $E\left[z_a^2 z_b\right]$, and $E\left[z_a z_b z_c\right]$. Anderson (2003) showed that all three types of expectation are zero because with multivariate normal data the third moment is equal to zero. Therefore, $\operatorname{cov}(\alpha' z, \tilde{z}) = 0$ □.

Letting $\alpha' = (\gamma' + x'\Delta)$ the covariance term from $\operatorname{var}\left(y_{(\mathrm{N\,x\,N})}\right)$ is

$\operatorname{cov}\left((\gamma' + x'\Delta)z, z'\Phi z\right) = \operatorname{cov}(\alpha' z, \tilde{z}) = 0$, by Theorem 1.

The resulting variance model is

$$\operatorname{var}\left(y_{(\mathrm{N}\times\mathrm{N})}\right) = (\gamma' + x'\Delta)\Sigma_z(\gamma' + x'\Delta)' + 2tr\left(\Phi\Sigma_z\right)^2 + \sigma^2$$

A.3 Control x Noise x Noise Model

$$y_{(\mathrm{C}\times\mathrm{N}\times\mathrm{N})} = \beta_0 + x'\beta + x'\mathrm{B}x + (\gamma' + x'\Delta)z + z'(\Phi + \Psi_x)z + \varepsilon$$

$$\begin{aligned}E\left(y_{(\mathrm{C}\times\mathrm{N}\times\mathrm{N})}\right) &= E\left(\beta_0 + x'\beta + x'\mathrm{B}x + (\gamma' + x'\Delta)z + z'(\Phi + \Psi_x)z + \varepsilon\right) \\ &= \beta_0 + x'\beta + x'\mathrm{B}x + tr\left((\Phi + \Psi_x)\Sigma_z\right)\end{aligned}$$

Additionally, from the noise by noise model

$$\begin{aligned}\operatorname{var}\left(y_{(\mathrm{C}\times\mathrm{N}\times\mathrm{N})}\right) &= \operatorname{var}\left(\beta_0 + x'\beta + x'\mathrm{B}x + \left(\gamma' + x'\Delta\right)z + z'\left(\Phi + \Psi_x\right)z + \varepsilon\right) \\ &= (\gamma' + x'\Delta)\Sigma_z(\gamma' + x'\Delta)' + 2tr\left(\left(\Phi + \Psi_x\right)\Sigma_z\right)^2 + \sigma^2 + \ldots \\ &\quad 2\operatorname{cov}\left(\left(\gamma' + x'\Delta\right)z, z'\left(\Phi + \Psi_x\right)z\right)\end{aligned}$$

Letting $\alpha' = (\gamma' + x'\Delta)$ and $\Gamma = \left(\Phi + \Psi_x\right)$, then by Theorem 1,

$$\operatorname{cov}\left(\left(\gamma' + x'\Delta\right)z, z'\left(\Phi + \Psi_x\right)z\right) = \operatorname{cov}\left(\alpha' z, z'\Gamma z\right) = \operatorname{cov}\left(\alpha' z, \tilde{z}\right) = 0$$

Therefore, the resulting variance model is,

$$\operatorname{var}\left(y_{(\mathrm{C}\times\mathrm{N}\times\mathrm{N})}\right) = (\gamma' + x'\Delta)\Sigma_z(\gamma' + x'\Delta)' + 2tr\left(\left(\Phi + \Psi_x\right)\Sigma_z\right)^2 + \sigma^2$$

A.4 Noise x Control x Control Model

$$y_{(\mathrm{N}\times\mathrm{C}\times\mathrm{C})} = \beta_0 + x'\beta + x'\mathrm{B}x + \left(\gamma' + x'\Delta + \tilde{x}\right)z + z'\left(\Phi + \Psi_x\right)z + \varepsilon$$

$$\begin{aligned}E\left(y_{(\mathrm{N}\times\mathrm{C}\times\mathrm{C})}\right) &= E\left(\beta_0 + x'\beta + x'\mathrm{B}x + \left(\gamma' + x'\Delta + \tilde{x}\right)z + z'\left(\Phi + \Psi_x\right)z + \varepsilon\right) \\ &= \beta_0 + x'\beta + x'\mathrm{B}x + tr\left(\left(\Phi + \Psi_x\right)\Sigma_z\right)\end{aligned}$$

$$\begin{aligned}\operatorname{var}\left(y_{(\mathrm{N}\times\mathrm{C}\times\mathrm{C})}\right) &= \operatorname{var}\left(\beta_0 + x'\beta + x'\mathrm{B}x + \left(\gamma' + x'\Delta + \tilde{x}\right)z + z'\left(\Phi + \Psi_x\right)z + \varepsilon\right) \\ &= \left(\gamma' + x'\Delta + \tilde{x}\right)\Sigma_z\left(\gamma' + x'\Delta + \tilde{x}\right)' + 2tr\left(\left(\Phi + \Psi_x\right)\Sigma_z\right)^2 + \sigma^2 + \ldots \\ &\quad 2\operatorname{cov}\left(\left(\gamma' + x'\Delta + \tilde{x}\right)z, z'\left(\Phi + \Psi_x\right)z\right)\end{aligned}$$

Letting $\alpha' = \left(\gamma' + x'\Delta + \tilde{x}\right)$ and $\Gamma = \left(\Phi + \Psi_x\right)$, them by Theorem 1,

$$\operatorname{cov}\left(\left(\gamma' + x'\Delta + \tilde{x}\right)z, z'\left(\Phi + \Psi_x\right)z\right) = \operatorname{cov}\left(\alpha' z, z'\Gamma z\right) = \operatorname{cov}\left(\alpha' z, \tilde{z}\right) = 0$$

Therefore, the resulting variance model is

$$\operatorname{var}\left(y_{(\mathrm{N}\times\mathrm{C}\times\mathrm{C})}\right) = \left(\gamma' + x'\Delta + \tilde{x}\right)\Sigma_z\left(\gamma' + x'\Delta + \tilde{x}\right)' + 2tr\left(\left(\Phi + \Psi_x\right)\Sigma_z\right)^2 + \sigma^2$$

Bibliography

Acito, N., Corsini, G., Diani, M., and Greco, M. "Reducing Computational Complexity in Hyperspectral Anomaly Detection: A Feature Level Fusion Approach," in *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, pp. 1804–1807, 2006.

Anderson, T.W. *An Introduction to Multivariate Statistical Analysis.* Hoboken, NJ: Wiley-Interscience, 2003.

Banerjee, A., Burlina, P., and Diehl, C. "A Support Vector Method for Anomaly Detection in Hyperspectral Imagery," *IEEE Transactions Geoscience Remote Sensing,* vol. 44, no. 8, pp. 2282-2291, Aug. 2006.

Banerjee, A., Burlina, P., and Meth, R. "Fast Hyperspectral Anomaly Detection Via SVDD," in *Proceedings of IEEE International Conference on Image Processing*, vol. 4, pp. 101-104, 2007.

Banks, J., Carson, J.S., Nelson, B.L., and Nicol, D.M. *Discrete-Event System Simulation*. Prentice Hall, New York, 2009.

Bellucci, J.P., Smetek, T.E., and Bauer, K.W. "Improved Hyperspectral Image Processing Algorithm Testing Using Synthetic Imagery and Factorial Designed Experiments," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48 no. 3, pp. 1211-1223, Mar. 2010.

Berk, A., Andersonb, G.P., Bernstein, L.S., Acharya, P.K., Dothe, H., Matthew, M.W., Adler-Golden, S.M., Chetwynd, J.H Jr., Richtsmeier, S.C., Pukall, B., Allred, C.L., Jeong, L.S., and Hoke, M.L. "MODTRAN4: Radiative Transfer Modeling for Atmospheric Correction," in *Proceeding of SPIE Optical Spectroscopic Techniques and Instrumentation for Atmospheric and Space Research III*, vol. 3756, Jul. 1999.

Billor, N., Hadi, A.S., and Velleman, P.F. "BACON: Blocked Adaptive Computationally Efficient Outlier Nominators," *Computational Statistics and Data Analysis*, vol. 34, pp. 279-298, Sep. 2000.

Chandola, V., Banerjee, A., and Kumar, V. "Anomaly Detection: A Survey." *ACM Computer Surveys*, vol. 41, no. 3, pp. 15:1-58, Jul. 2009.

Chang, C.-I. and Chiang, S.-S. "Anomaly Detection and Classification for Hyperspectral Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 6, pp. 1314-1325 May 2002.

Chang, C.-I., and Ren, H. "An Experiment-Based Quantitative and Comparative Analysis of Target Detection and Image Classification Algorithms for Hyperspectral Imagery*," IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 2, pp. 1044–1063, Mar. 2000.

Chen, W., Liu, L., Zhang, C., Wang, J., Wang, J., and Pan, Y. "Monitoring the Seasonal Bare Soil Areas in Beijing Using Multi-Temporal TM Images," in *Proceedings IEEE International Geoscience and Remote Sensing Symposium*, vol. 5, pp. 3379-3382, 2004.

Chiang, S-S., Chang C-I., and Ginsberg, I. "Unsupervised Target Detection in Hyperspectral Images Using Projection Pursuit," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 7, pp. 1380–1391, Jul. 2001.

Clare, P., Bernhardt, M., Oxford, W., Murphy, S., Godfree, P., and Wilkinson, V. "A New Approach to Anomaly Detection in Hyperspectral Images," in *Proceedings of SPIE Conference on Algorithms and Technology for Multispectral, Hyperspectral, and Ultraspectral Imagery IX*, vol. 5093, pp. 17–28, 2003.

Dillon, W.R. and Goldstein, M. *Multivariate Analysis: Methods and Applications.* New York, NY: John Wiley and Sons, 1984.

Duda, R.O., Hart, P.E., and Stork, D.G. *Pattern Classification.* New York, NY: Wiley-Interscience, 2001.

Eismann, M.T., Stocker, A.D., and Nasrabadi, N.M. "Automated Hyperspectral Cueing for Civilian Search and Rescue," *Proceeding of the IEEE*, vol. 97, no. 6, Jun. 2009.

Eismann, M.T. *Hyperspectral Remote Sensing*, Bellingham. WA: SPIE Press, 2012.

Farrell Jr., M.D. and Mersereau, R.M. "On the Impact of PCA Dimension Reduction for Hyperspectral Detection of Difficult Targets," *IEEE Geosciense Remote Sensing Letters*, vol. 2, no. 2, pp. 192–195, Apr. 2005.

Fawcett, T., "An Introduction to ROC Analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861-874, Jun. 2006.

Gaucel, J.M., Guillaume, M., and Bourennane, S. "Whitening Spatial Correlation Filtering for Hyperspectral Anomaly Detection," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 333–336, 2005.

Grossman, J.M., Bowles, J., Haas, D., Antoniades, J.A., Grunes, M.R., Palmadesso, P., Gillis, D., Tsang, K.Y., Baumbeck, M., Daniel, M., Fisher, J., and Triandaf, I. "Hyperspectral Analysis and Target Detection System for the Adaptive Spectral Reconnaissance Program," in *Proceedings of SPIE Conference on Algorithms for Multispectral and Hyperspectral Imagery IV*, vol. 3372, pp. 2–13, 1998.

Hsueh, M. and Chang, C.-I. "Adaptive Causal Anomaly Detection for Hyperspectral Imagery," in *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, vol. 5, pp. 3222–3224, 2004.

Hyvärinen, A. "Fast and Robust Fixed-Point Algorithms for Independent Component Analysis," *IEEE Transactions on Neural Networks*, vol. 10, vo. 3, pp. 626-634, May 1999.

Hyvärinen, A., Karhunen, J., and Oja, E. *Independent Component Analysis.* New York, NY: Wiley-Interscience, 2001.

Johnson, R.J., Williams, J.P., and Bauer, K.W. "AutoGAD: An Improved ICA Based Hyperspectral Anomaly Detection Algorithm," *IEEE Transactions on Geoscience and Remote Sensing*, in review, 2012.

Kay, S.M. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall PTR, 1993.

Köksoy, O. "Multiresponse Robust Design: Mean Square Error (MSE) Criterion," *Applied Mathematics and Computation*, vol. 175, no. 2, pp. 1716–1729, Apr. 2006.

Kwon, H., Der, S.Z., and Nasrabadi, N.M. "Adaptive Anomaly Detection Using Subspace Separation for Hyperspectral Imagery," *Optical Engineering*, vol. 42, no. 11, pp. 3342–3351, Nov. 2003.

Landgrebe, D.A. "Hyperspectral Image Data Analysis," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 17-28, Jan. 2002.

Landgrebe, D.A. *Signal Theory Methods in Multispectral Remote Sensing.* Hoboken, NJ: John Wiley & Sons, 2003.

Lim, J.S. *Two-Dimensional Signal and Image Processing.* Englewood Cliffs, NJ: Prentice Hall PTR, 1990.

Lin, D.K.J. and Tu, W. "Dual Response Surface Optimization," *Journal of Quality Technology*, vol. 27, no. 1, pp. 34–39, Jan. 1995.

Liu, W. and Chang, C.-I. "A Nested Spatial Window-Based Approach to Target Detection for Hyperspectral Imagery," in *Proceedings of IEEE International Geoscience and Remote Sensing Symposium,* vol. 5, pp. 266-268, 2004.

Liu, W. and Chang, C.-I. "Multiple-Window Anomaly Detection for Hyperspectral Imagery," in *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, vol. 2, pp. 41-44, 2008.

Manolakis D. and Shaw, G. "Detection Algorithms for Hyperspectral Imaging Applications," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 29-43, Jan. 2002.

Manolakis, D., Marden, D., and Shaw, G. "Hyperspectral Image Processing for Automatic Target Detection Applications," *Lincoln Laboratory Journal*, vol. 14, no. 1, pp. 79-116, 2003.

Manolakis, D., Zhang, D., Rossacci, M., Lockwood, R., Cooley, T., and Jacobson, J. "Maintaining CFAR Operation in Hyperspectral Target Detection Using Extreme Value Distributions," in *Proceedings SPIE Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIII*, vol. 6565, pp. 1-11, 2007.

Manolakis, D., Lockwood, R., Cooley, T., and Jacobson, J. "Is There a Best Hyperspectral Detection Algorithm?," in *Proceedings SPIE Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XV*, vol. 7334, pp. 1-16, 2009.

Mao, K.Z. "RBF Neural Network Center Selection Based on Fisher Ratio Class Separability Measure," *IEEE Transactions on Neural Networks*, vol. 13, no. 5, pp. 1211–1217, Sep. 2002.

Mindrup, F.M., Bihl, T.J, and Bauer, K.W. "Modeling Noise in a Framework to Optimize the Detection of Anomalies in Hyperspectral Imaging', in Dagli, C.H. (Ed.), *Intelligent Engineering Systems through Artificial Neural Networks: Computational Intelligence in Architecting Complex Engineering Systems*, vol. 20, pp. 517–524, 2010.

Mindrup, F.M., Bauer, K.W., and Friend, M.A. "Extending Robust Parameter Design to Noise by Noise Interactions with an Application to Hyperspectral Imagery," *International Journal of Quality Engineering and Technology*, vol. 3, no. 1, pp. 1–19, 2012.

Montgomery D.C. *Design and Analysis of Experiments.* Hoboken, NJ: John Wiley and Sons, 2008.

Myers, R.H, Montgomery D.C., and Anderson-Cook, C.M. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. Hoboken, NJ: John Wiley and Sons, 2009.

Nasrabadi, N.M. "A Nonlinear Kernel-Based Joint Fusion/Detection of Anomalies Using Hyperspectral and SAR Imagery," in *Proceedings of IEEE International. Conference on Image Processing*, pp. 1864–1867, 2008.

Pelleg, D. and Moore, A. "X-means: Extending K-means with Efficient Estimation of the Number of Clusters," in *Proceedings of the Seventeenth International Conference on Machine Learning*, pp. 727–734, 2000.

Reed, I.S. and Yu, X. "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, Oct. 1990.

Rickard, L.J., Basedow, R.W., Zalewski, E.F., Silverglate, P.R., and Landers, M. "HYDICE: An Airborne System for Hyperspectral Imaging," in *Proceedings of the SPIE*, vol. 1937, pp. 173-179, 1993.

Robinson, T.J., Borror, C.M., and Myers, R.H. "Robust Parameter Design: A Review," *International Quality and Reliability Engineering*, vol. 20, no. 1, pp. 81–101, Feb. 2004.

Rouse, J.W., Haas, R.H., Schell, J.A., and Deering, D.W. "Monitoring Vegetation Systems in the Great Plains with Third ERTS." in *Proceedings of ERTS Symposium*, NASA no. SP-351, pp. 309-317, 1973.

Schaum, A. "Joint Subspace Detection of Hyperspectral Targets', in *Proceedings of IEEE Aerospace Conference*, vol. 3, pp. 1818–1824, 2004.

Schott, J.R. *Remote Sensing: The Image Chain Approach.* New York, NY: Oxford University Press, 1997.

Searle, S.R. *Linear Models*. New York, NY: John Wiley and Sons, 1971.

Shaw, G. and Manolakis D. "Signal Processing for Hyperspectral Image Exploitation," *IEEE Signal Processing Magazin*e, vol. 19, no. 1, pp. 12–16, Jan. 2002.

Shi, M. and Healey, G. "Using Multiband Correlation Models for the Invariant Recognition of 3-D Hyperspectral Textures," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 5, pp. 1201–1209, May 2005.

Smetek, T.E., *Hyperspectral Imagery Target Detection Using Improved Anomaly Detection and Signature Matching Methods*. Dissertation, AFIT/DS/ENS/07-07, Department of Operation Sciences, Air Force Institute of Technology (AU), Wright-Patterson AFB, OH, Jun. 2007.

Smetek, T.E. and Bauer, K.W. "A Comparison of Multivariate Outlier Detection Methods for Finding Hyperspectral Anomalies," *Military Operations Research*, vol. 13, no. 4, pp. 19-43, Nov. 2008.

Stein, D.W.J., Beaven, S.G, Hoff, L.E., Winter, E.M., Schaum, A.P., and Stocker, A.D. "Anomaly Detection from Hyperspectral Imagery," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 58-69, Jan. 2002.

Stein, D., Stocker, A., and Beaven, S. "The Fusion of Quadratic Detection Statistics Applied to Hyperspectral Imagery," in *IRIA-IRIS Proceedings 2000 Meeting of the MSS Specialty Group on Camouflage, Concealment, and Deception*, pp. 271-280, 2000.

Stone, J.V. *Independent Component Analysis: A Tutorial Introduction*. Cambridge, MA: The MIT Press, 2004.

Taguchi, G. *Introduction to Quality Engineering: Designing Quality into Products and Processes*. White Plains, NY: Quality Resources, 1986.

Taguchi, G. *The System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Costs*. White Plains, NY: Kraus International Publications, 1987.

Taitano, Y.P., Geier, B.A., and Bauer, K.W. "A Locally Adaptable Iterative RX Detector," *EURASIP Journal on Advanced Signal Processing, Special Issue on Advanced Image Processing for Defense and Security Applications,* vol. 2010, 2010.

Tax, D.M.J. and Duin, R.P.W "Support Vector Domain Description," *Pattern Recognition Letters*, vol. 20, no. 11-13, pp. 1191-1199 Nov. 1999.

Tax, D.M.J. and Duin, R.P.W "Support Vector Data Description," *Machine Learning*, vol. 54, no. 1, pp. 45-66, Jan. 2004.

Williams, J.P., Bauer, K.W., and Friend M.A. "Clustering Hyperspectral Imagery for Robust Classification," *SPIE Journal of Applied Remote Sensing*, in review, 2012.

Williams, J.P., Bauer, K.W., and Friend M.A. "Further Extensions to Robust Parameter Design: Three Factor Interactions with an Application to Hyperspectral Imagery," *International Journal of Quality Engineering and Technology*, accepted, May 2012.

Williams, J.P., Bihl, T.J., and Bauer, K.W. "Mitigation of Correlation and Heterogeneity Effects in Hyperspectral Data," in Dagli, C.H. (Ed.), *Intelligent Engineering Systems through Artificial Neural Networks: Computational Intelligence in Architecting Complex Engineering Systems*, vol. 20, pp. 501–508, 2010.

Williams, J.P., Bihl, T.J., and Bauer, K.W. "Towards the Mitigation of Correlation Effects in Anomaly Detection for Hyperspectral Imagery," *Journal of Defense Modeling and Simulation*, in review, 2012.

Yanfeng, G., Ying, L., and Ye, Z. "A Selective Kernal PCA Algorithm for Anomaly Detection in Hyperspectral Imagery," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 725–728, 2006.

Vita

Captain Jason P. Williams graduated from the University of Illinois at Urbana-Champaign in December 2001 with a Bachelor of Science degree in Mathematics. He entered Officers Training School in November of 2002, and was commissioned into the Air Force on 21 February 2003. His first assignment was to Headquarters Air Force Operational Test and Evaluation Center (HQ AFOTEC) at Kirtland AFB, New Mexico, as a Test Capabilities Analyst. After a year working for HQ AFOTEC, he began work with AFOTEC Detachment 6 as the F/A-22 Air Combat Simulator Lead Analyst. In August of 2005, he entered the Graduate School of Engineering and Management, Air Force Institute of Technology at Wright-Patterson AFB, Ohio. Upon completion of his Masters of Science degree in Operations Research, he was be assigned to the Joint Mobile Network Operations (JMNO) Joint Test and Evaluation (JT&E) at MCB Quantico, Virginia, where he was the Chief of the Operations Research Analysis Branch. In January of 2009 he deployed to Al Udeid Air Base, Qatar for 126 days where he was an Iraq Assessments Analyst. In August 2009, he was assigned to the Graduate School of Engineering and Management, Air Force Institute of Technology at Wright-Patterson AFB, Ohio for a second time. After completion of his PhD in Operations Research he was assigned to United States Air Forces in Europe (USAFE) A9 Ramstein AB, Germany.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From – To)*
13-09-2012	Doctoral Dissertation	August 2009 - September 2012

4. TITLE AND SUBTITLE
Towards the Mitigation of Correlation Effects in the Analysis of Hyperspectral Imagery with Extensions to Robust Parameter Design

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Williams, Jason, P., Captain, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB, OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT/DS/ENS/12-07

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Intentionally left blank

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Standard anomaly detectors and classifiers assume data to be uncorrelated and homogeneous, which is not inherent in Hyperspectral Imagery (HSI). To address the detection difficulty, a new method termed Iterative Linear RX (ILRX) uses a line of pixels which shows an advantage over RX, in that it mitigates some of the effects of correlation due to spatial proximity; while the iterative adaptation from Iterative Linear RX (IRX) simultaneously eliminates outliers.

In this research, the application of classification algorithms using anomaly detectors to remove potential anomalies from mean vector and covariance matrix estimates and addressing non-homogeneity through cluster analysis, both of which are often ignored when detecting or classifying anomalies, are shown to improve algorithm performance.

Global anomaly detectors require the user to provide various parameters to analyze an image. These user-defined settings can be thought of as control variables and certain properties of the imagery can be employed as noise variables. The presence of these separate factors suggests the use of Robust Parameter Design (RPD) to locate optimal settings for an algorithm. This research extends the standard RPD model to include three factor interactions. These new models are then applied to the Autonomous Global Anomaly Detector (AutoGAD) to demonstrate improved setting combinations.

15. SUBJECT TERMS

Adaptive Matched Filter (AMF), Anomaly Classification, Anomaly Detection, Atmospheric Compensation, Autonomous Global Anomaly Detector (AutoGAD), Bare Soil Index (BI), Blocked Adaptive Computationally Efficient Nominators (BACON), Cluster Analysis, Hyperspectral Imagery (HSI), Iterative Linear RX (ILRX), Normalized Difference Vegetation Index (NDVI), RX Detector, Response Surface Methodology (RSM), Support Vector Data Description (SVDD), Robust Parameter Design (RPD)

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Kenneth W. Bauer (ENS)
U	**U**	**U**	**UU**	91	**19b. TELEPHONE NUMBER** *(Include area code)* (937) 255-3636, ext 4328; e-mail: kenneth.bauer@afit.edu

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18