Hindawi Publishing Corporation
Journal of Electrical and Computer Engineering
Volume 2012, Article ID 425947, 30 pages
doi:10.1155/2012/425947

Research Article

A Semiparametric Model for Hyperspectral Anomaly Detection

Dalton Rosario

SEDD/Image Processing Division, Army Research Laboratory, 2800 Powder Mill Road, Adelphi, MD 20783, USA

Correspondence should be addressed to Dalton Rosario, dalton.s.rosario.civ@mail.mil

Received 21 April 2012; Revised 19 July 2012; Accepted 1 August 2012

Academic Editor: James Theiler

Copyright © 2012 Dalton Rosario. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Using hyperspectral (HS) technology, this paper introduces an autonomous scene anomaly detection approach based on the asymptotic behavior of a semiparametric model under a multisample testing and minimum-order statistic scheme. Scene anomaly detection has a wide range of use in remote sensing applications, requiring no specific material signatures. Uniqueness of the approach includes the following: (i) only a small fraction of the HS cube is required to characterize the unknown clutter background, while existing global anomaly detectors require the entire cube; (ii) the utility of a semiparematric model, where underlying distributions of spectra are not assumed to be known but related through an exponential function; (iii) derivation of the asymptotic cumulative probability of the approach making mistakes, allowing the user some control of probabilistic errors. Results using real HS data are promising for autonomous manmade object detection in difficult natural clutter backgrounds from two viewing perspectives: nadir and forward looking.

1. Introduction

Hyperspectral (HS) sensors collect the radiation over a wide range of contiguous spectral bands, with each band corresponding to a unique spectral value. The field of view of the sensor is broken into hundreds of thousands of pixels, with each pixel representing from less than one to many squared meters of the region of interest depending on the spatial resolution of the sensor and the height of the sensor during the data collection. A collection of spatial-spectral images is put together resulting in an HS data cube, where the length and width represent the spatial dimension, and the depth represents the spectral dimension [1]. The resulting HS data cube consists of hundreds of thousands of pixels. Each pixel has tens or hundreds of data points, each point corresponding to a unique spectral value. In theory, the spectral signature of each pixel should uniquely characterize the physical material in that spatial land area. In practice, the recorded spectral signatures will never be identical for samples of the same material. Owing to the different illumination conditions, atmospheric effects, sensor noise, and so forth, the resulting spectral signatures for HS imagery pixels containing similar materials will exhibit spectral variability.

The discriminant capability, however, of spectral signatures has led to two major applications: object classification and target detection. The former aims to assign all pixels of the image to thematic classes, the latter searches the image for the presence of specific material (the target). As highlighted in [2], from a theoretical point of view, target detection can be considered as a binary classification problem, aiming at classifying the image into the target class and the background class. But, since targets are usually sparsely populated in the scene, while the background is abundant and heterogeneous, a major distinction between the approaches designed for target detection and classification is that target detectors cannot use criteria based on the minimization of classification error since that would lead to labeling every pixel as background. So, a typical solution proposed in the literature for target detection is to use the Neyman-Pearson approach, as discussed by Manolakis in [3], where maximizing the detection probability for a fixed false-alarm rate is the adopted criterion for the algorithm design.

Due to the availability of spectral signature libraries for a wide range of materials, spectral signature-based target detectors have been widely examined [3, 4]. These methods assume the target spectral signature is both reliable and known a priori and aim at finding highly correlated spectra

Report Documentation Page

Form Approved
OMB No. 0704-0188

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

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2012		**00-00-2012 to 00-00-2012**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
A Semiparametric Model for Hyperspectral Anomaly Detection	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Army Research Laboratory,SEDD/Image Processing Division,2800 Powder Mill Road,Adelphi,MD,20783**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	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

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**30**	

in the scene corresponding to the reference target spectra; these methods are also known as target matching.

Target matching approaches are complicated by the large number of possible objects of interest and uncertainty as to the reflectance/emission spectra of these objects. For example, the surface of an object of interest may consist of several materials, and the spectra may be affected by weathering or other unknown factors. One may be interested in a large number of possible objects each with several signatures. Thus, the multiplicity of possible spectra associated with the objects of interest and the complications of atmospheric compensation, as well spectral calibration, acquisition geometry, and contamination from adjacent objects (see, for instance, the discussion in [5, 6]), have led to the development and application of anomaly detectors that seek to distinguish observations of unusual materials from typical background materials without reference to target signatures.

Anomalies are defined with reference to a model of the background. Typically, background models are developed adaptively using reference data from either a local neighborhood of the test pixel or a large section of the image. Local and global spectral anomalies are defined as observations that deviate in some way from the neighboring clutter background or the image-wide clutter background, respectively. Both approaches have their merits.

Local anomaly detectors are typically designed under the assumption that an anomalous material (the target) is spectrally distinct from local neighborhood spectra, which are assumed to be controlled by a known multivariate distribution (Gaussian); also, it is assumed that the scales of targets are known a priori, or the viewing perspective is assumed to be nadir and the altitude of the flying platform carrying the sensor is available for target scale estimation. This kind of detectors is susceptible to unknown spectral mixtures that are often obtained by sampling spectra through a moving window in the imagery, as the window is placed at a transition of distinct regions, forcing the neighboring spectral mixture to be compared against spectra of one of the regions in that transition; this may significantly increase the false alarm rate. The local anomaly detector is also susceptible to false alarms that are isolated spectral anomalies. For example, consider a scene containing isolated trees on a grass plain. Each separate tree may be detected as a local spectral anomaly even if the image contains a separate region with many pixels of trees. The most popular local anomaly detector in the HS research community is based on maximum likelihood estimation under the multivariate normal distribution; this detector is commonly known as the Reed-Xi (RX) algorithm [7] and has become a benchmark for comparison. Kwon and Nasrabadi proposed a *kernelized* version of RX in [8], and Matteoli et al. proposed a segmented version of RX in [9], for the estimation of the local background covariance matrix from global background statistics to cope with the small sample size problem in estimating covariance from local patches in the image; a small sample size problem occurs when the number of spectral observations is lesser than the number of spectral bands (see examination of this problem in [2]). Banerjee et al. in [10] leveraged the employment of kernel methods and the method known as support vector data description (SVDD) to propose the kernel SVDD. But, it is widely known that the performance of kernel methods crucially depends on the kernel function and its parameter(s) [11]. More recently, Gurram and Kwon in [12] and Khazai et al. in [13] have also addressed the parameter sensitivity of kernel-SVDD based detectors, which is still an open area of research. In the statistical based arena, Stein et al. in [14] presented an overview of the statistical anomaly detectors derived from three background models: local Gaussian model, Gaussian mixture model (GMM), and linear-mixing model. More recently, these models were compared against others approaches using the same dataset [15, 16]. In [15], the algorithms RX, GMM, and a cluster-based one were examined. Matteoli et al. in [16] extended the comparison to include the orthogonal subspace projection (OSP) detector and a deterministic signal subspace processing detector. Other classic approaches have also been adapted to the local anomaly detection problem, for example, Fisher's linear discriminant (see an implementation in [17]).

Global anomaly detectors are based on a simple universal distribution, which is designed to represent the background process in the whole image, thus, the name *global*. Example of these detectors is the GMM [14] or a different version of the RX algorithm, which estimates its required parameters (mean and covariance) not locally, but using spectra from the entire HS data cube; this version is informally known as the Global RX. By design, these methods are more resistant to the small sample size problem mentioned earlier, but they have limited ability to adapt to all nuances of the background class (sometimes referred to as an underfitting problem), which may result in both high false alarm and low anomaly detection rates. Other earlier versions of global detectors require that an HS data cube is first segmented into its constituent material classes, so detection is achieved by applying a cutoff threshold and automatically locating pixel clusters with pixel values above the threshold, representing the outliers of these classes. These hybrid algorithms vary in the method of segmentation, but also tend to use maximum likelihood detection under the multivariate normal distribution. The stochastic expectation maximization clustering algorithm by Stocker in [18] is a related example; see also Masson and Pieczynski in [19]. But, since segmentation results are known to be also sensitive to algorithms' parameters (see, for instance, [1]), utility of segmentation algorithms in the context of anomaly detection has not met expectations for varying real world scenarios.

Independently of the kind of anomaly detector in use, the following is a key consideration that should not be ignored: susceptibility to unknown spectral mixtures of unknown distributions often observed by sampling spectra through a moving window in the imagery, where the spectra in test belong to one of the components in that mixture (for instance, a local patch of canopy being compared against neighboring spectral mixture of the same canopy type and a patch of soil). Rosario in [20] examined this particular problem using near infrared HS imagery, where he showed that even on simple real case scenarios (e.g., motor

vehicles parked at an open grassy field having also trees in the background on a sunny day) transitions of distinct regions may contribute to 18% of all of the locally observed spectra in the imagery, using a small moving window (10 by 10 pixels). Of course, with increased scene complexity (increased heterogeneity), this percentage reaches higher levels.

This paper presents a quasiglobal, semiparametric anomaly detection (QG-SemiP) approach that is less susceptible to problems and issues mentioned above for existing local or global anomaly detectors. In particular, the approach requires only a small fraction of the HS cube to characterize the unknown clutter background (hence, the term quasi-global), in contrast to existing global anomaly detectors, which require the entire cube. It does not use segmentation and it is less susceptible to spectral mixtures in local neighborhoods of the imagery.

The approach consists of three major parts: (i) a data dimension reduction method, which plays an important role on the overall approach, since it maps the data from their native multivariate space to a univariate domain in order to avoid the small sample size problem mentioned earlier, gaining in the process insensitiveness to illumination on objects, reducing the dominance of the blackbody response produced by Earth (if the longwave infrared region applies), among other benefits to be discussed later; (ii) a univariate semiparametric model [21], which is chosen because of its robustness to samples representing a mixture of material types; (iii) a parallel random sampling scheme that characterizes the unknown background clutter, using a binomial probability density function to model the likelihood of sampling targets by chance and erroneously labeling them as clutter, justifying multiple sampling trials in parallel in order to significantly decrease the labeling error probability.

The semiparametric model is neither parametric, since the specific distribution controlling the data is not assumed to be known, nor nonparametric, since other parameters must still be estimated relating two unknown distributions. The semiparametric model, however, assumes that the distributions of the samples to be tested are related to each other through an exponential function (a distortion), having two unknown parameters. As it will be shown later in detail, the model is appealing for many reasons, including the following: if two spectral samples under test belong to the same material type (i.e., they are not anomalous to each other), then the assumed exponential function relating both distributions is reduced to unity. If the two samples under test belong to different material types (i.e., they are anomalous to each other), then the exponential function will impose a significant weight relating both distributions, indicating that the samples are anomalous to each other; a key point here is that such an outcome is invariant to whether the assumption of exponential relationship between the distributions is satisfied or not—this will be discussed in more detail later. Finally, another benefit, although not recognized earlier by users of semiparametric models in other areas of study, is the model's natural ability to handle samples representing a mixture of different material types, which also will be shown later. As a note, samples representing a mixture of material types are known to significantly increase the false alarm rate in operational scenarios, requiring autonomous anomaly detection; they can produce dominant edges between spatial regions of different material classes, later to be detected as meaningless (false) anomalies [2].

The strength of the semiparametric model handling the mixture problem is attributed to the fact that a sample under test is expected to contribute to the estimation of the distribution function controlling the sample labeled as reference, where both samples are expected to *equally* contribute to the estimation (when both are in fact under the same distribution), only the reference sample will be able to contribute (when both samples are from different distributions), or both the reference sample and a portion of the test sample will contribute (when the reference is a mixture and the test sample represents a component in that mixture, or vice versa). These outcomes are produced naturally by the model because of the imposed exponential relationship between the two distributions, as shown later using simulated univariate data to make the point. Another appeal for using an exponential distortion assumption is that many of the known distribution functions can be expressed in terms of an exponential distortion of another distribution, including all of the exponential family of distributions [22].

Rosario in [23] published a much earlier and limited conference paper version of this work, where a two-step semiparametric detector (data transformation and semiparametric test statistic) was introduced to the limited task of local anomaly detection (where prior knowledge of targets' scales was required, imposing the limitation) and its performance was compared only against the RX algorithm. Relative to the previous work in [23], this paper significantly extends both the overarching methodology and presents additional results using the extended approach to test significantly more challenging scenarios from the ground to ground viewing perspective, where targets' scales are unknown a priori. In other words, this work enables capability rather than just offers an incremental improvement. The specific contributions in this paper (method extension and additional results) are as follows.

(i) The two-step anomaly detector (data transformation and semiparametric test statistic) is employed for the first time in a quasi-global framework (which is also proposed herein), where only a fraction of the entire data cube being represented by blocks of data are randomly selected from the imagery and used as reference sets of spectra to test the entire imagery. The results are later fused using order statistics, as the sampling scheme is modeled by the Binomial distribution.

(ii) An analytical cutoff threshold is derived from the approach's asymptotic cumulative probability of rejecting a null hypothesis, when either the null or the alternative hypothesis is true (given that the null

hypothesis is based on a multisample testing and order statistic scheme).

(iii) The extended approach is applied to additional real HS imagery to automatically find manmade objects in the scene, producing excellent results in difficult natural clutter scenarios viewed from both nadir and forward looking viewing perspectives.

(iv) Performance of the extended approach is compared via ROC curves against multiple methods found in the literature, for example, global methods (k-means, Gaussian mixture model, global RX), local methods (kernel RX, standard RX, Fisher's linear discriminant, and the local semiparametric detector discussed in [23]).

(v) An analytical study of the two-sample test framework, using the local semiparameter detector in [23] as the base detector.

(vi) An analytical study of the multi-sample fusion test framework, using the semiparametric model embedded in the quasi-global framework, which relaxes the prior knowledge requirement of target scales, hence, enabling scene anomaly detection tasks independently of the viewing perspective (nadir or forward looking).

(vii) A study of the semiparametric model's behavior in the presence of samples representing a mixture of two different material classes, which is the most common mixture case scenario given the sliding window sizes typically used in anomaly detection operational scenarios.

(viii) A subsection specially devoted to discuss the motivation and give a qualitative assessment of the data transformation used in the two-step semiparametric anomaly detection of [23].

For convenience, a list of notations is available after the appendix.

This paper is organized as follows: Section 2 describes spatial data window models for HS sensing, a semiparametric model, and a hypothesis test; Section 3 discusses the sampling method, its probabilistic model, and introduces QG-SemiP; Section 4 discusses performance of QG-SemiP testing nadir and forward looking HS imagery, consisting of manmade objects in difficult natural background scenarios; Section 5 concludes the paper.

2. Problem Formulation, Data Transformation, and Semiparametric Model Description

The main goal of anomaly detection algorithms testing incoming imagery is to detect objects that are spectrally anomalous to its surroundings, yielding in the process a tolerable number of *nuisance* detections. In many surveillance and reconnaissance applications, it is desired that manmade objects are detected as being anomalous to the surrounding natural clutter. Both format and model of the data play a significant role in attempting to achieve the intended goal.

2.1. Remote Sensing Data and Data Format. Experiment was carried out on data sets from two distinct sensors and viewing perspectives: (i) the hyperspectral digital imagery collection experiment (HYDICE) sensor, from a nadir looking perspective; (ii) the SOC-700 hyperspectral sensor, from a forward looking perspective. Data from these sensors will be referred to in this paper as nadir looking imagery and forward looking imagery, respectively.

The HYDICE sensor records 210 spectral bands in the visible to near infrared (VNIR: bands between 0.38 and 0.97 μm) and short-wave infrared (SWIR: bands between 1.0 and 2.50 μm). An extended description of this dataset can be found, for example, in Schweizer and Moura (2000). The results shown in this section for one data subcube are representative for other sub-cubes in the HYDICE (forest radiance) dataset. An illustrative subcube (shown as an average of 150 bands; 640×100 pixels) is depicted in Figure 1 (Cube 1, top). We discarded water absorption and low signal to noise ratio bands; the bands used are the 23rd–101st, 109th–136th, and 152nd–194th. The scene consists of 14 stationary motor vehicles (targets near a treeline) in the presence of natural background clutter (e.g., trees, dirt roads, grasses). Each target consists of about 7×4 pixels, and each pixel corresponds to an area of about 1.3×1.3 square meters at the given altitude.

The forward looking imagery used for this work was recorded using the SOC-700 VNIR HS spectral imager, which is commercially available off the shelf. The system produces HS data cubes of fixed dimensions $R = 640$ by $C = 640$ pixels by $K = 120$ spectral bands between 0.38 and 0.97 μm. Figure 1 (Cube 2 through 4, bottom) depicts examples of the forward looking imagery, where each pixel in any of the three cube examples corresponds to the average of 120 band values. Data cubes Cube 2 and Cube 3 were collected during the summer of 2004 in California, USA; Cube 4 was collected during the spring of 2008 in New Jersey, USA. From actual ground truth, it is known that the scene in Cube 2 (see Figure 1) contains three motor vehicles and a standing person in the center of the scene (i.e., two pick-up trucks to the left in proximity to each other, a man slightly forward from the vehicles in the center, and a sport utility to the right). Although the natural clutter in Cube 2 and Cube 3 is dominated by Californian valley-type trees and/or terrain at the same general geolocation, the data in Cube 3 depict a significantly more complex scenario. From actual ground truth, it is known that in Cube 3 a sport utility vehicle is inconspicuously deployed in the shades of a large cluster of trees. Portions of the shadowed vehicle can be observed near the center in Cube 3. Cube 4 was recorded in a wooded region at Picatinny, where (according to the available ground truth) a sport car is located behind multiple tree trunks, partially obscuring the vehicle; see Figure 1 (Cube 4).

The four data cubes in Figure 1 are independently displayed as intensity images after linear mapping the gray scale of each to the range 0–255. Pixel intensities shown in each individual surface is only relative to corresponding values in that surface; in other words, pixel values representing the same material (e.g., general terrain) may be displayed with different intensities in another surface. This fact explains,

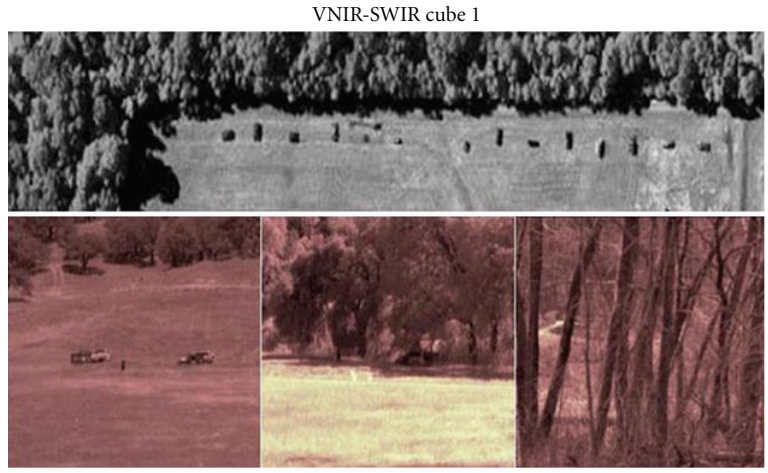


Figure 1: Examples of nadir looking imagery (Cube 1) and forward looking imagery (Cubes 2 through 4). An effective anomaly detection algorithm suite would allow a machine to automatically detect the presence of manmade objects (targets), while suppressing the cluttered environment, using no prior information about what constitutes clutter background or target in the imagery.

for instance, the difference in brightness between the same terrain under similar atmospheric conditions shown in Cube 2 and Cube 3. (The strong reflections from certain parts of the vehicles captured by the sensor in Cube 2 are not as dominant in Cube 3 because the vehicle in Cube 3 is in tree shades; the open field in Cube 3 is then the strongest reflector in the scene).

Next, we present a model of observed data using a sliding $n \times n$ window in $\mathbf{X}$ (a data cube). The data format of $\mathbf{X}$ is shown in (1), where r ($r = 1,\ldots,R$) and c ($c = 1,\ldots,C$) index pixels $\mathbf{x}_{rc}$ in the $R \times C$ spatial area $\mathbf{X}$, where $n \ll R$ and $n \ll C$. Pixels within a fixed $n \times n$ block of data in $\mathbf{X}$ are indexed from the upper left corner of this block using ij relative to rows and columns in $\mathbf{X}$, where $i = 1,\ldots,(R-n-1)$ and $j = 1,\ldots,(C-n-1)$. A representation of an $n \times n$ window at pixel location $(i, j) = (2, 2)$ in $\mathbf{X}$ is as follows:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_{11}, & \mathbf{x}_{12}, & \cdots, \mathbf{x}_{1C} \\ & \overbrace{\begin{matrix} \mathbf{x}_{ij}, \mathbf{x}_{i(j+1)}, \ldots, \mathbf{x}_{i(j+n-1)} \\ \mathbf{x}_{(i+1)j}, \mathbf{x}_{(i+1)(j+1)}, \ldots, \mathbf{x}_{(i+1)(j+n-1)} \\ \vdots \\ \mathbf{x}_{(i+n-1)j}, \mathbf{x}_{(i+n-1)(j+1)}, \ldots, \mathbf{x}_{(i+n-1)(j+n-1)} \end{matrix}}^{n \times n \text{ window, where in this case } i=j=2} & \\ \mathbf{x}_{21}, & & \cdots, \mathbf{x}_{2C} \\ \mathbf{x}_{(2+1)1}, & & \cdots, \mathbf{x}_{(2+1)C} \\ \vdots & & \vdots \\ \mathbf{x}_{(2+n-1)1}, & & \cdots, \mathbf{x}_{(2+n-1)C} \\ \vdots & & \\ \mathbf{x}_{R1}, & \mathbf{x}_{R2}, & \cdots, \mathbf{x}_{RC} \end{bmatrix}. \tag{1}$$

Before pixels within a block of data can be used by a detector, they need to be rearranged to a sequence of multivariate variables. The rearrangement is made by concatenating individual rows in the $n \times n$ window in (1) as follows:

$$\begin{aligned} \mathbf{W}_1 &= \left\lfloor \mathbf{x}_{ij},\ldots,\mathbf{x}_{i(j+n-1)},\mathbf{x}_{(i+1)j},\ldots,\mathbf{x}_{(i+1)(j+n-1)},\ldots, \right. \\ &\qquad \left. \mathbf{x}_{(i+n-1)j},\ldots,\mathbf{x}_{(i+n-1)(j+n-1)} \right\rfloor \\ &= [\mathbf{y}_{11},\ldots,\mathbf{y}_{1n_1}], \end{aligned} \tag{2}$$

where $\mathbf{W}_1 \in \mathbf{R}^{K \times n_1}$, $n_1 = n^2$, and $\mathbf{y}_{1h} \in \mathbf{R}^K$ ($h = 1,\ldots,n_1$), such that $\mathbf{y}_{11} = \mathbf{x}_{ij}$, $\mathbf{y}_{12} = \mathbf{x}_{i(j+1)}$, and so forth until finally $\mathbf{y}_{1n_1} = \mathbf{x}_{(i+n-1)(j+n-1)}$. Since a window can be anywhere in $\mathbf{X}$ and $\mathbf{X}$ represents any HS data cube, $\{\mathbf{y}_{1h}\}_{h=1}^{n_1}$ are considered random vectors and the entire set of spectra that constitutes $\mathbf{X}$ will be observed through the $n \times n$ window.

Using the assumption that the random vectors in $\mathbf{W}_1$ are independent and identically distributed (i.i.d.), the distribution of data within the window, using (2), can be simplified to the following:

$$\mathbf{y}_{11},\ldots,\mathbf{y}_{1n_1} \sim \text{i.i.d. } g_1(\mathbf{y} \mid \boldsymbol{\theta}), \tag{3}$$

where $g_1(\mathbf{y} \mid \boldsymbol{\theta})$ is a conditional multivariate probability density function (PDF) and $\boldsymbol{\theta}$ is its parameter set; both $g_1(\mathbf{y} \mid \boldsymbol{\theta})$ and $\boldsymbol{\theta}$ are typically unknown for real applications.

Normally, an anomaly detector requires two input sets of spectra ($\mathbf{W}_1 \in \mathbf{R}^{K\times n_1}$) and ($\mathbf{W}_0 \in \mathbf{R}^{K\times n_0}$) to perform its task on $\mathbf{X}$. The test sample ($\mathbf{W}_1$) is obtained at a fixed location ij in $\mathbf{X}$, as shown in (1) and (2); but, the source to obtain a reference sample ($\mathbf{W}_0$) will depend on the application, or viewing perspective.

For the nadir looking imagery, the most popular sampling method is to use pixel vectors surrounding a $n\times n$ block of data to construct $\mathbf{W}_0$, where $\mathbf{W}_1$ is constructed using the block of data to be tested. Notice that this sampling method is not suitable for the forward looking imagery because a priori knowledge of target scales in the imagery are required to set the size of a separation (guard) region between the block of data to be tested and locally surrounding samples.

For the forward looking imagery, the reference input set $\mathbf{W}_0$ could be made available from a spectral library, or be randomly selected from the testing data cube. In either case, $\mathbf{W}_0$ would be a rearranged version of a $n \times n$ block of data. The latter is addressed in Section 3, where (in order to make such a test useful for real applications) $\mathbf{W}_1$ is independently compared to multiple spectral sets $\mathbf{W}_0^{(f)} \in \mathbf{R}^{K\times n_0}$ ($f = 1,\ldots,N$); fusing thereafter these comparison results in order to yield a decision (output) surface, as described in Section 3.

Both input sets $\mathbf{W}_1$ and $\mathbf{W}_0$ feed the anomaly detector.

As mentioned earlier, whether the viewing perspective is nadir or forward looking, mixtures of different materials in $\mathbf{W}_1$ and/or in $\mathbf{W}_0$ can significantly degrade anomaly detectors' performances, as examined by Manolakis and Shaw in [2]. It is customary to assume normality in (3), or mixture of Gaussian distributions, but experience has shown (see [1]) that relaxation of these assumptions is needed.

We discuss next a two-step approach for anomaly detection, as introduced in [23], comprising of spectral transformation followed by the application of a univariate semiparametric model.

2.2. Data Transformation. This subsection discusses the employed method to transform spectra from their native multivariate space to a univariate domain, satisfying the univariate data requirement of the semiparametric model. We also provide justification for choosing the employed transformation and give some example cases to reinforce its use.

We consider a data transformation approach in two parts: (i) spectral differencing and (ii) angle mapping.

The rationale for (i) is twofold: (a) since HS samples are contiguous in the spectral domain (i.e., typical spectral resolution is of the order of 10 nanometers), more discriminant and independent information pertaining to a particular material type may be found between adjacent bands, which could augment the statistical power of detectors (this is specially the case in LWIR (longwave infrared) HS imagery where the radiance values observed in calibrated data (collected outdoor) are overwhelmingly influenced by the Planck's blackbody equation [1]), as the Earth's landscape (primarily, canopy and soil) behaves as a blackbody in the LWIR region of the electromagnetic spectrum (note: there is a whole topic of study in mathematical statistics on feature exploitation by *zero crossing*, which uses the output of random variable differencing as used herein for spectra, see [24]); (b) spectral differencing also puts significantly less weight in the importance of spectral magnitude (or bias) in anomaly detection applications, putting focus on the importance of spectral shape, instead. Spectral magnitude relates to the mean average of all measured radiance within a spectral sample, and spectral shape relates to the plotted curve of measured radiance as a function of wavelength. Existing classification and detection algorithms directly or indirectly exploit magnitude and/or shape of spectra in order to perform their tasks.

The benefit of (ii) is that it reduces the detection problem from a multivariate dimensional space to a univariate domain, avoiding the undesired problem of *singularity* issues during inverse estimations of covariance matrices. Singularity is known to occur when the sample size of a spectral sample is less than its number of spectral bands. Although there are approaches proposed in the literature to overcome this issue, the output of these approaches is not always desirable (see, for instance, [3]), since a typical HS sensor usually delivers between 120 and 1,000 bands, but targets may vary in number of pixels from as large as in the thousands to as small as 1 to 4 pixels, depending on the actual physical sizes of these targets and/or distance between the sensor and targets.

This paper is concerned about ensuring that the data transformation method can in fact reduce algorithm sensitivity to spectral magnitude (which can be achieved via (i)), so that a manmade object, for instance, deployed in tree shades can be considered as much as an anomaly relative to a dominant natural clutter background in the same way that the manmade object would have been if it were deployed, instead, out in the open field. All of this, while simultaneously preserving both a high sensitivity to spectral shape and the discriminant characteristics among spectra of distinct material types (which can be achieved via (ii)). If these requirements are matched, then the data transformation approach has achieved the overall desired goal—some example results are shown later in this subsection to reinforce those points.

Before those examples are presented, however, consider the following: let two spectra—having K spectral bands—be available for comparison, $\mathbf{y}_0 = [L_{01},\ldots,L_{0K}]$ and $\mathbf{y}_1 = [L_{11},\ldots,L_{1K}]$, where L_{ij} ($i = 0,1$; $j = 1,\ldots,K$) are nonnegative radiance values. Further assume that $\mathbf{y}_1$ is twenty percent stronger in magnitude than $\mathbf{y}_0$. One way to formalize the disparity in magnitude is to let $\{L_{0j} = \mu + \delta_{0j}\}_{j=1}^{K}$, $\{L_{1j} = 1.2\mu + \delta_{1j}\}_{j=1}^{K}$, and $\mu > 0$. Two key points are worth noticing: (a) the difference $L_{0(j+1)} - L_{0j} = (\mu + \delta_{0(j+1)}) - (\mu + \delta_{0j}) = \delta_{0(j+1)} - \delta_{0j}$ would provide more discriminant and independent information ($\delta_{0(j+1)} - \delta_{0j}$) than jointly using the highly correlated $L_{0(j+1)}$ and L_{0j}, ditto for $L_{1(j+1)} - L_{1j}$; (b) the difference $L_{1(j+1)} - L_{1j} = \delta_{1(j+1)} - \delta_{1j}$ would remove from consideration the 20 percent stronger average magnitude of L_{1j} over L_{0j}, if the $\{L_{0(j+1)} - L_{0j}\}_{j=1}^{K-1}$ were used instead for comparison against $\{L_{1(j+1)} - L_{1j}\}_{j=1}^{K-1}$.

In essence, (a) replaces the need for using—for instance—PCA (principal component analysis) to uncorrelate the data, as it is commonly employed in the HS community [2]); from (b), if $\mathbf{y}_1$ and $\mathbf{y}_0$ are observations of the same material under different illumination conditions (e.g., spectrum $\mathbf{y}_0$ representing a shaded object and spectrum $\mathbf{y}_1$ representing the same object but not shaded), then the average magnitude difference between the two spectra would not play a role in the comparison test, which is desired. For those readers who may have some concerns about what may be lost in the process of transforming the data from multivariate to univariate in the context of anomaly detection, as it will be shown shortly, the loss—although difficult to quantify—is not relevant to the anomaly detection problem, since an effective anomaly detector is not expected to distinguish a material type that is spectrally similar to another material type. If the detector is designed to be that sensitive, it would likely also produce an unacceptably high false alarm rate due to expected within class variability of the same material types dominating the scene (for instance, the expected within class variability of tree clusters across the scene).

The two-part transformation approach is described next.

Borrowing from the discussion in Section 2.1, the transformation approach requires two sets of spectra, a reference set ($\mathbf{W}_0$),

$$\begin{aligned}\mathbf{W}_0 &= [\mathbf{y}_{01},\ldots,\mathbf{y}_{0n_0}]\\ &= \begin{bmatrix} L_{011},\ldots,L_{01n_0}\\ \vdots\\ L_{0K1},\ldots,L_{0Kn_0}\end{bmatrix},\end{aligned} \tag{4}$$

where $\mathbf{y}_{0i} = [L_{01i},\ldots,L_{0Ki}]^t$ is calibrated spectra from a pixel-size location at the scene observed by the K-band sensor, during a particular set of atmospheric and illumination conditions; $(\cdot)^t$ is the vector transposed operator; L_{0ki} $(k = 1,\ldots,K)$ are radiance values, such that, adjacent radiance values are usually highly correlated; $i = 1,\ldots,n_0$ and n_0 is the sample size of $\mathbf{W}_0$; and an independent test set ($\mathbf{W}_1$) that most likely has the same atmospheric condition captured in (4), but not necessarily the same illumination condition captured in (4),

$$\begin{aligned}\mathbf{W}_1 &= [\mathbf{y}_{11},\ldots,\mathbf{y}_{1n_1}]\\ &= \begin{bmatrix} L_{111},\ldots,L_{11n_1}\\ \vdots\\ L_{1K1},\ldots,L_{1Kn_1}\end{bmatrix},\end{aligned} \tag{5}$$

where all of the qualifying comments made for (4) also apply to $\mathbf{y}_{1i} = [L_{11i},\ldots,L_{1Ki}]^t$ with $i = 1,\ldots,n_1$.

Letting u denote the index that distinguish both sets $\mathbf{W}_u$ $(u = 0, 1)$, the magnitude of L_{uki} in (4) and (5) depends on the amount of illumination (e.g., shaded or nonshaded objects) and the illumination environment, this dependence can be significantly reduced by applying the first order differentiation—an approximation of the first derivative—to the columns of $\mathbf{W}_u$, or

$$\begin{aligned}\nabla_0 &= \begin{bmatrix} (L_{021}-L_{011}),\ldots,(L_{02n_0}-L_{01n_0})\\ \vdots\\ (L_{0K1}-L_{0(K-1)1}),\ldots,(L_{0Kn_0}-L_{0(K-1)n_0})\end{bmatrix},\\ \nabla_1 &= \begin{bmatrix} (L_{121}-L_{111}),\ldots,(L_{12n_0}-L_{11n_1})\\ \vdots\\ (L_{1K1}-L_{1(K-1)1}),\ldots,(L_{1Kn_0}-L_{1(K-1)n_1})\end{bmatrix}.\end{aligned} \tag{6}$$

Notice in (6) that $\nabla_0 \in \mathbf{R}^{(K-1)\times n_0}$ and $\nabla_1 \in \mathbf{R}^{(K-1)\times n_1}$, and the sample means of ∇_0 and ∇_1 are, respectively,

$$\begin{aligned}\overline{\nabla}_0 &= \frac{1}{n_0}\nabla_0\,\mathbf{1}_{n_0\times 1},\\ \overline{\nabla}_1 &= \frac{1}{n_1}\nabla_1\,\mathbf{1}_{n_1\times 1},\end{aligned} \tag{7}$$

where $\mathbf{1}_{d\times 1}$ is a column vector of dimension d filled with real values of 1's.

Denoting the columns of ∇_0 (which corresponds to the reference set) as $\{\nabla_{0i} \in \mathbf{R}^{(K-1)}\}_{i=1}^{n_0}$, then the multivariate reference and test samples can be transformed to univariate reference and test samples through the following angle-mapping formulas:

$$x_{0i} = \frac{180}{\pi}\arccos\left(\frac{\nabla_{0i}^t\overline{\nabla}_0}{\|\nabla_{0i}\|\left\|\overline{\Delta}_0\right\|}\right), \tag{8}$$

$$x_{1i} = \frac{180}{\pi}\arccos\left(\frac{\nabla_{0i}^t\overline{\nabla}_1}{\|\nabla_{0i}\|\left\|\overline{\Delta}_1\right\|}\right), \tag{9}$$

where $0^\circ \le x_{0i} \le 90^\circ$, $0^\circ \le x_{1i} \le 90^\circ$, the operator $\|\cdot\|$ using a column vector $\mathbf{x}$ is the square root of $\mathbf{x}^t\mathbf{x}$ (note: although we prefer to use a metric that yields a number having a geometric interpretation, the reader who is concerned about algorithm speed may replace the angle mapper metric with any other comparable metric of choice, for instance, the correlation metric [2] or the normalized dot product showed inside the arccos operator in (8) and (9). The most important aspect about the employed metric is that it must be able to preserve the discriminant characteristics among spectra of different material types, as it will be shown shortly).

From (8) and (9), the following two univariate sequences are constructed:

$$\begin{aligned}x_0 &= (x_{01}, x_{02},\ldots,x_{0n_0}),\\ x_1 &= (x_{11}, x_{12},\ldots,x_{1n_0}),\end{aligned} \tag{10}$$

where x_0 (reference) and x_1 (test) are the input sequences to be used by the univariate based anomaly detection technique discussed in Section 2.3. Note that both samples end up having the same sample size, n_0.

As mentioned earlier, the employed data transformation was specifically chosen to offer a number of desired properties, including reduced sensitivity to spectral magnitude

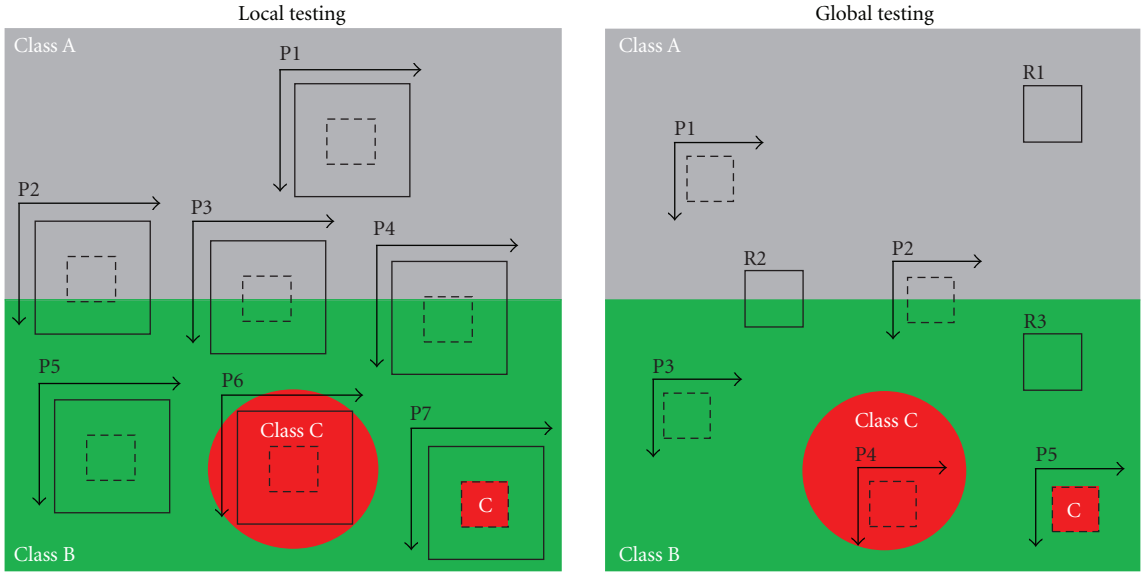


Figure 2: Most common window-based testing scenarios in anomaly detection problems, assuming for simplification that the scene background consists of two distinct material types (Class A and Class B) and a third material (Class C) also distinct from Class A and Class B depicts a anomalous material relative to the background.

and preservation of discriminant features among spectra of distinct material types. Regarding the latter, notice in (10) that, for the proposed transformation to work as advertised, when both multivariate samples $\mathbf{W}_0$ and $\mathbf{W}_1$ happen to be observations of the same material type, the component values in x_0 and x_1 are expected to be comparable and closer to 0° in the scale between 0° and 90°; however, when $\mathbf{W}_0$ and $\mathbf{W}_1$ are observations of distinct material types, the component values in x_0 and x_1 should be proportionally apart, where values in x_0 are expected to be closer to 0° while values in x_1 are closer to 90°. In addition, when the observation represents a mixture of two different material types, the proposed transformation should yield a univariate sample that is representative of the mixture.

Now, we will take a closer look at these expectations, using (8) and (9) to transform two sets of real HS spectra for a qualitative assessment, addressing first the most common sliding window-based testing scenarios naturally occurring in anomaly detection problems: *local testing*, which requires a priori knowledge of object scales (range dependent), and *global testing*, which does not require a priori knowledge of object scales (range invariant). Figure 2 depicts these scenarios (in this context, local testing means comparing clustered spectra against neighboring spectra, while global testing means comparing clustered spectra against spectra located elsewhere across the same imagery).

Figure 2 illustrates the same data-cube spatial representation under the two-test case scenarios, where for simplification the scene background is spatially dominated by only two distinct material types (Class A and Class B) and a third material (Class C—also distinct from Class A and Class B) illustrates the presence of an anomalous material relative to the background. Notice also that the two objects of Class C in the scene have significant size and shape differences, so that, the advantage of approaching the anomaly detection problem from a global rather than a local perspective can be pointed out.

The left hand side image in Figure 2 shows the overlaid sliding window locations of the standard approach to local anomaly detection in the HS research community (see [3]), where a sliding window consisting of an interior square (inner window) is concentrically located along with a larger square so that spectra observed through the inner window can be compared against spectra observed through the outer portion (outer window) of the larger square (i.e., the area of the larger square minus the inner window). Furthermore, the spectral set observed through the outer window is labeled as the *reference sample*, while the spectral set observed through the inner window is labeled as the *test sample*.

As the test window slides across the image one pixel location per algorithmic test, the labels P1 through P7 (left hand side image in Figure 2) highlight seven key test locations in the image. Table 1 summarizes a list of *plausible* anomaly declarations by an anomaly detector versus the *desired* declarations, using the known ground truth information about the scene.

The last column of Table 1 (desired anomaly) shows that out of the seven most distinct window locations for local testing, only two (P6 and P7) are desired, which may contradict declarations made by anomaly detection models currently found in the literature. In fairness, these detectors would be performing the job they are employed to do, as the plausible anomaly declarations shown in Table 1 are indeed correct in the strict sense. For instance, P2 shows a test between observations from Class A (test sample) and observations from a mixture (reference sample: Class A and Class B), while P6 shows a test between observations from

Table 1: Plausible versus desired declarations of *local anomalies*.

Window location	Plausible anomaly	Desired anomaly
P1	No	No
P2	Yes	No
P3	No	No
P4	Yes	No
P5	No	No
P6	No	Yes
P7	Yes	Yes

Table 2: Plausible versus desired declarations of *global anomalies* (Using R1, R2, and R3 as reference samples, see Figure 2).

Window location	Plausible anomaly	Desired anomaly
P1	No (R1)	No (R1)
	Yes (R2)	No (R2)
	Yes (R3)	Yes (R3)
P2	Yes (R1)	No (R1)
	No (R2)	No (R2)
	Yes (R3)	No (R3)
P3	Yes (R1)	Yes (R1)
	Yes (R2)	No (R2)
	No (R3)	No (R3)
P4	Yes (R1)	Yes (R1)
	Yes (R2)	Yes (R2)
	Yes (R3)	Yes (R3)
P5	Yes (R1)	Yes (R1)
	Yes (R2)	Yes (R2)
	Yes (R3)	Yes (R3)

the same material type (the intended anomaly: Class C). The reason, however, these declarations are not desired is that locations similar to P2 will likely accentuate edges as anomalies across the image, increasing the probability of false alarms, and locations similar to P6 will suppress the intended anomaly, decreasing the probability of correct detections. Location P6 also emphasizes the lack of robustness using the local testing approach to find anomalies in the scene, as a priori knowledge of object scales (consisting of the anomalous material) are not always available.

Regarding global testing shown in Figure 2 (the illustration at the right hand side), the window locations denoted as P1 through P5 depict distinct testing locations (observed test samples), while locations denoted as R1 through R3 depict distinct observations (fixed reference samples) that may have been randomly sampled from the image (prior to testing) and used to test every possible window-sized regions across the image, including P1 through P5. Table 2 shows the plausible declarations of anomalies versus the desired declarations, using the same ground truth information about the scene.

The last column of Table 2 (desired anomaly) in essence shows that any test that involves a mixture of classes and a component of that mixture should not be declared as an anomaly so that only truly anomalous material types (in this case Class C) relative to the background will be accentuated. An implementation scheme for the global testing approach, which requires fusion of declarations, will be discussed later. For now consider the following: the final declaration for any given window location is to retain the declaration NO, if it is there, out of all of the results produced by the particular testing location. Using this rule, locations P1, P2, and P3 would produce a final declaration of NO, and P4 and P5 would produce a final declaration of YES, as it would be desired by a global detection scheme. Notice also that P4 ensures that the circular object (Class C) would be accentuated as an anomaly, using the same test window size that would have detected the other anomaly of different scale shown in location P5; this is also desired.

Using real spectral samples, we will now qualitatively demonstrate the behavior of transformation output equations (10), when exposed to the key window positions shown in Figure 2. In anticipation, we would like to see that the data transformation will preserve distinction between samples of two different classes and produce results corresponding to a mixture of classes, when such a mixture is being observed. If successful, those results would give us some level of confidence that not much is being lost in terms of class distinction and that examples of mixtures would be shown as mixtures by the data transformation (demonstration of the desired anomaly detection declarations will be shown later, when the semiparametric model is discussed in Section 2.3. We also defer answering questions about what would happen when other possible window location cases appear, besides the ones shown in Figure 2, until discussion of results from testing real HS imagery in Section 4).

Figure 3 shows spectral transformation results using two sets of real spectra (Class A and Class B), where in this case spectral band differencing was not used as input for angle estimation among spectra and spectral means (see (6)), the actual individual radiance values (L's) were used instead of the difference between adjacent radiance values (L's). Figure 4 shows results using the same two sets of spectra but this time around using band differencing, following exactly the path from (6) to (10).

Class A in Figure 3 consists of 200 real spectra representing a grassy area in an open field; Class B consists of 200 real spectra representing a cluster of tree leaves in the same geographic location of Class A. The employed HS sensor operates in the VNIR (visible to near infrared) region, producing 120 spectral bands per spectrum. To observe the behavior of the proposed data transformation as it processes spectra equivalent to the window location examples shown in Figure 2, Class A is denoted as the reference sample and processed with an independent test set representing Class A (200 spectra), it was also processed with a second test set representing Class B (200 spectra). The spectral sample means of both spectral sets are also shown in Figure 3.

Using both the reference and test samples from the same class (Class A) exemplifies window locations P1, P3, P5,

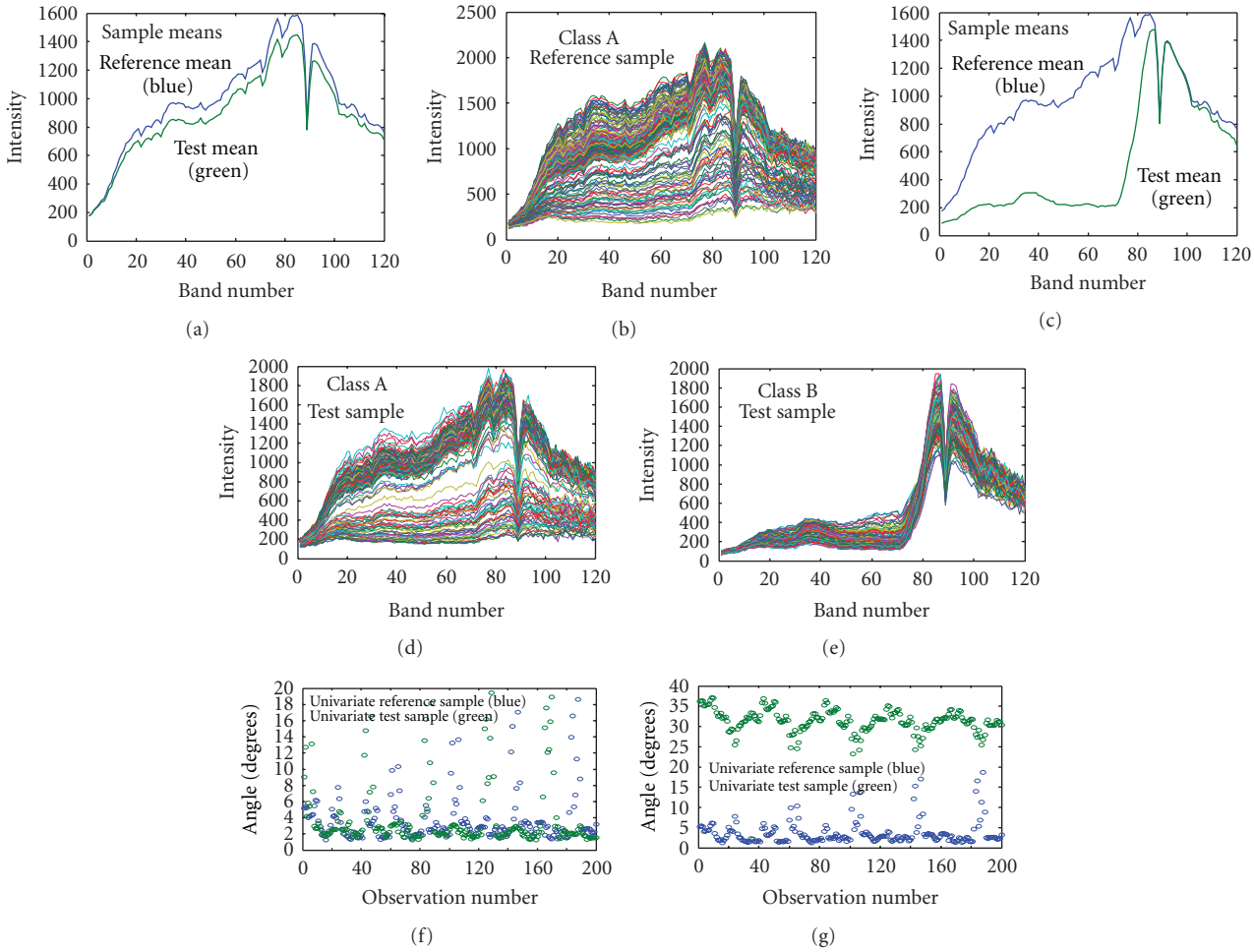


Figure 3: Spectral transformation from multivariate to univariate sample using the actual spectral values instead of spectral differencing.

and P6 in the local testing image in Figure 2; for the global testing illustration in Figure 2, it also represents the cases when spectra from R1 are processed with spectra from P1 and when spectra from R3 are processed with spectra from P3. Using spectra Class A as the reference and spectra from Class B as the test exemplifies the local testing locations P7 (i.e., the presence of an anomaly) and global testing location duals (R1, P4), (R1, P5), (R3, P4), and (R3, P5). The key point to notice in Figure 3 is the angle mapped plots shown in the bottom portion of the figure, where the plot at the left hand side shows that both the univariate reference sample (blue bubbles) and the univariate test sample (green bubbles) are comparable, preserving the fact that both samples belong to the same class. On the other hand, the angle mapped plot shown on the right hand side shows that the spectral transformation preserves the distinction between the samples from Class A and Class B. Both results are desired and would be passed as input to the employed detector for a decision.

The same experiment was held, but instead differentiating data in the spectral direction in order to check whether such a step would change the desired outcome from the spectral transformation shown in Figure 3. Figures 4(a) and 4(b) shows the band differencing means using (7) choosing samples from Class A to be both the reference and test sets (Figure 4(a)), and choosing a sample from Class A as reference and a sample of Class B as the test (Figure 4(b)).

The angle plots shown in Figures 4 (c) and 4(d) affirm that the differentiation step shown in (6) and (7), followed by output transformation from multivariate to univariate data, do preserve the lack of distinction between samples from the same class (Figure 4(c)) and a strong distinction between samples from Class A (reference) and samples from Class B (test), see plots in (Figure 4(d)) . This example provides a direct assertion that what is lost due to the transformation is not relevant to the anomaly detection problem, since the goal of an anomaly detector is to find those material types that are truly distinct from the material types spatially dominating the background scene, as other factors will conspire against the detector's effectiveness (e.g., mixture of distinct material classes, within material class variability). In other words, if a specific material type (target) is desired by the user to be considered as an anomaly relative to a natural environment but the target is not sufficiently distinct spectrally from

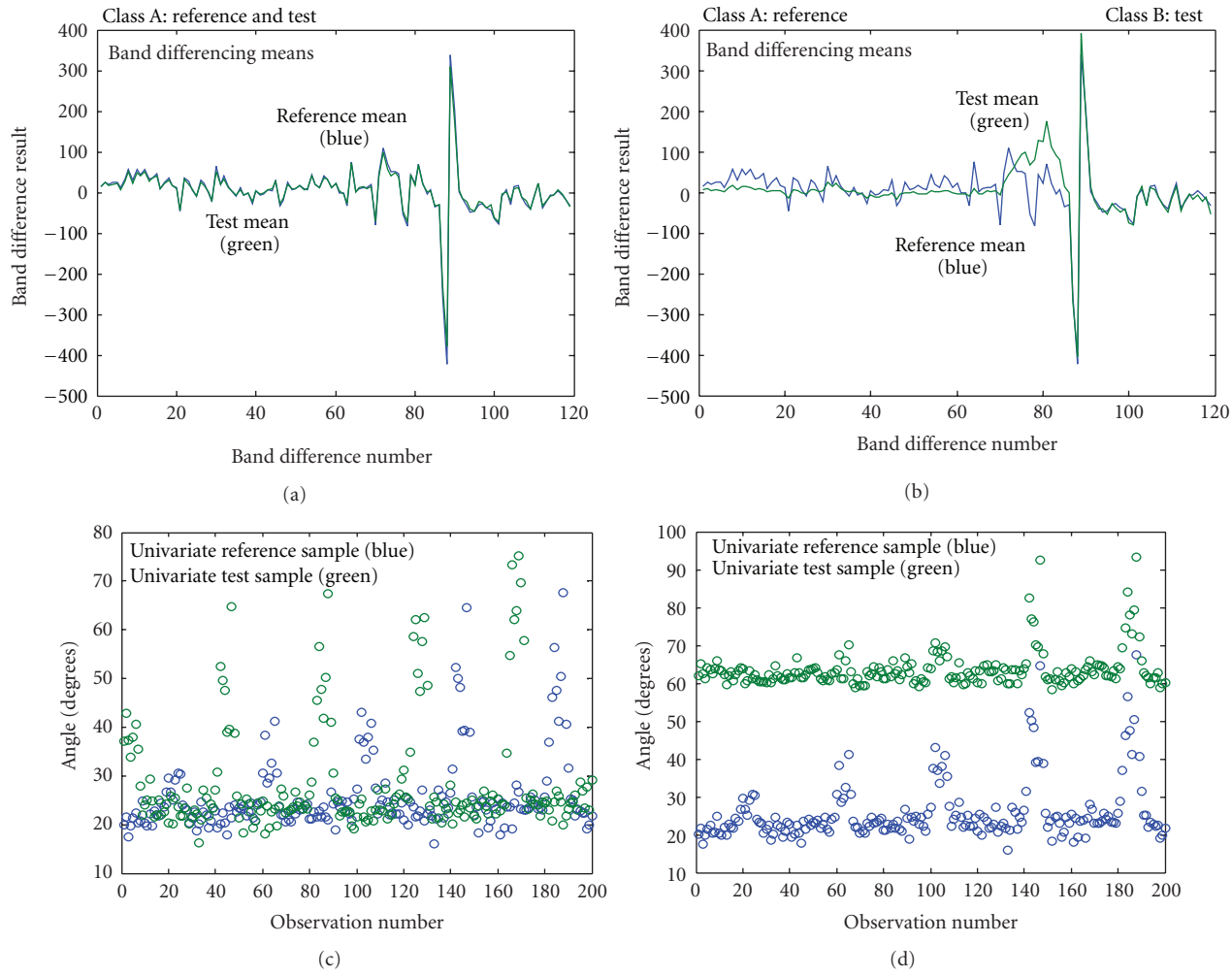


Figure 4: Spectral transformation from a multivariate to univariate sample using spectral differencing.

the background, then an effective anomaly detector would most likely not declare the target as an anomaly. Ironically, as mentioned earlier, this is a virtue because a detector that is too sensitive in distinguishing two spectrally similar material types is also likely to produce an unacceptably high number of false alarms due to expected within class variability of dominant background material types in the scene.

Next, we would like to check how a spectral set, representing a mixture, is altered by the data transformation. For this demonstration, we constructed a reference sample by combining 100 spectra of Class A with 100 spectra of Class B, so that, the reference sample represented a mixture of two classes, and arbitrarily chose the test set to be represented by 200 spectra of Class B (the latter were independent from the ones used to construct the reference mixture). Figures 5(a) and 5(b) show both the mixture (reference construct) and the component of that mixture (Class B: test sample), and the resulting angle mapped plots using spectra without band differencing (Figure 5(c)) and spectra after band differencing (Figure 5(d)).

The key message from Figures 5(c) and 5(d) is that the data transformation with or without the band differencing step does preserve in the univariate domain the fact that the material class of the multivariate test spectra is a component of the mixture of classes represented by the multivariate reference set of spectra. In this particular case, the univariate reference sample (blue bubbles) clearly shows the presence of two classes (i.e., half of the observations has lower angle values and the other half has higher angle values), while the test univariate sample (green bubbles) shows observation values commensurate to the one of the mixture class component (lower angle values). Although this fact is not necessarily desired for anomaly detection (see, for instance, P2 and P4 in Figure 2 (local testing, left image), the data transformation at least does not seem to alter such a scenario involving a mixture, which is fine as long as the employed anomaly detector is designed to handle similar challenging cases.

In summary, the proposed data transformation does preserve distinctions or similarities that exist between spectral sets and also offers additional benefits, as highlighted earlier in this subsection. What is needed now is a model that is more effective finding anomalies, while simultaneously managing the expected negative-impact nuisances naturally occurring in real operational scenarios (e.g., the presence

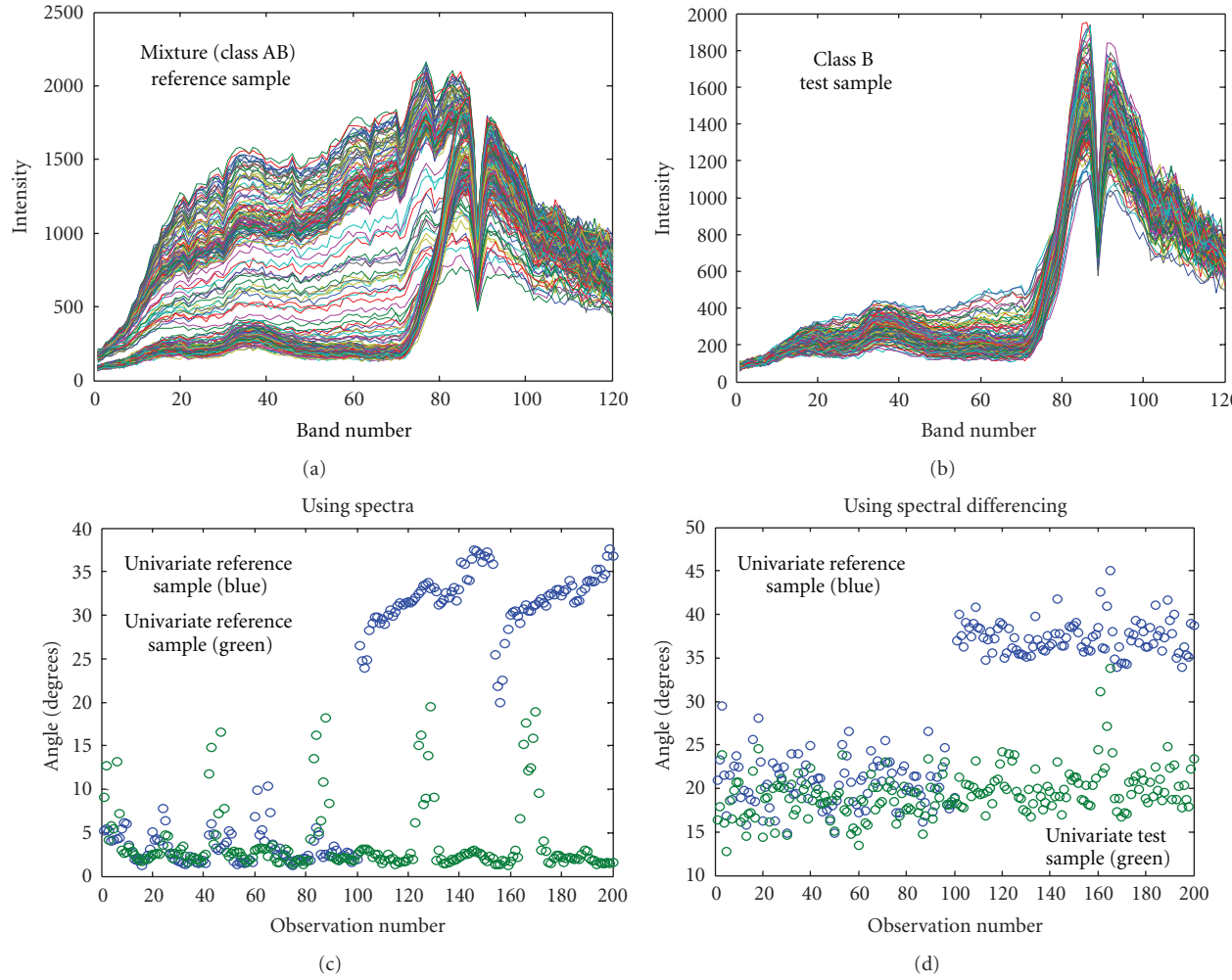


Figure 5: Spectral transformation from a multivariate to univariate sample for a mixture and a component of that mixture.

of mixtures). A semiparametric model is described next for that purpose.

2.3. Univariate Semiparametric Model. Let two multivariate samples $\mathbf{W}_1$ and $\mathbf{W}_0$ be transformed to $x_0 = (x_{01}, x_{02}, \ldots, x_{0n_0}) \sim f_0(x)$ and $x_1 = (x_{11}, x_{12}, \ldots, x_{1n_1}) \sim f_1(x)$, respectively (using, for instance, (4) through (9)), where $f_0(x)$ and $f_1(x)$ are unknown joint PDFs.

To simplify the anomaly detection problem using the transformed data, suppose the two random samples (real valued, not vector valued) x_0 (reference) and x_1 (test) are independent, consisting of i.i.d. (see mathematical notation list after the appendix) random variables controlled by unknown marginal PDFs g_0 and g_1, respectively. Moreover, let g_0 and g_1 be related through the following model:

$$\begin{aligned} x_1 &= (x_{11}, \ldots, x_{1n_1}) \text{ i.i.d} \sim g_1(x) \\ x_0 &= (x_{01}, \ldots, x_{0n_0}) \text{ i.i.d} \sim g_0(x), \end{aligned} \tag{11}$$

$$\frac{g_1(x)}{g_0(x)} = \exp(\alpha + \beta x). \tag{12}$$

The model in (11)-(12) is appealing for many reasons, consider the following examples.

If x_0 and x_1 are samples of the same distribution, then the assumed exponential relationship is reduced to *unity* so that $g_1 = g_0$ (whether g_1 or g_0 is known or not), indicating that x_0 is not anomalous to x_1. If x_0 and x_1 are samples of different distributions, then the exponential function will impose a significant weight relating both distributions, indicating that x_1 and x_0 are anomalous to each other. The key point here is that the latter outcome is invariant to and independent of whether the assumption of exponential relationship between the distributions is satisfied or not; that is, if the exponential relationship assumption is satisfied, then x_1 is anomalous to x_0, but if this assumption is not satisfied, x_1 is still anomalous to x_0. Since the application of interest only requires a determination of whether x_1 is anomalous to x_0, satisfying the imposed assumption of (12) is inconsequential. So, a hypothesis test could just be designed to check whether $\exp(\alpha + \beta x) = 1$ in order to determine the presence of anomalies.

However, the relationship assumption in (12) plays a major role in the mathematical development of the statistical

test and, more importantly for the application in context, it also plays an important role in determining whether a portion or the entire test sample (x_1) can contribute to the estimation of the reference distribution (g_0). The latter is a subtle feature never before recognized in other areas of study, for example, pharmaceutical [22], where semiparametric models are more commonly employed for their utility. The implication of this subtlety is that when one of the samples represents a mixture of different material types and the other represents a component of that mixture, the model in (11)-(12) allows both samples, through the assumed relationship, to contribute to the distribution estimation of the chosen reference sample; in essence stating that the assumption may also be partially satisfied as long as a portion of the test sample belongs to the reference distribution. This is manifested when the test produces an estimation of $\exp(\alpha+\beta x)$ that is significantly closer to unity than it would produce when the test sample has absolutely no relationship with the reference sample. In the practical scenario this is the difference between having a detector capable or not of naturally handling samples representing a mixture, as it will be shown later in this subsection. As mentioned earlier, samples representing a mixture of material types are known to significantly increase the false alarm rate in operational scenarios for autonomous anomaly detection—they can produce dominant edges between regions of different material classes, later to be detected as false alarms [2].

Notice in (12) that, since g_1 is a density, $\beta = 0$ must imply that $\alpha = 0$, as α merely functions as a normalizing parameter, following from the requirement that a PDF (in this case g_1) must integrate to unity, see PDF properties in [25]. Notice also that if $\beta = 0$ then x_0 and x_1 must belong to the same population (i.e., $g_1 = g_0$). Using this fact, a test statistic can be designed to test the following hypotheses:

$$\begin{aligned} H_0 &: \beta = 0 \quad (g_1 = g_0) \quad \text{anomaly absent,} \\ H_1 &: \beta \neq 0 \quad (g_1 \neq g_0) \quad \text{anomaly present.} \end{aligned} \tag{13}$$

In order to estimate β, denote the union of x_0 and x_1 (combined data) by t,

$$t = (x_{11},\ldots,x_{1n_1},x_{01},\ldots,x_{0n_0}) \equiv (t_1,\ldots,t_n), \tag{14}$$

and following the construction by Qin and Zhang in [21] and Fokianos et al. in [22], a maximum likelihood estimator of $G_0(x)$—the continuous cumulative distribution function (CDF) corresponding to the reference $g_0(x)$, can be obtained by maximizing the likelihood over the class of step CDF's with jumps at the observed values $t_1,\ldots,t_n$. Accordingly, if $\widetilde{g}_0(t_i) = dG(t_i)$, where $d(\cdot)$ denotes the differentiation operator, $i = 1,\ldots,n$, the likelihood becomes,

$$\begin{aligned} \zeta(\alpha,\beta,\widetilde{g}_0) &= \prod_{i=1}^{n_0}\widetilde{g}_0(x_{0i})\prod_{j=1}^{n_1}\exp\left(\alpha+\beta x_{1j}\right)\widetilde{g}_0\left(x_{1j}\right) \\ &= \prod_{i=1}^{n=n_1+n_0}\widetilde{g}_0(t_i)\prod_{j=1}^{n_1}\exp\left(\alpha+\beta x_{1j}\right). \end{aligned} \tag{15}$$

One can now express each $\widetilde{g}_0(t_i)$ in terms of α and β and then substitute the terms $\widetilde{g}_0(t_i)$ back into the likelihood to produce a function of α and β only. When α and β are fixed, (15) is maximized by maximizing only the product term $\prod_{i=1}^{n}\widetilde{g}_0(t_i)$, subject to the constraints

$$\sum_{i=1}^{n}\widetilde{g}_0(t_i) = 1, \qquad \sum_{i=1}^{n}\widetilde{g}_0(t_i)[\exp(\alpha+\beta t_i) - 1] = 0. \tag{16}$$

Denoting $\rho = n_1/n_0$, Qin and Zhang in [21] showed that

$$\widetilde{g}_0(t_i) = \frac{1}{n_0}\frac{1}{1+\rho\exp(\alpha+\beta t_i)}, \tag{17}$$

and that the value of the profile log-likelihood $\log[\zeta(\alpha,\beta,\widetilde{g}_0]$ up to a constant can be expressed as a function of α and β only, or

$$l(\alpha,\beta) = \sum_{j=1}^{n_1}(\alpha+\beta x_{1i}) - \sum_{i=1}^{n}\log[1+\rho\exp(\alpha+\beta t_i)]. \tag{18}$$

A system of score equations that maximizes (18) over (α,β) is shown below [21],

$$\begin{aligned} \frac{\partial l(\alpha,\beta)}{\partial\alpha} &= -\sum_{i=1}^{n}\frac{\rho\exp[\alpha+\beta t_i]}{1+\rho\exp[\alpha+\beta t_i]} + n_1 = 0, \\ \frac{\partial l(\alpha,\beta)}{\partial\beta} &= -\sum_{i=1}^{n}\frac{t_i\rho\exp(\alpha+\beta t_i)}{1+\rho\exp(\alpha+\beta t_i)} + \sum_{j=1}^{n_1}x_{1j} = 0. \end{aligned} \tag{19}$$

The solution of the score equations yields the maximum likelihood estimators $(\hat{\alpha},\hat{\beta})$ and consequently by substitution also yields an estimator of $\widetilde{g}_0(t_i)$, or [21]

$$\widehat{\widetilde{g}}_0(t_i) = \frac{1}{n_0}\frac{1}{1+\rho\exp\left(\hat{\alpha}+\hat{\beta}t_i\right)}. \tag{20}$$

So, in summary, by using profiling, an estimator (20) for $\widetilde{g}_0(t_i)$ is attained in addition to score equations (19), where both the reference and test samples as shown in (14) are used to estimate g_0 (the reference PDF). This is only possible because the model in (12) implies that g_1 can be expressed in terms of g_0. This feature allows this model to be robust when either g_0 or g_1 is bimodal or multimodal (representing a sample mixture) while the other represents a component of the same mixture—a key factor for handling transitions of distinct regions in the anomaly detection application, as it will be shown later.

Using results from Fokianos et al. in [22], the estimator $\hat{\beta}$ has the normal asymptotic behavior, as the sample size tends to infinity ($n \to \infty$), or

$$\sqrt{n}\left(\hat{\beta}-\beta_0\right) \xrightarrow[n\to\infty]{} N\left(0,\frac{\rho^{-1}(1+\rho)^2}{v^2}\right), \tag{21}$$

where β_0 denotes the true parameter, v^2 is the variance (a scalor) using components from the combined sequence t, $\rho = n_1/n_0$, $n = n_1+n_0$, and $\to$ means *converges to*—in this case to a normal distribution having 0 mean and variance equals to $\rho^{-1}(1+\rho)^2/v^2$.

Both estimators $\hat{\alpha}$ and $\hat{\beta}$ are required to estimate v^2 in (21) via $\hat{\tilde{g}}_0(t_i)$. Denoting $\hat{v}^2$ the estimator of v^2 and using results from [22],

$$\hat{v}^2 = \sum_i t_i^2 \hat{\tilde{g}}_0(t_i) - \left(\sum_i t_i \hat{\tilde{g}}_0(t_i) \right)^2, \tag{22}$$

where $\hat{\tilde{g}}_0(t_i)$ is shown in (20).

Normalizing the left side of (21) with $\rho^{-1}(1+\rho)^2/\hat{v}^2$, setting $\beta_o = 0$ and squaring the final result, and using known properties of the family of Chi square distributions [25], one can test H_0 in (13) with the following expression:

$$Z_{\text{SemiP}} = n\rho(1+\rho)^{-2}\hat{\beta}^2\hat{v}^2 \xrightarrow[n\to\infty]{} \chi_1^2, \tag{23}$$

which has the Chi square distribution asymptotic behavior with 1 degree of freedom, χ_1^2. Under the idealized assumptions of model (11) and (12), a decision can be based on the value of Z_{SemiP} in (23), that is, high values of Z_{SemiP} reject hypothesis H_o, declaring the presence of local anomalies (note: Regarding (23), as typical from any asymptotic expression, the larger the value of n is the more accurate and precise the approximation of the expression will be. Since n in this context coincides with twice the sample size of the reference sample, sample sizes typically used in anomaly detection applications will suffice (greater than 100, yielding n greater than 200). Practitioners in statistics usually recommend that for univariate variables, asymptotic expressions should use at least thirty two observations, indicating that n in this case should be at least 32 or greater).

The test statistic in (23) will be referred to from here forward as the SemiP test statistic or SemiP anomaly detector, which has two steps: data transformation and test statistic estimation.

2.4. Implementation Notes for the Standing Alone SemiP Detector

2.4.1. Function Maximization. In order to implement (23), we perform an unconstrained maximization of the log maximum likelihood function in (18), or alternatively one could minimize the negative version of (18), to obtain the extremum estimators $\hat{\alpha}$ and $\hat{\beta}$. We used one of the conventional unconstrained nonlinear optimization algorithms—the simplex method [26], which is available in Matlab software (Release 13) under the function name fminsearch. The simplex method is a direct search method that does not use numerical or analytic gradients. If n is the length of x, a simplex in n-dimensional space is characterized by the $n+1$ distinct vectors that are its vertices. For instance, in two-space, a simplex is a triangle; in three-space, it is a pyramid. At each step of the search, a new point in or near the current simplex is generated. The function value at the new point is compared with the function's values at the vertices of the simplex and, usually, one of the vertices is replaced by the new point, giving a new simplex. This step is repeated until the diameter of the simplex is less than the specified tolerance. A limitation using such a method is that it may find a local extremum, so the choices of initial parameters may prove to be critical in some cases; however, we found in practice that by setting the initial values to ($\alpha = 0$, $\beta = 0$), the method converges reasonably fast and works very well for all of the cases that we have observed, independently of whether anomalies were present or not in the tests.

The term $\hat{v}^2$ in (23) is computed using $\hat{v}^2 = \hat{E}(t^2) - \hat{E}^2(t)$, where $\hat{E}(t^k) = \sum_i t_i^k \hat{\tilde{g}}_0(t_i)$ and $\hat{\tilde{g}}_0(t_i)$ is shown in (20).

2.4.3. Decision Threshold. As mentioned earlier, using (23), high values of Z_{SemiP} reject hypothesis H_o in (13), declaring that x_1 is an anomaly relative to x_0. Using this detector as a standing alone unit, one could set a decision threshold based on the *type I error*, that is, based on the probability of rejecting H_o given that H_o is true. Using a standard integral table for the Chi square distribution, with 1 degree of freedom, find a threshold that yields an acceptable probability of error (e.g., 0.001).

2.5. Model Behavior in the Presence of Sample Mixtures. We show in this subsection the robustness of the semiparametric model toward an asymmetric test, that is, when a sample of a mixture is compared against a component of that mixture, which is found locally across the image in the form of spatial transitions. More specifically, we would like to show that $\hat{\beta}$ (estimator for β in (13)) is significantly closer to *ZERO* when, for two PDFs $g_A(y) \neq g_B(y)$, $y_1 \sim g_B(y)$ and $y_0 \sim g_B(y)$ or $y_0 \sim (g_A(y) \cup g_B(y))$ (representing the union $\cup$ or a mixture of two PDFs) than when $y_1 \sim g_B(y)$ and $y_0 \sim g_A(y)$. We illustrate this fact using simulated data and focusing on three specific case studies.

Case 1. $y_0 \sim g_A(y)$ versus $y_1 \sim g_B(y)$.

Case 2. $y_0 \sim (g_A(y) \cup g_B(y))$ versus $y_1 \sim g_B(y)$.

Case 3. $y_0 \sim g_B(y)$ versus $y_1 \sim g_B(y)$.

According to [20], Case 2, which represents a transition of distinct regions in the image, appears some 20% of the entire image, or higher, as local patches are observed through a small moving window across typical images. Therefore, Case 2 is a major source of false alarms that could be avoided using a more robust model of the background than the typical models used in the target community. In anomaly detection applications, it is desired that the employed detector declares the presence of an anomaly for Case 1 but *no* anomalies for Cases 2 and 3; a declaration of *no* anomalies present is also desired, if Case 2 were reversed, that is, $y_0 \sim g_B(y)$ versus $y_1 \sim (g_A(y) \cup g_B(y))$, although this case is not shown here, its results are consistent with Table 3.

The results shown in Table 3 were computed using 100 simulated random samples from an i.i.d. Gaussian distribution to represent the reference sequence y_0 and another 100 samples to represent the test sequence y_1. A sequence representing a mixture of two classes consists of

Table 3: Behavior of $\hat{\beta}$ on different case studies.

Case studies	Simulated samples Parameters $\mu_A = 2000;\ \sigma_A^2 = 200$ $\mu_B = 1000;\ \sigma_B^2 = 100$	$\hat{\beta}$ ($\hat{\alpha}$)	Mean estimates $\hat{\mu} = \sum_{i=1}^{n} t_i \widehat{\widetilde{g}}_0(t_i)$ $\hat{\mu}_2 = (1/n_0) \sum_{i=1}^{n_0} y_{0i}$ $\hat{\mu}_3 = (1/n) \sum_{i=1}^{n} t_i$	Variance estimates $\hat{v}^2 = \sum_{i=1}^{n} t_i^2 \widehat{\widetilde{g}}_0(t_i) - \hat{\mu}^2$ $\hat{v}_2^2 = (1/n_0) \sum_{i=1}^{n_0} (y_{0i} - \hat{\mu}_2)^2$ $\hat{v}_3^2 = (1/n) \sum_{i=1}^{n} (t_i - \hat{\mu}_3)^2$
Case 1	$y_0 \sim$ i.i.d $N(\mu_A, \sigma_A^2)$ $y_1 \sim$ i.i.d $N(\mu_B, \sigma_B^2)$	−0.7500 (848.75)	$1.9967e+003$ $1.9967e+003$ $1.4983e+003$	151.9466 153.4815 $2.4982e+005$
Case 2	$y_0 \sim$ i.i.d $N(\mu_A, \sigma_A^2)$ or i.i.d $N(\mu_B, \sigma_B^2)$ $y_1 \sim$ i.i.d $N(\mu_B, \sigma_B^2)$	−0.0073 (8.0110)	$1.4990e+003$ $1.4990e+003$ $1.2494e+003$	$2.4997e+005$ $2.5316e+005$ $1.8859e+005$
Case 3	$y_0 \sim$ i.i.d $N(\mu_B, \sigma_B^2)$ $y_1 \sim$ i.i.d $N(\mu_B, \sigma_B^2)$	−0.0046 (4.5900)	999.8392 999.8392 999.6346	89.2284 89.2574 89.5693

50 samples for each class resulting in a total of 100 samples. The formulation and parameters used to generate these sequences are shown in Table 3 for different case studies, where the samples in row 2 (starting from the left upper corner in Table 3) simulates a local test between a genuine isolated object (y_1) and its homogeneous surrounding background (y_0)—*Case 1*, the samples in row 3 simulates a local test at a transition between two classes, where the test sample belongs to one of these classes—*Case 2*; the samples in row 4 simulates a local test within a homogeneous region—*Case 3*. Practical implementation details of the SemiP detector, which includes the estimation of (α, β), are shown in Section 2.3. The parameters (α, β) were estimated by maximizing the log likelihood function, using an optimization subroutine initialized to (0,0), (0,0), and (0,0) for *Cases 1, 2,* and *3*, respectively, so that convergence to a solution down to a tolerable error could be achieved using the subroutine.

Since the solution of the semiparametric model uses the union of samples t and estimators $\hat{\alpha}$ and $\hat{\beta}$ to estimate $\widetilde{g}_0$ (which itself is an estimator of g_0), we also included in Table 3 the mean estimates $\hat{\mu}$, $\hat{\mu}_2$, and $\hat{\mu}_3$ and the variance estimates $\hat{v}^2$, $\hat{v}_2^2$, $\hat{v}_3^2$; where $\hat{v}^2$ estimates variance from the solution of the semiparametric model using the union of samples $t = (y_0, y_1)$ and $\widehat{\widetilde{g}}_0$; $\hat{v}_2^2$ estimates variance using only the reference sample y_0; and $\hat{v}_3^2$ is the sample variance using t. The mean estimates were computed accordingly, see Table 3.

In reference to results shown in Table 3, recall that the null hypothesis is $\beta = 0$, and notice that the value of $\hat{\beta}$ in Table 3 are significantly closer to *zero* for *Case 3* (homogeneous region) and *Case 2* (a transition of two different region) than for *Case 1* (genuine local anomaly), where in *Case 1* y_0 and y_1 do belong to different classes. Notice also that the disparity between the values of $\hat{v}_2^2$ and $\hat{v}_3^2$ for each case study also reflects how close $\hat{\beta}$ is to *zero*. For instance, the disparity between $\hat{v}_2^2$, and $\hat{v}_3^2$ for *Case 1* is quite large compared to corresponding disparities for *Cases 2* and *3*.

The semiparametric model handles mixture by showing sensitivity on the estimation of β and α, such that when the test sample has strong statistical information about one of the subclasses in the reference sample, the semiparametric method responds by keeping both $\hat{\beta}$ and $\hat{\alpha}$ relatively close to *zero* in order to maximize the log likelihood function in (18). The estimates $\hat{\alpha}$, $\hat{\beta}$ affect directly the computation of $\widehat{\widetilde{g}}_0$, which in turn is used to compute $\hat{v}^2$.

To shed some light on the effect of $\hat{\beta}$ and $\hat{\alpha}$ on $\widehat{\widetilde{g}}_0$, we present some results in Figure 2. The plots shown in Figure 2 (row 1) corresponds to *Case* 1; where, the plots on the left depict the values of t_i as a function of index i, for convenience we have marked where the sequences y_0 and y_1 are relative to each other within t; and the plots on the right depict $\widehat{\widetilde{g}}_0(t_i)$ versus i. Likewise for *Case* 2 (row 2) and *Case* 3 (row 3).

Let us consider first *Cases* 3 and 1. As mentioned earlier, because of the semiparametric model in (11) and (12), the union of samples $t = (y_0, y_1)$, where y_0 is the reference sequence, is used to estimate the reference PDF estimator $\widetilde{g}_0$. Circumstances when both samples belong to the same population (*Case* 3), the estimated cumulative *weight* for the test sample y_1, that is, $w_1 = \sum_{i=1}^{n_1} \widehat{\widetilde{g}}_0(y_{1i})$, is expected to be approximately equal to the estimated cumulative weight for the reference sample y_0, that is, $w_0 = \sum_{i=n_1+1}^{n_1+n_0} \widehat{\widetilde{g}}_0(y_{0(i-n_1)})$; because the constraint $\sum_{i=1}^{n=n_1+n_0} \widetilde{g}_0(t_i) = 1$ was used in the profiling method to attain $\widetilde{g}_0(t_i)$ in terms of α, β, we would expect both w_1 and w_0 to be near 0.500. Using simulated samples for the normal distributions and parameters shown in Table 3 for *Case* 3, we obtained $w_1 = 0.4998$ and $w_0 = 0.5002$, which are very close to our expectations, see Figure 6. We interpret the actual values of $\widetilde{g}_0(t_i)$ shown for *Case* 3 in Figure 6(f) to be an indication that the semiparametric method regards the test sample y_1 to be as *important* in the estimation of $\widetilde{g}_0$ as the reference sample y_0 is in that estimation, for the right justification, as both y_0 and y_1 happens to be governed by the same distribution.

In contrast, when both y_0 and y_1 belong to clearly different classes (see *Case* 1 in Table 3), the semiparametric method *recognizes* this difference and virtually *shuts down* the contribution of y_1 in the estimation of $\widetilde{g}_0$. The way it shuts down the contribution of y_1 is by maximizing the log likelihood function with values of $\hat{\beta}$ and $\hat{\alpha}$ that allow the estimates of atomic exponential distortions $\exp(\hat{\alpha} + \hat{\beta} t_i)$

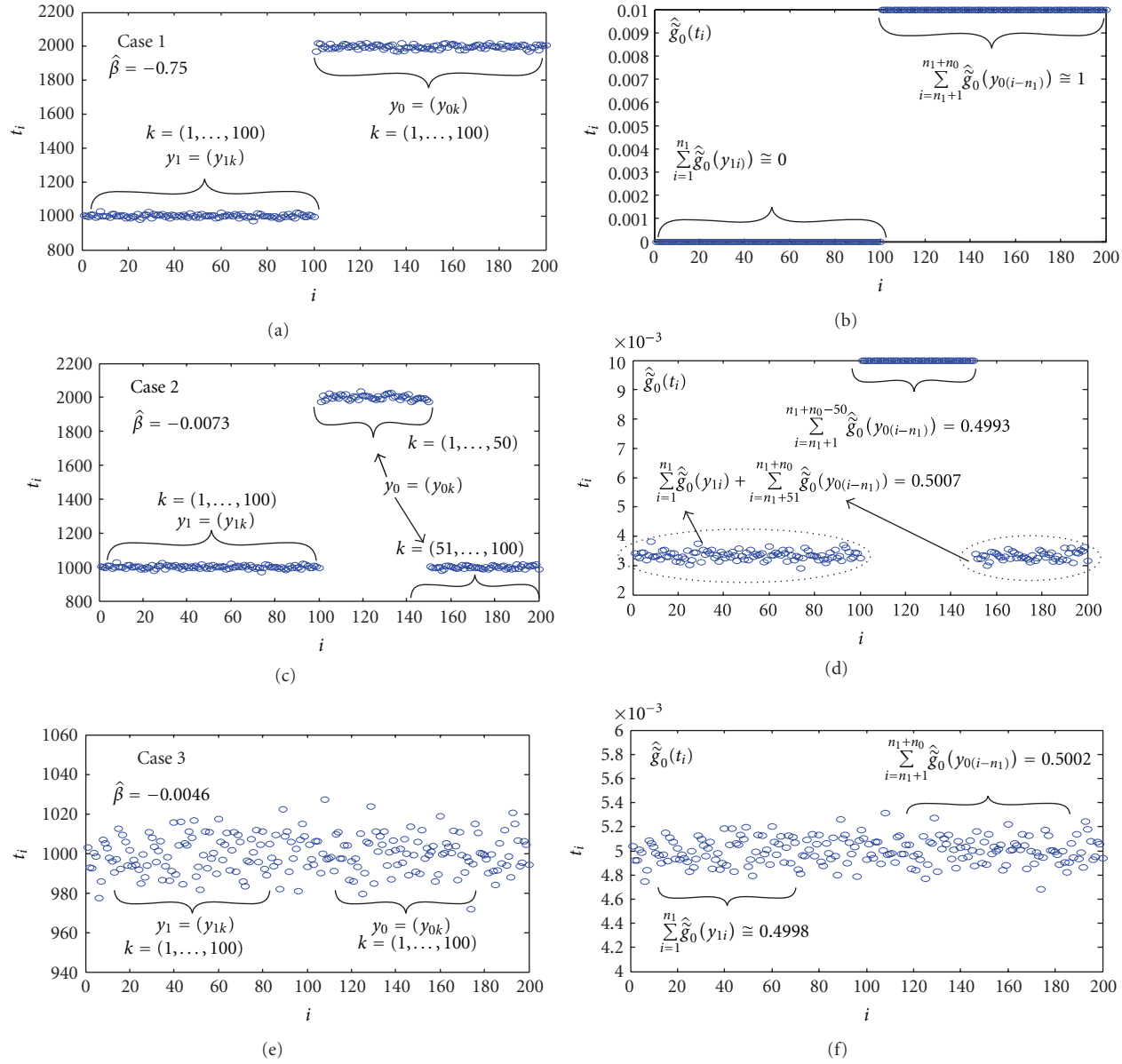


Figure 6: Effect of estimators $\hat{\alpha}$, $\hat{\beta}$ on the computation of $\hat{\tilde{g}}_0$.

to be relatively *high* when t_i corresponds to components of y_1.

And since $\exp(\hat{\alpha}+\hat{\beta}t_i)$ are inversely proportional to $\hat{\tilde{g}}_0(t_i)$, see (20), the contributions of corresponding components of y_1 in t_i estimating $\tilde{g}_0$ are shut down as an indication of *nonimportance* to this estimation. The implication of this shut down is that the value of $\hat{\beta}$ is relatively away from *zero* (relative to *Case* 3, for instance), which rejects the null hypothesis as desired. Figures 6(a) and 6(b) show the combined samples for *Case* 1 and the resulting cumulative weights for the test sample $w_1 = \sum_{i=1}^{n_1} \hat{\tilde{g}}_0(y_{1i}) \cong 0.0$ and for the reference sample $w_0 = \sum_{i=n_1+1}^{n_1+n_0} \hat{\tilde{g}}_0(y_{0(i-n_1)}) \cong 1.0$, where $\cong$ denotes *approximately equal to*. The shutdown is reflected in the results for w_1.

In reference to *Case 2*, where the information carried in y_1 is also contained in y_0, as half of the random components in y_0 are governed by the same distribution of the random components in y_1—see Table 3 and Figures 6(c) and 6(d), the semiparametric method *recognizes* this fact by holding the value of $\hat{\beta}$ at near *zero*, and interestingly by allowing the contributions of y_1 estimating $\tilde{g}_0$ to be comparable to those contributions of the portion of y_0 that are similar to y_1. In other words, since both y_1 and y_0 are used to estimate $\tilde{g}_0$ and half of the random components in y_0 are governed by the same distribution of all of the random components in y_1 and the other half are governed by a different distribution, the semiparametric method will not *discriminate* between y_1 and the portion of y_0 that is similar to y_1. The outcome of

this behavior is that, in order to maximize the log likelihood function, the values of $\hat{\beta}$ and $\hat{\alpha}$ are kept statistically close to *zero* (in this case $\hat{\beta} = -0.0073$) to reflect that the information in y_1 is *important* in the estimation of $\tilde{g}_0$. The way this method explicitly shows this nondiscrimination is by yielding the cumulative weight using y_1 and the portion of y_0 that is similar to y_1 to hold half of the power, while the other half of the power is allocated to the portion of y_0 that is dissimilar to y_1. In other words, using results from Figure 6(d) and this claim, we would expect a value of 0.5 (half power) for the cumulative weight using y_1 and the portion of y_0 that is *similar* to y_1; we obtained $\sum_{i=1}^{n_1} \hat{\tilde{g}}_0(y_{1i}) + \sum_{i=n_1+51}^{n_1+n_0} \hat{\tilde{g}}_0(y_{0(i-n_1)}) = 0.5007$. And we would expect the other half of the power to be in the cumulative weight using the portion of y_0 that is *dissimilar* to y_1; we obtained $\sum_{i=n_1+1}^{n_1+n_0+50} \hat{\tilde{g}}_0(y_{0(i-n_1)}) = 0.4993$. Notice that adding 0.5007 and 0.4993 yields exactly 1.0 as expected because the constraint $\sum_{i=1}^{n=n_1+n_0} \tilde{g}_0(t_i) = 1$ was used in the profiling method in order to attain a representation of $\tilde{g}_0$ in terms of free parameters β and α.

A conclusion that we can draw from this discussion is that the semiparametric method will indirectly compare two samples y_0 (reference) and y_1 (test) by assuming that the distribution of $y_1(g_1)$ and the distribution of $y_0(g_0)$ are related (exponentially) to each other and that, therefore, the information content in both samples can be used to estimate one of these distributions, in particular, g_0.

We found this indirect comparison method to be highly sensitive to the cumulative contribution of y_1 estimating g_0. This sensitivity has an important practical value in the anomaly detection application for three reasons.

First, if $g_1 = g_0$, both samples y_0 and y_1 are expected to equally contribute to the estimation of g_0, which in fact would improve that estimation due to the increase of sample size. Result: y_1 would be labeled as *not* being anomalous to y_0 in this application.

Second, if $g_1 \neq g_0$, sample y_1 is *not* expected to be allowed to contribute to the estimation of g_0, thus, this estimation would solely rely on the cumulative contribution of y_0. Result: y_1 would be labeled as being anomalous to y_0 in this application.

And third, if g_0 is a mixture of densities, such that, g_1 is a component in that mixture, we found that the contribution of y_1 would not be suppressed, but proportional to the weight of g_1 in that mixture (see Figure 6). Result: y_1 would be labeled as *not* being anomalous to y_0 in this application.

This behavior of the semiparametric test statistic is highly desired in the target community because it conforms with be behavior of a human analyst performing the same task in the target application, and it separates this method from existing ones performing the same task.

3. Quasiglobal Semiparametric Approach

The semiparametric test statistic is used as the primary scoring method for the quasi-global anomaly detection approach. As mentioned in Section 1, the quasi-global anomaly detection approach was designed to tackle the forward looking anomaly detection problem, although the application of the quasi-global algorithm using nadir looking imagery are also considered in Section 4.

We start by describing the background sampling method and its probabilistic model, followed by description of the quasi-global algorithm framework using the semiparametric test statistic.

3.1. Sampling Method and Its Probabilistic Model. Assume that target pixels are present in the $R \times C$ spatial area of a $R \times C \times K$ HS data cube $\mathbf{X}$, denote a the total number of target pixels in $\mathbf{X}$, q the probability of a pixel in $\mathbf{X}$ belonging to the target, and the relationship $q = a/A$, where $A = RC$ (or all pixels in $\mathbf{X}$) (in most applications q is unknown, and if multiple targets are present in the imagery, a will be the total number of all pixels belonging to all targets present in the imagery; also, these targets may or may not have the same material type). In order to represent the unknown clutter background in the imagery, let N blocks of data—all having a fixed small area $(n \times n) \ll (R \times C)$—be randomly selected from the $R \times C$ area, see one of the data cubes in Figure 1. In theory, for $(n \times n) = (1 \times 1)$ and using the assumption that target pixels in $\mathbf{X}$ are disjoint and randomly located across the $R \times C$ imagery area (note that in practice, this assumption is not satisfied when targets are present in the scene, but we will use this assumption to establish a guideline), the probability P that at least one block of data has a target pixel is

$$P(m \geq 1) = 1 - p(m = 0), \tag{24}$$

where p is the binomial density function [27], given parameters q and N, and $m \in \{0, 1, \ldots, N\}$ is the number of blocks of data containing a target pixel, or

$$p(m \mid q, N) = \frac{N!}{m!(N-m)!} q^m (1-q)^{N-m}, \tag{25}$$

(symbols $\mid$ and ! denote *given parameters* and the *factorial operator*, resp.).

For convenience, we will refer to $P(m \geq 1)$ as the *probability of contamination* and m the number of *contaminated* blocks of data.

The implementation of this contamination model to the autonomous background sampling problem requires that each one of the $N(n \times n)$ blocks of data be regarded as an independent reference set $\mathbf{W}_0^{(f)}$ $(f = 1, 2, \ldots, N)$ representing clutter spectra, where $\mathbf{W}_0^{(f)} \in \mathbf{R}^{K \times n_0}$ is a rearranged sequence version of the fth block of data having $n_0 = n^2$ spectra. By necessity, n_0 must be significantly greater than *one*—for statistical purposes—but yet significantly smaller than $A = RC$ (e.g., $n_0/A = 20^2/640^2 = 0.000977$) in order to reasonably satisfy the assumption that a $n \times n$ block of data has an unit area at the $R \times C$ imagery area. A contaminated block of data, then, will be treated qualitatively as a block having target pixels covering a large portion of the block's area (e.g., greater than 0.70). In addition—when targets are present, since pixels representing a single target are expected to be clustered in the imagery, the assumption that each target pixel is randomly located

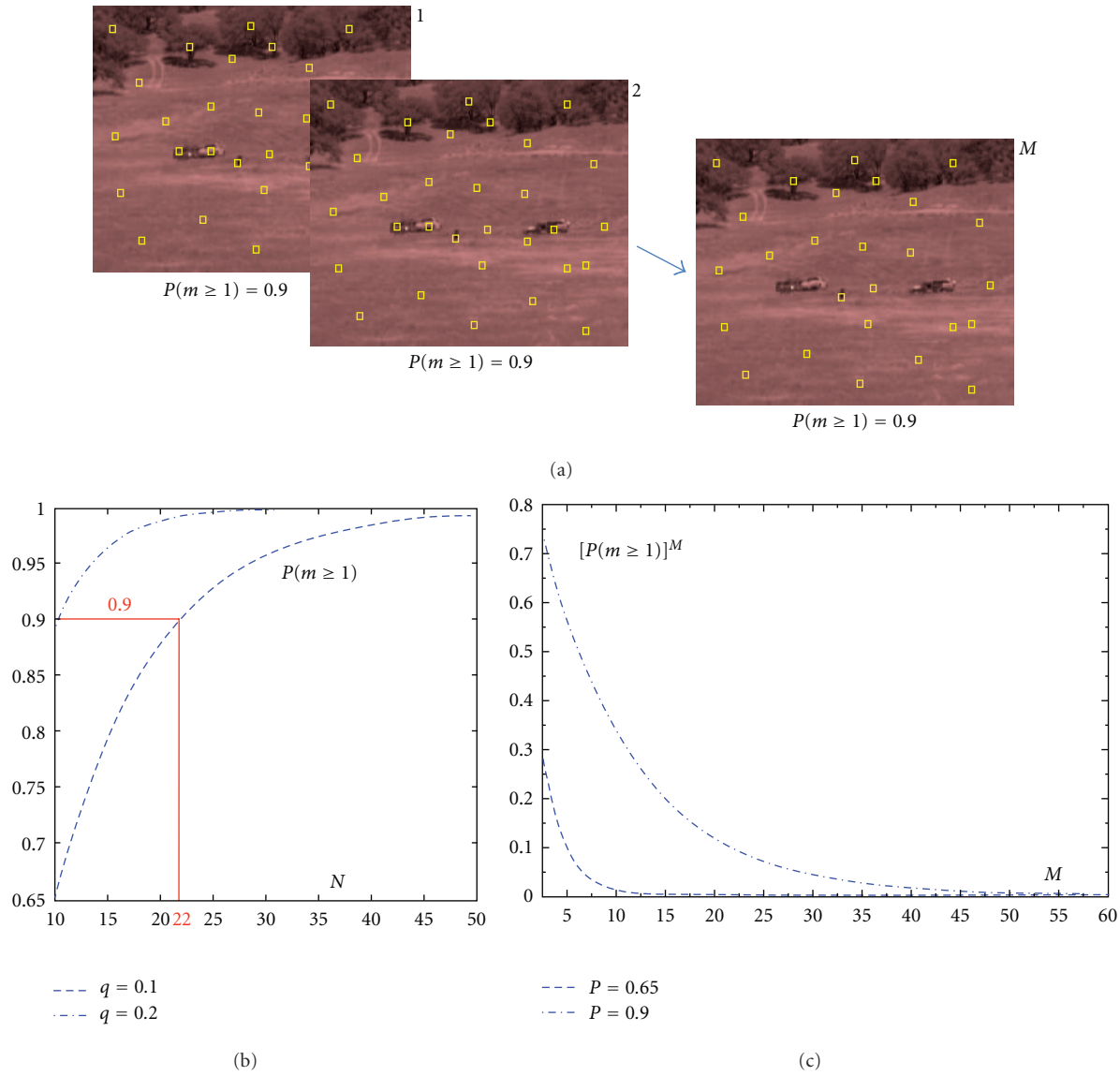


Figure 7: The relationship between N(the number of randomly selected blocks of data, shown as yellow squares on the imagery) and the contamination probability $P_g(m \geq 1) = P(m \geq 1)$ is shown in (b) for a given q (e.g., $q = 0.10$), which is an upper bound guess representing the maximum ratio between target pixels over the $R \times C$ area. To better characterize the unknown clutter background, a high $N$is most desired, but at a high cost, that is, an undesirably high $P_g(m \geq 1)$. The overall contamination probability, however, can significantly decrease by independently repeating the random sampling process Mnumber of times, as shown (c) of figure, and then fusing results using a suitable method.

across the imagery area will be ignored. Using (24), while ignoring the nonclustered target pixel assumption, implies that the probability of contamination will be overestimated, as blocks of data are less likely to be randomly selected from the same cluster of target pixels (for the autonomous background sampling problem, it is more conservative to overestimate the probability of contamination than to underestimate).

Figure 7(b) shows a plot of the probability of contamination $P(m \geq 1)$ versus N, for two values of q (0.1 and 0.2). It is highlighted in the plot in reference that, for instance, if parameters are set to $(q, N) = (0.10, 22)$ then $P(m \geq 1) = 0.90$. Notice that for $N = 22$, if target pixels are present but cover less than $q = 0.10$ of the imagery area, $P(m \geq 1) = 0.90$ is overestimated by two fronts: (i) pixels from a single target are not randomly spread across the imagery area, but clustered, and (ii) the cumulative number of target pixels covers less than 0.10 of the imagery area. So, (24) provides an upper bound (conservative) approximation of the probability of contamination, given parameters q and N.

Figure 7(b) also shows the tradeoff between having a larger number of spectral sets (increasing N) in order to adequately represent the clutter background, which is desired, and the cost of increasing probability of contamination,

which is not desired (in particular, contamination implies that once target pixels are randomly selected by chance from the imagery area, they will be used by a detector as reference set to test the entire imagery, in which case targets would be suppressed).

Since the presence of target pixels in the imagery is unknown a priori, finding a way to decrease the probability of contamination becomes a necessity. In order to decrease this probability, using an adequately large N and a sensible value for q, we propose to independently repeat the random sampling process described in this subsection M number of times. Figure 7(c) illustrates the outcome of M repetitions. If we denote the probability of contamination of the gth random sampling process (or repetition) as $P_g(m \geq 1)$, $1 \leq g \leq M$, for a fixed q and N, note that each $P_g(m \geq 1) = P(m \geq 1)$ and, since $0.0 \leq P(m \geq 1) \leq 1.0$ and these processes will be repeated independently from each other, the overall probability $\widetilde{P}$ that *all* of the processes will be contaminated with at least a contaminated block of data will decrease as a function of increasing M, or

$$\widetilde{P} = P_1(m \geq 1)P_2(m \geq 1) \cdots P_M(m \geq 1) = [P(m \geq 1)]^M. \tag{26}$$

Figure 7(c) plots $\widetilde{P}$ as a function of increasing M, for $P(m \geq 1) = 0.90$ and $P(m \geq 1) = 0.65$. Taking, as an example, the $\widetilde{P}$ curve in Figure 7 corresponding to using $P(m \geq 1) = 0.90$ in (24), notice that for $M > 40$, $[P(m \geq 1)]^M$ decreases to virtually *zero*. This outcome implies that at least one out of the $M > 40$ processes has an extremely high probability of not being contaminated, given that $N = 22$ and target pixels do not cover significantly more than 10% of the imagery area ($q = 0.10$).

3.2. Algorithmic Fusion. A framework for the quasi-global semiparametric anomaly detection algorithm can now be developed using (i) the repeated random sampling model discussed in Section 3.1 (needed to characterize the unknown clutter background in the imagery), (ii) the semiparametric anomaly detector discussed in Section 2.3 (needed to test reference data against the entire imagery), (iii) a way to fuse the results from testing N randomly chosen blocks of data against the entire imagery using small windows (this will produce a 2-dim output surface per repetition), and (iv) a way to fuse M independently produced 2-dim output surfaces into a single 2-dim decision surface.

Start by letting a HS data $(R \times C \times K)$ cube $\mathbf{X}$ be available for autonomous testing. Let also N blocks $(n \times n)$ of data be randomly selected from the $\mathbf{X}$'s $R \times C$ spatial area and used as a reference library set $\mathbf{W}_0^{(f)}$ $(f = 1, 2, \ldots, N)$ representing clutter background spectra, where $\mathbf{W}_0^{(f)} = (\mathbf{y}_{01}^{(f)}, \ldots, \mathbf{y}_{0n_0}^{(f)})$ is a rearranged sequence version of the fth block of data having $n_0 = n^2$ spectra, where $\{\mathbf{y}_{0u}^{(f)}\}_{u=1}^{n_0} \in \mathbf{R}^K$ are K-dim column vectors. Let $\mathbf{W}_1 = (\mathbf{y}_{11}, \ldots, \mathbf{y}_{1n_1})$ be the rearranged version of a $(n \times n)$ window of test data at location ij in $\mathbf{X}$—see (1) for column vectors $\{\mathbf{y}_{1h}\}_{h=1}^{n_1} \in \mathbf{R}^K$; first, we would like to automatically test $\mathbf{W}_1$ against all $\{\mathbf{W}_0^{(f)}\}_{f=1}^{N}$, and produce a single output (scalar) value $\widetilde{Z}_{\text{SemiP}}^{(ij)} \geq 0.0$ from these N test results.

For better clarity in this subsection, we repeat the data transformation steps discussed in Section 2.2, but with the inclusion of index $f = 1, \ldots, N$ and letting $\mathbf{y}_{ij} = L_{i1j}, \ldots, L_{iKj}$, where L_{ikj} is the kth radiance value in $\mathbf{y}_{ij}$, $k = 1, \ldots, K$, and

$$\mathbf{W}_0^{(f)} = \left[\mathbf{y}_{01}^{(f)}, \ldots, \mathbf{y}_{0n_0}^{(f)}\right] = \begin{bmatrix} L_{011}^{(f)}, \ldots, L_{01n_0}^{(f)} \\ \vdots \\ L_{0K1}^{(f)}, \ldots, L_{0Kn_0}^{(f)} \end{bmatrix}, \tag{27}$$

$$\nabla_0^{(f)} = \begin{bmatrix} \left(L_{021}^{(f)} - L_{011}^{(f)}\right), \ldots, \left(L_{02n_0}^{(f)} - L_{01n_0}^{(f)}\right) \\ \vdots \\ \left(L_{0K1}^{(f)} - L_{0(K-1)1}^{(f)}\right), \ldots, \left(L_{0Kn_0}^{(f)} - L_{0(K-1)n_0}^{(f)}\right) \end{bmatrix}, \tag{28}$$

$$\overline{\nabla}_0^{(f)} = \frac{1}{n_0} \nabla_0^{(f)} \mathbf{1}_{n_0 \times 1} \tag{29}$$

and, denoting the columns of $\nabla_0^{(f)}$ as $\{\nabla_{0u}^{(f)}\}_{u=1}^{n_0}$,

$$\left\{ x_{0u}^{(f)} = \frac{180}{\pi} \arccos\left(\frac{\left(\nabla_{0u}^{(f)}\right)^t \overline{\nabla}_0^{(f)}}{\left\|\nabla_{0u}^{(f)}\right\| \left\|\overline{\Delta}_0^{(f)}\right\|} \right) \right\}_{u=1}^{n_0}. \tag{30}$$

And equivalently for $\mathbf{W}_1 = (\mathbf{y}_{11}, \ldots, \mathbf{y}_{1n_1})$—the rearranged version of a $(n \times n)$ window of test data at location ij in $\mathbf{X}$ and the columns of $\nabla_0^{(f)}$ in (28)—$\{\nabla_{0u}^{(f)}\}_{u=1}^{n_0}$, one has

$$\left\{ x_{1u}^{(f)} = \frac{180}{\pi} \arccos\left(\frac{\left(\nabla_{0u}^{(f)}\right)^t \overline{\nabla}_1}{\left\|\nabla_{0u}^{(f)}\right\| \left\|\overline{\Delta}_1\right\|} \right) \right\}_{u=1}^{n_0}. \tag{31}$$

From (30) and (31), the following two univariate sequences will be used as inputs to the SemiP detector:

$$x_0^{(f)} = \left(x_{01}^{(f)}, x_{02}^{(f)}, \ldots, x_{0n_0}^{(f)}\right), \tag{32}$$

$$x_1^{(f)} = \left(x_{11}^{(f)}, x_{12}^{(f)}, \ldots, x_{1n_0}^{(f)}\right), \tag{33}$$

where $1 \leq f \leq N$.

Following the discussion that led to (33), results from the semiparametric test statistic can be used (or fused) as following:

$$\widetilde{Z}_{\text{SemiP}}^{(ij)} = \min_{1 \leq f \leq N} Z_{\text{SemiP}}^{(ij)(f)}, \tag{34}$$

where

$$Z_{\text{SemiP}}^{(ij)(f)} = n\rho(1+\rho)^{-2}\left(\widehat{\beta}^{(f)}\right)^2 \widehat{\nu}^{2(f)}, \tag{35}$$

$\{Z_{\text{SemiP}}^{(ij)(f)}\}_{f=1}^{N} \geq 0.0$, $n_1 = n_0 = n^2$, $\rho = n_1/n_0$, $(i = 1,\ldots,R-n-1)$ and $(j = 1,\ldots,C-n-1)$ index the left-upper corner pixel of an $n \times n$ window in $\mathbf{X}$; using (32) and (33),

$$t^{(f)} = \left(x_{01}^{(f)},\ldots,x_{0n_0}^{(f)},x_{11}^{(f)},\ldots,x_{1n_0}^{(f)}\right) = \left(t_1^{(f)},\ldots,t_{n_1+n_0}^{(f)}\right), \tag{36}$$

$$\hat{v}^{2(f)} = \sum_{u=1}^{n_1+n_o} t_u^{2(f)}\hat{\tilde{g}}_0\left(t_u{}^{(f)}\right) - \left(\sum_{u=1}^{n_1+n_0} t_u{}^{(f)}\hat{\tilde{g}}_0\left(t_u{}^{(f)}\right)\right)^2, \tag{37}$$

$$\hat{\tilde{g}}_0\left(t_u{}^{(f)}\right) = \frac{1}{n_0}\frac{1}{1+\rho\exp\left(\hat{\alpha}^{(f)}+\hat{\beta}^{(f)}t_u{}^{(f)}\right)}, \tag{38}$$

and estimates $\hat{\alpha}^{(f)}$ and $\hat{\beta}^{(f)}$ can be obtained by replacing (α,β) with $(\hat{\alpha},\hat{\beta})$ in (18) and then performing an unconstrained maximization of $l(\alpha,\beta)$; for this paper, a standard unconstrained minimization routines available in Matlab software (i.e., *fminsearch*) was used, setting initial values to $(\hat{\alpha},\hat{\beta}) = (0,0)$.

Notice that if $Z_{\text{SemiP}}^{(ij)(1)}, Z_{\text{SemiP}}^{(ij)(2)},\ldots,Z_{\text{SemiP}}^{(ij)(N)}$ are placed in ascending order and denoted by $Z_{\text{SemiP}(1)}^{(ij)}$, $Z_{\text{SemiP}(2)}^{(ij)},\ldots,$ $Z_{\text{SemiP}(N)}^{(ij)}$, such that $Z_{\text{SemiP}(1)}^{(ij)} \leq Z_{\text{SemiP}(2)}^{(ij)} \leq \ldots \leq Z_{\text{SemiP}(N)}^{(ij)}$, then $\tilde{Z}_{\text{SemiP}}^{(ij)} = Z_{\text{SemiP}(1)}^{(ij)}$ according to (34)—the lowest order statistics (see, for instance, [28]). This fact will be used in estimating the asymptotic behavior of the overall quasi-global semiparametric algorithm, shown in the Appendix.

Notice also that if $\mathbf{W}_1$ is significantly different from all $\{\mathbf{W}_0^{(f)}\}_{f=1}^{N}$, then all of the corresponding results $\{Z_{\text{SemiP}}^{(ij)(f)}\}_{f=1}^{N}$ in (35) would yield high values; this outcome would guarantee the lowest order statistics $\tilde{Z}_{\text{SemiP}}^{(ij)}$ in (34) to hold a high value. Otherwise, if $\mathbf{W}_1$ is significantly similar to at least one of the samples in $\{\mathbf{W}_2^{(f)}\}_{f=1}^{N}$, then at least one of the corresponding results in $\{Z_{\text{SemiP}}^{(ij)(f)}\}_{f=1}^{N}$ would yield a low value; this low value would be assigned to $\tilde{Z}_{\text{SemiP}}^{(ij)}$, according to (34).

Since it is unknown a priori whether target spectra are present in $\mathbf{X}$, the entire $\mathbf{X}$ needs to be tested. In order to achieve it, all $\{\tilde{Z}_{\text{SemiP}}^{(ij)}\}_{i=1,j=1}^{R-n-1,C-n-1}$ must be computed according to (34), producing a 2-dim output surface $\tilde{\mathbf{Z}}_{\text{SemiP}}^{(g)}$. The index $g(1 \leq g \leq M)$ has been introduced to results produced by (34) in order to denote the repetition number discussed in Section 3.1, yielding

$$\tilde{\mathbf{Z}}_{\text{SemiP}}^{(g)} = \begin{bmatrix} \tilde{Z}_{\text{SemiP}}^{(11)(g)},\ldots,\tilde{Z}_{\text{SemiP}}^{[1(C-n-1)](g)} \\ \vdots\vdots\vdots \\ \tilde{Z}_{\text{SemiP}}^{[(R-n-1)1](g)},\ldots,\tilde{Z}_{\text{SemiP}}^{[(R-n-1)(C-n-1)](g)} \end{bmatrix}. \tag{39}$$

The computation leading to (39) will be independently repeated M number of times in order to significantly reduce the probability of contamination (i.e., samples of targets labeled as clutter background). Applying a cutoff threshold to all pixel values $\tilde{Z}_{\text{SemiP}}^{(ij)(g)}$ in $\tilde{\mathbf{Z}}_{\text{SemiP}}^{(g)}$, such that, pixel values that are above or equal to the threshold become 1 and values below become 0, yielding a binary surface (a probabilistic cutoff threshold for this algorithm is presented in Section 4.3). The M binary surfaces are fused using the logic *OR* operator $\oplus$, leading to the algorithm's final output surface $\mathbf{Z}_{\text{SemiP}}$, or

$$\mathbf{Z}_{\text{SemiP}} = \begin{bmatrix} \left(\tilde{Z}_{\text{SemiP}}^{(11)(1)} \oplus \cdots \oplus \tilde{Z}_{\text{SemiP}}^{(11)(M)}\right),\ldots,\left(\tilde{Z}_{\text{SemiP}}^{[1(C-n-1)](1)} \oplus \cdots \oplus \tilde{Z}_{\text{SemiP}}^{[1(C-n-1)](M)}\right) \\ \vdots\vdots\vdots \\ \tilde{Z}_{\text{SemiP}}^{[(R-n-1)1](1)} \oplus \cdots \oplus \tilde{Z}_{\text{SemiP}}^{[(R-n-1)1](M)},\ldots,\tilde{Z}_{\text{SemiP}}^{[(R-n-1)(C-n-1)](1)} \oplus \cdots \oplus \tilde{Z}_{\text{SemiP}}^{[(R-n-1)(C-n-1)](M)} \end{bmatrix}. \tag{40}$$

Figure 8 illustrates $\tilde{\mathbf{Z}}_{\text{SemiP}}^{(g)}$ (39) and $\mathbf{Z}_{\text{SemiP}}$ (40) through a repeated random sampling and result fusion diagram. The diagram shows M independent paths, where, in each path, independent blocks of data are randomly selected from the input HS data cube so that the entire data cube can be tested against these blocks of data, using a testing window of the same block size. Each path, which is indexed by $g(1 \leq g \leq M)$, produces a 2-dim output surface $(\tilde{\mathbf{Z}}_{\text{SemiP}}^{(g)})$, where, at the backend of the diagram, all $\{\tilde{\mathbf{Z}}_{\text{SemiP}}^{(g)}\}_{g=1}^{M}$ passes through a logical OR operator on a pixelwise fashion (i.e., only the pixel values at the same pixel location are logically OR'ed), producing a final 2-dim surface $\mathbf{Z}_{\text{SemiP}}$, as shown in (40).

The motivation and functionality shown in Figure 8 are summarized as follows: for a given repetition g $(1 \leq g \leq M)$, assume that the realization of $\mathbf{W}_1$ from a window location ij in $\mathbf{X}$ is a spectral sample of a target, and the realizations of $\{\mathbf{W}_0^{(f)}\}_{f=1}^{N}$ are samples of various materials composing the clutter background in $\mathbf{X}$, that is, the randomly selected blocks of data are not contaminated with target spectra. The semiparametric order statistics in (34) is expected to yield a high value at that ij location. Moreover, if the target scale in $\mathbf{X}$ is larger than $n \times n$, then the target will be represented by multiple pixels in $\tilde{\mathbf{Z}}_{\text{SemiP}}^{(g)}$ (39), having high values. These pixels are expected to be clustered, hence, accentuating the target spatial location in $\tilde{\mathbf{Z}}_{\text{SemiP}}^{(g)}$. However, as discussed in Section 3.1, the contamination probability

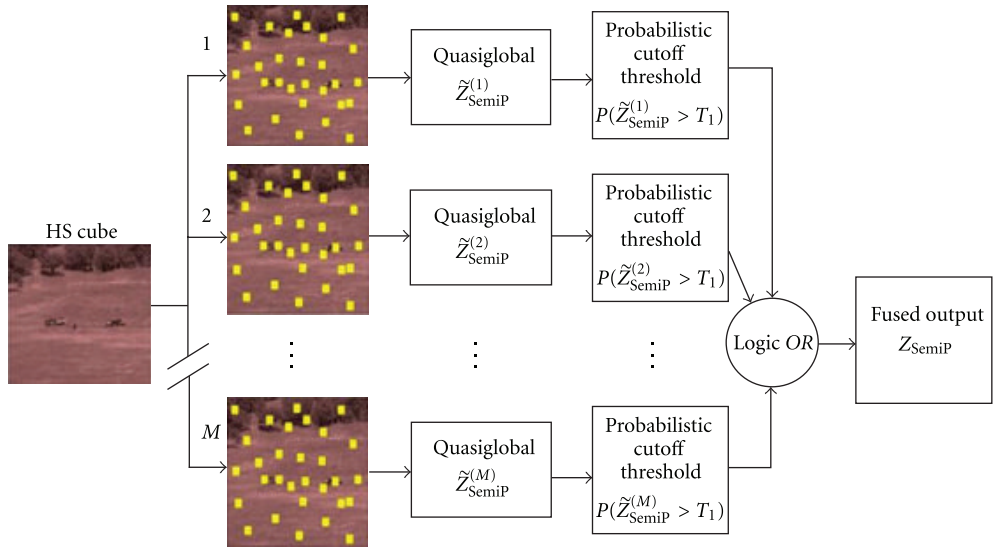


Figure 8: Quasiglobal, semiparametric anomaly detection algorithm.

$P(m \geq 1)$, for a given g, increases as a function of increasing N, see Figure 7. Figure 7 shows further that for a fixed q, N and an adequately large M, if (for instance) results $\widetilde{Z}^{(22)(1)}_{\text{SemiP}}, \widetilde{Z}^{(22)(2)}_{\text{SemiP}}, \ldots, \widetilde{Z}^{(22)(M)}_{\text{SemiP}}$ correspond to the same portion of the target at a testing window location ($i = 2$, $j = 2$), then (40) give us the confidence that at least one term in $\widetilde{Z}^{(22)(1)}_{\text{SemiP}}, \widetilde{Z}^{(22)(2)}_{\text{SemiP}}, \ldots, \widetilde{Z}^{(22)(\text{M})}_{\text{Semip}}$ will have a high value with high probability $[1.0 - \widetilde{P}(\widetilde{m} = M)]$; after application of a cutoff threshold to results in (39), the high value(s) in reference would be captured by the logic OR operator, for example, $(\widetilde{Z}^{(22)(1)}_{\text{SemiP}} \oplus \cdots \oplus \widetilde{Z}^{(22)(\text{M})}_{\text{SemiP}})$, as shown in (40) for all ij locations. Notice that a target may also be represented by multiple (clustered) pixel locations in $\mathbf{Z}_{\text{SemiP}}$ (40).

3.3. Setting the Cutoff Threshold and Other Parameters. For autonomous remote sensing applications, properly setting the algorithm's parameters is a critical step. This subsection presents a guideline to address this step. For the quasi-global, semiparametric anomaly detection algorithm, the parameters of main concern are T_1 (the probabilistic cutoff threshold), N (the number of randomly selected blocks of data), M (the number of testing repetitions), and q (the upper bound ratio of target pixels in the data cube over the spatial area of this cube).

Using the asymptotic behavior shown in (A.11) in the Appendix, a cutoff threshold is attained as follow:

$$T_1 = T(a) = \widehat{\mu}_{\widetilde{g}} + a\widehat{\sigma}_{\widetilde{g}}, \tag{41}$$

where $a = \widetilde{g}^{-1}(1 - \varepsilon_1)$ is the $1 - \varepsilon_1$ quantile of $\widetilde{g}(z) = Lg(z)[1 - G(z)]^{L-1}$ (see Appendix),

$$\widehat{\mu}_{\widetilde{g}} = \sum_{u=1}^{\widetilde{n}} z_u \widetilde{g}(z_u),$$
$$\widehat{\sigma}_{\widetilde{g}} = \sqrt{\sum_{u=1}^{\widetilde{n}} z_u^2 \widetilde{g}(z_u) - \widehat{\mu}_{\widetilde{g}}^2} \tag{42}$$

are estimates of the mean and standard deviation, respectively, of the known distribution $\widetilde{g}(z)$—these estimates can be attained a priori through a simulation experiment using $\widetilde{n}$ samples of $\widetilde{g}(z)$, and $0 \leq \varepsilon_1 \leq 1$.

For setting parameters N and M, as discussed in Section 3.1, the quasi-global semiparametric algorithm—ideally—requires an adequately large, which undesirably increases the contamination probability $P(m \geq 1)$ per repetition, and an adequately large M, which desirably decreases the overall cumulative contamination probability $\widetilde{P}(\widetilde{m} = M)$. From (24), (25), and (26) and using the *log* of base 10, a direct transformation leads to

$$N = \frac{\log[1 - P(m \geq 1)]}{\log(1 - q)}, \tag{43}$$

$$\text{M} = \frac{\log\left[\widetilde{P}(\widetilde{m} = M)\right]}{\log\left[1 - (1 - q)^N\right]}. \tag{44}$$

For a given q, we can fix the values of $P(m \geq 1)$, $\widetilde{P}(\widetilde{m} = M)$ and obtainand M directly using (43) and (44), respectively. As a guideline, $P(m \geq 1)$ should be set high, but less than 1.0, so that N can also be relatively high and $\widetilde{P}(\widetilde{m} = M) < 1.0$; $\widetilde{P}(\widetilde{m} = M)$ should be set very low, near zero. As long as the guideline is followed, interestingly, the actual values of $P(m \geq 1)$ and $\widetilde{P}(\widetilde{m} = M)$ are unimportant.

(a) (b) (c)

Figure 9: QG-SemiP results testing Cube 2 (a) for scene anomalies; output surface (b) was produced using ($q = 0.10$; $T_1 = T(20)$; $N = 3$; $M = 3$); output surface (c) was produced using ($q = 0.10$; $T_1 = T(20)$; $N = 22$; $M = 40$). Bright pixel values (white) in the output surfaces correspond to values above the probabilistic cutoff threshold T_1—depicting the highest confidence level of anomaly presence in the imagery, relative to N randomly selected blocks of data. Testing procedure was independently repeated M times, as highlighted in Figure 8. Using the available ground truth information of the scene, the white clusters in the far right figure cover about 90% of the motor vehicles (the targets) and no false alarms.

For example, we could fix $P(m \geq 1) = 0.90$ and $\widetilde{P}(\widetilde{m} = M) = 0.01$, and for $q = 0.10$, we obtain directly from (43) and (44) parameter values $N \approx 45$ and $M \approx 44$ (Since N and M are defined as integers, these numbers are rounded off $\approx$). For the results shown in Section 4, we fixed at once $q = 0.10$, $P(m \geq 1) = 0.90$, and $P_R(m_R = M) = 0.015$, which by using (43) and (44) yield $N \approx 22$ and $M \approx 40$.

4. Results

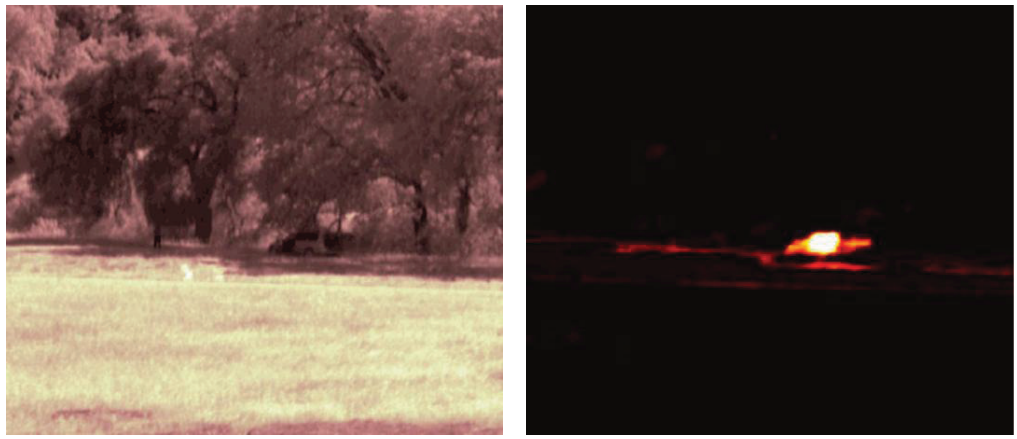
The QG-SemiP algorithm was applied to the HS imagery shown in Figure 1, that is, Cube 1 (nadir looking imagery) and Cube 2 through Cube 4 (forward looking imagery), to test for scene (spectral) anomalies. This subsection presents performance summary using results and guideline discussed in Section 3.3 to set algorithm parameters (q, T_1, N, M). Results using forward looking imagery will be discussed first.

4.1. Forward Looking Imagery. Results testing Cube 2 are shown in Figure 9. For display purposes, the output surface $\mathbf{Z}_{\text{SemiP}}$ (Figure 9, center and right hand side surfaces) is *not* shown as a binary surface; instead, each $\widetilde{\mathbf{Z}}^{(g)}_{\text{SemiP}}$ $(g = 1, \ldots, M)$ output surface is mapped using a pseudocolor map, such that, the brightest pixel values in those surfaces (*white* colored pixels, representing strongest evidence of anomalies) show the locations of results above or equal to the cutoff threshold T_1; while other colors (yellow, red, brown, and black) show lesser evidence of anomalies at the corresponding pixel locations, relative to randomly selected blocks of data. (The color *black* represents no evidence of anomalies.) All of the M surfaces $\widetilde{\mathbf{Z}}^{(g)}_{\text{SemiP}}$, using the threshold based colormap in reference, are then summed to yield the output surface shown in Figure 9 through Figure 11. Additional details follow.

Figures 9(b) and 9(c) show two different outcomes of the quasi-global, semiparametric anomaly detection algorithm, where $n \times n$ was fixed at once to 20×20 (for all data blocks and testing window sizes) and algorithm's parameters were set to ($q = 0.10$; $T_1 = T(20)$; $N = 3$; $M = 3$)—center display—and ($q = 0.10$; $T_1 = T(20)$; $N = 22$; $M = 40$)—right hand side display. The center output surface depicts an example when N is not set sufficiently high, hence, obtaining an inadequate representation of the clutter background. In this case, three blocks of data were randomly selected from the scene (most likely from the open field area, since it is the largest area in the scene), and used by the QG-SemiP detector to suppress (according to $\widetilde{\mathbf{Z}}^{(g)}_{\text{SemiP}}$ $(g = 1, \ldots, M)$ in (39)) the open field in Cube 2, not only once, but most likely $M = 3$ times. As a result, the three motor vehicles and the canopy area on the upper portion of that scene were accentuated relative to the open field. For this initial experiment, we ignored the binomial distribution model and set parameters N and M intentionally low and tested Cube 2 to show the undesired result in Figure 9(b).

For parameters accordingly set to ($q = 0.10$; $T_1 = T(20)$; $N = 22$; $M = 40$) yielded a significantly better result by detecting only the three motor vehicles in the scene, while suppressing the unknown clutter background, see Figure 9(c). Using the available ground truth information of the scene, the white clusters cover about 90 percent of the motor vehicles' spatial area (the targets) and no false alarms. As discussed earlier, for many remote sensing applications, targets (if present in the scene) will not cover more than 10 percent of the imagery spatial area. For instance, the motor vehicle shown at broadside in Cube 2 has 25,000 pixels, covering 6.1% of the imagery area (25000/409600). Note that $q = 0.1$ is robust, because it is independent of targets' aspect or depression angles, relative to the sensor; independent of the number of targets in the scene; and independent of targets' scales, relative to other objects in the scene.

The output surface shown in Figure 9(c) shows three manmade objects (motor vehicles) clearly accentuated (pixel values above T_1) relative to the unknown cluttered environment. It is an achievement, given that no prior information is used about the materials composing the clutter background,



(a) (b)

Figure 10: Results testing Cube 3 (a) and corresponding output surface (b). Parameters were set to ($q = 0.10$; $T_1 = T(20)$; $N = 22$; $M = 40$). A motor vehicle is parked in tree shades—around center of the scene. Using the available ground truth information of the scene, the white clusters cover about 73% of the motor vehicle's spatial area and no false alarms.

or about whether targets are present in the scene, or about targets' scales relative to other objects in the imagery. But notice in Figure 9 that the standing person in the scene center is not detected, possibly because the window size might be too large and/or there must have some materials in that background (randomly selected) spectrally similar to the materials representing that person (e.g., pants, shirt, skin). Figures 10 and 11 show additional results.

Figures 10 and 11 show results for Cube 3 and Cube 4, respective; both HS cubes particularly represent difficult cases of clutter suppression. Parameters were also set to ($q = 0.10$; $T_1 = T(20)$; $N = 22$; $M = 40$) for both HS cubes.

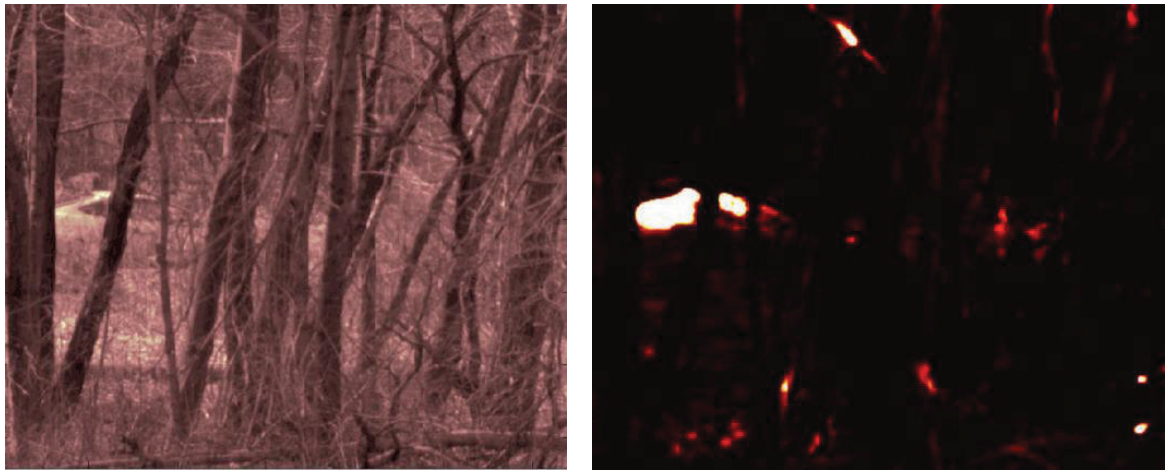
The guideline described in Section 3.3 for setting parameters also worked very well for both complex scenes depicted in Figures 10 and 11. Both output surfaces clearly accentuates the presence of a motor vehicle—one in tree shades (center in Cube 3) and another motor vehicle parked behind a heavily cluttered environment (center left hand side in Cube 4); see *white* pixels, or pixel values greater than or equal to T_1, in both output surfaces in Figures 10 and 11. Setting $T_1 = T(20)$, in essence, means setting the cutoff threshold at 20 standard deviations above the mean of distribution $\tilde{g}(z)$ in (A.10).

Using the available ground truth information of the scenes in Figure 9 through Figure 11, quantitative comparative performances were obtained via receiver's operating characteristic (ROC) curves (vertical axis show Pd for probability of detection, and horizontal axis shows Pfa for probability of false alarms) for some of the anomaly detectors mentioned in Section 1. The data cubes were processed using the global RX, k means, GMM, and QG-SemiP. In essence we used the k mean and GMM in place of SemiP in the context of the parallel random sample, but had to take into consideration some of the inherent constraints of these methods. For instance, for k means and GMM, we recorded ROC curves for parameter N set to $N = 3, 5, 10, 20,$ and 50 and repetition parameter $M = 50$, while for QG-SemiP N was set to 20, 50, 75, and 100 with $M = 50$. Global RX estimates the mean and covariance from the entire data cube, as discussed in Section 1. Figure 12 shows performance of these detectors using ABC to label QG-SemiP, global to label global RX, KM to label k means, and GMM.

Although the parameter values for k means and GMM start at a lower value than QG-SemiP, Figure 12 shows that at lower N the detectors k means and GMM actually perform much better than for N set at higher values, as Nmay be interpreted by these algorithms as the number of distinct classes in the scene. Such performance degradation occurs with large N because the spatial area that corresponds to each individual N block of data is now smaller and samples of the targets, required by the algorithm, need to be included into one of the classes. This outcome contaminated the distribution of the background clutter forcing it to be closer to the distribution of the targets, resulting in performance degradation. The global RX performed reasonably well, as expected since the scene in Figure 9 is relatively less complex. The QG-SemiP detector, also as expected, improved performance as N increased.

Performance of these algorithms on the scenes in Figures 10 and 11 are shown in Figures 13 and 14, respectively. Performance degradation of k means, GMM, and global RX are evident from Figure 13, since the target is on tree shades. QG-SemiP performs well for $N > 20$. In Figure 14, the performance of the k means performed poorly at $N = 20$, since the target is partially blocked by tree trunks, while GMM performed poorly at $N = 3$ and $N = 5$ and was unable to converge at higher N values. The global RX surprisingly worked reasonably well, but completely underperforming the QG-SemiP detector.

4.2. Nadir Looking Imagery. For the nadir looking imagery, Cube 1 in Figure 1 (top), ROC curves are also used as a means to quantitatively compare five anomaly detection approaches: local RX, local KRX, local FLD, local SemiP, and QG-SemiP; see Section 1.



(a) (b)

Figure 11: Results testing Cube 4 (a) and corresponding output surface (b). Parameters were set to ($q = 0.10$; $T_1 = T(20)$; $N = 22$; $M = 40$). A motor vehicle is shown on the left hand side, between top and bottom, behind tree trunks—a sport car. Cube 4 exemplifies a hard case of autonomous clutter suppression. Using the available ground truth information of the scene, the white clusters cover about 44% of the motor vehicle's spatial area and less than 2% of false alarms.

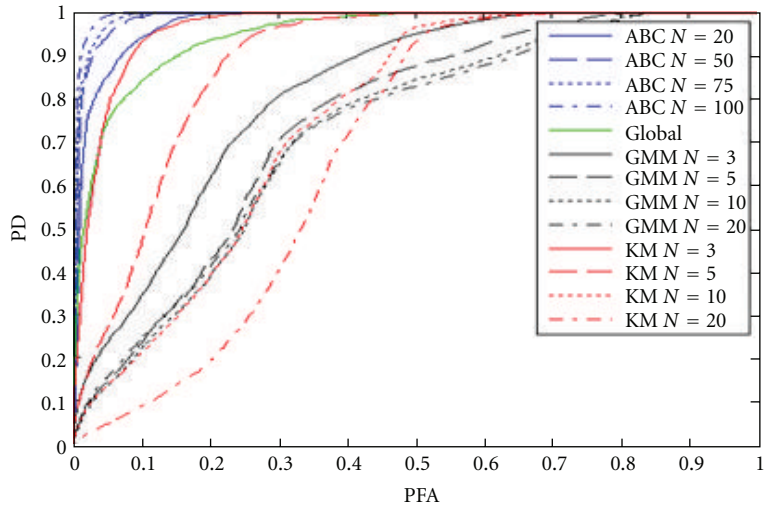


Figure 12: ROC curve performance testing QG-SemiP (ABC), global RX (Global), k means (KM), and GMM detectors on scene shown in Figure 9.

Local anomaly detectors process small ($n \times n$) windows of the HS data cube $\mathbf{X}$, where all the $\mathbf{x}_{rc}$ ($r = 1, \ldots, R$; $c = 1, \ldots, C$) in $\mathbf{X}$ are used; modeling is only done at the level of the $n \times n$ windows, where $n \ll R$ and $n \ll C$ ($\ll$ denoting *many orders of magnitude smaller than*); at the level of the pixel area surrounding these windows. Blocks of data ($n \times n$ windows) that are spectrally different from pixels surrounding them score high using an effective detector in contrast to blocks of data that are not spectrally different from their surrounding pixels. After the detector scores the entire $\mathbf{X}$, it yields a 2-dim surface $\mathbf{Z}$ $[a(R-n-1) \times (C-n-1)$ array of scalars], where a cutoff threshold is then compared to the pixel values in $\mathbf{Z}$. Pixels having values greater than the threshold are labeled local anomalies (notice that the SemiP detector will be used in both local and quasi-global versions).

As described in Section 2.1, the set of *14* ground vehicles near the treeline in Cube 1 (Figure 1) constitutes the target set, but, since anomaly detectors are not designed to detect a particular target set, the meaning of false alarms is not absolutely clear in this context. For instance, a genuine local anomaly not belonging to the target set would be incorrectly labeled as a false alarm. Nevertheless, it does add some value to our analysis to compare detections of targets versus nontargets among the different algorithmic ap-proaches.

Figure 15 shows the ROC curves produced by the output of the five algorithms testing Cube 1 for local or scene

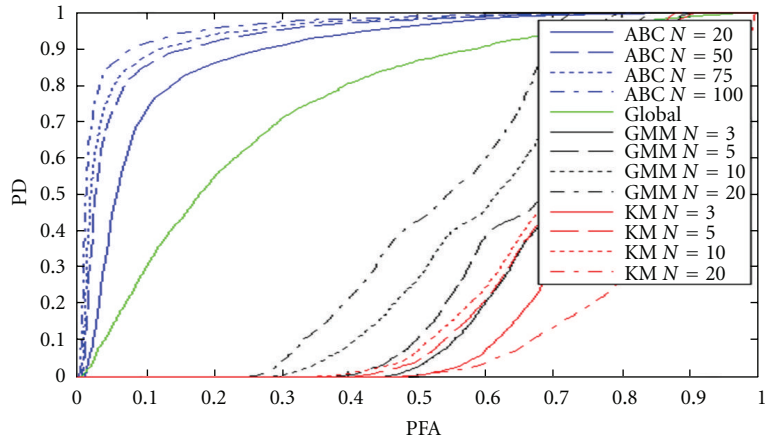


Figure 13: ROC curve performance testing QG-SemiP (ABC), global RX (global), k means (KM), and GMM detectors on scene shown in Figure 10.

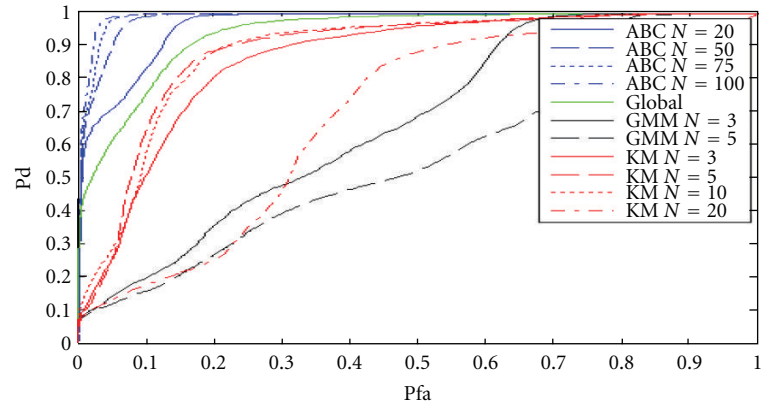


Figure 14: ROC curve performance testing QG-SemiP (ABC), global RX (global), k means (KM), and GMM detectors on scene shown in Figure 11.

anomalies. Detection performance was measured using the ground truth information for the HYDICE imagery. We used the coordinates of all the rectangular target regions and their shadows to represent the ground truth target set. As it can be readily assessed from Figure 15, the local SemiP and QG-SemiP anomaly detection approaches outperform the other three techniques on the tested scene, followed by KRX's performance. The significant display of performance shown in Figure 15 by the semiparamtric algorithms can be further appreciated by taking a closer look at the output surfaces produced by all five detectors, as they show evidences of candidate local and scene anomalies. The intensity of local peaks shown in Figure 16 reflects the strength of the detectors' evidences. Figure 16 shows that the surfaces produced by FLD, RX, and KRX detectors are expected to be significantly more sensitive (producing *false alarms*) to changing cutoff thresholds then the ones produced by the local SemiP and QG-SemiP approaches.

Spatial areas shown in Cube 1 containing the presence of clutter mixtures (e.g., edge of terrain, edge of tree clusters), where FLD, RX, and KRX yield a high number of potential false alarms (false anomalies), are suppressed by the SemiP approach, local, and quasi-global. The reason for this suppression is that, as part of the overall comparison strategy, the semiparametric model combines both the test and reference samples in order to estimate the underlying PDF of the reference sample, as shown by simulation earlier. As such, the semiparametric test statistic ensures that a component of a mixture (e.g., shadow) will not be detected as a local anomaly when it is tested against the mixture itself (e.g., trees and shadows). Performances of such cases are represented in Figure 16 in the form of softer anomalies (significantly less-dominant peaks) in the local SemiP's and QG-SemiP's output surfaces. It is evident from Figure 16 that both versions of the SemiP detectors perform remarkably well accentuating the presence of dominant local or scene anomalies (e.g., targets) from softer anomalies (e.g., transitions of distinct regions). The natural ability of the semiparametric model to manage spectral mixture can best explain the local SemiP and QG-SemiP superior ROC-curve performances shown in Figure 15.

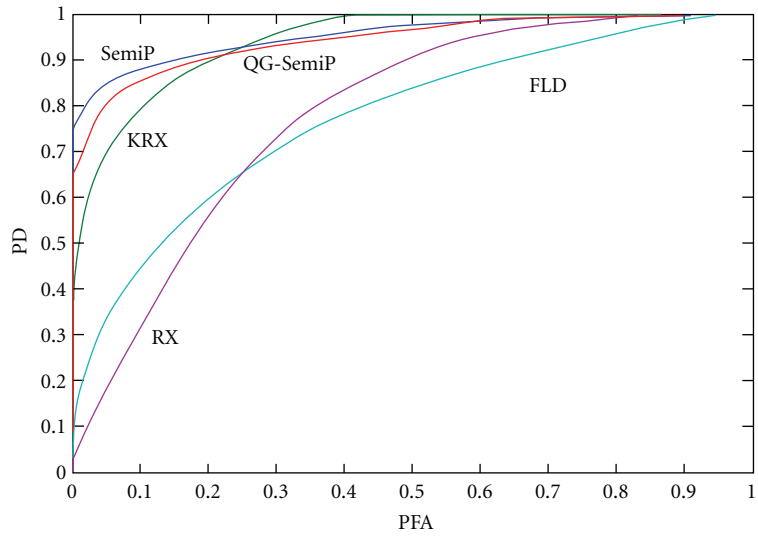


Figure 15: ROC curves for the nadir looking imagery (Cube 1) data scene shown in Figure 1. The semiparametric method, being used in both local and quasiglobal versions, are noticeably less sensitive to different cutoff thresholds; their performances almost achieve an ideal ROC curve for that scene, that is, a step function starting at point (PFA = 0, PD = 1).

5. Conclusion

This paper introduced an adaptive quasi-global, semiparametric anomaly detection algorithm and evaluated the approach using real HS imagery, where targets (manmade objects) are found in difficult natural clutter backgrounds viewed from two different perspectives—nadir and forward looking. The algorithm features a semiparametric test statistic, which has been recently found to be robust against spectral mixture, and applies random sampling of the imagery to test for anomalies. Random sampling and testing are repeated a number of times in order to mitigate the probability of contamination (spectral samples of candidate targets being sampled and used as clutter reference samples). As such, the algorithm requires no prior information (e.g., a spectral library of the clutter background and/or target, target size, or shape). The algorithm is free from training requirements. We found that the semiparametric model has a natural ability to handle mixtures, although an exhaustive survey of the literature reveals that this fact has never been noticed before by practitioners of the model in other fields of study (e.g., biotechnology).

The repeated sampling and testing procedure was modeled by the binomial family of distributions, where the only target related parameter q (the upper bound proportion of target pixels potentially covering the spatial area of the imagery) is robust—thus invariant—to different sizes and shapes of targets, number of targets present in the scene, target aspect angle, partially obscured targets, or sensor viewing perspective. The paper also discussed how other parameters (N, the total number of sampled data blocks to take from the HS imagery, and M, the number of process repetitions) can be automatically set using a simple guideline.

The algorithm fuses intermediate results through the application of minimum order statistics and logic *OR* operation. The paper presented the algorithm's asymptotic behavior under the null hypothesis, when either the null or the alternative hypothesis is true, for the two-sample test case and the multi-sample test case, where the cumulative probability of the algorithm making mistakes was derived. Using the cumulative probability, a cutoff threshold can be determined from a user specified probability of error. This is a desired feature giving the user some control of predicted errors.

The inherent challenges in adequately modeling spectral variability of targets, while managing spectral variability between targets, have prompted the introduction of a more robust family of algorithms that attempts to detect targets as being anomalous to an unknown natural clutter background. This paper presented an approach that inherently addresses many of the problems and issues pertaining to anomaly detection applications. The advantages of using anomaly detection algorithms have been discussed in this paper; however, it should be emphasized that targets would not be detected as specific manmade objects; they would be detected as anomalies. This is a limitation.

Appendix

Asymptotic Behavior of the Quasiglobal Semiparametric Algorithm

This section shows an analytical asymptotic analysis of the quasi-global semiparametric anomaly detection algorithm. In particular, we would like to investigate the algorithm's cumulative probability of rejecting the null hypothesis, when either the null or the alternative hypothesis is true; that is, the algorithm's probability of making mistakes. We will look first at the asymptotic behavior of the two sample test case, where

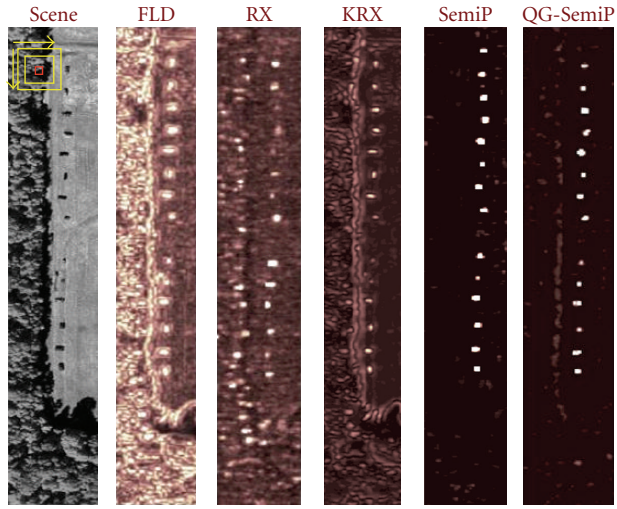


Figure 16: Detection algorithms' output surfaces for Cube 1 (far left). The intensity of local peaks reflects the strength of evidences as *seen* by different anomaly detection approaches. Boundary issues were ignored in this test; surfaces were magnified to about the size of the original image only for the purpose of visual comparison. FLD, RX, KRX, and SemiP performed local anomaly detection by testing spectra within a testing window (red square shown in the scene display—far left, top) to spectra surrounding the testing window (outer window bounded by yellow lines). QG-SemiP performed global anomaly detection, as presented in this paper.

the detector tests a reference sample against a test sample; then we will look at the multisample test case, which uses order statistic to reduce N results to a single result.

Two-Sample Test (2ST). In order to declare an anomaly, a decision threshold T must be chosen; hopefully, separating without errors the null and the alternative hypotheses in some decision space. In the paper's context, values of Z_{SemiP} in (23) greater than T are automatically labeled as anomalies. And since, in real world applications, decision errors are unavoidable, we would like to know whether the asymptotic behaviors of these errors can be determined, and whether they are favorable. The *power function* can provide those answers for the two-sample test (2ST) case. Using (13), the power function of the semiparmetric test statistic for the two-sample test (2ST) case is as follows:

$$\psi(\beta) = \begin{cases} P_{\beta=0}(Z_{\text{SemiP}} > T) \\ P_{\beta\neq 0}(Z_{\text{SemiP}} > T). \end{cases} \tag{A.1}$$

In essence, the power function ψ yields the cumulative probability P of rejecting the null hypothesis H_0 in (13) when either H_0 ($\beta = 0$) or the alternative H_1 ($\beta \neq 0$) is true. This rejection region is $Z_{\text{SemiP}} > T$, where Z_{SemiP} is defined in (23) and T is the decision threshold. Notice in (A.1) that ψ under H_0 corresponds to the well known type I error, or the probability of missing (i.e., the probability of rejecting H_0, given that H_0 is true) and that ψ under H_1 corresponds to the complement of the type II error, or probability of false alarms (i.e., one minus the probability of rejecting H_1, given that H_1 is true). The type I and type II errors constitute the only error types encountered in the context of our discussion. In the ideal case, ψ yields 0.0 when H_0 is true and 1.0 when H_1 is true. Except in trivial situations, this ideal cannot be attained. So, one of our goals is to show that ψ tends in probability to ε (a scalar controlled by the user), when H_0 is true, and that ψ tends in probability to 1.0, when the alternative hypothesis H_1 is true.

If H_0 in (13) is true, the semiparametric detector has the asymptotic behavior shown in (23), and the type-I error is readily obtained by the following:

$$P_{\beta=0}(Z_{\text{SemiP}} > T) \xrightarrow[n\to\infty]{} P(\xi > T) = \varepsilon, \tag{A.2}$$

where ξ is a Chi square distributed random variable with 1 degree of freedom, Z_{SemiP} as defined in (23), ε and T are nonnegative real values.

Setting $\psi(\beta) = P_{\beta=0}(Z_{\text{SemiP}} > T)$, ψ is indeed an asymptotic size ε test, which is controlled by the user.

Now consider an alternative parameter value, such that $\beta \neq 0$, and let ζ^2 be the true variance in (21) and $\hat{\zeta}^2$ be a estimator of ζ^2, or

$$\begin{aligned} \zeta^2 &= \frac{\rho^{-1}(1+\rho)^2}{\nu^2}, \\ \hat{\zeta}^2 &= \frac{\rho^{-1}(1+\rho)^2}{\hat{\nu}^2}, \end{aligned} \tag{A.3}$$

where $\rho = n_1/n_0$ and $\hat{\nu}^2$ is defined in (22).

From (23), we can now write

$$Z_{\text{SemiP}} = \left\{ \left[\underbrace{\left(\frac{\hat{\beta}-\beta}{\sqrt{\zeta^2/n}} \right)}_{A} + \underbrace{\left(\frac{\beta}{\sqrt{\zeta^2/n}} \right)}_{B} \right]^2 \underbrace{\left(\frac{\zeta^2}{\hat{\zeta}^2} \right)}_{C} \right\}. \tag{A.4}$$

Notice in (A.4) that, as n_1 and n_0 go to $+\infty$, $\rho = n_1/n_0$ tends to 1 and ζ^2 tends to ζ^2 for $\nu^2 > 0$. According to (21),

the term A in (A.4) converges in distribution to the standard Normal, $N(0,1)$, as n_1 and n_0 (hence, n) go to $+\infty$, no matter what the values of β or ζ^2 are. Note also that the term B converges to $+\infty$ or $-\infty$ in probability, as n goes to $+\infty$, depending on whether β is positive or negative. Since the estimator $\hat{\tilde{g}}_0$ in (38) has been shown [21] to be biased, as n_1 and n_0 go to $+\infty$, $\hat{\zeta}^2$ tends to a constant, see definition of $\hat{v}^2$ in (22); leading the term C also to a constant, no matter what values of ζ^2 is. Thus, Z_{SemiP} converges to $+\infty$ in probability and

$$P_{\beta\neq 0}(\text{reject } H_0) = P_{\beta\neq 0}(Z_{\text{SemiP}} > T) \xrightarrow[n\to\infty]{} 1. \tag{A.5}$$

In this way, the semiparametric test statistic shown in (23) also has the properties of asymptotic size ε and asymptotic power 1, which is highly desired.

Multisample Test (mST). The discussions in Sections 2.3 and 4.1 ensure that the output of the semiparametric 2ST has two asymptotic outcomes: $Z_{\text{SemiP}} \xrightarrow[n\to\infty]{} \chi_1^2$ in distribution, if H_0 in (13) is true, or $Z_{\text{SemiP}} \xrightarrow[n\to\infty]{} +\infty$ in probability, if H_1 is true.

Using results leading to (13) and the order statistics $\tilde{Z}_{\text{SemiP}}^{(ij)} = \min_{1\le f\le N} Z_{\text{SemiP}}^{(ij)(f)}$ in (34), for the multi-sample test (mST) case, we propose the following null H_2 and alternative H_3 hypotheses

$$\begin{aligned} H_2 &: \text{at least one } \left\{\beta^{(f)}\right\}_{f=1}^{N} = 0, \\ H_3 &: \text{all } \left\{\beta^{(f)}\right\}_{f=1}^{N} \neq 0, \end{aligned} \tag{A.6}$$

where $\beta^{(f)}$ is the true logistic function parameter—see (12)—corresponding to the fth randomly selected block of data from a HS data cube $\mathbf{X}$.

Now, consider the following: for a given spatial location in $\mathbf{X}$, let $Z_{\text{SemiP}}^{(f)}$ be the semiparametric detector's output for the fth block of data, and assume, without loss of generality, that each one of the first L outputs in the independent sequence of results ($1 \le L \le N$, where N is the total number of randomly selected blocks of data in $\mathbf{X}$) has the asymptotic chi-square behavior shown in (23), and that each one of the remainder results has the asymptotic behavior tending to $+\infty$, or

$$\tilde{Z}_{\text{SemiP}}\Big|_{H_2} = \min\left\{\begin{array}{l} Z_{\text{SemiP}}^{(1)} \xrightarrow{n\to\infty} \chi_1^2 \\ Z_{\text{SemiP}}^{(2)} \xrightarrow{n\to\infty} \chi_1^2 \\ \vdots \\ Z_{\text{SemiP}}^{(L)} \xrightarrow{n\to\infty} \chi_1^2 \\ Z_{\text{SemiP}}^{(L+1)} \xrightarrow{n\to\infty} +\infty \\ \vdots \\ Z_{\text{SemiP}}^{(N)} \xrightarrow{n\to\infty} +\infty \end{array}\right\}. \tag{A.7}$$

Under the null hypothesis H_2 in (A.6), $\tilde{Z}_{\text{SemiP}}$ in (A.7) is bounded because, as $n \to \infty$, $\tilde{Z}_{\text{SemiP}}$ will converge in law to the distribution of the lowest order statistics. (The order statistics of a random sample $Z_1,\ldots,Z_N$ are the sample values placed in ascending order. They are often denoted by $Z_{(1)},\ldots,Z_{(N)}$, where $Z_{(1)} = \min_{f\le f\le N} X_f$ and $Z_{(N)} = \max_{f\le f\le N} X_f$.) To attain an approximation of the type I error using (A.7), we first ignore all the components in (A.7) that converge in probability to $+\infty$, then we consider only the components that converge in distribution, that is, $(Z_{\text{SemiP}}^{(1)}, Z_{\text{SemiP}}^{(2)},\ldots,Z_{\text{SemiP}}^{(L)})$. The distribution of

$$Z_{\text{SemiP}(1)} = \min_{f\le f\le L} Z_{\text{SemiP}}^{(f)} \tag{A.8}$$

from the culled sequence can be attained with the application of Theorem 1.

Theorem 1. *Let $X_{(1)},\ldots,X_{(n)}$ denote the order statistics of a random sample from a continuous population with cumulative distribution function (cdf) $F(x)$ and pdf $f(x)$. Then the pdf of $X_{(j)}$ is*

$$\tilde{f}(x) = \frac{n!}{(j-1)!(n-j)!} f(x)[F(x)]^{j-1}[1-F(x)]^{n-j}, \tag{A.9}$$

where $(\cdot)!$ denotes the factorial operator.

The proof of Theorem 1 can be found in [28].

Using Theorem 1 with $j = 1$ and $n = L$, the pdf of $\tilde{Z}_{\text{SemiP}} = Z_{\text{SemiP}(1)}$ under H_2 in (A.6) is

$$\tilde{g}(z) = Lg(z)[1-G(z)]^{L-1}, \tag{A.10}$$

where $g(z)$ is the Chi square pdf with 1 degree of freedom and $G(z)$ is the corresponding cdf.

Denote the kth logistic function parameter $\beta^{(k)}$ in (A.6) to correspond to the one of the minimum order statistics $Z_{\text{SemiP}(1)}$. As the sample size increases in $Z_{\text{SemiP}(1)}$, that is, $n = n^{(k)} \to \infty$, the probability of rejecting the null hypothesis H_2 in (A.6), when $\beta^{(k)} = 0$, converges to

$$\hat{\psi}\left[\beta^{(k)}\right] = P_{\beta^{(k)}=0}\left(\tilde{Z}_{\text{SemiP}} > T_1\right) \xrightarrow[n\to\infty]{} P(\xi > T_1) = \varepsilon_1, \tag{A.11}$$

where ξ is a random variable distributed by $\tilde{g}(z)$, as defined in (A.10); T_1 a nonnegative real value; ε_1 is a positive real value, controlled by the user.

The variable $\hat{\psi}$ in (A.11) is the type I error under H_2 for the mST case, and it is indeed an asymptotically size ε_1 test.

Now consider the alternative hypothesis H_3 in (A.6), where all $\{\beta^{(f)}\}_{f=1}^{N} \neq 0$. From (A.7) one can write

$$\tilde{Z}_{\text{SemiP}}\Big|_{H_3} = \min\left\{\begin{array}{l} Z_{\text{SemiP}}^{(1)} \xrightarrow{n\to\infty} +\infty \\ \vdots \\ Z_{\text{SemiP}}^{(N)} \xrightarrow{n\to\infty} +\infty \end{array}\right\}. \tag{A.12}$$

From (A.12), $\tilde{Z}_{\text{SemiP}}$ will converge in probability to $+\infty$, hence, the probability P of rejecting the null hypothesis H_2, given that H_3 is true, tends to 1.0, or

$$P_{\beta^{(k)}\neq 0}(\text{rejecting } H_2) = P_{\beta^{(k)}\neq 0}(Z_{\text{SemiP}(1)} > T_1) \xrightarrow[n\to\infty]{} 1. \tag{A.13}$$

In this way, the quasi-global semiparametric anomaly detection algorithm has the desired properties of asymptotic size ε_1, which is controlled by the user, and asymptotic power 1.0.

Notations

Bold upper case letters may denote a data cube (3 dimensions) or a matrix (e.g., $\mathbf{X}, \mathbf{W}_0$), where the specific case in use is defined in the text.

Lower cases letters denote vectors (bold) or sequences (not bold) (e.g., $\mathbf{x}$, $x_1 = (x_{11}, \ldots, x_{1n_1})$).

PDF or pdf: Probability density function.

IID or iid: Independent and identically distributed.

$\mathbf{R}^{d_1 \times d_2}$-$d_1$ by d_2 dimensional set of real numbers.

$\in$ denotes set belonging

HS: Hyperspectral

$\mathbf{X}$: Observed hyperspectral data cube with dimensions of R rows, C columns, and K bands.

$\mathbf{x}_{rc}$: Observed spectrum contained in $\mathbf{X}$ with spatial indexes $(r = 1, \ldots, R)$ and $(c = 1, \ldots, C)$.

A slideing $n \times n$ window is a 3-dim subset of $\mathbf{X}$, containing $n \cdot n$ spectra.

$\mathbf{W}_1 \in \mathbf{R}^{K \times n_1}$: A matrix representing a hyperspectral sample being observed from a sliding $n \times n$ window in $\mathbf{X}$ (also referred to herein as a test sample); this is a rearranged version of a 3-dim subdata cube, where vertical direction is the dimension K of bands and the horizontal direction is the dimension of countable samples with sample size $n_1 = n^2$.

$\mathbf{y}_{1h} \in \mathbf{R}^K$ $(h = 1, \ldots, n_1)$: An observed spectrum of K bands contained in $\mathbf{W}_1$.

$g_1(\mathbf{y}|\boldsymbol{\theta})$: Multivariate joint PDF of $\mathbf{y}_{11}, \ldots, \mathbf{y}_{1n_1}$

$\mathbf{W}_0 \in \mathbf{R}^{K \times n_0}$: A matrix representing a hyperspectral sample labled as a reference sample of sample size n_0, having the same specifications of $\mathbf{W}_1$ except perhaps the sample size (n_0 may be different from n_1).

$\mathbf{y}_{0h} \in \mathbf{R}^K (h = 1, \ldots, n_0)$: An observed spectrum of K bands contained in $\mathbf{W}_0$.

$g_0(\mathbf{y}|\boldsymbol{\theta})$: Multivariate joint PDF of $\mathbf{y}_{01}, \ldots, \mathbf{y}_{0n_0}$.

$\nabla_0 \in \mathbf{R}^{(K-1) \times n_0}$: Output from differentiating $\mathbf{W}_0$.

$\nabla_1 \in \mathbf{R}^{(K-1) \times n_1}$: Output from differentiating $\mathbf{W}_1$.

$x_0 = (x_{01}, x_{02}, \ldots, x_{0n_0})$: Univariate sequence used as the reference sample.

$x_1 = (x_{11}, x_{12}, \ldots, x_{1n_1})$: Univariate sequence used as the test sample.

$g_0(x)$: Univariate PDF labeled as reference.

$g_1(x)$: Univariate PDF labeled as test

$t = (x_{11}, \ldots, x_{1n_1}, x_{01}, \ldots, x_{0n_0}) \equiv (t_1, \ldots, t_n)$: Sample concatenation, combining samples.

$\widetilde{g}_0(t)$: estimator of $g_0(x)$.

$\widehat{\widetilde{g}}_0(t)$: estimator of $\widetilde{g}_0(t)$.

H_i: Statistical hypothesis i.

Z_{SemiP}: Univariate output of the semiparametric detector.

P: Cumulative probability function, using the binomial family of PDFs as base PDF.

N: The number of randomly selected blocks of data used to represent background objects.

Trial (or process): Take N random blocks of data from the data cube under test, label them as reference background objects, and—using the semiparametric detector and a sliding window across the data cube—test the entire data cube against the same set of N reference blocks of data.

M: The number of trials (repetitions or parallel processes).

$P_g(m \geq 1)$: Cumulative probability of contamination, that is, probability of labeling a randomly selected target sample as a background sample at the gth trial or process.

$\widetilde{P}$: Overall cumulative probability that all of the trials (or processes) are contaminated with at least a contaminated sample from the randomly selected set of reference samples $\{\mathbf{W}_0^{(f)}\}_{f=1}^N$.

$\widetilde{Z}_{\text{SemiP}}^{(ij)}$: Retains the lowest order statistics from a set of N semiparametric detector's results.

$\widetilde{\mathbf{Z}}_{\text{SemiP}}^{(g)}$: The 2-dimensional output surface, consisting of $\widetilde{Z}_{\text{SemiP}}^{(ij)}$ values, from the gth trial.

T_1: Adaptive cutoff threshold for $\widetilde{\mathbf{Z}}_{\text{SemiP}}^{(g)}$.

$\mathbf{Z}_{\text{SemiP}}$: A final binary 2-dimensional output surface of the quasi-global semiparametric detector.

References

[1] R. Shcowengerdt, *Remote Sensing, Models and Methods for Image Processing*, Academic, San Diego, Calif, USA, 2nd edition, 1997.

[2] D. G. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 29–43, 2002.

[3] D. Manolakis, "Taxonomy of detection algorithms for hyperspectral imaging applications," *Optical Engineering*, vol. 44, no. 6, pp. 1–11, 2005.

[4] D. G. Manolakis and G. Shaw, "Hyperspectral image processing for automatic target detection applications," *IEEE Signal Processing Magazine*, vol. 14, pp. 79–116, 2003.

[5] P. H. Suen, G. Healey, and D. Slater, "The impact of viewing geometry on material discriminability in hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 7, pp. 1352–1359, 2001.

[6] G. Healey, "Models and methods for automated material identification in hyperspectral imagery acquired under unknown illumination and atmospheric conditions," *IEEE Transactions*

on Geoscience and Remote Sensing, vol. 37, no. 6, pp. 2706–2717, 1999.

[7] X. Yu, L. E. Hoff, I. S. Reed, A. M. Chen, and L. B. Stotts, "Automatic target detection and recognition in multiband imagery: a unified ML detection and estimation approach," *IEEE Transactions on Image Processing*, vol. 6, no. 1, pp. 143–156, 1997.

[8] H. Kwon and N. M. Nasrabadi, "Kernel RX-algorithm: a nonlinear anomaly detector for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 2, pp. 388–397, 2005.

[9] S. Matteoli, M. Diani, and G. Corsini, "Different approaches for improved covariance matrix estimation in hyperspectral anomaly detection," in *Riunione Annuale dell'Associazione Gruppo Nazionale Telecomunicazioni e Tecnologie dell'Informazione*, pp. 1–8, 2009.

[10] A. Banerjee, P. Burlina, and C. Diehl, "A support vector method for anomaly detection in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 8, pp. 2282–2291, 2006.

[11] Y. Tan and J. Wang, "A support vector machine with a hybrid kernel and minimal vapnik-chervonenkis dimension," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 4, pp. 385–395, 2004.

[12] P. Gurram and H. Kwon, "Support-vector-based hyperspectral anomaly detection using optimized kernel parameters," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 6, pp. 1060–1064, 2011.

[13] S. Khazai, S. Homayouni, A. Safari, and B. Mojaradi, "Anomaly detection in hyperspectral images based on an adaptive support vector method," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 4, pp. 646–650, 2011.

[14] D. W. J. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 58–69, 2002.

[15] P. Hytla, R. C. Hardie, M. T. Eismann, and J. Meola, "Anomaly detection in hyperspectral imagery: a comparison of methods using seasonal data," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIII*, vol. 6565 of *Proceedings of SPIE*, September 2007.

[16] S. Matteoli, F. Carnesecchi, M. Diani, G. Corsini, and L. Chiarantini, "Comparative analysis of hyperspectral anomaly detection strategies on a new high spatial and spectral resolution data set," in *Image and Signal Processing for Remote Sensing XIII*, vol. 6748 of *Proceedings of SPIE*, September 2007.

[17] H. Kwon, S. Z. Der, and N. M. Nasrabadi, "Adaptive anomaly detection using subspace separation for hyperspectral imagery," *Optical Engineering*, vol. 42, no. 11, pp. 3342–3351, 2003.

[18] A. Stocker, "Stochastic expectation maximization (SEM) algorithm," in *Proceedings of the DARPA Adaptive Spectral Reconnaissance Algorithm Workshop*, 1999.

[19] P. Masson and W. Pieczynski, "SEM algorithm and unsupervised statistical segmentation of satellite images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 31, no. 3, pp. 618–633, 1993.

[20] D. Rosario, *Algorithm development for hyperspectral anomaly detection [Ph.D. dissertation]*, The University of Maryland, College Park, Md, USA, 2008.

[21] J. Qin and B. Zhang, "A goodness-of-fit test for logistic regression models based on case-control data," *Biometrika*, vol. 84, no. 3, pp. 609–618, 1997.

[22] K. Fokianos, J. Qin, B. Kedem, and D. A. Short, "Semiparametric approach to the one-way layout," *Technometrics*, vol. 43, no. 1, pp. 56–65, 2001.

[23] D. Rosario, "A semiparametric approach using the discriminant metric SAM (spectral angle mapper)," in *Automatic Target Recognition XIV*, vol. 5426 of *Proceedings of SPIE*, pp. 58–66, September 2004.

[24] B. Kedem, *Time Series Analysis by Higher Order Crossings*, IEEE Press, New York, NY, USA, 1994.

[25] E. L. Lehmann, *Testing Statistical Hypotheses*, Chapman & Hall, New York, NY, USA, 2nd edition, 1993.

[26] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the Nelder-Mead simplex method in low dimensions," *SIAM Journal on Optimization*, vol. 9, no. 1, pp. 112–147, 1999.

[27] A. M. Law and W. D. Kelton, *Simulation Modeling and Analysis*, McGraw -Hill, Boston, Mass, USA, 3rd edition, 2000.

[28] G. Casella and R. L. Berger, *Statistical Inference*, Dexbury Press, Belmont, Calif, USA, 1990.