

# Efficient Detection in Hyperspectral Imagery

Susan M. Schweizer, *Member, IEEE*, and José M. F. Moura, *Fellow, IEEE*

**Abstract**—Hyperspectral sensors collect hundreds of narrow and contiguously spaced spectral bands of data. Such sensors provide fully registered high resolution spatial and spectral images that are invaluable in discriminating between man-made objects and natural clutter backgrounds. The price paid for this high resolution data is extremely large data sets, several hundred of Mbytes for a single scene, that make storage and transmission difficult, thus requiring fast onboard processing techniques to reduce the data being transmitted. Attempts to apply traditional maximum likelihood detection techniques for in-flight processing of these massive amounts of hyperspectral data suffer from two limitations: first, they neglect the spatial correlation of the clutter by treating it as spatially white noise; second, their computational cost renders them prohibitive without significant data reduction like by grouping the spectral bands into clusters, with a consequent loss of spectral resolution.

This paper presents a maximum likelihood detector that successfully confronts both problems: rather than ignoring the spatial and spectral correlations, our detector exploits them to its advantage; and it is computationally expedient, its complexity increasing only linearly with the number of spectral bands available. Our approach is based on a Gauss–Markov random field (GMRF) modeling of the clutter, which has the advantage of providing a direct parameterization of the inverse of the clutter covariance, the quantity of interest in the test statistic. We discuss in detail two alternative GMRF detectors: one based on a binary hypothesis approach, and the other on a ‘single’ hypothesis formulation. We analyze extensively with real hyperspectral imagery data (HYDICE and SEBASS) the performance of the detectors, comparing them to a benchmark detector, the RX-algorithm. Our results show that the GMRF ‘single’ hypothesis detector outperforms significantly in computational cost the RX-algorithm, while delivering noticeable detection performance improvement.

**Index Terms**—Gauss–Markov random field, hyperspectral sensor imagery, maximum-likelihood detection, ‘single’ hypothesis test.

## I. INTRODUCTION

THE use of hyperspectral sensor imagery (HSI) for automatic target detection and recognition (ATD/R) is a relatively new and exciting area of research. Hyperspectral sensors are passive sensors that simultaneously record hundreds of narrow bands from the electromagnetic spectrum, and group the bands in what is called a hyperspectral data cube. Our focus is on using hyperspectral sensor data for the detection of anomalous man-made objects in natural clutter backgrounds. Hyperspectral sensor data shows great potential for use in automatic

target detection and recognition because it provides both spatial and spectral features about the targets and backgrounds in the imagery. It has been shown that the spectral characteristics of natural clutter differ in significant ways from the spectral characteristics of man-made objects [1], thus, they can be a tremendous aid in discriminating between the two classes.

*Hyperspectral Imagery: Computational Challenges:* Hyperspectral imagery collected by airborne sensors pose the challenge of efficiently<sup>1</sup> processing the massive amount of data that results from the combination of spatial and spectral information acquired by the sensors. The hyperspectral data sets, with hundreds of spectral bands collected in a few minutes, run in the hundreds of megabytes, often in excess of gigabytes. Such high data rates necessitate the implementation of onboard processing, since they push the limits of today’s technology. Much of the current research focuses on the development of computationally efficient detection algorithms that can be used to determine regions of interest (ROIs) in the data. One idea for processing HSI data is the “directed vision concept,” where HSI data cues a high resolution electrooptical (EO) sensor, and high spatial resolution single-band image chips of only the regions of interest are transmitted to the ground, thus reducing the initially overwhelming amount of data collected by the hyperspectral sensor. The single-band chips that are transmitted down to ground stations can be evaluated further by an image analyst (IA) for final classification of the detected objects. In this directed vision processing scenario, there are two important requirements with which any successful detection algorithm must comply. First, the algorithm should be computationally efficient in order to meet the real-time nature of the onboard processing system. Second, the algorithm should perform effectively with respect to the false alarm rate, so that the image analyst receiving the detection cues is not inundated with false positive detections.

*Anomaly Detection: The GMRF Approach:* A general consideration for any detector is the type of *a priori* information that is assumed about the targets and clutter. We can consider three cases: known target model parameters (or statistics) and known clutter statistics; known target model and unknown clutter statistics; or both unknown target model and unknown clutter statistics. The third set of assumptions is the most general, and is more realistic when performing onboard processing of real hyperspectral data, since the impact of the atmosphere on spectral content, as well as the variability of target signatures, makes it difficult to *a priori* train on the data. In this paper, we present a new *anomaly* detection algorithm for hyperspectral data, i.e., a detector that assumes no prior knowledge about the target and the clutter statistics. The detector models the clutter by a Gauss–Markov random

Manuscript received March 18, 1999; revised December 15, 2000. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Mita Desai.

S. M. Schweizer was with the Electrical and Computer Engineering Department, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213-3890 USA. She is now with Mediasite, Inc., Pittsburgh, PA 15222 USA.

J. M. F. Moura is with the Electrical and Computer Engineering Department, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213-3890 USA (e-mail: moura@ece.cmu.edu).

Publisher Item Identifier S 1057-7149(01)02474-5.

<sup>1</sup>In the paper, when we use the word efficient, we are referring to simplicity in the computational complexity of the processing algorithms, not statistical efficiency in the sense of the Cramér–Rao bound [2].

field (GMRF), which presents several advantages. We often refer to the detector as the GMRF algorithm. A GMRF is a spatially and spectrally colored random field, so accounting for both spatial and spectral correlation of the clutter. Secondly, because GMRF's parameterize directly the inverse of the covariance matrix, the GMRF algorithm does not necessitate the inversion of a large data covariance matrix that hinders other maximum likelihood (ML) detector approaches. This impacts the computational complexity of the GMRF detector, which increases linearly with the number of spectral bands. This efficiency permits using as many bands as necessary when processing hyperspectral data; thus, the GMRF detector has room to adapt to the increased capabilities of future hyperspectral and ultraspectral sensors. This is a significant advantage over other maximum likelihood detection algorithms that are computationally severely limited to a small number of bands.

*Goals of the Work:* Our goal is to design an adaptive anomaly detection algorithm that is computationally efficient and exhibits a low false alarm rate with high detection probability. We develop two adaptive maximum likelihood GMRF anomaly detection algorithms: a binary hypothesis detector and a 'single' hypothesis detector. These detectors adapt to the unknown clutter statistics by using the approximate maximum likelihood estimation technique, one of three studied in [3]. We show that the GMRF adaptive 'single' hypothesis anomaly detector outperforms the GMRF binary hypothesis detector and the RX-algorithm [4]. The RX-algorithm is the benchmark anomaly detection algorithm, originally developed for multispectral imagery but that, at the time of this study, was considered a prime candidate for local anomaly detection with hyperspectral imagery. Our results illustrate that the GMRF 'single' hypothesis anomaly detector provides a significant computational advantage over the RX-algorithm, while achieving better detection performance. These characteristics led this GMRF anomaly detector to be chosen as one of the processing algorithm candidates for the U.S. Defense Advanced Research Projects Agency's (DARPA) Adaptive Spectral Reconnaissance Program (ASRP).

*Paper Survey:* In the following sections, we provide an overview of the available detection algorithms for hyperspectral imagery and highlight the need for a new computationally efficient algorithm geared toward true hyperspectral data. In Section III we present, in detail, the new modeling framework for hyperspectral clutter backgrounds, and, in Section IV, use the model to develop both binary and 'single' hypothesis formulations of the maximum likelihood detector. In Section V, using real hyperspectral sensor imagery, HYDICE and SEBASS, we show the significant computational advantage and the improved detection performance provided by the GMRF detector algorithm over the RX-algorithm. Finally, Section VI concludes the paper.

## II. LITERATURE SURVEY

Due to the large amounts of data that are collected with hyperspectral sensors much of the prior work has focused strictly on compression of the data sets for storage and transmission. More recently, work has been published in the context of detection and classification. Much of it has actually emphasized applying algorithms intended for multispectral data to hyperspectral data

by reducing the number of spectral bands used for processing. These algorithms fall into two basic categories: spectral-only, and spatial-spectral algorithms. The spectral-only algorithms almost all rely on a known spectral signature for the target or targets of interest. Basically, they are classification rather than detection algorithms. Algorithms that fall into this category are the spectral matched filter [5], the spectral angle mapper [6], and linear mixture models [7]–[9]. For a good overview of spectral-only algorithms see [10]. The main limitation of these spectral-only algorithms is that, in addition to ignoring the available spatial information, they require a known target signature. Reliable target signatures are difficult to ascertain due to variations in the target signature that result from atmospheric and illumination effects.

Principal component analysis (PCA), related to the Karhunen-Loève or Hotelling transform, is most often used prior to another detection or classification algorithm for purposes of reducing the dimensionality of the hyperspectral data sets, thus making the applied detection and classification algorithms more efficient computationally. However, principal component analysis is, itself, a computationally undesirable task. In addition, the reduction of redundant information with principal component analysis is based on reconstructing the data using a subset of the principal components. Most often, the components used for reconstruction are those associated with the largest eigenvalues. Researchers have shown [5], however, that components associated with lower order eigenvalues often contain important features for target discrimination. Thus, there is ambiguity as to what are the "appropriate" principal components to use for data reduction.

We are interested in *anomaly* detection algorithms. By definition, anomaly detection algorithms apply when there is no known target spectral signature, [10]. Basically, in our problem, we are simply attempting to locate anything that displays different spatial and/or spectral characteristics from its surroundings. This leaves us with the class of algorithms that use both spatial and spectral features from the hyperspectral imagery. In general, spatial-spectral algorithms can be further divided into local anomaly and global anomaly detectors.

### A. Local Anomaly Detectors

Local anomaly detectors process small windows of the HSI in order to compare the spatial and spectral properties of the centrally located pixels in the window (*target region*) with the properties of the perimeter pixels (*clutter region*). Those pixels that are spatially-spectrally different from their surrounding backgrounds are considered detections.

In [4], an algorithm, commonly referred to as the RX-algorithm, is derived. This is the benchmark anomaly detection algorithm for multispectral data, which, in contrast with HSI, is characterized by less than 20 spectral bands. The RX-algorithm is a maximum likelihood (ML) anomaly detection procedure that simplifies the clutter to being spatially white. The RX-algorithm uses a binary hypothesis approach to detection, and implements a generalized likelihood ratio test (GLRT).<sup>2</sup> Evalua-

<sup>2</sup>A GLRT replaces unknown parameters in the likelihood ratio by their maximum likelihood estimates.

tion of the ML-detection statistic requires full spectral sample covariance matrices to be estimated and then inverted, or the evaluation of their determinants. Extending application of the RX-algorithm from multispectral to hyperspectral imagery suffers from two major limitations. First, the clutter model implemented in the RX-algorithm is restricted to being spatially uncorrelated, or spatially white. This model neglects the potentially valuable spatial correlation information of the clutter. The second limitation is its computational cost arising from the expensive inversion or determinant evaluation of the covariance matrix of the HSI data under each of the hypotheses. Even under the simplifying spatially white clutter assumption, the computational complexity of this detector increases as  $N_b^3$ , where  $N_b$  is the number of spectral bands. For HSI, where the number of spectral bands  $N_b$  runs into the hundreds, the RX-algorithm rapidly becomes unfeasible. However, due to the algorithm success with multispectral data, it has been incorporated into the DARPA ASR program (see end of Section I) as one of the baseline processing algorithms. To run the RX-algorithm on the hyperspectral data collected by ASRP, either a subset of bands is used, or the bands are aggregated.

In [11], an adaptive spatial/spectral detection method is presented in which it is originally assumed that the clutter is fully spatially and spectrally correlated. However, in evaluating the spatial-spectral covariance matrix the cross-covariance terms are neglected in order to improve the computational cost of the algorithm. By eliminating these cross-covariance terms, the correlation between bands, i.e., the spectral correlation, is basically ignored. Computationally, the algorithm in [11] suffers from the same problem as the RX-algorithm when applied to hyperspectral imagery: It requires taking the inverse of a spectral covariance matrix that has dimensions equal to the number of spectral bands used for processing.

### B. Global Anomaly Detectors

In global anomaly detection, the image scene is first segmented into its constituent classes, then detection is achieved by determining the outliers of these classes. In general, the algorithms vary in the method of segmentation, but tend to use ML-detection once the classes are determined.

One approach to global anomaly detection, which has been incorporated into the DARPA's ASR program, is stochastic expectation maximization (SEM) coupled with ML-detection [12]. This algorithm uses the stochastic expectation maximization clustering algorithm, as presented in [13], as a preprocessing stage to the detector. The number of classes is assumed to be known *a priori*. The clustering algorithm models each class with an ensemble of Gaussian random variables, using only spectral information. Basically, once the image has been segmented, the RX-algorithm, discussed in Section II.A, is used on each class to determine anomalous pixels. A similar approach using  $K$ -means clustering in conjunction with the RX-algorithm is presented in [14]. Although the stochastic expectation maximization and the  $K$ -means algorithms both provide performance improvements over direct application of the RX-algorithm, they are not only challenged computationally due to the need to invert a large data covariance matrix,

but they are limited by a need to know the number of classes in the scene.

The majority of global anomaly detectors employ two-dimensional (2-D) Markov random field (MRF) modeling in order to incorporate spatial features into the segmentation process, since MRFs have been proven to be quite powerful models in the classification of 2-D images. For instance, the Bayesian clustering algorithm presented in [14] uses causal MRF's to incorporate contextual information, and is developed for the purpose of subpixel anomaly detection. Contextual information refers to *a priori* knowledge that adjacent pixels tend to belong to the same class. Although it is a simple matter to compute the probability of each pixel with respect to its classification once the model has been formed, formulation of the model requires that the Hotelling Transform [15], [16] be applied to the data. For hyperspectral data, evaluation of the Hotelling transform is in itself significantly challenging computationally.

The algorithms described in this section represent the most prominent methods for detection being applied to hyperspectral imagery today. The algorithm that has received the most attention in the hyperspectral detection community is the local anomaly detection RX-algorithm, which, as mentioned before, is the benchmark algorithm for multispectral data. As discussed in Section II-B, several of the global detection algorithms have incorporated the RX-algorithm for use after segmentation. However, the RX-algorithm is limited by its assumption of spatially uncorrelated clutter, and, like most of the algorithms described above, it involves computationally demanding matrix inversions (or determinant evaluations). Application of the RX-algorithm to hyperspectral data can only be made possible by preprocessing the data to reduce the number of spectral bands.

There is a need for computationally efficient detectors for hyperspectral sensors that can jointly process all the available spectral bands, and that can exploit simultaneously the spatial and spectral correlation properties of the clutter. Our GMRF approach to anomaly detection addresses both of these concerns. We describe the GMRF detector in the remainder of the paper, and we use the RX-algorithm as the baseline to which we contrast the GMRF algorithm.

### III. GMRF CLUTTER MODEL

We capture the highly correlated spatial and spectral nature of the background clutter in hyperspectral sensor imagery using a noncausal, GMRF. GMRF models are desirable because they possess two important properties that are intrinsic to most spatial phenomenon: noncausality and Markovianity. Noncausality refers to the notion that the field at any pixel is influenced by the field in all directions around it; there is no preferred direction of dependence. Markovianity is the statistical formalization of the notion of locality, i.e., that the field at a pixel is regressed on the values of the field at neighboring pixels. The GMRF model that we implement extends these properties to a three-dimensional (3-D) field that is appropriate for hyperspectral imagery. As first suggested in [17], we assume that the Gaussian process describing the dominant image background has a slowly varying

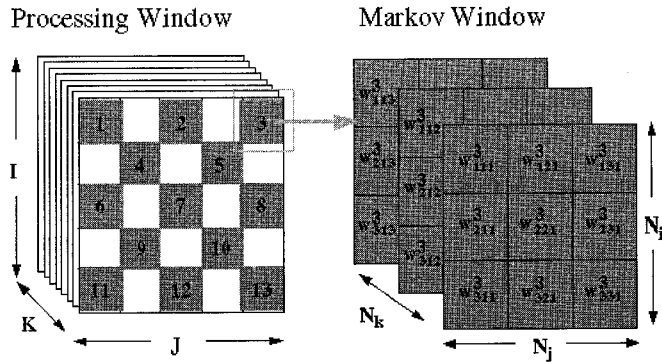


Fig. 1. Sectioning of the hyperspectral image cube. Processing is done on small homogeneous windows of data (left). The processing window is further divided into a set of subcubes, called Markov windows. Within each Markov window, a 3-D finite lattice field is defined (right).

covariance structure. Consequently, processing is done on sub-blocks of data for which the clutter is assumed to be statistically homogeneous. In this section, we develop, in detail, the 3-D GMRF clutter model.

#### A. Notation

We fix notation. Considering the hyperspectral cube as a 3-D finite lattice, the intensity at each pixel is referenced by the variable  $w$  and three subindices  $i, j$ , and  $k$ , which indicate the spatial location and the particular spectral band in which the pixel lies, see Fig. 1. Processing is done on small regions of the data set in which the total number of rows, columns, and spectral bands are represented by the variables  $I, J$ , and  $K$ , respectively. The spatial dimensions,  $I$  and  $J$ , are chosen to validate the assumption of homogeneous clutter. We divide the  $I \times J \times K$  processing window lattice into subcubes of size  $N_i \times N_j \times N_k$  as shown in Fig. 1. We refer to these subcubes as Markov windows. Typically, we use all the available spectral bands,  $N_k = K$ . Pixels within the same Markov window are assumed to have significant spatial correlation with each other, while the spatial correlation between pixels of one Markov window and pixels of another Markov window is assumed to be negligible. This assumption allows us to use the different Markov windows as independent realizations of the clutter field, which is important to provide statistical significance when estimating the field parameters as we will discuss later in the paper. The pixels in a Markov window are compiled in vector form by the vec operator explained below. As a result, the processing window is divided into a set of  $n$  independent data vectors. These independent vectors are distinguished by the superscript  $m$  where  $1 \leq m \leq n$ .

#### B. Kronecker Product

To handle the representation of the GMRF clutter model, we use the matrix operator known as the Kronecker product [18]. The Kronecker product, which is useful in mapping low-order matrices into high-order matrices, is mathematically defined as  $A \otimes B \equiv [a(m, n)B]$ . If  $A$  is a size  $M \times M$  matrix and  $B$  is a size  $N \times N$  matrix, the resulting matrix is size  $MN \times MN$ . The Kronecker product provides a simple means of representing certain large structured matrices. In addition, the Kronecker product

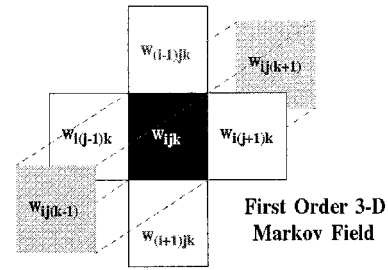


Fig. 2. Three-dimensional Markov random field. In a first order field, a pixel is dependent on its six nearest neighbors: four spatial and two spectral neighbors.

possesses several properties which enable us to derive explicit expressions for the eigenvalues, eigenvectors, and the determinant of the inverse of the clutter covariance matrix. An extensive list of these properties can be found in [18] and [15].

#### C. Vec Operator

The vec operator stacks the transposed rows of a matrix one on top of the other to form one long vector [15]. This is the lexicographic ordering of the pixels, or row major form. If the matrix is  $N_i \times N_j$ , the vector resulting from the vec operation will be of dimension  $N_i N_j \times 1$ .

#### D. Spatially-Spectrally Correlated Clutter

Within the Markov window  $m$ ,  $w_{ijk}^m$ ,  $1 \leq i \leq N_i$ ,  $1 \leq j \leq N_j$ ,  $1 \leq k \leq N_k$  represents the 3-D finite lattice field modeling the clutter described by an extension of the minimum mean square error (MMSE) representation of Woods [19]

$$\begin{aligned} w_{ijk}^m = & \beta_h \left( w_{i(j-1)k}^m + w_{i(j+1)k}^m \right) \\ & + \beta_v \left( w_{(i-1)jk}^m + w_{(i+1)jk}^m \right) \\ & + \beta_s \left( w_{ij(k-1)}^m + w_{ij(k+1)}^m \right) + \epsilon_{ijk}^m. \end{aligned} \quad (1)$$

The parameters  $\beta_h, \beta_v$ , and  $\beta_s$  are the predictor coefficients for the spatial and spectral dimensions, respectively, and  $\epsilon_{ijk}^m$  is the prediction error that has a particular correlation structure discussed below. We are assuming a first order 3-D Markov model with zero Dirichlet boundary conditions. It is assumed, without loss of generality, that the clutter is zero mean. In practice, the spatially varying mean is locally estimated and then removed from the data. In a first order 3-D Markov model, the intensity of a pixel is described in terms of its six nearest neighbors: four spatial and two spectral neighbors, see Fig. 2. Higher order Markov models can be similarly defined. For the sake of simplicity, we will restrict the discussion to first order fields.

We get a clutter vector,  $\underline{W}_m = [\underline{w}^m(1)' \dots \underline{w}^m(N_k)']'$ , by stacking the  $N_k$  vectors  $\underline{w}^m(k)$ , each of size  $N_i N_j \times 1$ . The vectors  $\underline{w}^m(k)$  result from applying the vec operator to the  $N_i N_j$  pixels in the Markov window  $m$  for each of the  $N_k$  consecutive spectral bands. The clutter vector  $\underline{W}_m$  is of size  $N_i N_j N_k \times 1$ . This process is repeated for each Markov window within the processing window leading to a set of  $n$  data vectors. An error vector,  $\underline{\epsilon}$ , is formed in the same manner by beginning with the 3-D lattice of prediction errors,  $\epsilon_{ijk}^m$ .

Incorporating the clutter vector notation into (1), the data can be compactly represented by the matrix-vector equation

$$A\mathbf{W}_m = \boldsymbol{\varepsilon}_m, \quad (2)$$

$$A = \begin{bmatrix} A_1 & A_2 & & & \\ A_2 & \ddots & \ddots & & \mathbf{0} \\ & \ddots & \ddots & \ddots & \\ \mathbf{0} & & \ddots & \ddots & A_2 \\ & & & A_2 & A_1 \end{bmatrix}. \quad (3)$$

In Kronecker notation,  $A = I_{N_k} \otimes A_1 + H_{N_k}^1 \otimes A_2$ . The matrices  $A_1, A_2, B, C$ , and  $D$  are themselves structured and defined as

$$\begin{aligned} A_1 &= I_{N_i} \otimes B + H_{N_i}^1 \otimes C, & A_2 &= I_{N_i} \otimes D, \\ B &= -\beta_h H_{N_j}^1 + I_{N_j}, & C &= -\beta_v I_{N_j}, \\ D &= -\beta_s I_{N_j}. \end{aligned}$$

The symbols  $I_{N_k}, I_{N_j}$ , and  $I_{N_i}$  are identity matrices, while  $H_{N_k}^1, H_{N_j}^1$ , and  $H_{N_i}^1$  are Toeplitz matrices which have zeros everywhere except for the first upper and lower diagonals which are composed of all 1's. The subscript denotes the size of the matrices.

The matrix  $A$ , referred to as the potential matrix, is sparse block tridiagonal, and contains the relevant information regarding the GMRF structure [20]. The subcomponents of the potential matrix also display a highly sparse structure. The blocks  $A_1$  and  $A_2$  are block tridiagonal and block diagonal, respectively. The subcomponents  $B, C$ , and  $D$  are either tridiagonal or diagonal.

It is shown in [19], [20] that the set of error vectors,  $\boldsymbol{\varepsilon}_m$ , are samples from a zero-mean Gaussian colored noise process that has covariance  $\Sigma_\varepsilon = \sigma^2 A$ . A main advantage of Gauss and Markov models is their parameterization of the inverse clutter covariance matrix,  $\Sigma_w^{-1}$ . Starting from (2), [21], the inverse clutter covariance matrix is simply a scaled version of the potential matrix,  $\Sigma_w^{-1} = (1/\sigma^2)A$ . In Kronecker notation, the inverse covariance matrix is expressed as

$$\begin{aligned} \Sigma_w^{-1} &= \frac{1}{\sigma^2} I_{N_k} \otimes I_{N_i} \otimes I_{N_j} - \frac{\beta_h}{\sigma^2} I_{N_k} \otimes I_{N_i} \otimes H_{N_j}^1 \\ &\quad - \frac{\beta_v}{\sigma^2} I_{N_k} \otimes H_{N_i}^1 \otimes I_{N_j} - \frac{\beta_s}{\sigma^2} H_{N_k}^1 \otimes I_{N_i} \otimes I_{N_j}. \end{aligned} \quad (4)$$

As an example, we present the inverse clutter covariance matrix using  $N_i = N_j = N_k = 2$ . The  $B, C$ , and  $D$  components of  $\Sigma_w^{-1}$  are

$$\begin{aligned} B &= \begin{bmatrix} 1 & -\beta_h \\ -\beta_h & 1 \end{bmatrix}, & C &= \begin{bmatrix} -\beta_v & 0 \\ 0 & -\beta_v \end{bmatrix}, \\ D &= \begin{bmatrix} -\beta_s & 0 \\ 0 & -\beta_s \end{bmatrix} \end{aligned} \quad (5)$$

and the overall inverse covariance matrix is

$$\Sigma_w^{-1} = \frac{1}{\sigma^2} \begin{bmatrix} B & C & D & \mathbf{0} \\ C & B & \mathbf{0} & D \\ D & \mathbf{0} & B & C \\ \mathbf{0} & D & C & B \end{bmatrix}. \quad (6)$$

$\Sigma_w^{-1}$  is a highly structured sparse matrix that is completely defined by the four scalar parameters,  $\sigma^2, \beta_h, \beta_v$ , and  $\beta_s$ . When using real data, these four parameters are to be estimated from the data. Although we address the issue of parameter estimation in another paper [3], we briefly present the results of that work in the next section.

### E. Approximate Maximum-Likelihood Estimates

In [3] we develop several parameter estimators for 3-D GMRFs: maximum likelihood (ML), least squares (LS), and approximate maximum likelihood (AML). The results in [3] suggest that the AML technique is a good cost/performance compromise, and we adopt it for our GMRF fully adaptive anomaly detector. In this paper, we refer the reader to [3], simply present the AML estimates and use them in our GMRF detector. The Markov parameter AML estimates, assuming spatially-spectrally correlated clutter, are shown in (7)–(9) at the bottom of the page where the quantities  $\chi_h, \chi_v$ , and  $\chi_s$ , called the one-step-ahead correlations, are

$$\begin{aligned} \chi_h &= \sum_{m=1}^n \mathbf{W}'_m \left( I_{N_k} \otimes I_{N_i} \otimes H_{N_j}^1 \right) \mathbf{W}_m \\ &= \sum_{m=1}^n \sum_{i=1}^{N_i} \sum_{j=1}^{N_j-1} \sum_{k=1}^{N_k} w_{ijk}^m w_{(j+1)k}^m \end{aligned} \quad (10)$$

$$\begin{aligned} \chi_v &= \sum_{m=1}^n \mathbf{W}'_m \left( I_{N_k} \otimes H_{N_i}^1 \otimes I_{N_j} \right) \mathbf{W}_m \\ &= \sum_{m=1}^n \sum_{i=1}^{N_i-1} \sum_{j=1}^{N_j} \sum_{k=1}^{N_k} w_{ijk}^m w_{(i+1)jk}^m \end{aligned} \quad (11)$$

$$\widehat{\beta}_h = \frac{\xi \chi_h}{|\chi_h| \cos\left(\frac{\pi}{N_j+1}\right) + |\chi_v| \cos\left(\frac{\pi}{N_i+1}\right) + \alpha |\chi_s| \cos\left(\frac{\pi}{N_k+1}\right)} \quad (7)$$

$$\widehat{\beta}_v = \frac{\xi \chi_v}{|\chi_h| \cos\left(\frac{\pi}{N_j+1}\right) + |\chi_v| \cos\left(\frac{\pi}{N_i+1}\right) + \alpha |\chi_s| \cos\left(\frac{\pi}{N_k+1}\right)} \quad (8)$$

$$\widehat{\beta}_s = \frac{\alpha \xi \chi_s}{|\chi_h| \cos\left(\frac{\pi}{N_j+1}\right) + |\chi_v| \cos\left(\frac{\pi}{N_i+1}\right) + \alpha |\chi_s| \cos\left(\frac{\pi}{N_k+1}\right)} \quad (9)$$

$$\begin{aligned}\chi_s &= \sum_{m=1}^n \underline{W}'_m (H_{N_k}^1 \otimes I_{N_i} \otimes I_{N_j}) \underline{W}_m \\ &= \sum_{m=1}^n \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \sum_{k=1}^{N_k-1} w_{ijk}^m w_{ij(k+1)}^m\end{aligned}\quad (12)$$

and  $\alpha = (N_k(N_j - 1)/N_j(N_k - 1))$ ,  $\xi = 0.5 - \delta$ , and  $\delta$  is a small number (for example,  $\delta = .01$ ) included to ensure that the parameter estimates are within the parameter space.

It can be shown that the estimate of the scaling parameter,  $\sigma^2$ , is a linear combination of the estimates  $\widehat{\beta}_h$ ,  $\widehat{\beta}_v$ , and  $\widehat{\beta}_s$

$$\widehat{\sigma}^2 = \frac{1}{nN_iN_jN_k} (S_x - 2\widehat{\beta}_h\chi_h - 2\widehat{\beta}_v\chi_v - 2\widehat{\beta}_s\chi_s) \quad (13)$$

$$\begin{aligned}S_x &= \sum_{m=1}^n \underline{W}'_m \underline{W}_m \\ &= \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \sum_{k=1}^{N_k} (w_{ijk}^m)^2.\end{aligned}\quad (14)$$

The quantity  $S_x$  is the power of the field.

#### IV. DETECTOR STRUCTURE

We present two approaches to detection: the binary hypothesis technique using a deterministic target model, and the ‘single’ hypothesis method using an implied GMRF target model. In both cases, the direct parameterization of the inverse clutter covariance matrix that is provided by our 3-D GMRF modeling framework [see (4)] reduces the final detection statistic to an evaluation of scalar quantities; no matrix inversion is required. The computational advantages of the ‘single’ hypothesis test, as well as its ability to model the target stochastically, make it a promising alternative.

We discuss the general setup of the detection problem in Section IV.A, assuming that the parameters defining the GMRF model have been estimated using the AML method summarized in Section III-E. In Section IV-B, the binary hypothesis approach is derived, and the deterministic model that is associated with the targets is presented. In Section IV-C, we present the ‘single’ hypothesis paradigm. For both detection methods, we incorporate the GMRF modeling framework and present the resulting simplifications.

##### A. Problem Formulation

Regardless of which detection approach is used, we first separate the processing window into two regions: a perimeter *clutter* region and an interior *unknown* region of dimension  $N_t \times N_t$ , as is done in [22] (see Fig. 3). Since we realistically assume that the first and second order statistics of the clutter data are unknown, this sectioning of the processing window is necessary for estimating all the parameters that will completely define the clutter statistics.

Each of the regions shown in Fig. 3 is further divided into independent Markov windows as discussed in Section III-A (see Fig. 1). By applying the vec operator to each of the Markov windows, in each region, we obtain two sets of independent vectors, referred to as the *clutter* and *observation* vectors. The *clutter* vectors,  $\underline{X}_1, \dots, \underline{X}_n$ , where  $n = ((I^2 - N_t^2)/N_iN_j)$ , are from

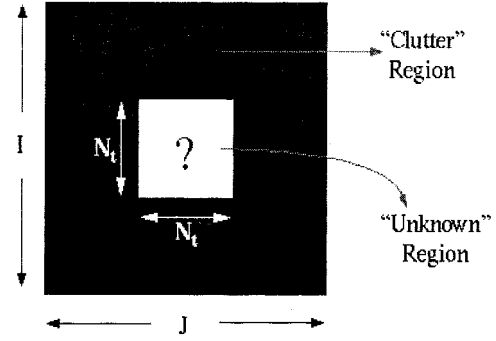


Fig. 3. Sectioning of the processing window for parameter estimation. The window is divided into 2 regions: A *clutter* region and an *unknown* region of size  $N_t \times N_t$ .

the *clutter* region, and the *observation* vectors,  $\underline{Y}_1, \dots, \underline{Y}_m$ , where  $m = (N_t^2/N_iN_j)$ , are those vectors from the *unknown* region. Recall, the dimension of the processing window is  $I \times J$ , and the dimension of the Markov windows are  $N_i \times N_j$ . Although we only need to estimate four scalar parameters, the use of all the Markov windows is necessary since our assumption about homogeneous clutter is, in general, not valid. By averaging the statistics over the entire processing window, false detections due to nonhomogeneous clutter are reduced. An entire processing window of data produces the detection statistic for one pixel; the center pixel of the window. An output statistic for every pixel in the image is obtained by moving the processing window throughout the image. Using this window formulation, we assume that, when a target is centrally located in the processing window, the detection statistic will be maximized.

The Markov parameters for the clutter background are estimated using the *clutter* vectors  $\underline{X}_p$ . The GMRF model assumes that the clutter is zero mean, thus, prior to estimating the Markov parameters, the mean  $\underline{m}_x$  of the clutter vectors is removed

$$\underline{m}_x = \frac{1}{n} \sum_{p=1}^n \underline{X}_p. \quad (15)$$

##### B. Binary Hypothesis Testing

In the binary hypothesis testing paradigm, the detection problem is formulated as a decision at each pixel as to whether only clutter is present or clutter plus target. Formally, the binary hypothesis testing problem is described by

$$\begin{aligned}H_0 : \underline{Z}_q &= \underline{Y}_q - \underline{m}_x = \underline{W}, \quad 1 \leq q \leq m \\ H_1 : \underline{Z}_q &= \underline{Y}_q - \underline{m}_x = \underline{W} + \underline{R}_q.\end{aligned}\quad (16)$$

The  $\underline{Y}_q$ 's are the independent observation vectors from the *unknown* region of the processing window, the  $\underline{Z}_q$ 's are the observation vectors with the clutter mean removed,  $\underline{W}$  is the clutter, a zero-mean GMRF process defined by the second order statistics derived in Section III-D, and the variable  $\underline{R}_q$  represents the target spatial-spectral signature associated with the observation  $q$ .

1) *Target Signature*: We use a deterministic, additive, model for the target spatial-spectral signature, similar to the one used in the RX-algorithm [4], [23]. In this model, the target signal is represented by the matrix  $R_q^{\text{tmp}} = \underline{b}_s' \underline{b}_q$ , where the vector  $\underline{b}_s$  of

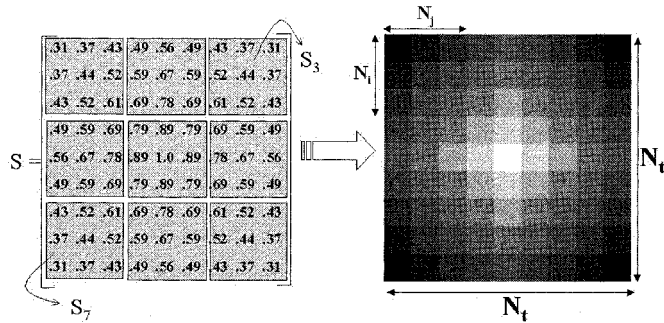


Fig. 4. Gaussian 2-D target spatial pattern.

dimension  $N_k \times 1$ , contains the additive spectral signature due to the target in each of the spectral bands, and  $\underline{s}_q = \text{vec}(S_q)$  where  $S_q$  is an  $N_i \times N_j$  matrix indicating the target spatial pattern within a Markov window. The  $S_q$  matrices are subsets of a larger  $N_t \times N_t$  matrix,  $S$ , that contains the spatial pattern describing the entire target. The  $S_q$  matrices are formed by subdividing the full target spatial pattern into blocks of the same dimension as the Markov window subdivisions made in the *unknown* region of the processing window (see Figs. 1 and 4). The interpretation of the model is that, for any pixel location,  $(i, j)$  in which target energy is present, the overall energy contribution to each spectral band, due to the target, is proportional to the value of the target spectral signature in band  $k$ ,  $b(k)$ . The amount by which the spectral signature is weighted depends on the spatial location [24].

The general target spatial pattern is assumed to be known, and, in practice, is often taken to be a 2-D Gaussian, as shown in Fig. 4 [22]. Patterns, such as the 2-D Gaussian, that do not rely on detailed shape or orientation information, only require *a priori* knowledge of the approximate size of the targets of interest. The value at each location in the pattern matrix indicates the proportion of the energy reflected, or emitted, at that location due to the target. The remainder of the energy recorded for that location is due to the reflection or emission properties of the clutter. A value of 1 indicates that the pixel location is purely target, while a value of 0.5 indicates that half the recorded energy is due to the target, and half is due to clutter.

Although we assume that the target spatial pattern is known *a priori*, the target spectral signature is assumed to be unknown, and must be estimated using data from the *unknown* region of the processing window. For computational simplicity, we use a LS approach to estimating the spectral signature,  $\hat{b}$ , see [25]. The LS estimates of the  $N_k$  elements of the spectral signature are defined by

$$\hat{b}(k) = \frac{\sum_{i=1}^{N_t} \sum_{j=1}^{N_t} S(i, j) z_{ijk}}{\sum_{i=1}^{N_t} \sum_{j=1}^{N_t} S(i, j)^2} \quad (17)$$

where  $1 \leq k \leq N_k$ . Equation (17) is a weighted average of the pixels in the *observation* window: those pixels for which the target is a stronger reflector, i.e., for which the target pattern values are closer to 1, are weighted more heavily than those pixels for which the target is a weak reflector.

The target spatial-spectral signature,  $\underline{R}_q$  is a vector of dimension  $N_i N_j N_k \times 1$ . It is obtained by applying the *vec* operator to the matrix  $\underline{R}_q^{\text{tmp}}$  as shown in (18) at the bottom of the page.

2) *Generalized Likelihood Ratio Test*: We use the GLRT to decide between hypotheses  $H_0$  and  $H_1$  at each pixel. Letting  $\underline{\theta}$  be a vector of the unknown Markov parameters, the procedure is to estimate  $\underline{\theta}$  assuming  $H_1$  is true, then estimate  $\underline{\theta}$  assuming  $H_0$  is true, and use these estimates in a likelihood ratio test as if they were correct [2]. Mathematically, the GLRT is

$$\Lambda(\underline{Z}_1, \dots, \underline{Z}_m) = \left( \frac{\max_{\underline{\theta}_1 | H_1} p(\underline{Z}_1, \dots, \underline{Z}_m | \underline{\theta}_1, H_1)}{\max_{\underline{\theta}_0 | H_0} p(\underline{Z}_1, \dots, \underline{Z}_m | \underline{\theta}_0, H_0)} \right) \underset{H_0}{\overset{H_1}{>}} \eta \quad (19)$$

where

$p(\underline{Z}_1, \dots, \underline{Z}_m | \underline{\theta}_i, H_i)$  joint probability density function (pdf) of the set of observation vectors;

$\underline{Z}_1, \dots, \underline{Z}_m$ , observation vectors with the Markov parameters estimated assuming hypothesis  $H_i$  is true;

$\eta$  nonnegative threshold that, ideally, is determined using the Neyman–Pearson Criterion [2].

The probability density functions for a set of  $m$  independent observation vectors factor as

$$\begin{aligned} H_0 : p(\underline{Z}_1, \dots, \underline{Z}_m | \underline{\theta}, H_0) &= \frac{|\Sigma_w^{-1}|^{\frac{m}{2}}}{2\pi^{\frac{mN_i N_j N_k}{2}}} \exp\left(-\frac{1}{2} \sum_{q=1}^m (\underline{Z}_q' \Sigma_w^{-1} \underline{Z}_q)\right), \\ H_1 : p(\underline{Z}_1, \dots, \underline{Z}_m | \underline{\theta}, H_1) &= \frac{|\Sigma_w^{-1}|^{\frac{m}{2}}}{2\pi^{\frac{mN_i N_j N_k}{2}}} \\ &\times \exp\left(-\frac{1}{2} \sum_{q=1}^m (\underline{Z}_q - \underline{R}_q)' \Sigma_w^{-1} (\underline{Z}_q - \underline{R}_q)\right) \end{aligned}$$

where  $\Sigma_w^{-1}$  is defined by (4), and  $\underline{Z}_q$  is defined by (16).

The general detector structure, after parameter estimation, reduces to

$$\Lambda(\underline{Z}_1, \dots, \underline{Z}_m) = \frac{p(\underline{Z}_1, \dots, \underline{Z}_m | H_1, \hat{\underline{\theta}}_1)}{p(\underline{Z}_1, \dots, \underline{Z}_m | H_0, \hat{\underline{\theta}}_0)} = \left| \frac{\Sigma_w^{-1} | \hat{\underline{\theta}}_1}{\Sigma_w^{-1} | \hat{\underline{\theta}}_0} \right| \underset{H_0}{\overset{H_1}{>}} \eta \quad (20)$$

$$\begin{aligned} \underline{R}_q &= \text{vec}(\underline{R}_q^{\text{tmp}}) \\ &= [\hat{b}(1)s_q(1), \dots, \hat{b}(1)s_q(N_i N_j), \dots, \hat{b}(N_k)s_q(1), \dots, \hat{b}(N_k)s_q(N_i N_j)]'. \end{aligned} \quad (18)$$

where  $|\Sigma_w^{-1}|$  is the determinant of the inverse covariance matrix evaluated with the estimates of the potentials under hypothesis  $H_i$ . The ratio of determinants is straightforward to calculate since the GMRF model provides us with a direct parameterization of the inverse covariance matrix from which explicit expressions for the eigenvalues can be derived. The eigenvalues of the inverse covariance matrix  $\Sigma_w^{-1}$  are obtained from the eigenvalues of the potential matrix since  $A \sim \Sigma_w^{-1}$ , and are a function of the estimates of the Markov parameters (see [3]).

For a first-order spatially–spectrally correlated GMRF, the eigenvalues of  $A$  are shown in (21) at the bottom of the page where  $1 \leq j \leq N_j$ ,  $1 \leq i \leq N_i$ , and  $1 \leq k \leq N_k$ . The quantities  $\widehat{\beta}_h^{H_i}$ ,  $\widehat{\beta}_v^{H_i}$ , and  $\widehat{\beta}_s^{H_i}$  are the estimates of the first-order GMRF parameters assuming hypothesis  $H_i$  is true. Substituting the product of the eigenvalues for the determinant terms in (20), and taking the logarithm, the GLRT detector, in its final form, is

$$\begin{aligned} & \ln \Lambda(\underline{Z}_1, \dots, \underline{Z}_m) \\ &= \frac{n N_i N_j N_k}{2} \ln \frac{\widehat{\sigma}_0^2}{\widehat{\sigma}_1^2} \\ &+ \frac{n}{2} \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \sum_{k=1}^{N_k} \left[ \ln \left( \lambda_{ijk}^{H_1}(A) \right) - \ln \left( \lambda_{ijk}^{H_0}(A) \right) \right]. \end{aligned} \quad (22)$$

As an alternative to the deterministic target model presented in this section, we consider using a stochastic target model due to the variation in target signature that results from atmospheric and illumination effects. This presents difficulties with the binary hypothesis problem. In particular, if we use a 3-D non-causal GMRF to model the target, in addition to the clutter, and assume that the GMRF parameters defining the target are different from those that describe the clutter, it eliminates the main advantage of the model: its ability to simply, and directly, parameterize the inverse clutter covariance matrix. Basically, the Markovianity of the data is lost since  $(\Sigma_w + \Sigma_r)^{-1} \neq \Sigma_w^{-1} + \Sigma_r^{-1}$ . So, to develop an anomaly detector that can efficiently handle both the clutter and target being modeled by GMRF's, we look to an alternative detection approach: ‘single’ hypothesis testing.

### C. ‘Single’ Hypothesis Testing

A ‘single’ hypothesis test, as described by Fukunaga in [26], is useful for situations in which one class is well-defined and the others are not. Although various types of natural clutter are present in the imagery, we work with the underlying assumption that within the  $I \times J$  processing block, see Fig. 1, there is only one type of clutter present, i.e., the clutter is locally homogeneous. Section III-D presented a 3-D noncausal GMRF model

for this local clutter background that is adapted to fit the data through estimation of the Markov parameters.

The ‘single’ hypothesis test statistic measures the distance between the data vectors in the *target* window and the *clutter*, to decide if the target vectors belong to the clutter Gaussian distribution. The data vectors  $\underline{Y}_q$  are the target vectors derived from the central *unknown* region discussed in Section IV-A. Now, in addition to the clutter, the target is also modeled as a 3-D-GMRF, albeit with different Markov parameters. The clutter model is fit according to the approximate maximum-likelihood estimates given in Section III-E, and detailed in [3]. Mathematically, the distance measure for the set of  $m$  independent target vectors is

$$\begin{aligned} \frac{1}{m} \sum_{q=1}^m d_q^2 &= \frac{1}{m} \sum_{q=1}^m (\underline{Y}_q - \underline{m}_x)' \widehat{\Sigma}_w^{-1} (\underline{Y}_q - \underline{m}_x) \\ &= \text{tr} \widehat{\Sigma}_w^{-1} \tilde{S} \end{aligned} \quad (23)$$

where  $\underline{m}_x$  and  $\widehat{\Sigma}_w^{-1}$  are the the mean and inverse covariance matrix estimates for the clutter, and

$$\tilde{S} = (1/m) \sum_{q=1}^m (\underline{Y}_q - \underline{m}_x)(\underline{Y}_q - \underline{m}_x)'$$

is the sample covariance of the data vectors in the target region, assuming they have the mean of the clutter. The distance defined by (23) reveals two computational advantages of the ‘single’ hypothesis test over the binary hypothesis method. First, the ‘single’ hypothesis paradigm avoids the estimation and removal of the target spectral signature,  $\underline{R}_q$ . Second, the computation of the determinant of  $\Sigma_w^{-1}$  is not required.

Substituting the spatially–spectrally correlated Kronecker parameterization from (4) for  $\widehat{\Sigma}_w^{-1}$  into (23), and letting  $\underline{Z}_q = \underline{Y}_q - \underline{m}_x$ , the statistical distance between the observation vectors and the clutter vectors becomes

$$\begin{aligned} \frac{1}{m} \sum_{q=1}^m d_q^2 &= \frac{1}{m \sigma^2} \sum_{q=1}^m (\underline{Z}_q' \hat{A} \underline{Z}_q) \\ &= \frac{1}{m \widehat{\sigma}^2} \sum_{q=1}^m [\underline{Z}_q' (T_4 - \beta_h T_1 - \beta_v T_2 - \beta_s T_3) \underline{Z}_q] \end{aligned} \quad (24)$$

where

$$\begin{aligned} T_1 &= I_{N_k} \otimes I_{N_i} \otimes H_{N_j}^1; \\ T_2 &= I_{N_k} \otimes H_{N_i}^1 \otimes I_{N_j}; \\ T_3 &= H_{N_k}^1 \otimes I_{N_i} \otimes I_{N_j}; \\ T_4 &= I_{N_k} \otimes I_{N_i} \otimes I_{N_j}; \end{aligned}$$

are used to represent the four Kronecker terms in the parameterization of  $\widehat{\Sigma}_w^{-1}$  [see (4)].

$$\lambda_{ijk}^{H_i}(A) = \left[ 1 - 2\widehat{\beta}_h^{H_i} \cos\left(\frac{j\pi}{N_j + 1}\right) - 2\widehat{\beta}_v^{H_i} \cos\left(\frac{i\pi}{N_i + 1}\right) - 2\widehat{\beta}_s^{H_i} \cos\left(\frac{k\pi}{N_k + 1}\right) \right] \quad (21)$$



By simplifying terms, (24) can be written as

$$\frac{1}{m} \sum_{q=1}^m d_q^2 = \frac{1}{m\widehat{\sigma}^2} \sum_{q=1}^m (S_y - \widehat{\beta}_h \mathcal{Y}_h - \widehat{\beta}_v \mathcal{Y}_v - \widehat{\beta}_s \mathcal{Y}_s) \quad (25)$$

where  $\widehat{\beta}_h, \widehat{\beta}_v, \widehat{\beta}_s$ , and  $\widehat{\sigma}^2$  are the estimates of the Markov parameters using the data in the *clutter* region, and the quantities  $\mathcal{Y}_h, \mathcal{Y}_v, \mathcal{Y}_s$ , and  $S_y$  are the one-step-ahead correlations and the autocorrelation for the data in the unknown region. These quantities are computed as in (10)–(12) and (14), respectively, with the field  $w_{ijk}^m$  replaced by  $z_{ijk}^q$  that is the clutter-mean-removed intensity of a pixel in the *unknown* region at spatial location  $i, j$ , spectral band  $k$ , and within the  $q$ th Markov window.

## V. PERFORMANCE RESULTS

We study the GMRF anomaly detection algorithm in terms of computational and detection performance. For this evaluation, we compare the GMRF ‘Single’ Hypothesis (GMRF-SH) and Binary Hypothesis (GMRF-BH) algorithms to the RX-algorithm developed by Reed and Yu [4].

### A. Computational Performance

A goal in developing an anomaly detector for hyperspectral imagery is to have a computationally efficient algorithm, possibly running in real-time, as the sensor is being flown over the area of interest. Computational considerations are, thus, of the utmost importance.

To evaluate the computational effectiveness of the algorithms, we derive from the C-code implementation of each algorithm, an expression relating the number of floating point operations (FLOPS) required to compute the detection statistic for one pixel in the image set to the number of spectral bands,  $N_b$ , used for processing. This number of FLOPS is also dependent on the sizes of the processing, *target*, and Markov windows that are represented by  $N_w, N_t$ , and  $N_m$ , respectively. We obtain (26)–(28) shown at the bottom of the page. The *target* window is the same as the *unknown* window illustrated in Fig. 3.

The number of FLOPS as a function of the number of spectral bands  $N_b$  used for processing is shown in Fig. 5(a). We show the results for two different window combinations:  $N_w = 9, N_t =$

$3, N_m = 3$  and  $N_w = 27, N_t = 9, N_m = 3$ . These represent the smallest and largest window combinations that are appropriate for detecting the targets of interest in 1 m spatial resolution data. To more easily see where the cross-over points between the algorithms are, we include in Fig. 5(b) a zoomed in portion of the left bottom part of the plot that appears in Fig. 5(a). Most noticeably, the GMRF algorithms have the advantage that the total number of FLOPS increases linearly with the number of spectral bands  $N_b$ . This contrasts with the  $N_b^3$  dependence of the RX-algorithm. The three algorithms show a similar dependence with the window sizes.

Fig. 5(b) shows that the GMRF-SH algorithm is always computationally superior to the GMRF-BH algorithm and computationally more efficient than the benchmark RX-algorithm when using more than approximately 15 spectral bands of data. Under 15 bands, the GMRF-SH algorithm is more computationally intense than RX due to the overhead involved in estimating the Markov parameters. However, the GMRF algorithms can potentially perform even better, computationally, than indicated by Fig. 5(b) because their C-code implementations have not been fully optimized. The code for the RX-algorithm, on the other hand, is a mature implementation that is distributed for use in industry.

Another means of reducing the computational load of the GMRF algorithms is by using a subset of the data within the processing window for parameter estimation. The justification for this approach is in the fact that there are only three scalar parameters that need to be estimated from the data. Future research efforts will focus on determining a minimal subset of data that can be used for estimation without degrading detection performance.

### B. Detection Performance

We have shown that the application of a 3-D GMRF clutter model to hyperspectral imagery produces a significant computational improvement over the RX-algorithm. We now verify that this computational gain is coupled with an increase in detection performance. In this section, we present detection performance comparisons between the GMRF algorithms and the RX-algorithm using receiver operating characteristic (ROC) curves. The

$$\begin{aligned} \text{FLOPs(GMRF-SH)} &= N_b \left( 2 + 1.5 \frac{N_w}{N_m} + 2.5 \frac{N_t}{N_m} + 2.5 \frac{N_w^2}{N_m^2} + 0.5 \frac{N_w N_t}{N_m^2} - 2 \frac{N_w^2}{N_m} - \frac{N_w N_t}{N_m} \right) \\ &\quad + N_b \left( 19.5 N_w^2 + 0.5 N_w N_t + \frac{N_t^2}{N_m^2} + 2 \frac{N_t^2}{N_m} - N_t^2 + 2 N_m^2 \right) - 3 N_w^2 + 74 \end{aligned} \quad (26)$$

$$\begin{aligned} \text{FLOPs(GMRF-BH)} &= N_b \left( 6 + 1.5 \frac{N_w}{N_m} + 2.5 \frac{N_t}{N_m} + 3.5 \frac{N_w^2}{N_m^2} + 0.5 \frac{N_w N_t}{N_m} - 7 \frac{N_w^2}{N_m} - \frac{N_w N_t}{N_m} \right) \\ &\quad + N_b \left( 36.5 N_w^2 + 0.5 N_w N_t + 11 \frac{N_t^2}{N_m^2} - 32 N_t^2 + 12 N_m^2 + 2 N_m \right) \\ &\quad - 6 (N_w^2 - N_t^2) + 103 \end{aligned} \quad (27)$$

$$\begin{aligned} \text{FLOPs(RX)} &= N_b^3 + N_b^2 (4 + 2 N_w^2 - 2 N_t^2) + 12 \\ &\quad + N_b \left( 2 + 4.5 N_w - 3.5 N_t + 3 N_w^2 + 3 N_w N_t - 3 N_t^2 + 3 \frac{N_t^2}{N_w} \right) \end{aligned} \quad (28)$$

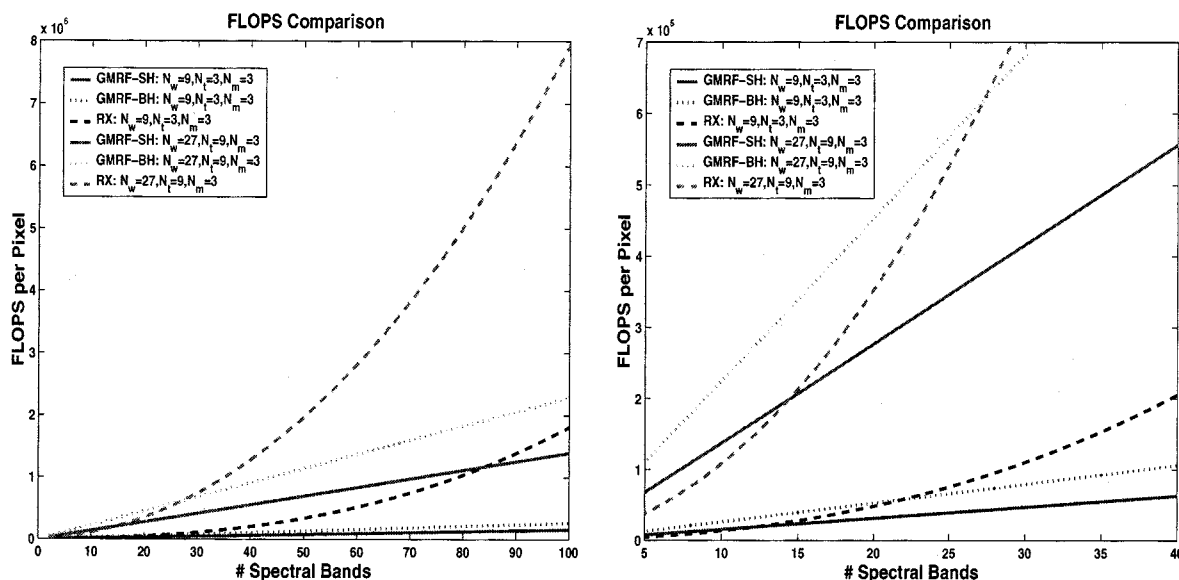


Fig. 5. (a) Computational comparison: GMRF versus RX for two window combinations. GMRF algorithms depend linearly on  $N_b$  rather than  $N_b^3$ . The GMRF-SH algorithm has a computational advantage over the GMRF-BH algorithm. (b) Magnified view of the left-bottom region of the plot in (a).

receiver operating characteristic curves plot the detection probability versus the false alarm rate per  $\text{km}^2$ . This conforms with standard practice in these applications.

We carried out experiments on several different hyperspectral data sets from both the hyperspectral digital imagery collection experiment (HYDICE) [27] sensor and the spatially enhanced broadband array spectrograph system (SEBASS) [28] sensor. The data have a spatial resolution of approximately 1 meter. Due to space limitations, we show the results for a subset of the data processed. We refer to four data sets: DsetA, DsetB, DsetD, and DsetE. In some cases, the data sets were broken down into specific portions of the electromagnetic spectrum such as the visible-to-near infrared (VNIR) and the short-wave infrared (SWIR). The HYDICE sensor records 210 spectral bands in the VNIR and SWIR, and the SEBASS sensor records 256 bands in the mid-wave and long-wave infrared. For more information on hyperspectral imagery see [29]. DsetA and DsetB are HYDICE data sets, while DsetD and DsetE are SEBASS data sets. In all our examples, we have aggregated the total number of bands in order to be able to use the RX-algorithm for comparison purposes. This of course means that our comparison results do not show the full performance potential of the GMRF algorithm. It will be important to investigate the impact of the total number of bands used for processing on the detection performance of the RX and GMRF algorithms. We intend to pursue this in future work.

We first compare the performance results of the binary hypothesis and ‘single’ hypothesis formulations of the GMRF detector. Fig. 6 shows the receiver operating characteristic curves for the two algorithms on four different data sets. In each case, the GMRF-SH algorithm significantly outperforms the GMRF-BH algorithm. The number of targets in these data sets varies between five and 20, and the number of spectral bands used for processing varies between 19 and 30. Since the GMRF-SH algorithm has also been shown to be computationally superior to the GMRF-BH algorithm, we conclude that

the ‘single’ hypothesis formulation, which uses a stochastic target model, is the better detector choice for our GMRF fully adaptive anomaly detector on data of approximately 1 m spatial resolution.

We now compare the performance of the GMRF-SH detector to that of the RX-algorithm. Fig. 7 shows the receiver operating characteristic curves for these detectors. The results illustrate that the GMRF-SH algorithm provides, at a minimum, a slight performance gain over the benchmark RX-algorithm. More importantly, coupled with this improvement in performance is a reduced processing time, even for as few as 19 spectral bands. Table I shows the average processing times for the RX and GMRF algorithms for various numbers of spectral bands on the different data sets. The times shown in the table are for C-Code running on a 250 MHz UltraSparc server. All times would be significantly reduced by implementing on parallel processors, since the GMRF algorithms, as well as the RX-algorithm, process on a small window of the data that is moved, pixel-by-pixel, throughout the image. The processing within each window is completely independent and could be implemented using separate processors.

Although the RX-algorithm processes slightly faster when using ten spectral bands, the processing times for all data sets with over 19 spectral bands are far superior for the GMRF algorithm. When using 105 bands, the GMRF algorithm takes less than one tenth the amount of time to process as the RX-algorithm. When using hyperspectral imagery, it is most likely that the number of bands being used for processing will be more than the 15 bands required to make the GMRF algorithm computationally less expensive than the benchmark algorithm. Combining this significant computational gain with the improved detection performance, clearly makes the GMRF-SH algorithm the better overall performer in these cases. To generalize these results, it will be important to test the GMRF-SH algorithm on a larger sampling of hyperspectral imagery, as it becomes available to the research community.

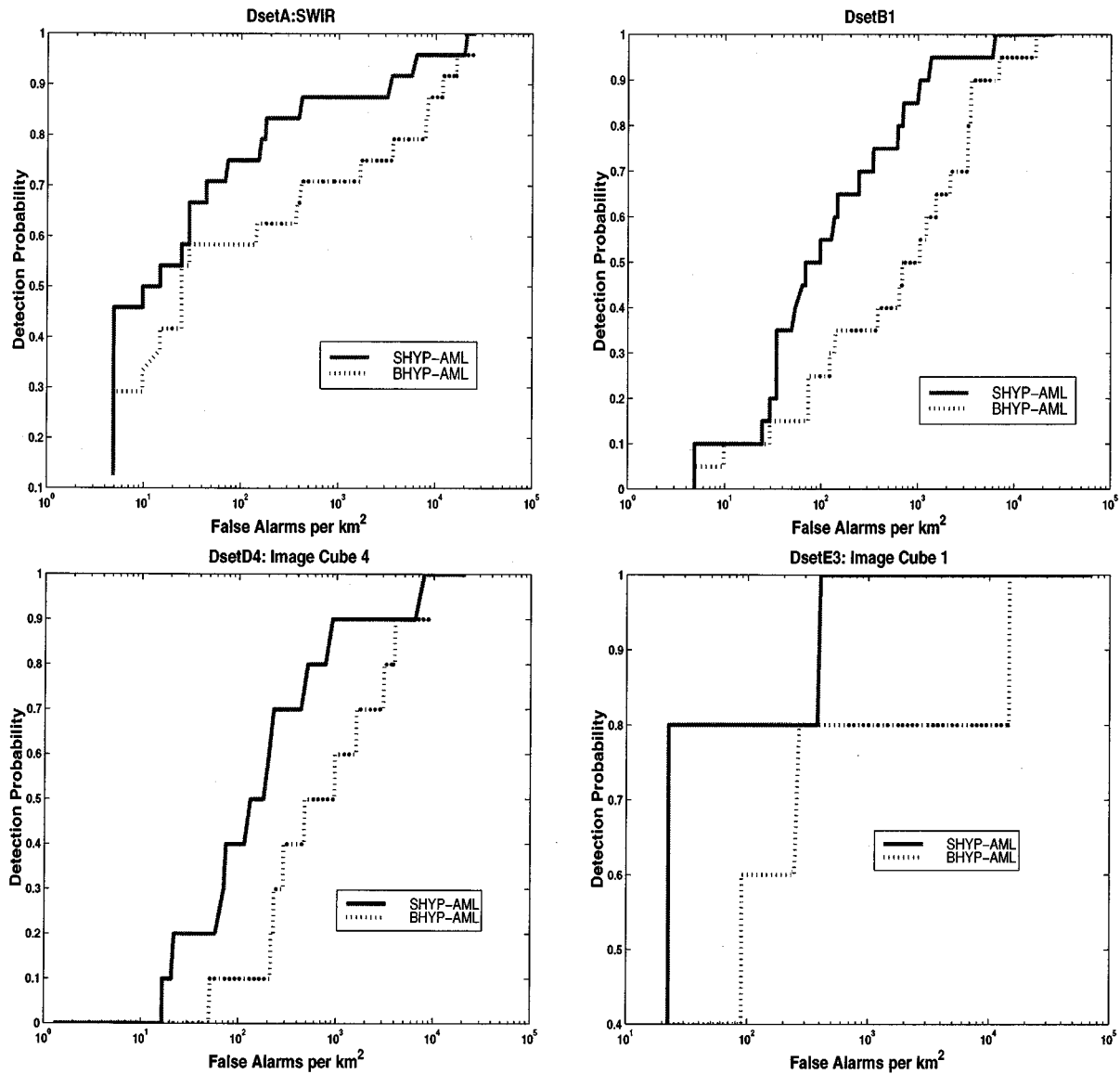


Fig. 6. Detection performance: GMRF-SH (solid) versus GMRF-BH (dotted). (a) DsetA:SWIR, 22 spectral bands; (b) DsetB, 30 spectral bands; (c) DsetD4, 19 spectral bands; and (d) DsetE, 31 spectral bands.

### C. Effect of Window Size on Performance

We now investigate the impact of the size of the processing window on the performance of the GMRF and RX algorithms. The size of the processing window is controlled by two factors: statistically reliable parameter estimation, and the assumption of homogeneous background clutter. There is a tradeoff between these two factors since reliable parameter estimation requires larger window sizes, while a homogeneous clutter assumption dictates smaller window sizes. Larger window sizes increase the likelihood of there being different types of clutter in the window, as well as the likelihood of there being a target in the perimeter region of the processing window; in either case, there is a significant impact on the computed clutter statistics.

The GMRF algorithm has an advantage over RX. For a first-order GMRF, there are only four parameters to estimate rather than the  $N_b N_{b+1}/2$  elements of the unstructured covariance matrix that must be estimated by the RX-algorithm. The small

number of parameters in the GMRF model requires less data for accurate and reliable estimation, thus allowing smaller processing window sizes. The smaller window sizes also help in validating the assumption of homogeneous clutter, which leads to improvements in detection performance.

Fig. 8 shows the receiver operating characteristic curves for the RX and GMRF-SH algorithms on DsetA:SWIR and DsetA:VNIR for various processing window sizes. For the SWIR hyperspectral images the target and Markov windows remain fixed at  $9 \times 9$  and  $3 \times 3$ , respectively, while the size of the processing window is switched from  $27 \times 27$  to  $21 \times 21$ , and, finally, to  $15 \times 15$ . With the VNIR image, we use target and Markov windows both of dimension  $3 \times 3$ , and vary the processing window from  $21 \times 21$  to  $15 \times 15$ , and, finally, to  $9 \times 9$ . For the SWIR data, the performance for the RX-algorithm remains relatively constant as the processing window size is decreased. Basically, gains in performance due to a more homogeneous window of data are eliminated by performance

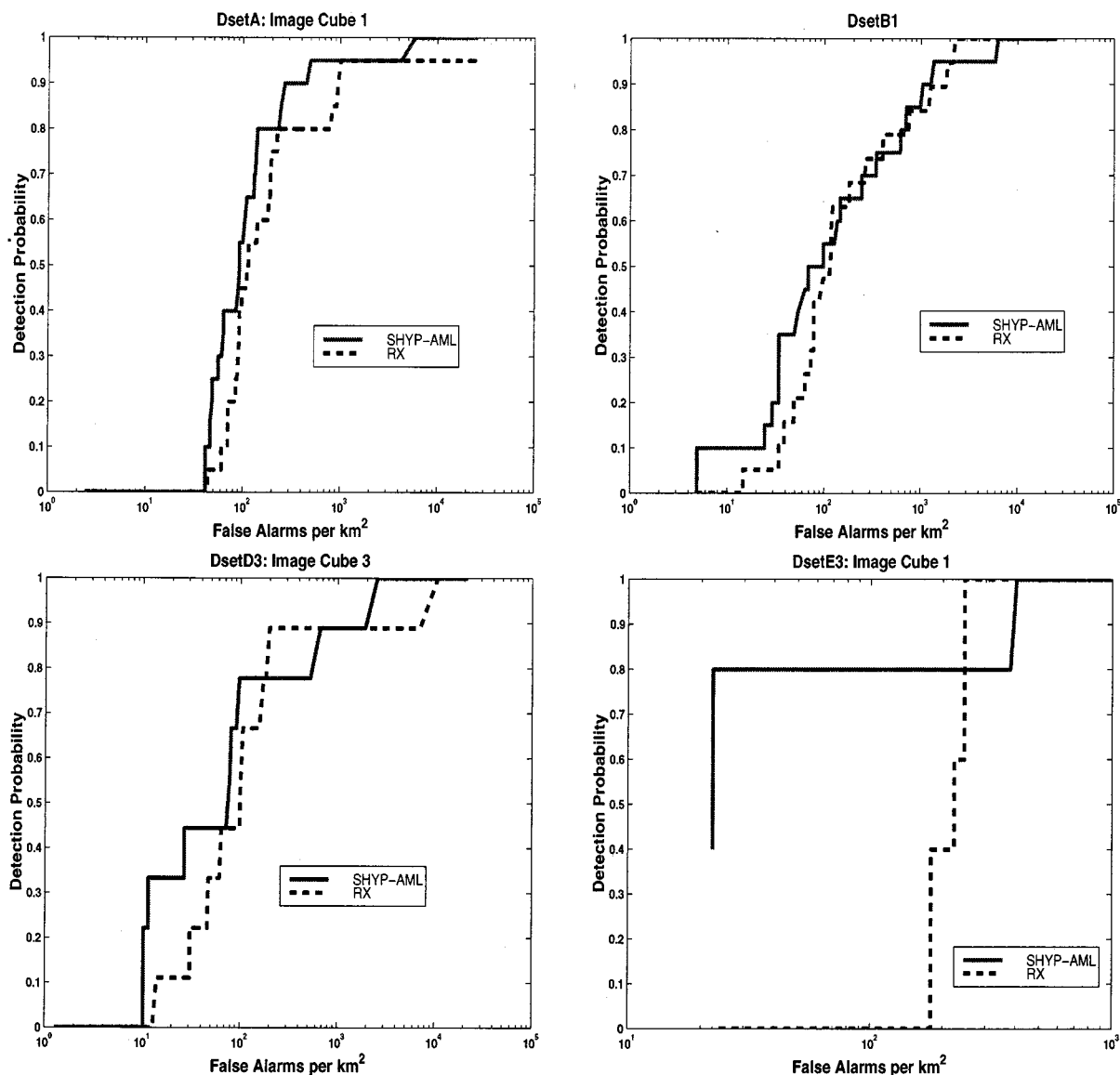


Fig. 7. Detection performance: GMRF-SH (solid) versus RX (dotted). (a) DsetA, 41 spectral bands; (b) DsetB, 30 spectral bands; (c) DsetD3, 19 spectral bands; and (d) DsetE, 41 spectral bands.

TABLE I  
ALGORITHM PROCESSING TIME COMPARISON BETWEEN GMRF AND RX

Algorithm Processing Times			
Data Set/# Bands	Image Size	GMRF	RX
DsetA/41	640 x 320	21.3 min	127.3 min
DsetB/10	960 x 320	25 min	22 min
DsetB/30	960 x 320	80 min	165 min
DsetB/105	960 x 320	4.5 hrs	44.5 hrs
DsetD/19	6100 x 128	61 min	99 min
DsetE/31	350 x 128	4.3 min	13.9 min

degradations resulting from insufficient amount of data in the processing window for reliable parameter estimation. The performance of the GMRF algorithm, on the other hand, improves with decreasing window size, because there is not the degradation associated with unreliable parameter estimates. Similar results are seen with the VNIR data. In this case, the performance of the RX-algorithm steadily degrades as the size of the processing window is made smaller. The performance of

the GMRF algorithm improves when moving from a 21 point to a 15 point processing window, and begins to degrade with a nine-point window, but not to the degree that the RX-algorithm does. These preliminary results support the idea that the GMRF algorithm performs better on smaller processing window sizes than the RX-algorithm. Using smaller processing windows reduces the overall computation time of the algorithm.

## VI. SUMMARY

In this paper we presented two 3-D noncausal GMRF anomaly detectors for natural clutter backgrounds in hyperspectral imagery. The GMRF model focuses on capturing the high levels of spatial and spectral correlation that exist in the data.

We showed that GMRF modeling provides several major contributions to the challenging problem of detection in hyperspectral imagery. First, and most importantly, the model leads to a direct parameterization of the inverse of the clutter covariance

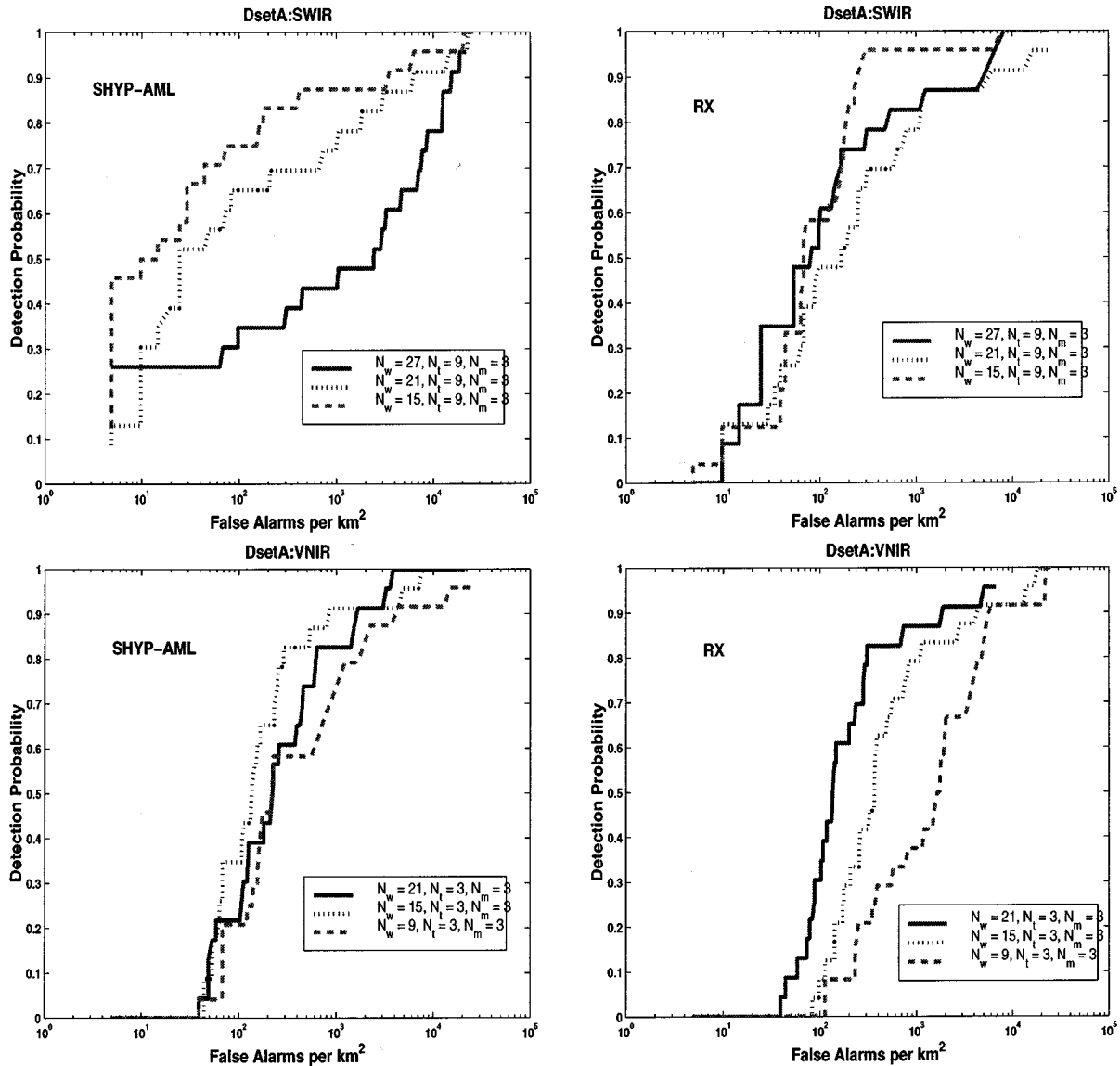


Fig. 8. Detection performance for decreasing processing window size on (a) GMRF-SH algorithm on DsetA:SWIR (top left); (b) RX-algorithm on DsetA:SWIR (top right); (c) GMRF-SH algorithm on DsetA:VNIR (bottom left); and (d) RX-algorithm on DsetA:VNIR (bottom right).

matrix. This parameterization avoids the expensive matrix inversions or determinant computations that challenge many other available detection routines. In addition, the parameterization of the inverse clutter covariance matrix is completely characterized by only four parameters,  $\sigma^2$ ,  $\beta_h$ ,  $\beta_v$ , and  $\beta_s$ , which makes it possible to use small sized processing and target windows, while still obtaining reliable parameter estimates. Finally, unlike the spatially white assumption of the benchmark multispectral RX-algorithm, the GMRF clutter model accurately assumes that the clutter is both spatially and spectrally correlated: an assumption that we predict will become increasingly significant with better spatial resolution data.

We formulated the ML-detector using the GMRF clutter model in conjunction with both binary (GMRF-BH) and 'single' (GMRF-SH) hypothesis testing paradigms. The GMRF-BH models the target as deterministic, while the GMRF-SH describes it through a stochastic model. Our performance results show that the GMRF-SH provides better detection and computational performance than both the binary

hypothesis GMRF detector and the benchmark RX-algorithm. The computational cost of the GMRF-SH algorithm increases linearly with the number of spectral bands, while the RX-algorithm grows with the third power of the number of spectral bands. This leads to a smaller number of FLOPS for our GMRF-SH algorithm when using more than approximately 15 bands for processing.

Finally, we analyzed the effect of the processing window size on the overall performance of the RX and GMRF detectors. The results indicate that the GMRF detector performs well with smaller window sizes, in contrast with the RX-algorithm, which leads to further reductions in the overall processing time for the GMRF algorithm.

#### ACKNOWLEDGMENT

The authors would like to thank Dr. R. Johnson, Dr. W. Summers, and Dr. B. Best from Science Applications International

Corporation, Arlington, VA, as well as the Multi-Sensor Exploitation Testbed and Adaptive Spectral Reconnaissance Programs in DARPA/STO, run by LtCol. B. Tousley, for providing the real hyperspectral data sets used for analysis.

#### REFERENCES

- [1] A. D. Stocker, A. Oshagan, J. H. Seldin, J. N. Cederquist, and C. R. Schwartz, "Analysis of infrared multi-spectral target/background field measurements," *Proc. SPIE*, vol. 2235, pp. 148–161, 1994.
- [2] H. L. Van Trees, *Part I: Detection, Estimation, and Modulation Theory*. New York: Wiley, 1968. Ch. 2: Classical Detection and Estimation Theory.
- [3] S. M. Schweizer and J. M. F. Moura, "Hyperspectral imagery: Clutter adaptation in anomaly detection," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1855–1871, Aug. 2000.
- [4] 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 Trans. Image Processing*, vol. 6, pp. 143–156, Jan. 1997.
- [5] E. Crist, C. Schwartz, and A. Stocker, "Pairwise adaptive linear matched-filter algorithm," in *Proc. DARPA Adaptive Spectral Reconnaissance Algorithm Workshop*, Jan. 1999.
- [6] H. T. Haskett and A. K. Sood, "Adaptive real-time endmember selection algorithm for sub-pixel target detection using hyperspectral data," in *Proc. 1997 IRIS Specialty Group Camouflage, Concealment, Deception*, Oct. 1997.
- [7] J. Grossmann, J. Bowles, D. Haas, J. Antoniadis, M. Grunes, P. Palmadesso, D. Gillis, K. Tsang, M. Baumbach, M. Daniel, J. Fisher, and I. Triandaf, "Hyperspectral analysis and target detection system for the adaptive-spectral reconnaissance program (ASRP)," *Proc. SPIE*, vol. 3372, pp. 2–13, Apr. 1998.
- [8] C.-I. Chang, X.-L. Zhao, M. L. G. Althouse, and J. J. Pan, "Least squares subspace projection approach to mixed pixel classification for hyperspectral images," *IEEE Trans. Geosci. Remote Sensing*, vol. 36, pp. 898–912, May 1998.
- [9] D. Slater and G. Healey, "Exploiting an atmospheric model for automated invariant material identification in hyperspectral imagery," *Proc. SPIE*, vol. 3372, pp. 60–71, Apr. 1998.
- [10] D. Manolakis, G. Shaw, and N. Keshava, "Comparative analysis of hyperspectral adaptive matched filter detectors," *Proc. SPIE*, Apr. 2000.
- [11] C. F. Ferrara, "Adaptive spatial/spectral detection of subpixel targets with unknown spectral characteristics," *Proc. SPIE*, vol. 2235, pp. 82–93, 1994.
- [12] A. Stocker, "Stochastic expectation maximization (SEM) algorithm," in *Proc. DARPA Adaptive Spectral Reconnaissance Algorithm Workshop*, Jan. 1999.
- [13] P. Masson and W. Pieczynski, "SEM algorithm and unsupervised statistical segmentation of satellite images," *IEEE Trans. Geosci. Remote Sensing*, vol. 31, pp. 618–633, May 1993.
- [14] E. Ashton, "Detection of subpixel anomalies in multispectral infrared imagery using an adaptive Bayesian classifier," *IEEE Trans. Geosci. Remote Sensing*, vol. 36, pp. 506–517, Mar. 1998.
- [15] A. K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989. Ch. 2: Two Dimensional Systems and Mathematical Preliminaries.
- [16] R. A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*. New York: Academic, 1997.
- [17] B. R. Hunt and T. M. Cannon, "Nonstationary assumptions for Gaussian models of images," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-6, pp. 876–882, Dec. 1976.
- [18] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus*. New York: Wiley, 1988. Ch. 2: Kronecker products, the vec operator and the Moore–Penrose inverse.
- [19] J. W. Woods, "Two-dimensional discrete Markovian fields," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 232–240, 1972.
- [20] J. M. F. Moura and N. Balram, "Recursive structure of noncausal Gauss-Markov random fields," *IEEE Trans. Inform. Theory*, vol. 38, pp. 334–354, Mar. 1992.
- [21] —, "Statistical algorithms for numerical Gauss-Markov fields," in *Handbook of Statistics*. Amsterdam, The Netherlands: North-Holland, 1993, ch. 15, pp. 623–691.
- [22] S. Beaven, L. Hoff, and D. Ly, "R-X detection algorithm definition," in *Proc. DARPA Adaptive Spectral Reconnaissance Workshop*, July 1997.

- [23] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1760–1770, Oct. 1990.
- [24] A. Stocker, X. Yu, E. Winter, and L. Hoff, "Adaptive detection of subpixel targets using multi-band frame sequences," *Proc. SPIE*, vol. 1481, pp. 156–169, Apr. 1991.
- [25] S. M. Schweizer, "The GMRF Detector for hyperspectral imagery: An efficient fully-adaptive maximum likelihood detector," Ph.D. dissertation, Carnegie Mellon Univ., Pittsburgh, PA, Dec. 1999.
- [26] K. Fukunaga, *Statistical Pattern Recognition*, 2nd ed. New York: Academic, 1990.
- [27] <http://rsd-www.nrl.navy.mil/hydice/>, Naval Res. Lab (NRL) Hydice Web site.
- [28] *Hydra data collection data distribution manual*, Nov. 1998.
- [29] K. Belokon *et al.*, *Multispectral Imagery Reference Guide*: LOGICON Geodynamics, Inc., 1997.



**Susan M. Schweizer** (S'91–M'99) received the B.S.E.E. degree in 1993 from the University of Notre Dame, Notre Dame, IN, and the M.Sc. and the Ph.D. degrees in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, in 1995 and 1999, respectively.

She is currently a Senior Research Scientist with MediaSite, Inc., Pittsburgh, PA. MediaSite, Inc., is a leader in providing technology, software, and services for enabling the indexing, search, and retrieval of video content over the Internet. Her research inter-

ests are in the areas of signal, image, and video processing, statistical signal processing, and pattern recognition.

Dr. Schweizer was the recipient of a National Science Foundation Fellowship Award (1993–1996) and winner of an Amelia Earhart Fellowship Award (1998–1999) from Zonta International. She is a member of SPIE.



**José M. F. Moura** (S'71–M'75–SM'90–F'94) received the Engenheiro Electrotecnico degree in 1969 from Instituto Superior Tecnico (IST), Lisbon, Portugal, and the M.Sc., E.E., and D.Sc. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 1973 and 1975, respectively.

He has been a Professor of electrical and computer engineering at Carnegie Mellon University, Pittsburgh, PA, since 1986. In 1999–2000, he was a Visiting Professor of electrical engineering at

MIT. Prior to this, he was on the faculty of IST (1975–1984). He was Genrad Associate Professor (Visiting) of electrical engineering and computer science at MIT (1984–1986), and a Visiting Research Scholar with the Department of Aerospace Engineering, University of Southern California, Los Angeles, during the Summers of 1978–1981. His research interests include statistical signal processing and telecommunications, image processing, and video representations. He has published more than 230 technical contributions, is the co-editor of two books, holds four U.S. patents on image and video processing, and digital communications, and has given numerous invited seminars at U.S. and European universities and laboratories.

Dr. Moura serves as Vice President for Publications for the IEEE Signal Processing Society (SPS), and is a member of the Board of Governors of the SPS. He is also Vice President for Publications for the IEEE Sensors Council. He is on the Editorial Board of the PROCEEDINGS OF THE IEEE. He chairs the IEEE Signal Processing Society Publications Board. He was Editor-in-Chief for the IEEE TRANSACTIONS IN SIGNAL PROCESSING (1975–1999). He is currently a member of the IEEE Multimedia Signal Processing technical committee and a founding member of the IEEE Sensor Array and Multichannel Technical Committee, and a past member of the IEEE Underwater Acoustics Technical Committee. He was on the IEEE Press Board (1991–1995). He Guest Co-edited the August 2000 issue of the IEEE TRANSACTIONS ON INFORMATION THEORY. He is a Fellow of the IEEE, a corresponding member of the Academy of Sciences of Portugal, a member of SIAM, AMS, AAAS, and Sigma Xi. He received the IEEE Millenium Medal in 2000.