Characterizing non-Gaussian clutter and detecting weak gaseous plumes in hyperspectral imagery

James Theiler, Bernard R. Foy, and Andrew M. Fraser

Los Alamos National Laboratory, Los Alamos, NM 87545

ABSTRACT

To detect weak signals on cluttered backgrounds in high dimensional spaces (such as gaseous plumes in hyperspectral imagery) without excessive false alarms requires that the background clutter be effectively characterized. If the clutter is Gaussian, the well-known linear matched filter optimizes the sensitivity to a given plume signal while suppressing the effect of the background clutter. In practice, the background clutter is rarely Gaussian. Here we illustrate non-linear corrections to the matched filter that are optimal for two non-Gaussian clutter models and we report on parametric and nonparametric characterizations of background clutter.

Keywords: imagery, hyperspectral, clutter, non-Gaussian, matched filter, chemical plume

1. INTRODUCTION

Imaging spectrometers have been developed and fielded in recent years with high spectral resolution and low sensor noise. The sensitivity and specificity of these sensors enables them to detect weak signals in cluttered backgrounds. We are interested in the detection of thin chemical plumes in hyperspectral imagery, and we in particular seek to improve the detection performance by more carefully characterizing the variability of the background. We refer to this background as "clutter" to emphasize that it arises not so much from noise in the sensor, but from the unknown layout of whatever materials happen to be in the scene. In particular, the clutter is not something that one would particularly expect to be Gaussian in its distribution. Our interest is twofold: how can we effectively characterize the clutter, and how can we exploit that characterization to better detect plumes?

We begin with a review of the physical formulation of the plume problem in Section 2 where we identify the linearization that we use as our starting point. This formulation is approximate and does not cover all aspects of the problem, but it allows us to concentrate on one of the most important aspects of the problem: the characterization of the background clutter. In Section 3, we describe a variety of non-Gaussian models and statistics that can be used to characterize the clutter. We follow in Section 4 with methods of using these statistics and models to effectively detect weak plumes. In Section 5 we describe statistics of actual images, and in Section 6 we conclude with an outline of progress to date and areas for future work.

2. A BRIEF FORMULATION OF THE PLUME PROBLEM

Radiative emission, absorption, and transport through the atmosphere are all *nonlinear* physical processes.[1] The hyperspectral sensor itself will introduce further nonlinearities. Nonetheless, a *linear* signal-in-noise model can often be adequate for modelling the effect of *weak* plumes with characteristic spectral signatures. Note, however, that if the background clutter is non-Gaussian, then a linear signal model does not necessarily imply that optimal detectors are linear. Physically motivated nonlinear preprocessing steps (*e.g.*, conversion of radiance to brightness temperature, or compensation of atmospheric distortions) can sometimes improve the quality of the linear signal-in-noise model.

In each pixel, imaging sensors measure an at-sensor radiance ($\mathbf{L}^{\text{at}-\text{sensor}}$). Off-plume contributions to $\mathbf{L}^{\text{at}-\text{sensor}}$ include thermal emission from atmosphere ($\mathbf{L}^{\text{path}}$) and instrument* ($\mathbf{L}^{\text{sensor}}$), and radiance from the ground ($\mathbf{L}^{\text{gnd}}$), attenuated by atmospheric absorption ($\boldsymbol{\tau}^{\text{atm}}$). That is

$$\mathbf{L}^{\text{at}-\text{sensor}-\text{off}-\text{plume}} = \mathbf{L}^{\text{path}} + \mathbf{L}^{\text{sensor}} + \mathbf{L}^{\text{gnd}}\boldsymbol{\tau}^{\text{atm}} \tag{1}$$

Author contact information: {jt,bfoy,afraser}@lanl.gov

*This includes both sensor noise and thermal radiation from the sensor itself onto its own focal plane.

where

$$\mathbf{L}^{\mathrm{gnd}} = \mathbf{B}(T^{\mathrm{gnd}})\boldsymbol{\varepsilon}^{\mathrm{gnd}} + \mathbf{L}^{\mathrm{downwelling}}(1 - \boldsymbol{\varepsilon}^{\mathrm{gnd}}) \tag{2}$$

and $\mathbf{B}(T)$ is blackbody radiation at temperature T, and $\boldsymbol{\varepsilon}^{\mathrm{gnd}}$ is the ground emissivity. Here, $\mathbf{L}^{\mathrm{downwelling}}$ includes radiation impinging on the ground, a fraction $1 - \boldsymbol{\varepsilon}^{\mathrm{gnd}}$ of which is then reflected back toward the sensor.[1]

The on-plume radiance includes two other effects: thermal emission from the plume ($\mathbf{L}^{\mathrm{plume}}$), and absorption of ground emission by the plume ($\boldsymbol{\tau}^{\mathrm{plume}}$). Thus:

$$\mathbf{L}^{\mathrm{at-sensor}} = \mathbf{L}^{\mathrm{path}} + \mathbf{L}^{\mathrm{sensor}} + \underbrace{\mathbf{L}^{\mathrm{plume}}\boldsymbol{\tau}^{\mathrm{atm}}} + \mathbf{L}^{\mathrm{gnd}}\underbrace{\boldsymbol{\tau}^{\mathrm{plume}}}\boldsymbol{\tau}^{\mathrm{atm}} \tag{3}$$

where the underbraces identify the terms that are due to the presence of a plume. At each wavelength λ, the weak plume transmissivity depends on column density n_o and the plume signature b_λ with

$$\tau_\lambda^{\mathrm{plume}} = e^{-n_o b_\lambda} \approx 1 - n_o b_\lambda \tag{4}$$

The vector emissivity is given by

$$\boldsymbol{\varepsilon}^{\mathrm{plume}} = \mathbf{1} - \boldsymbol{\tau}^{\mathrm{plume}} \approx n_o \mathbf{b}, \tag{5}$$

and the radiance due to the plume is given by

$$\mathbf{L}^{\mathrm{plume}} = \mathbf{B}(T^{\mathrm{plume}})\boldsymbol{\varepsilon}^{\mathrm{plume}} \approx \mathbf{B}(T^{\mathrm{plume}})n_o \mathbf{b}. \tag{6}$$

So the full at-sensor radiance, due to the background and the plume, is given by

$$\mathbf{L}^{\mathrm{at-sensor}} \approx \left[\mathbf{L}^{\mathrm{path}} + \mathbf{L}^{\mathrm{sensor}} + \mathbf{L}^{\mathrm{gnd}}\boldsymbol{\tau}^{\mathrm{atm}}\right] + n_o \mathbf{b}\boldsymbol{\tau}^{\mathrm{atm}}\left[\mathbf{B}(T^{\mathrm{plume}}) - \mathbf{L}^{\mathrm{gnd}}\right] \tag{7}$$

We remark that all quantities vary from pixel to pixel; $\mathbf{L}^{\mathrm{path}}$ and $\boldsymbol{\tau}^{\mathrm{atm}}$ are usually nearly constant from pixel to pixel . All but n_o are wavelength dependent. We can re-express this as

$$\mathbf{r} = \epsilon\, \mathbf{b} + \mathbf{x} \tag{8}$$

where $\mathbf{r} = \mathbf{L}^{\mathrm{at-sensor-on}}$ is the radiance measured at the sensor, $\mathbf{x} = \left[\mathbf{L}^{\mathrm{path}} + \mathbf{L}^{\mathrm{gnd}}\boldsymbol{\tau}^{\mathrm{atm}}\right] + \mathbf{L}^{\mathrm{sensor}}$ is the background plus noise (what we are calling "clutter"), $\mathbf{b}$ is the plume signature, and $\epsilon = n_o \boldsymbol{\tau}^{\mathrm{atm}}\left[\mathbf{B}(T^{\mathrm{plume}}) - \mathbf{L}^{\mathrm{gnd}}\right]$ is the plume "strength" which includes as well as column density, factors for atmospheric transmission, and the thermal contrast between the plume and the ground. The part in square brackets can have a complicated wavelength dependence, but is often constant across the spectrum of interest.[2] We remark that although $\mathbf{r}$ corresponds to what we *measure*, $\mathbf{x}$ corresponds to the clutter that we *model*.

This fairly simple formulation motivates our interest in studying Eq. (8). Alternative physical derivations appear in Refs. [2–4]. One may reformulate this expression to incorporate some atmospheric compensation;[5] to the extent that the compensation is accurate, this provides a number of advantages, from simplifying the ground clutter, to reducing the wavelength dependence of the plume strength.

3. CHARACTERIZING CLUTTER

3.1. Gaussian model

A multivariate Gaussian distribution provides a simple model for the clutter, with a number of useful advantages. It is parameterized by a mean $\boldsymbol{\mu} \in \mathcal{R}^d$ and a symmetric positive semi-definite covariance matrix $K \in \mathcal{R}^{d \times d}$:

$$P(\mathbf{x}) = (2\pi)^{-d/2} |K|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T K^{-1}(\mathbf{x} - \boldsymbol{\mu})\right] \tag{9}$$

The covariance matrix $K = \langle(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T\rangle$ expresses the various band-to-band correlations. Given the covariance matrix, one can identify the directions in $\mathcal{R}^d$ in which the clutter has low variance. These low-variance directions are useful because they are more noticeably influenced by weak signals.

Since $\boldsymbol{\mu}$ and K are not known *a priori*, they are usually estimated from the data using

$$\hat{\boldsymbol{\mu}} \;=\; (1/N)\sum_i \mathbf{x}_i \tag{10}$$

$$\hat{K} \;=\; (1/N)\sum_i (\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T \tag{11}$$

With $O(d^2)$ free parameters in the $\hat{K}$, it is easy to overfit the model, and in particular to underestimate the variance in the "thin" directions.[6]

3.2. Non-Gaussian models

3.2.1. Elliptically Contoured (EC) distributions

Manolakis *et al.*[7] proposed using elliptically contoured (EC) distributions to describe hyperspectral data. One very useful property that EC distributions share with Gaussian distributions is that they are characterized by a covariance matrix K. In general, an EC distribution is given by a formula

$$P(\mathbf{x}) = (2\pi)^{-d/2}\,|K|^{-1/2}\,h(r^2) \tag{12}$$

where $r^2 = (\mathbf{x} - \hat{\mathbf{x}})^T K^{-1}(\mathbf{x} - \boldsymbol{\mu})$ is the Mahalanobis distance, and $h(r^2)$ is a positive monotonically decreasing function of r^2. The Gaussian is a special case of the EC family, given by $h(r^2) = \exp(-r^2/2)$.

Since the EC distributions are characterized by a covariance matrix K, they also identify the useful "thin" directions. But by generalizing $h(r^2)$, the tails of the distribution can be more faithfully modeled. Marden and Manolakis[8, 9] recommend a multivariate t-distribution, for which (up to normalization)

$$h(r^2) = \left[1 + \frac{1}{\nu}\, r^2\right]^{-(d+\nu)/2} \tag{13}$$

Here, the parameter ν characterizes the shape of the distribution; as $\nu \to \infty$, the distribution approaches a Gaussian, but for smaller values of ν the distributions have heavier tails.

3.2.2. Mixture models

Mixture models are distributions that can be expressed as positive linear combinations of other distributions; *e.g.*,

$$P(\mathbf{x}) = p_1 P_1(\mathbf{x}) + p_2 P_2(\mathbf{x}) + \cdots + p_K P_K(\mathbf{x}) \tag{14}$$

where $0 \le p_i \le 1$, and $\sum_i p_i = 1$, and each P_i is a normalized probability distribution. In practice, K is usually small (to keep a reasonable number of fitted parameters), and each P_i is a relatively simple distribution, such as a Gaussian or an EC distribution.

Although mixture models are flexible and physically plausible – each mixture component corresponds to a physically distinct material on the ground – there are costs to using them. Because the likelihood function has singularities, direct maximum likelihood parameter estimation can fail outright. Even for regularized criterion functions, there are not closed form expressions for maxima; simultaneous optimization of all the parameters in all the models can be problematic and iterative procedures such as expectation-maximization are typically employed, but these are not guaranteed to converge to a global optimum. In general, these models have a lot of free parameters, and with limited data to fit them, the variance of the estimates can be unduly large.

But despite these pitfalls, mixture models can still be useful. Funk *et al.*[10] used k-means to fit data to a mixture of Gaussians which were effectively assumed not to overlap and found improved detection performance. Marden and Manolakis[8] used an expectation-maximization (EM) algorithm to fit multiple EC distribution clusters to hyperspectral data. In another paper, Manolakis and Marden[11] described a local principal components approach using vector quantization that reduces Euclidean distance errors.

3.2.3. Parzen window model

The Parzen window[12] (see also Ref. [13, pp. 164–171]) model provides a nonparametric estimate of the distribution $P(\mathbf{x})$ from a sample of N points taken from that distribution:

$$\hat{P}(\mathbf{x}) = (1/N) \sum_{i=1}^{N} \sigma^{-d} g\left(\frac{|\mathbf{x} - \mathbf{x}_i|}{\sigma}\right) \tag{15}$$

where $g(r)$ is a "kernel" function that decreases monotonically with increasing r and is normalized so that a d-dimensional radially symmetric distribution has unit volume. A typical choice is

$$g(r) = (2\pi)^{-d/2} \exp(-r^2/2). \tag{16}$$

The tail of a Parzen window estimator depends on the kernel $g(r)$, and is Gaussian if the kernels are Gaussian. Thus, this is not a very good way to model fat-tailed distributions.

Also, all directions are only as thin as the Gaussian kernel $g(r)$. To achieve a large dynamic range in the variance of the distribution in different directions, a very large number of kernels are necessary.

What the Parzen window does have that is potentially advantageous is that it is very data adaptive; in fact, with appropriate choice of σ as a function of N, it can in the $N \to \infty$ limit model virtually any continuous distribution.

3.2.4. Endmembers

Endmembers[14] model the data in a way that does not provide an explicit probability distribution function $P(\mathbf{x})$ for the background clutter. In its simplest form, the endmember model provides a small number k of distinct material spectra, $\{\mathbf{e}_1, \ldots, \mathbf{e}_k\}$, and then asserts that most of the variance in the clutter can be accounted for in terms of positive linear combinations of these spectra. That is, each pixel $\mathbf{x}_i$ is modelled as

$$\mathbf{x}_i = \sum_{j=1}^{k} p_{ij} \mathbf{e}_j + \mathbf{n}_i \tag{17}$$

subject to $\sum_j p_{ij} = 1$. Here, $0 \le p_{ij} \le 1$ is interpreted as the abundance of the jth material in the ith pixel, and $\mathbf{n}_i$ is the residual which is usually modelled as some kind of Gaussian noise process.

The endmembers are sometimes referred to as "interferers" and detection in a scene that is modeled by endmembers is usually achieved by first "projecting out" the subspace spanned by the endmembers, and then employing an ordinary matched filter on the lower dimensional $(d-k)$ space that is orthogonal to the endmembers. See, for instance, Refs. [15, 16].

4. LINEAR AND NONLINEAR PLUME DETECTORS

Usually detector performance is graded probabilistically in terms of a *receiver operating characteristic* (ROC) that describes the trade off between detection probability and false alarm rate. Given probabilistic descriptions of the possible signals, i.e., P_{H_0} and P_{H_1} for measurements of a pixel with and without a plume respectively and a radiance measurement $\mathbf{r}$, the familiar likelihood ratio test is optimal.[13, 17] Here the test is given by

$$\mathcal{H} = \begin{cases} H_0 & \mathcal{L}(\mathbf{r}) < \gamma \\ H_1 & \mathcal{L}(\mathbf{r}) \ge \gamma \end{cases} \tag{18}$$

where

$$\mathcal{L} = \frac{P_{H_1}(\mathbf{r})}{P_{H_0}(\mathbf{r})} \tag{19}$$

is the likelihood ratio. The threshold γ determines a single point on the ROC curve, and is often specified to yield a particular false alarm rate.
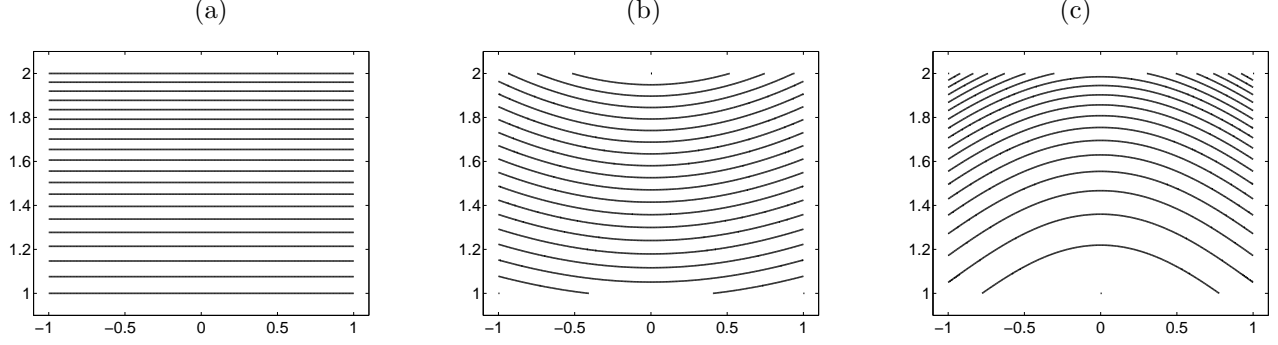
Figure 1. Decision boundaries of the GLRT detector. The lines follow level sets of $\mathcal{L}(\mathbf{r})$ from Eq. (21). A whitened distribution in $d = 2$ dimensional space is centered at the origin; these contours correspond to detection of a plume with signature $\mathbf{b} = [0, 1]$ in the vertical direction. **(a)** A Gaussian distribution leads to linear contours, corresponding to constant values of the matched filter $\mathbf{q} = K^{-1}\mathbf{b}$. Since this is whitened space, $K = I$ and $\mathbf{q} = \mathbf{b}$ and the contours are horizontal lines. **(b)** A leptokurtic (fat-tailed) distribution is provided by a multivariate t-distribution with $\nu = 5.0$. Here the detection contours bend away from the origin, relative to how they behave for a Gaussian distribution. **(c)** A platykurtic (thin-tailed) distribution is given by $P(\mathbf{r}) = C\exp(-|\mathbf{r}|^4)$ where C is just a normalization constant. Here, the contours bend back toward the origin.

When probabilistic descriptions are unavailable or incomplete, designing optimal or even good detectors is more challenging. In this section, we will describe and illustrate a variety of detectors and discuss conditions for which they are optimal or useful.

If the plume strength ϵ were known, then the likelihood ratio could be applied directly, and the optimal test for plume versus no-plume would be in terms of the ratio

$$\mathcal{L}(\mathbf{r}) = \frac{P(\mathbf{r} - \epsilon\,\mathbf{b})}{P(\mathbf{r})}. \tag{20}$$

But since ϵ is unknown, this approach needs to be modified. Two approaches for doing this follow.

4.1. Generalized Likelihood Ratio Test (GLRT)

To treat the case of unknown signal strength, Kelly[18] introduced the generalized likelihood ratio test (GLRT). The idea behind the GLRT is to replace all nuisance parameters (such as ϵ) with their maximum likelihood estimates. Here,

$$\mathcal{L}(\mathbf{r}) = \frac{\max_\epsilon P(\mathbf{r} - \epsilon\,\mathbf{b})}{P(\mathbf{r})}. \tag{21}$$

Fig. 1 shows contours of this likelihood ratio for three different distributions $P(\mathbf{x})$: one is Gaussian, one is fat-tailed, and one is thin-tailed. The GLRT is not known in general to be optimal, but Scharf and Friedlander[19] have shown that it does possess optimality properties in some special cases (in particular, when the distribution is Gaussian).

4.1.1. GLRT for Gaussian

If $P(\mathbf{r})$ is Gaussian, then the formula in Eq. (21) can be used directly. We will for simplicity assume $\boldsymbol{\mu} = 0$ and that K is known. Then $P(\mathbf{r}) = C\exp(-\frac{1}{2}\mathbf{r}^T K^{-1}\mathbf{r})$ for constant $C = (2\pi)^{-d/2}|K|^{-1/2}$, and the GLRT becomes

$$
\begin{aligned}
\mathcal{L}(\mathbf{r}) &= \frac{\max_\epsilon C\exp\left[-\frac{1}{2}(\mathbf{r} - \epsilon\,\mathbf{b})^T K^{-1}(\mathbf{r} - \epsilon\,\mathbf{b})\right]}{C\exp\left[-\frac{1}{2}\mathbf{r}^T K^{-1}\mathbf{r}\right]} \\
&= \max_\epsilon \exp\left[-\frac{1}{2}(\mathbf{r} - \epsilon\,\mathbf{b})^T K^{-1}(\mathbf{r} - \epsilon\,\mathbf{b}) + \frac{1}{2}\mathbf{r}^T K^{-1}\mathbf{r}\right]
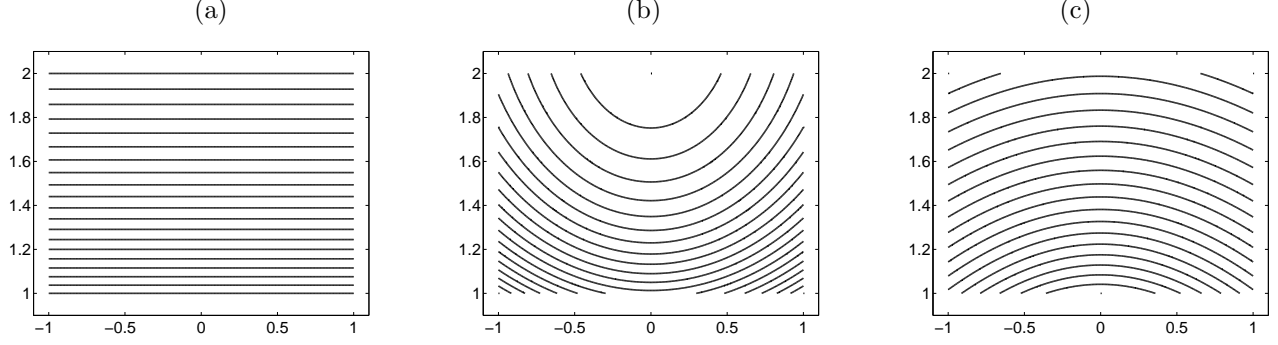\end{aligned}
\tag{22}
$$

Figure 2. Decision boundaries for the local derivative detector. This figure is the same as Fig. 1, except that it uses the local derivative method instead of the generalized likelihood ratio test. The lines follow solutions of Eq. (28) **(a)**. A Gaussian distribution leads to linear contours. The contours in this figure (like those in Fig. 1) are not calibrated, so provide information only on the "shape" of the contours. **(b)** A leptokurtic (fat-tailed) distribution leads to detection contours that bend away from the origin. **(c)** A platykurtic (thin-tailed) distribution leads to contours that bend back toward the origin.

Taking the logarithm,

$$\ell(\mathbf{r}) = \log \mathcal{L}(\mathbf{r}) \quad = \quad \max_\epsilon \left(-(\mathbf{r} - \epsilon\,\mathbf{b})^T K^{-1}(\mathbf{r} - \epsilon\,\mathbf{b}) + \mathbf{r}^T K^{-1}\mathbf{r} \right)$$

$$= \quad \max_\epsilon \left(\epsilon\,\mathbf{b}^T K^{-1}\mathbf{r} + \epsilon\mathbf{r}^T K^{-1}\mathbf{b} - \epsilon^2 \mathbf{b}^T K^{-1}\mathbf{b} \right) \tag{23}$$

The maximum occurs at

$$\epsilon = \frac{\mathbf{b}^T K^{-1}\mathbf{r}}{\mathbf{b}^T K^{-1}\mathbf{b}} \tag{24}$$

which leads to

$$\ell(\mathbf{r}) = \frac{\left(\mathbf{b}^T K^{-1}\mathbf{r}\right)^2}{\mathbf{b}^T K^{-1}\mathbf{b}}. \tag{25}$$

Recall that this derivation assumed that the Gaussian distribution $P(\mathbf{r})$, and in particular the covariance K, is known precisely. In practice, K must be estimated from data. The natural choice, advocated by Reed *et al.*,[20] is to use the estimator for $\hat{K}$ in Eq. (11) in place of the exact K in Eq. (25). In Kelly's derivation,[18] a more careful approach, in which $\hat{\boldsymbol{\mu}}$, $\hat{K}$, and ϵ are simultaneously estimated in the numerator of Eq. (21), leads to a discriminant function of the form

$$\frac{|\mathbf{b}^T \hat{K}^{-1}\mathbf{r}|^2}{\mathbf{b}^T \hat{K}^{-1}\mathbf{b}\left(1 + \frac{1}{N}\mathbf{r}^T \hat{K}^{-1}\mathbf{r}\right)}, \tag{26}$$

although a later publication[21] suggested that the simpler matched filter in Eq. (25) was more practical.

Note that Eq. (25) implies that the contours of constant likelihood ratio occur at constant values of $\mathbf{q}^T\mathbf{r}$ where $\mathbf{q} = K^{-1}\mathbf{b}$. These are the contours of the optimal linear matched filter that one can obtain more simply by maximizing a signal to clutter ratio. Although we can derive the filter without appealing to Gaussian distributions at all, if the data are not Gaussian, the probabilistic performance of the detector may not be optimal.

4.2. Local derivative method (small ϵ limit)

Since we do not know the magnitude of ϵ, we cannot use the likelihood ratio in Eq. (20). In the GLRT, we use the maximum likelihood estimate of ϵ, but that can lead to values of ϵ that might not accord with what is known

about the physics of the plume. We will consider here the situation in which we don't know the magnitude of ϵ, but we do know that it is small. We begin with the likelihood ratio in Eq. (20), and take the small ϵ limit.

$$\mathcal{L}(\mathbf{r}) = \frac{P(\mathbf{r} - \epsilon\,\mathbf{b})}{P(\mathbf{r})} \approx \frac{P(\mathbf{r}) - \epsilon\,\mathbf{b}\cdot\nabla P(\mathbf{r})}{P(\mathbf{r})} \tag{27}$$

Since we are interested in contours of constant $\mathcal{L}(\mathbf{r})$, these occur when

$$\frac{\mathbf{b}\cdot\nabla P(\mathbf{x})}{P(\mathbf{x})} = c \tag{28}$$

for some constant c which depends on p. We remark that the constant c is like the threshold γ in Eq. (18), and needs to be "calibrated" against, for instance, a desired false alarm rate. Fig. 2 illustrates what contours specified by Eq. (28) look like for three different distributions $P(\mathbf{x})$.

4.2.1. Gaussian $P(\mathbf{r})$

Note that for Gaussian $P(\mathbf{r}) = C \exp(-\frac{1}{2}\mathbf{r}^T K^{-1}\mathbf{r})$, we have

$$\frac{\mathbf{b}\cdot\nabla P}{P} = \mathbf{b}\cdot\nabla(\log P) = \mathbf{b}\cdot\nabla\left(\log C - \frac{1}{2}\mathbf{r}^T K^{-1}\mathbf{r}\right) = \mathbf{b}^T K^{-1}\mathbf{r} \tag{29}$$

which implies that contours are along lines of constant $\mathbf{q}^T\mathbf{r}$ where $\mathbf{q} = K^{-1}\mathbf{b}$.

5. STATISTICS OF HYPERSPECTRAL IMAGERY

It is evident by inspection that background clutter in hyperspectral imagery is non-Gaussian. To illustrate this, we provide two small 128×128 chips from AVIRIS[22, 23] dataset `f970620t01p02_r03_sc01`; this is a reflectance image over the Moffett Field area in California. One chip is from a predominantly rural area, and the other is an urban scene. In both cases, as seen in Fig. 3, scatterplots of the first two principal components of the hyperspectral image show decidedly non-Gaussian distributions.

But quantitative characterization of multivariate distributions (especially the high-dimensional distributions in hyperspectral imagery) is difficult – this *is* the curse of dimensionality problem – so we characterize distributions of various one-dimensional projections of the data.

5.1. Mahalanobis distances

Given a covariance matrix K, the Mahalanobis distance is Euclidean distance in the whitened space: in particular, the Mahalanobis distance between two points $\mathbf{r}_1$ and $\mathbf{r}_2$ is given by

$$r^2 = (\mathbf{r}_1 - \mathbf{r}_2)^T K^{-1}(\mathbf{r}_1 - \mathbf{r}_2). \tag{30}$$

If the data were Gaussian, the Mahalanobis distances would have a chi-squared distribution with d degrees of freedom, where d is the dimension of the space (*i.e.,* the number of channels in the hyperspectral imagery). For more general EC distributions, the distribution of Mahalanobis distances will be different. For real hyperspectral images, the distribution of Mahalanobis distances is usually reported to be fat-tailed,[7] and we see that as well in Fig. 4. If the data are distributed according to the multivariate t-distribution in Eqs. (12,13), then the Mahalanobis distances are distributed with an $F_{d,\nu}$ distribution.[8]

5.2. Distribution of matched-filter projections

A more direct alternative to Mahalanobis distances are projections of the data onto one-dimensional axes. This provides scalars whose distributions can be characterized by statistics, such as the kurtosis, which is defined

$$\kappa_{\mathbf{q}} = \frac{\langle(\mathbf{q}^T\mathbf{r} - \mathbf{q}^T\hat{\boldsymbol{\mu}})^4\rangle}{\langle(\mathbf{q}^T\mathbf{r} - \mathbf{q}^T\hat{\boldsymbol{\mu}})^2\rangle^2} \tag{31}$$

Unlike the Mahalanobis distances, direct projections provide statistics which do not depend on estimating the covariance matrix. The projection provides a different set of scalars for each projection direction $\mathbf{b}$. As well as providing a statistic that is direction-dependent, this approach provides a statistic that is particularly appropriate for characterizing the "important" directions of a distribution – these are the directions that a matched filter would employ in a linear plume detection scenario.
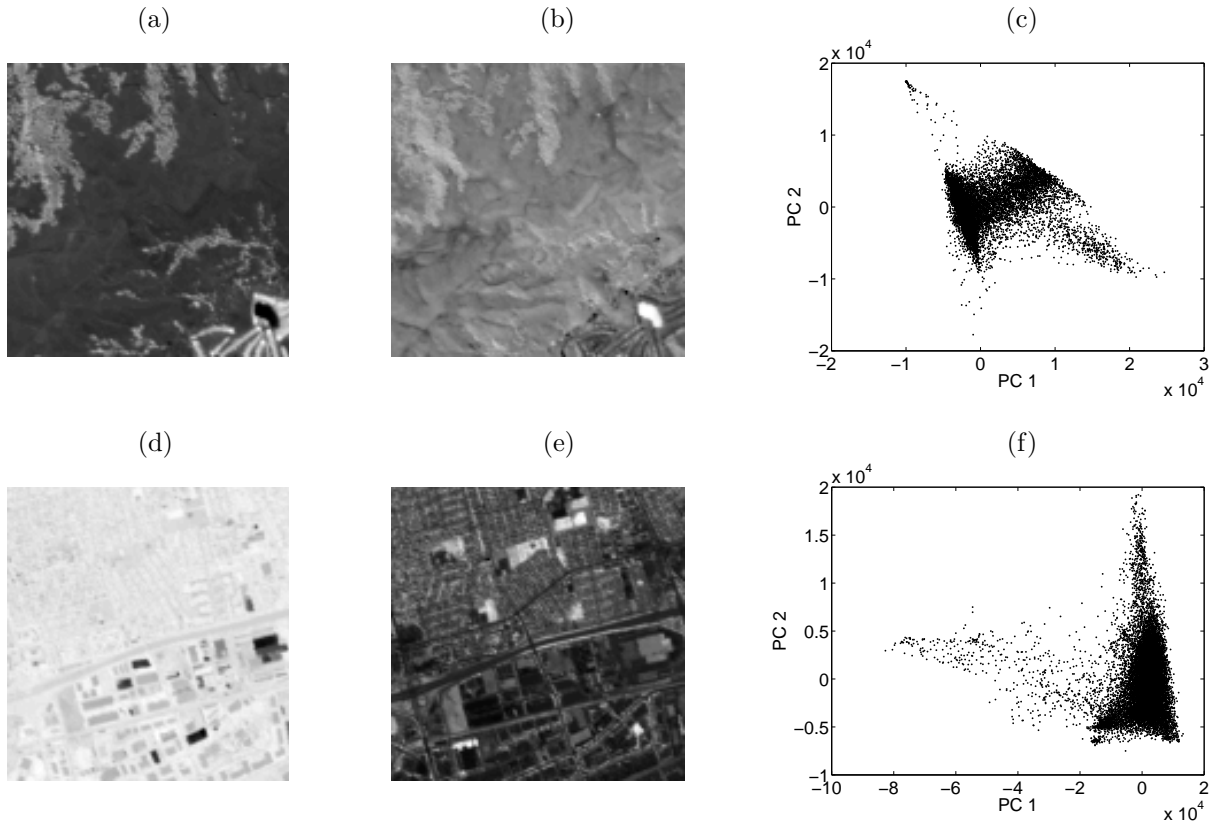
Figure 3. Two small 128×128 chips from an AVIRIS hyperspectral image. The upper panels **(a-c)** correspond to a chip from a mountainous part of the scene, while the lower panels **(d-f)** correspond to an image of the urban area. Here, **(a,d)** are images of the first principal component, and **(b,e)** are images of the second principal component. Scatterplots of the first and second principal components appear in **(c,f)**. It is evident in these scatterplots that the data are not well described as Gaussian.

5.2.1. Finite sample effects in estimates of kurtosis and variance

For an EC distribution, the kurtosis will be the same for all projections. We do know that the variance of the data along a projection *does* vary with direction; indeed, the variance in direction $\mathbf{b}$ is given by $\mathbf{b}^T K \mathbf{b}$. So one way to investigate whether the kurtosis varies with direction is to test whether it varies with variance. For a given data set, we examined both the kurtosis and the variance in a number of random directions. Since we are particularly interested in directions associated with matched filters, we will for every direction $\mathbf{b}$ also consider the matched-filter direction $\mathbf{q} = K^{-1}\mathbf{b}$. If we do this for simulated data from an EC distribution, we should find that the kurtosis is independent of direction, and therefore, independent of variance. Fig. 5(a) shows what this plot looks like when the experiment is performed naively, on a dataset with a small number of pixels. It appears that there is a correlation between variance and kurtosis, with the low-variance directions exhibiting low kurtosis. That this is a finite-sample effect is evident in Fig. 5(b) which shows that for a data set with a lot more samples, the effect is much smaller. To alleviate this finite-sample bias, we perform a more careful computation, in which the hyperspectral image is partitioned into three disjoint regions. The first region is used to estimate the covariance matrix, the second region is used for estimating variance, and the third region is used for estimating kurtosis. As Fig. 5(c) shows, this more careful approach does not show a correlation between variance and kurtosis for simulated EC data.

5.2.2. Kurtosis and variance in real hyperspectral imagery

In investigating whether such a correlation exists in real data, we will henceforth use the more careful approach with the partition of the image into disjoint regions. In Fig. 6, we compute kurtosis-variance plots for the two
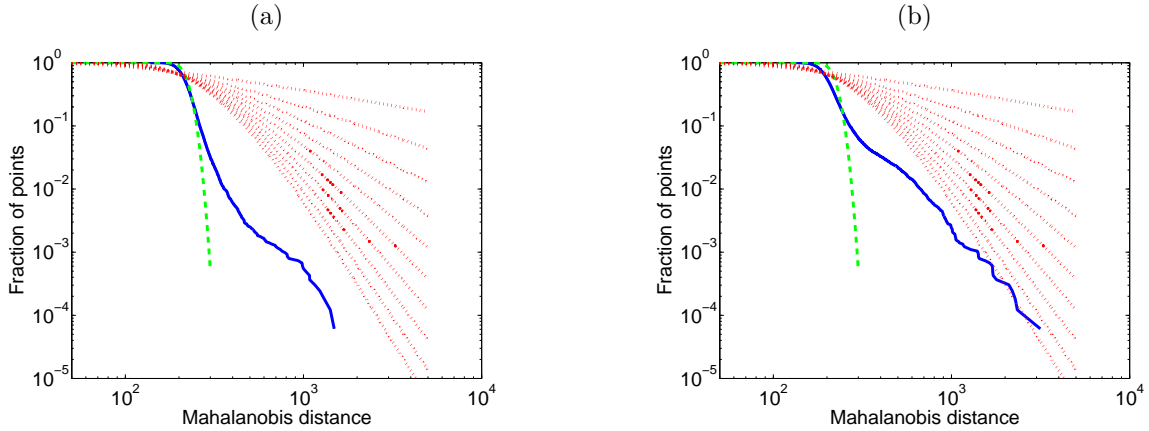
Figure 4. Cumulative histograms of Mahalanobis distances for points in the two AVIRIS images shown in Fig. 3. Panel **(a)** is the rural scene, and panel **(b)** is the urban scene. The dashed line corresponds to the χ^2_d distribution with $d = 224$, and the dotted lines correspond to F distributions with $d = 224$ and $\nu = 1, 2, \ldots, 10$. If the data were Gaussian, then the (dashed) chi-squared curve would characterize the Mahalanobis distances. Although no single F distribution fully characterizes the Mahalanobis distances, it is evident that these distances are distributed with tails much fatter than that of the chi-squared distribution. The urban scene in **(b)** is somewhat fatter-tailed than the rural scene in **(a)**.
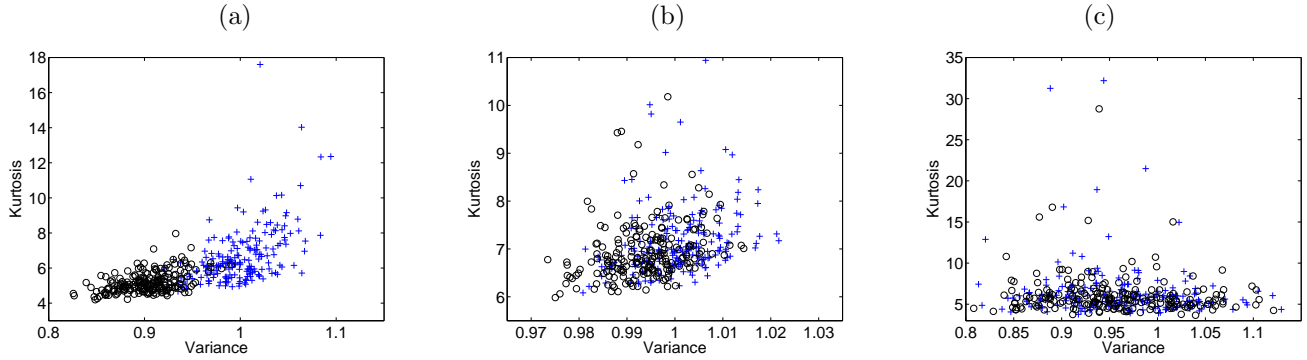

Figure 5. Variance-Kurtosis plots for simulated EC data. Variance and kurtosis associated with random directions are shown with the plus (+) symbol, and the matched-filter directions are indicated with open circles (○). **(a)** Here the data is a simulated 64×64 image, with $d = 100$ channels and the multivariate t-distribution is used with $\nu = 5.0$. The covariance matrix used for the simulated data is the identity matrix: $K = I$. Here, the same 64×64×100 data is used for the covariance matrix estimation, for the variance computation, and for the kurtosis computation. Two phenomena are evident in this plot: kurtosis is positively correlated with variance, and the matched-filter directions have both smaller variance and smaller kurtosis. **(b)** The same naive computation is performed in this panel as was done in panel (a), but the difference is that a much larger sample of data is used: a simulated 256×256 image. The same phenomena that were observed in (a) are observed in this panel, but with a much larger number of samples, the effects are decidedly less pronounced. **(c)** Same data as in panel (a), but the computation is based on a more careful scheme. The data is divided into three distinct regions. One region (half of the pixels) is used to estimate the covariance matrix $\hat{K}$ that is used to generate matched-filter directions $\mathbf{q} = \hat{K}^{-1}\mathbf{b}$. The remaining half of the data is divided into two quarters. For each direction, one quarter is used for the variance estimate, and the other quarter is used for the kurtosis estimate. This plot shows that this more careful computation yields a scatterplot which does not exhibit either of the anomalous phenomena seen in the more naive computation in panel (a).
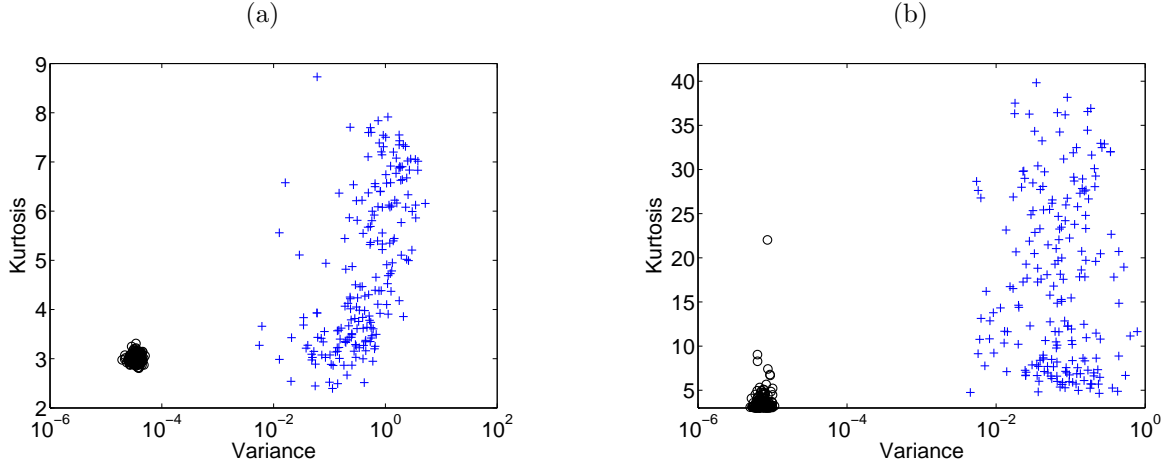
(a)   (b)

Figure 6. Kurtosis-variance plots for the **(a)** rural and **(b)** urban scenes of the AVIRIS data described in Fig. 3. Variance and kurtosis associated with random directions are shown with the plus (+) symbol, and the matched-filter directions are indicated with open circles (∘). It is not surprising that the variance would be much smaller for the matched-filter directions, indeed many orders of magnitude smaller, since the matched-filter directions are explicitly designed to minimize the magnitude of the background clutter. But what is striking is that the kurtosis in the low-variance matched-filter directions is much smaller than the kurtosis in the higher-variance random directions; indeed, the matched-filter directions exhibit a kurtosis roughly equal to three, the same value exhibited by a Gaussian distribution. Also, as observed in Fig. 4, the urban scene in (b) exhibits considerably fatter tails (more kurtosis) than does the rural scene in (a).
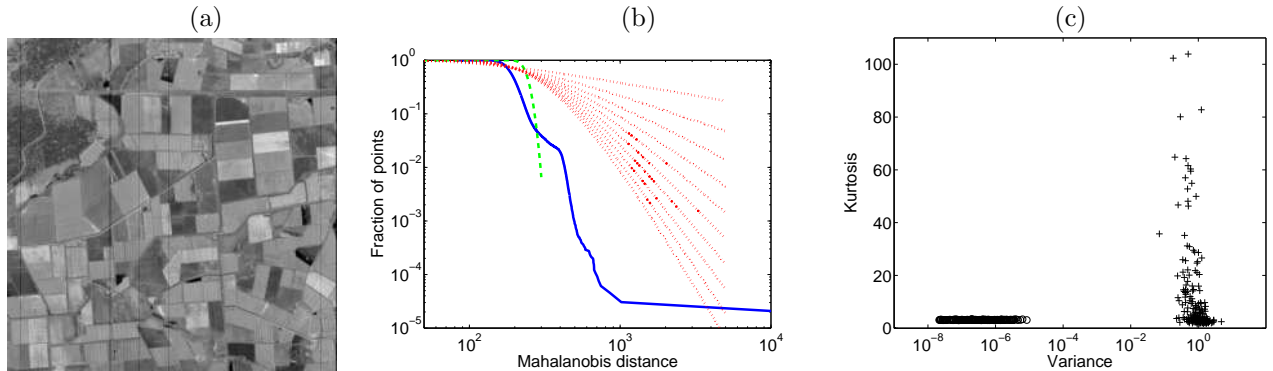

(a)   (b)   (c)

Figure 7. (a) Broadband image of the Hyperion image, obtained by adding the signal in all the channels. Some evidence of striping can be seen even in this broadband image. Particular channels show the effect much more strikingly. **(b)** Mahalanobis histogram, like those shown in Fig. 4 indicates a very fat-tailed distribution. **(c)** Kurtosis-variance plot exhibits very large kurtosis values for the random directions, although the match-filter directions show, by comparison, much smaller kurtosis values.

AVIRIS datasets, and find that there is a strong correlation. Even accounting for the finite sample effects, lower variance directions in these datasets exhibit smaller kurtosis (and are more Gaussian) than random directions which tend to have higher variance. This suggests that the elliptically contoured models do not adequately capture this aspect of the tails of the distributions of real data.

5.2.3. Remark on image artifacts

The effect seen in the AVIRIS data has been observed in other hyperspectral datasets as well. We exhibit here a dataset of the Collembally Irrigation Area in Australia, taken from the spaceborne Hyperion sensor.[24]

    Fig. 7 illustrates this Mahalanobis and Kurtosis-variance analysis of the original Hyperion data. Evidence for a strongly leptokurtic (heavy-tailed) distribution is seen in both the Mahalanobis histogram and in the kurtosis-
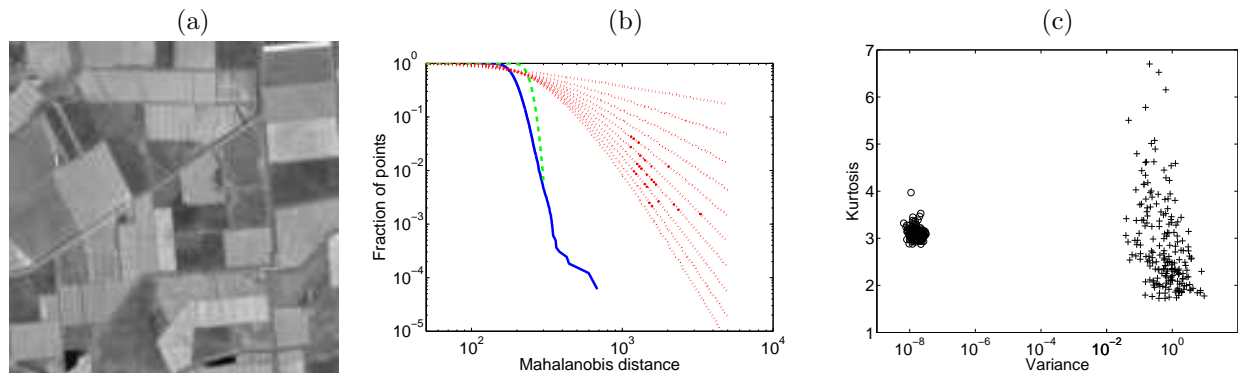
Figure 8. **(a)** Broadband image of the cleaned-up Hyperion image. Comparison to Fig. 7(a) shows that this is a sub-image of the original; the sub-image was chosen to avoid some of the artifacts that were evident in the original. Also, some of the more problematic bands in the original (specifically, bands 94, 99, 116, 168, 169, 190, and 203) were eliminated in this image. **(b)** The Mahalanobis histogram shows that the fat-tail seen in Fig. 7(b) is hardly evident in the cleaned-up image. **(c)** Kurtosis-variance plot shows that the kurtosis for random directions (+) in the cleaned-up image is considerably smaller than for the original image, as seen in Fig. 7(c). In fact, in this data, the kurtosis for the random direction averages near 3.0, the kurtosis of a Gaussian distribution. The kurtosis for the matched-filter directions are much more tightly clustered around the Gaussian kurtosis value of 3.0.

variance plot. However, a cursory examination of the data indicates that there are a number of imaging artifacts. Indeed some of these can be seen in Fig. 7(a) in the form of vertical stripes.

By eliminating some of the problematic spectral bands, and by choosing a chip of the image that avoids the edges of the data and some of the striping artifacts, a "cleaned-up" dataset was produced. This is a very unsophisticated approach to data clean-up, but it illustrates the point, evident in Fig. 8, that cleaning up an image can go a long way to eliminating, or at least reducing, the fat-tailed distribution of the data.

6. CONCLUSION

We observe, as others have (*e.g.*, Ref. [7]), that hyperspectral image data is not always well-modelled as Gaussian, and usually exhibits heavier tails than Gaussian. In the presence of non-Gaussian clutter, the optimal detectors of weak signatures were shown to be nonlinear. A local derivative method was introduced as a small ϵ alternative to the generalized likelihood ratio test (GLRT). The local derivative method is equivalent to the GLRT for Gaussian data, and gives results for non-Gaussian distributions that are qualitatively similar to the GLRT.

Our investigation of the non-Gaussian characteristics of hyperspectral imagery emphasized distributions of projected data, instead of Mahalanobis distances. This enabled us to compare the structure of the hyperspectral clutter in different directions, and we found that projections on directions with small variance are closer to Gaussian than projections on directions with larger variance. Since matched-filter directions have low variance by design, we might expect matched-filter detectors to be more nearly optimal than would be predicted by fat-tailed models of clutter in hyperspectral imagery.

These results contrast with those of McVey *et al.*,[25] who report that false positives produced by a matched filter are distributed with fatter-than-Gaussian tails. Similar results are reported by Manolakis *et al.*[7] Perhaps image artifacts, as depicted in Fig. 7 and Fig. 8, play a role in the discrepancy between these reports and our observations. It is also possible that actual gaseous plume signatures are not well-modeled as "random" directions (*i.e.*, there may be systematic correlation between the target and the ground scene signatures).

ACKNOWLEDGMENTS

REFERENCES

1. J. Schott, *Remote Sensing: the Image Chain Approach*, Oxford University Press, New York, 1997.
2. B. R. Foy, R. R. Petrin, C. R. Quick, T. Shimada, and J. J. Tiee, "Comparisons between hyperspectral passive and multispectral active sensor measurements.," *Proc. SPIE* **4722**, pp. 98–109, 2002.
3. A. Hayden, E. Niple, and B. Boyce, "Determination of trace-gas amounts in plumes by the use of orthogonal digital filtering of thermal-emission spectra," *Applied Optics* **35**, pp. 2803–2809, 1996.
4. S. J. Young, "Detection and quantification of gases in industrial-stack plumes using thermal-infrared hyperspectral imaging," Tech. Rep. ATR-2002(8407)-1, The Aerospace Corporation, 2002.
5. S. J. Young, B. R. Johnson, and J. A. Hackwell, "An in-scene method for atmospheric compensation of thermal hyperspectral data," *J. Geophys. Res. Atm.* **107**, pp. 4774–4774, 2002.
6. P. V. Villeneuve, H. A. Fry, J. Theiler, B. W. Smith, and A. D. Stocker, "Improved matched-filter detection techniques," *Proc. SPIE* **3753**, pp. 278–285, 1999.
7. D. Manolakis, D. Marden, J. Kerekes, and G. Shaw, "On the statistics of hyperspectral imaging data," *Proc. SPIE* **4381**, pp. 308–316, 2001.
8. D. B. Marden and D. Manolakis, "Modeling hyperspectral imaging data," *Proc. SPIE* **5093**, pp. 253–262, 2003.
9. D. B. Marden and D. Manolakis, "Using elliptically contoured distributions to model hyperspectral imaging data and generate statistically similar synthetic data," *Proc. SPIE* **5425**, pp. 558–572, 2004.
10. C. C. Funk, J. Theiler, D. A. Roberts, and C. C. Borel, "Clustering to improve matched filter detection of weak gas plumes in hyperspectral imagery," *IEEE Trans. Geoscience and Remote Sensing* **39**, pp. 1410–1420, 2001.
11. D. Manolakis and D. Marden, "Dimensionality reduction of hyperspectral imaging data using local principal components transforms," *Proc. SPIE* **5425**, pp. 393–401, 2004.
12. E. Parzen, "On estimation of probability density function and mode," *Ann. Math. Stat.* **33**, pp. 1065–1076, 1962.
13. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, New York, 2001.
14. J. W. Boardman, "Automating spectral unmixing of AVIRIS data using convex geometry concepts," in *Summaries of the Fourth Annual JPL Airborne Geoscience Workshop*, R. O. Green, ed., pp. 11–14, 1994.
15. J. C. Harsanyi and C.-I. Chang, "Hyperspectral image classification and dimensionality reduction: and orthogonal subspace projection approach," *IEEE Trans. Geoscience and Remote Sensing* **32**, pp. 779–785, 1994.
16. J. J. Settle, "On the relationship between spectral unmixing and subspace projection," *IEEE Trans. Geoscience and Remote Sensing* **34**, pp. 1045–1046, 1996.
17. K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, San Diego, second ed., 1990.
18. E. J. Kelly, "An adaptive detection algorithm," *IEEE Trans. Aerospace and Electronic Systems* **22**, pp. 115–127, 1986.
19. L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Processing* **42**, pp. 2146–2156, 1994.
20. I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerospace and Electronic Systems* **10**, pp. 853–863, 1974.
21. F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. Aerospace and Electronic Systems* **28**, pp. 208–216, 1992.
22. G. Vane, R. O. Green, T. G. Chrien, H. T. Enmark, E. G. Hansen, and W. M. Porter, "The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)," *Remote Sensing of the Environment* **44**, pp. 127–143, 1993.
23. AVIRIS Free Standard Data Products, Jet Propulsion Laboratory (JPL), National Aeronautics and Space Administration (NASA). `http://aviris.jpl.nasa.gov/html/aviris.freedata.html`.
24. `http://eo1.gsfc.nasa.gov/Technology/Hyperion.html`.
25. B. McVey, T. Burr, and H. Fry, "Distribution of chemical false positives for hyperspectral image data," Tech. Rep. LA-CP-02-521, Los Alamos National Laboratory, 2003.