

# Elliptically-contoured distributions for anomalous change detection in hyperspectral imagery

James Theiler, Clint Scovel, Brendt Wohlberg, and Bernard R. Foy  
Los Alamos National Laboratory, Los Alamos, NM 87545, USA

**Abstract**—We derive a class of algorithms for detecting anomalous changes in hyperspectral image pairs by modeling the data with elliptically-contoured (EC) distributions. These algorithms are generalizations of well-known detectors that are obtained when the EC function is Gaussian. The performance of these EC-based anomalous change detectors is assessed on real data using both real and simulated changes. In these experiments, the EC-based detectors substantially outperform their Gaussian counterparts.

**Index Terms**—Adaptive signal detection, Algorithms, Covariance matrices, Data models Ellipsoids, Gaussian distributions, Image analysis, Remote sensing, Pattern recognition

## I. INTRODUCTION

**G**IVEN two images of the same scene, taken at different times and under different conditions, the aim of anomalous change detection (ACD) is to find interesting changes that occurred in the scene. The emphasis on *anomalous* change recognizes that pervasive differences between the two images, *i.e.*, those differences that occur throughout the scene, are typically either uninteresting, or else are large enough that the analyst can readily find them without the aid of the change detection algorithm. (Delineating pervasive differences, *e.g.*, to outline the extent of flooding [1], or to characterize the range of forest fire damage, is still an interesting and important remote sensing problem, but it is different from the anomalous change detection problem.) These pervasive differences may be due to alterations in calibration, illumination, look angle, and even the choice of remote sensing platform. They can also be due to misregistration ([2], [3]) of the images, or to diurnal and seasonal variations [4] in the scene.

ACD algorithms attempt to make the distinction between incidental differences and anomalous changes by assuming that most of the differences that are observed in the image are incidental; that is, that the incidental differences are pervasive. From the image data, one can learn the patterns of these pervasive differences, and then the changes which do not fit the patterns are identified as anomalous.

An example of an effective ACD algorithm is the “chronochrome,” so called because it identifies changes in color (‘chromo’) with time (‘chrono’) [5]. For this algorithm, the “pattern” is a least-squares linear fit between the two images, and large residuals from that fit identify the anomalous changes. For hyperspectral imagery, there is a lot of color information; the linear fit has a large number of parameters, and is quite flexible at fitting pervasive differences. An extension of this idea to nonlinear fits was developed by Clifton [6], who used a neural net to do the fit. Further

variations include covariance equalization [7], which is based on a whitening/dewhitening principle [8], and multivariate alteration detection [9], which is based on canonical components analysis [10]. The notion of distinguishing pervasive differences from anomalous changes led to a more formal machine learning framework [11]. Although the framework considers arbitrary distributions, a new detector with a hyperbolic boundary is obtained when the data distribution is assumed to be Gaussian. Indeed, this Gaussian assumption, sometimes explicit and sometimes implicit, is common for many of these algorithms, and they all belong to a family of quadratic covariance-based ACD algorithms [12].

While the effectiveness of algorithms based on the Gaussian assumption has been well demonstrated, it is widely appreciated that data collected in the field is often far from Gaussian. The class of elliptically-contoured (EC) distributions provides a generalization of the Gaussian that has found utility both for radar signals [13] and, more recently, for hyperspectral imagery [14], [15], [16]. Like the Gaussian, which is a special case, these distributions characterize all the pairwise correlations between bands of a hyperspectral image with a covariance matrix, but unlike the Gaussian, they do not necessarily exhibit the rapidly decaying  $\exp(-r^2)$  tail. The freedom to model fatter tails is particularly useful for detections of anomalies and anomalous changes, because it is on the tail of these distributions that the distinctions between regular and anomalous data are made.

## II. DEVELOPMENT

In this section, we will derive anomalous change detectors for arbitrary data distributions. We will show that when these distributions are Gaussian, the detectors take on simple and in some cases well-known forms. Then we will derive the detectors obtained when the distribution is elliptically contoured.

### A. Arbitrary distributions

We start with a pair of co-registered images which we will call the x-image and the y-image. Let  $\mathbf{x} \in \mathbb{R}^{d_x}$  be the spectrum of a pixel in the x-image, and  $\mathbf{y} \in \mathbb{R}^{d_y}$  the spectrum of the corresponding pixel in the y-image. If we treat  $\mathbf{x}$  and  $\mathbf{y}$  as random variables, then we can write  $P(\mathbf{x}, \mathbf{y})$  as a joint probability distribution over  $\mathbf{x}$  and  $\mathbf{y}$ , and remark that  $P(\mathbf{x}, \mathbf{y})$  models what we mean by regular or *pervasive* differences. This leads to a natural way of identifying “irregular” differences (or *anomalous* changes): these are the pixels  $(\mathbf{x}, \mathbf{y})$  for which  $P(\mathbf{x}, \mathbf{y})$  is smallest.

TABLE I  
A FAMILY OF ACD ALGORITHMS PARAMETRIZED WITH  $\beta_x$  AND  $\beta_y$ .

| Likelihood Ratio   | ACD algorithm                            | $\beta_x$ | $\beta_y$ |
|--|--|-----------|-----------|
| $P(\mathbf{x}, \mathbf{y})/U(\mathbf{x}, \mathbf{y})$    | RX                                       | 0         | 0         |
| $P(\mathbf{x}, \mathbf{y})/[P(\mathbf{x})U(\mathbf{y})]$ | Chronochrome ( $\mathbf{y} \mathbf{x}$ ) | 0         | 1         |
| $P(\mathbf{x}, \mathbf{y})/[U(\mathbf{x})P(\mathbf{y})]$ | Chronochrome ( $\mathbf{x} \mathbf{y}$ ) | 1         | 0         |
| $P(\mathbf{x}, \mathbf{y})/[P(\mathbf{x})P(\mathbf{y})]$ | Hyperbolic ACD                           | 1         | 1         |

Following Refs. [17], [18], [19], we remark that anomaly detection can be recast as binary classification, where the second class corresponds to a uniform measure  $U$ , and the resulting likelihood ratio  $P(\mathbf{x}, \mathbf{y})/U(\mathbf{x}, \mathbf{y})$  is equivalent to the density  $P(\mathbf{x}, \mathbf{y})$ .

For change detection, the notion of *conditional* anomalousness is useful. Instead of the *joint* distribution  $P(\mathbf{x}, \mathbf{y})$ , use the *conditional* distribution  $P(\mathbf{y}|\mathbf{x}) = P(\mathbf{x}, \mathbf{y})/P(\mathbf{x})$ . When the pixel  $\mathbf{y}$  has an unusual value given the value of  $\mathbf{x}$ , then the conditional distribution will be small. For the multivariate Gaussian case, it can be shown [20] that this formalism leads to the chronochrome detector [5]. Note that there is an asymmetry in this formalism; the conditional distribution  $P(\mathbf{x}|\mathbf{y}) = P(\mathbf{x}, \mathbf{y})/P(\mathbf{y})$  leads to a different detector.

A machine learning framework for anomalous change, proposed in Ref. [11], leads to the symmetric likelihood ratio

$$\frac{P(\mathbf{x}, \mathbf{y})}{P(\mathbf{x})P(\mathbf{y})}. \quad (1)$$

When  $P(\mathbf{x}, \mathbf{y})$  is Gaussian, this ratio produces the Hyperbolic Anomalous Change Detector (HACD), so named for the hyperbolic boundary separating regular from anomalous.

These change detection algorithms have different origins, but Table I illustrates the close ties between them. The negative logarithm of the likelihood ratio is large when the likelihood ratio is small, and provides an anomalousness measure

$$\mathcal{A}(\mathbf{x}, \mathbf{y}) = -\log P(\mathbf{x}, \mathbf{y}) + \beta_x \log P(\mathbf{x}) + \beta_y \log P(\mathbf{y}) \quad (2)$$

parametrized by  $\beta_x$  and  $\beta_y$ . Which detector is better depends not on the distribution  $P(\mathbf{x}, \mathbf{y})$ , but on the kind of anomalies that are sought. Our aim here is not to choose among them, but to consider all of these algorithms, and to investigate how the choice of model for  $P(\mathbf{x}, \mathbf{y})$ , specifically the choice of EC distribution instead of Gaussian, affects their performance.

In what follows, it is convenient to introduce

$$\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \in \mathbb{R}^d, \quad (3)$$

with  $d = d_x + d_y$ , as the pixel in the “stacked” image.

### B. Gaussian distributions

It is natural, and often very useful, to model  $P(\mathbf{x}, \mathbf{y})$  as a multivariate Gaussian. We will assume without loss of generality that the mean of this distribution is the origin, so that  $\langle \mathbf{x} \rangle = \mathbf{0}$  and  $\langle \mathbf{y} \rangle = \mathbf{0}$ . Then, we can write the covariance matrix in terms of the stacked pixel  $\mathbf{z}$  defined in Eq. (3).

$$Z = \langle \mathbf{z}\mathbf{z}^T \rangle = \begin{bmatrix} X & C^T \\ C & Y \end{bmatrix} \quad (4)$$

where  $X = \langle \mathbf{x}\mathbf{x}^T \rangle$ ,  $Y = \langle \mathbf{y}\mathbf{y}^T \rangle$ , and  $C = \langle \mathbf{y}\mathbf{x}^T \rangle$ . The Gaussian model for the distribution of  $\mathbf{z}$  is given by

$$P(\mathbf{z}) = (2\pi)^{-d/2} |Z|^{-1/2} \exp \left[ -\frac{1}{2} \mathbf{z}^T Z^{-1} \mathbf{z} \right]. \quad (5)$$

For Gaussian distributions, small density corresponds to large Mahalanobis distance from the mean, so we can write  $\mathcal{A}(\mathbf{z}) = \mathbf{z}^T Z^{-1} \mathbf{z}$  as a measure of anomalousness. This expression<sup>1</sup> is often referred to as the RX anomaly detector [21].

In the Gaussian case, we can write all of the detectors in Table I with the expression  $\mathcal{A}(\mathbf{z}) = \mathbf{z}^T Q \mathbf{z}$  where the quadratic coefficient matrix is given by

$$Q = \begin{bmatrix} X & C^T \\ C & Y \end{bmatrix}^{-1} - \beta_x \begin{bmatrix} X^{-1} & 0 \\ 0 & 0 \end{bmatrix} - \beta_y \begin{bmatrix} 0 & 0 \\ 0 & Y^{-1} \end{bmatrix}. \quad (6)$$

Again, the different detectors are given by different choices of  $\beta_x$  and  $\beta_y$ . These are not the only quadratic ACD algorithms that have been proposed [12], but they correspond to a selection that can be derived as Gaussian special cases of detectors that are defined for general distributions. This makes it straightforward to generalize them to non-Gaussian EC distributions.

As a convenience, define the following three scalars for the pixel  $(\mathbf{x}, \mathbf{y})$ :

$$\begin{aligned} \xi_x &= \mathbf{x}^T X^{-1} \mathbf{x} \\ \xi_y &= \mathbf{y}^T Y^{-1} \mathbf{y} \\ \xi_z &= \mathbf{z}^T Z^{-1} \mathbf{z}. \end{aligned} \quad (7)$$

Then the anomalousness of change at the pixel  $(\mathbf{x}, \mathbf{y})$  can be expressed as

$$\mathcal{A}(\mathbf{x}, \mathbf{y}) = \xi_z - \beta_x \xi_x - \beta_y \xi_y. \quad (8)$$

where, again, the choice of  $\beta$  parameters specifies the algorithm (as shown in Table I).

### C. Elliptically-contoured (EC) distributions

For elliptically-contoured distributions [22],  $P(\mathbf{z})$  depends on the covariance matrix  $Z$  and can be written

$$P(\mathbf{z}) = |Z|^{-1/2} H(d, \xi_z) \quad (9)$$

where  $|Z|$  is the determinant of  $Z$ ,  $d$  is the dimension of  $\mathbf{z}$ ,  $\xi_z = \mathbf{z}^T Z^{-1} \mathbf{z}$  is a scalar that corresponds to the squared Mahalanobis distance of  $\mathbf{z}$  to the origin, and  $H$  is a positive scalar function. For example,  $H(d, \xi) = (2\pi)^{-d/2} \exp(-\xi/2)$  corresponds to the Gaussian distribution.

If we model our data with an EC distribution, then the anomalousness at pixel  $(\mathbf{x}, \mathbf{y})$  will depend on  $\mathbf{x}$  and  $\mathbf{y}$  only through the scalar values of  $\xi_x$ ,  $\xi_y$ , and  $\xi_z$  defined in Eq. (7). In particular, we can write

$$\mathcal{A}(\mathbf{x}, \mathbf{y}) = h(d, \xi_z) - \beta_x h(d_x, \xi_x) - \beta_y h(d_y, \xi_y) \quad (10)$$

where  $h(d, \xi) = -\log H(d, \xi)$ .

<sup>1</sup>RX is sometimes used to refer to an anomaly detection scheme in which the covariance  $Z$  is estimated locally rather than over the full image. This can be a useful extension, but for the purposes of the present exposition, we will consider statistics estimated from the entire image.

Following Kano [23], we define a *consistent family* of EC distributions as a set of functions  $H(d, \xi)$ , defined for all positive integers  $d$ , with the following property: if  $P(\mathbf{z}) = |Z|^{-1/2} H(d, \xi_z)$ , where  $\mathbf{z} \in \mathbb{R}^d$  is the stacked vector in Eq. (3), and  $\xi_z$  is the scalar defined in Eq. (7); then  $P(\mathbf{x}) = |X|^{-1/2} H(d_x, \xi_x)$  is the marginal distribution associated with the projection of  $\mathbf{z}$  onto the  $d_x < d$  dimensional subspace corresponding to  $\mathbf{x}$ .

In general, if the distribution of  $\mathbf{z}$  is EC, then the marginal distributions of  $\mathbf{x}$  and  $\mathbf{y}$  will also be EC [22], and one can write an explicit expression for these lower dimensional EC functions. Specifically, given  $H(d, \xi)$  for a given  $d$ , and  $d' < d$ , one can write [24]

$$H(d', \xi) = c(d', d) \int_0^\infty w^{(d-d')/2-1} H(d, w + \xi) dw, \quad (11)$$

where  $c(d', d)$  is a normalization constant. However, there is no guarantee that this integral will lead to a closed-form solution. Further, an arbitrary EC distribution is not necessarily expressible as the projection from a higher-dimensional EC distribution.

The Gaussian is an example of a consistent family, and as already seen in Eq. (8), leads to a simple anomalous change detector.

Not all families are consistent. For instance, a popular choice of EC distribution is given by the generalized Gaussian:

$$H(d, \alpha, \gamma, \xi) = c(d, \alpha, \gamma) \exp(-\gamma \xi^\alpha) \quad (12)$$

with  $c(d, \alpha, \gamma)$  the normalization constant. Here  $\alpha = 1$  produces the Gaussian distribution, and  $\alpha < 1$  is a fatter tailed distribution. However the projection of a generalized Gaussian to lower dimension does not produce a generalized Gaussian – that is, it does not satisfy the condition in Eq. (11), and it is *not* a consistent family [23].

Another generalization of the Gaussian which *is* a consistent family is the multivariate- $t$  distribution [14], [23], [24]:

$$H(d, \nu, \xi) = \frac{\Gamma(\frac{d+\nu}{2})}{\Gamma(\frac{\nu}{2}) \pi^{d/2} (\nu-2)^{d/2}} \left(1 + \frac{\xi}{\nu-2}\right)^{-(d+\nu)/2}. \quad (13)$$

This is a fatter tailed distribution than the Gaussian, and it gets fatter as  $\nu$  gets smaller. In fact, as  $\nu \rightarrow 2$ , the variance diverges. The limit  $\nu \rightarrow \infty$  recovers the Gaussian distribution. Not only is Eq. (13) consistent, it is also convenient. It provides a simple closed form expression for all positive integers  $d$ . By substituting the above multivariate- $t$  form into Eq. (10), and dropping unimportant additive constants, we obtain an expression for anomalousness of change that constitutes the main result in this paper.

$$\begin{aligned} \mathcal{A}(\mathbf{x}, \mathbf{y}) &= (d_x + d_y + \nu) \log \left(1 + \frac{\xi_z}{\nu-2}\right) \\ &\quad - \beta_x (d_x + \nu) \log \left(1 + \frac{\xi_x}{\nu-2}\right) \\ &\quad - \beta_y (d_y + \nu) \log \left(1 + \frac{\xi_y}{\nu-2}\right). \end{aligned} \quad (14)$$

Note that as  $\nu \rightarrow \infty$  (and in particular for  $\nu \gg d_x + d_y$ ), and dividing out an irrelevant factor of  $\nu$ , this expression reduces to the Gaussian limit in Eq. (8).



Fig. 1. Broadband image of the 224-channel hyperspectral AVIRIS data used in these experiments. The image is 150×500 pixels, and was obtained from flight f960323t01p02\_r04\_sc01 over the Florida coast.

### III. NUMERICAL EXPERIMENTS

#### A. Simulated pervasive differences and anomalous changes

For our first experiment to compare these EC-based algorithms and their Gaussian-based counterparts, we employed the data and simulation framework used in Ref. [12]. The data set (see Fig. 1) is a 224-channel image of the Florida coast, taken with the AVIRIS sensor [25]. A pervasive change is applied to all the pixels in the image to produce a second image, and principal components analysis is used to reduce the images to  $d = 10$  bands each. (We also considered canonical correlation analysis for dimension reduction, and obtained similar results.) Then, single-pixel anomalous changes are produced by scrambling the pixels in one of the images; this produces anomalous change pixels  $(\mathbf{x}, \mathbf{y})$  whose components  $(x$  and  $y)$  are not individually anomalous [12]. We modify this scheme slightly by randomly partitioning the pixels into training and testing sets, with an equal number of pixels in each set. The training set is used to estimate the covariance matrix  $Z$  and the EC parameter  $\nu$  (see the Appendix for details), and the testing set uses these values to evaluate performance of the algorithm.

Fig. 2(a) plots the resulting ROC curves on semi-logarithmic axes to emphasize the low false alarm rate regime; we remark that for any kind of anomaly or anomalous change detection, the low false alarm rate is usually the regime of most operational interest. Fig. 2(a) confirms previously reported observations [12] that among the Gaussian algorithms, Hyperbolic Anomalous Change Detection (HACD) outperforms the Chronochrome (CC) which outperforms the straight anomaly detection approach (RX). But our main experimental result is that the EC variants of these algorithms outperform their Gaussian counterparts.

#### B. Real pervasive differences, simulated anomalous changes

In an experiment that ran over many months, Eismann *et al.* [4] made a series of co-registered hyperspectral images of the same scene (see Fig. 3). In addition to a grassy field with trees in the background, some panels were also present in the scene. These panels exhibited spectra unlike what was in the rest of the scene and might be considered anomalous, but because they are in both images, they are not anomalous *changes*. A pair of images from this experiment, shown in Fig. 3(a,b), provides an example with real pervasive differences due to seasonal changes from August to October. Following Refs. [2], [4], the data were reduced to  $d = 10$  bands by taking principal

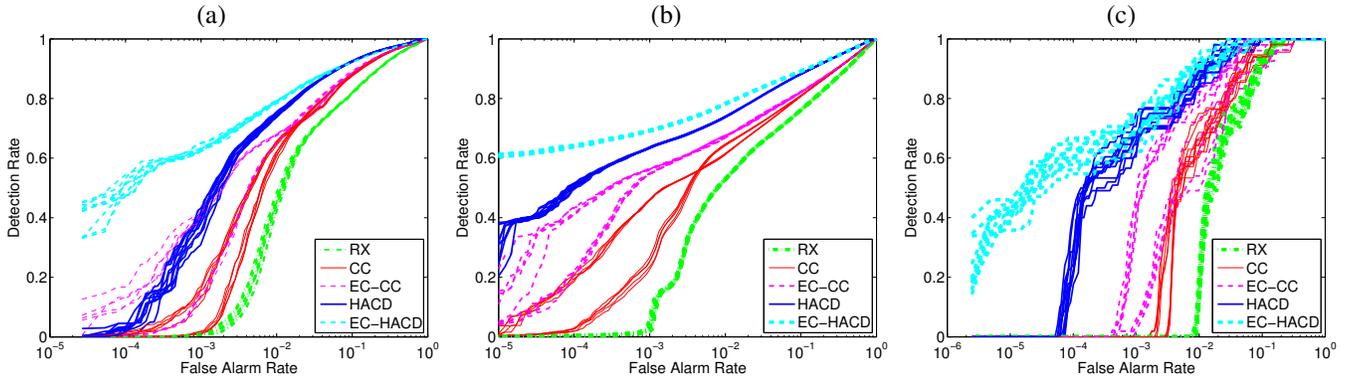


Fig. 2. ROC curves compare the performance of Gaussian-based algorithms (Hyperbolic ACD, Chronochrome (both variants), and RX) with the modification of these algorithms for elliptically contoured distributions. Note that EC-based RX is equivalent to standard RX. For each detector, ten ROC curves are plotted corresponding to different partitions of training and testing data. For the two chronochrome algorithms,  $(y|x)$  and  $(x|y)$ , we plot five ROC curves each, and we see in the CC and EC-CC plots that the ten curves break into two groups, corresponding to the two variants of the chronochrome. (a) Simulated pervasive differences correspond to a smoothed misregistration by one pixel of the AVIRIS data in Fig. 1; simulated anomalous changes are obtained by scrambling the pixels. (b) Real pervasive differences are obtained from separate images taken two months apart, and shown in Fig. 3(a,b); anomalous changes are simulated. (c) Real pervasive differences *and* real anomalous changes are from images shown in Fig. 3(a,c).

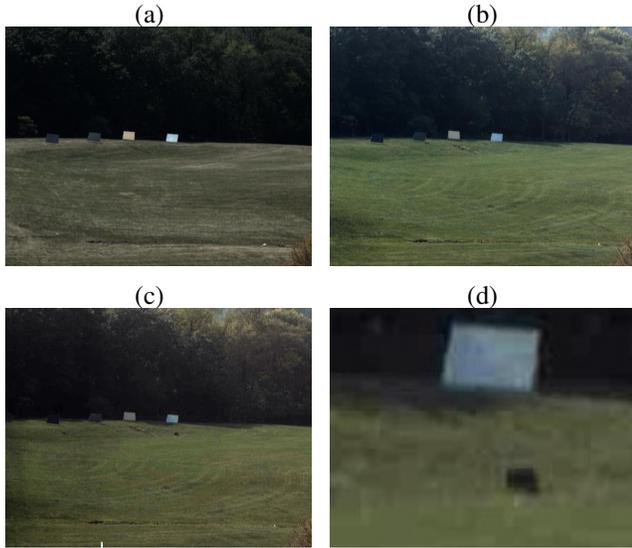


Fig. 3. Image of hyperspectral data described in Refs. [2], [4]. (a) Image taken August 25, 2005. (b) Image taken October 14, 2005. (c) Image taken October 14, 2005; after placing two dark tarps on the grass. (d) Inset from panel (c), showing one of the stationary panels and one of the emplaced tarps.

components from the data shown in Fig. 3(a). We used the simulation framework to introduce anomalous changes and computed ROC curves, as seen in Fig. 2(b). The results agree with those seen in the full simulation in Section III-A: HACD beat CC beat RX, but more to the point, EC-HACD beat HACD and EC-CC beat CC.

### C. Real pervasive differences, real anomalous changes

The advantage of the simulation framework is that it produces enough anomalous changes to permit good statistical comparisons. The disadvantage is that those anomalous changes are simulated. Fig. 3(a,c) shows two images, taken two months apart, but with anomalous changes (two folded tarps) introduced into the second image. Although results

with real anomalies (shown in Fig. 2(c)) are necessarily anecdotal, they roughly confirm what was observed for the simulated anomalies: again, in the low false alarm rate regime, the EC-based change detectors outperformed their Gaussian counterparts.

## IV. CONCLUSION

Because anomalies are rare, it is the tail of the distribution that matters most in modeling data for anomaly and anomalous change detection. While EC distributions share many of the useful properties of Gaussian distributions (specifically, that they are characterized primarily by a covariance matrix), they can have fatter tails than Gaussian distributions, and this accords more with what is observed in hyperspectral imagery. Substantially better anomalous change detection performance was obtained by replacing the Gaussian distribution with the multivariate- $t$  distribution.

## ACKNOWLEDGMENTS

We are grateful to Joseph Meola and Micheal Eismann for generously providing the hyperspectral change data in Fig. 3. This work was supported by the Los Alamos Laboratory Directed Research and Development (LDRD) program.

## REFERENCES

- [1] Remote Sensing Data Fusion Contest, 2009; <http://tlclab.unipv.it/dftc/home.do>.
- [2] J. Meola and M. T. Eismann, "Image misregistration effects on hyperspectral change detection," *Proc. SPIE*, vol. 6966, p. 69660Y, 2008.
- [3] J. Theiler, "Sensitivity of anomalous change detection to small misregistration errors," *Proc. SPIE*, vol. 6966, p. 69660X, 2008.
- [4] M. T. Eismann, J. Meola, and R. Hardie, "Hyperspectral change detection in the presence of diurnal and seasonal variations," *IEEE Trans. Geoscience and Remote Sensing*, vol. 46, pp. 237–249, 2008.
- [5] A. Schaum and A. Stocker, "Long-interval chronochrome target detection," *Proc. 1997 International Symposium on Spectral Sensing Research*, 1998.
- [6] C. Clifton, "Change detection in overhead imagery using neural networks," *Applied Intelligence*, vol. 18, pp. 215–234, 2003.
- [7] A. Schaum and A. Stocker, "Linear chromodynamics models for hyperspectral target detection," *Proc. IEEE Aerospace Conference*, pp. 1879–1885, 2003.

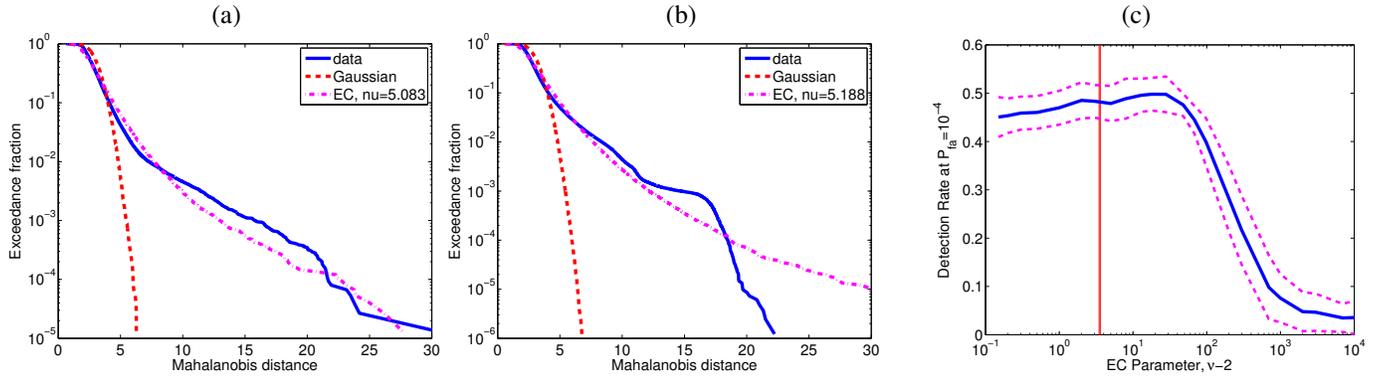


Fig. 4. **(a,b)** Exceedance plot shows the fraction of points in the data set whose Mahalanobis distance  $r$  is larger than the indicated value. This is essentially a cumulative histogram of Mahalanobis distance values, and it provides a useful way to visualize the fat tails of multivariate distributions. The solid line indicates the exceedance of the actual data; the dashed line shows the exceedance plot for Gaussian distributed data of the same dimension, and the dashed-dotted line is for data from a multivariate  $t$  distribution with  $\nu$  fit to the data using the formula in Eq. (16). It is clear that the tail of the distribution is more accurately characterized by the heavy-tailed EC distribution than by the Gaussian distribution. **(a)** Exceedance for AVIRIS data shown in Fig. 1, reduced to  $d = 10$  by principal components analysis. **(b)** Exceedance for  $d = 10$  dimensional data shown in Fig. 3. **(c)** Relative insensitivity to the parameter  $\nu$  of EC-HACD performance, shown here for the AVIRIS data in Fig. 2(a), with the detection rate  $P_d$  obtained when the false alarm rate is  $10^{-4}$ . Based on 100 trials at each value of  $\nu$ , shown is the mean  $P_d$  as a solid line, and one standard deviation shown as dashed lines. The performance does not appreciably change over a wide range of  $\nu$ , but it is much better than the  $\nu \rightarrow \infty$  limit that corresponds to the Gaussian HACD algorithm. The vertical line corresponds to the value of  $\nu = 5.54 \pm 0.08$  estimated by Eq. (16).

- [8] R. Mayer, F. Bucholtz, and D. Scribner, "Object detection by using "whitening/dewhiting" to transform target signatures in multitemporal hyperspectral and multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, pp. 1136–1142, 2003.
- [9] A. A. Nielsen, K. Conradsen, and J. J. Simpson, "Multivariate alteration detection (MAD) and MAF post-processing in multispectral bi-temporal image data: new approaches to change detection studies," *Remote Sensing of the Environment*, vol. 64, pp. 1–19, 1998.
- [10] H. Hotelling, "Relations between two sets of variables," *Biometrika*, vol. 28, pp. 321–327, 1936.
- [11] J. Theiler and S. Perkins, "Proposed framework for anomalous change detection," *ICML Workshop on Machine Learning Algorithms for Surveillance and Event Detection*, pp. 7–14, 2006.
- [12] J. Theiler, "Quantitative comparison of quadratic covariance-based anomalous change detectors," *Applied Optics*, vol. 47, pp. F12–F26, 2008.
- [13] E. Conte and M. Longo, "Characterisation of radar clutter as a spherically invariant random process," *IEE Proc. F (Communications, Radar and Signal Processing)*, vol. 134, pp. 191–197, 1987.
- [14] D. B. Marden and D. Manolakis, "Modeling hyperspectral imaging data," *Proc. SPIE*, vol. 5093, pp. 253–262, 2003.
- [15] A. Schaum, E. Allman, J. Kershenstein, and D. Alexa, "Hyperspectral change detection in high clutter using elliptically contoured distributions," *Proc. SPIE*, vol. 6565, p. 656515, 2007.
- [16] J. Theiler and B. R. Foy, "EC-GLRT: Detecting weak plumes in non-Gaussian hyperspectral clutter using an elliptically-contoured generalized likelihood ratio test," *Proc. IGARSS*, pp. 1–221, 2008.
- [17] T. Hastie, R. Tibshirani, and J. Friedman, *Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer-Verlag, 2001, this anomaly detection approach is developed in Chapter 14.2.4, and neatly illustrated in Fig 14.3.
- [18] J. Theiler and D. M. Cai, "Resampling approach for anomaly detection in multispectral images," *Proc. SPIE*, vol. 5093, pp. 230–240, 2003.
- [19] I. Steinwart, D. Hush, and C. Scovel, "A classification framework for anomaly detection," *J. Machine Learning Research*, vol. 6, pp. 211–232, 2005.
- [20] J. Theiler and S. Perkins, "Resampling approach for anomalous change detection," *Proc. SPIE*, vol. 6565, p. 65651U, 2007.
- [21] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 38, pp. 1760–1770, 1990.
- [22] S. Cambanis, S. Huang, and G. Simons, "On the theory of elliptically contoured distributions," *J. Multivariate Analysis*, vol. 11, pp. 368–385, 1981.
- [23] Y. Kano, "Consistency property of elliptical probability density functions," *J. Multivariate Analysis*, vol. 51, pp. 139–147, 1994.
- [24] E. Gomez-Sanchez-Manzano, M. A. Gomez-Villegas, and J. M. Marin, "Sequences of elliptical distributions and mixtures of normal distributions," *J. Multivariate Analysis*, vol. 97, pp. 295–310, 2006.
- [25] G. Vane, R. O. Green, T. G. Chrien, H. T. Enmark, E. G. Hansen, and W. M. Porter, "The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)," *Remote Sensing of the Environment*, vol. 44, pp. 127–143, 1993.
- [26] D. B. Marden and D. Manolakis, "Using elliptically contoured distributions to model hyperspectral imaging data and generate statistically similar synthetic data," *Proc. SPIE*, vol. 5425, pp. 558–572, 2004.

## APPENDIX

In addition to estimating the covariances  $X$ ,  $Y$ , and  $C$ , use of the change detection formula in Eq. (14) requires an estimate of  $\nu$  as well. Marden and Manolakis [26] considered various goodness-of-fit statistics applied to a histogram of Mahalanobis distances, such as shown in Fig. 4(a,b), and described a sophisticated (and expensive) approach that involved fitting a mixture of concentric EC distributions to the data. We have found that a precise estimate is not necessary, as seen in Fig. 4(c), and that a quick estimate is readily obtained by taking a simple ratio of moments of the distribution. In principle,  $\mathbf{x}$  or  $\mathbf{y}$  might exhibit different values of  $\nu$ ; our practical compromise was to take moments of the distribution of the stacked variable  $\mathbf{z}$ . Let  $r = \xi_z^{1/2} = |\mathbf{Z}^{-1/2}\mathbf{z}| = (\mathbf{z}^T \mathbf{Z}^{-1} \mathbf{z})^{1/2}$  be the scalar magnitude of the whitened pixel value. Define

$$\kappa_m = \langle r^{m+2} \rangle / \langle r^m \rangle \quad (15)$$

and note that

$$\nu = 2 + \frac{m\kappa_m}{\kappa_m - (d+m)} \quad (16)$$

is the value of  $\nu$  that is consistent with this  $\kappa_m$ .

As a practical matter, we use  $m = 1$  as this is the least sensitive to outliers. (Also, the moment  $\langle r^{m+2} \rangle$  is bounded only for  $\nu > m + 2$ .) For Gaussian distributions,  $\kappa_m = d + m$  and the denominator vanishes. In our implementation, if we observe  $\kappa_m \leq d + m$ , then we use the Gaussian anomalousness in Eq. (8) instead of Eq. (14). For the experiments described here, however, that condition was never observed.