LA-UR- 08-8009

Approved for public release;
distribution is unlimited.

Title:	Uncorrelated versus Independent Elliptically-Contoured Distributions for Anomalous Change Detection in Hyperspectral Imagery
Author(s):	James Theiler Clint Scovel Los Alamos National Laboratory
Intended for:	IS&T/SPIE Electronic Imaging conference San Jose, CA 18-22 January 2009

Los Alamos
NATIONAL LABORATORY
EST. 1943

Form 836 (7/06)

Uncorrelated versus Independent Elliptically-Contoured Distributions for Anomalous Change Detection in Hyperspectral Imagery

James Theiler and Clint Scovel

Los Alamos National Laboratory, Los Alamos, NM, USA

ABSTRACT

The detection of actual changes in a pair of images is confounded by the inadvertent but pervasive differences that inevitably arise whenever two pictures are taken of the same scene, but at different times and under different conditions. These differences include effects due to illumination, calibration, misregistration, etc. If the actual changes are assumed to be rare, then one can "learn" what the pervasive differences are, and can identify the deviations from this pattern as the anomalous changes. A recently proposed framework for anomalous change detection recasts the problem as one of binary classification between pixel pairs in the data and pixel pairs that are independently chosen from the two images. When an elliptically-contoured (EC) distribution is assumed for the data, then analytical expressions can be derived for the measure of anomalousness of change. However, these expression are only available for a limited class of EC distributions. By replacing independent pixel pairs with uncorrelated pixel pairs, an approximate solution can be found for a much broader class of EC distributions. The performance of this approximation is investigated analytically and empirically, and includes experiments comparing the detection of real changes in real data.

Keywords: change, hyperspectral, elliptically-distributed, covariance

1. INTRODUCTION

Each pixel in a hyperspectral image consists of a radiance (or, with some processing, reflectance) spectrum with typically hundreds of high signal-to-noise-ratio channels, each channel corresponding to a very narrow wavelength band. Such exquisitely detailed datasets provide an opportunity for precise discrimination of constituent materials from remote platforms. The precision also permits the detection of weak signals, as from subpixel targets or gaseous plumes, in broad area surveys.

Although researchers often bemoan the *deluge* of data provided by hyperspectral imagery, the fact that there is so much information *in each pixel* can actually simplify the analysis. In fact, most algorithms for hyperspectral image analysis effectively treat the image as a "bag of pixels" – with each pixel treated as an independent sample. Although the spatial correlations are nontrival and not inconsiderable, useful analysis can can often be obtained even while neglecting these correlations.

Not only are the pixels treated as if they were independent samples from a common distribution, that distribution is usually assumed to be a multivariate Gaussian. This is simplistic, but it does capture some important aspects of the data. The covariance matrix that characterizes the Gaussian encapsulates the (linear) correlations between every pair of image channels. The dynamical range between the largest and smallest eigenvalues of the covariance matrix can span many orders of magnitude.

One important aspect of hyperspectral image data which is *poorly* captured by the Gaussian model is its behavior on the tails of the distribution. It is widely recognized that the tails of most hyperspectral datasets are much fatter than Gaussian. And since detection of rare signals requires comparison out at the tails of the distribution, this is particularly important for anomaly detection applications.

The problem of "white balance" that bedevils amateur (and professional) photographers is particularly onerous in hyperspectral imagery. The problem is that the observed spectrum for a given material (whose "actual color" is fixed) will be different when viewed under even slightly different conditions (of illumination, sensor calibration, atmospheric distortion, *etc.*). For target detection applications, this means that the effective target signatures vary from image to image.[1] And for the change detection problem, it confounds the search for actual changes because under different conditions, the spectrum of every pixel changes.

2. MACHINE LEARNING FRAMEWORK FOR ANOMALOUS CHANGE DETECTION

Given two images, call them the x-image and the y-image, the aim is to find those few pixels for which the x-to-y change is unusual compared to the changes exhibited by the rest of the pixels.

Let $\mathbf{x} \in \mathcal{R}^{d_x}$ denote a pixel* in the x-image, and $\mathbf{y} \in \mathcal{R}^{d_y}$ be the corresponding pixel in the y-image. It is assumed that the images are registered (*i.e.,* that corresponding pixels $\mathbf{x}$ and $\mathbf{y}$ correspond to the same location in the scene), but it is acknowledged that this registration is not always precise.[2,3] Here, we write d_x and d_y as the number of spectral channels in the x-image and y-image respectively. We seek a function $\mathcal{A}(\mathbf{x}, \mathbf{y})$ which quantifies the "anomalousness" of the change that has occurred at this pixel location.

In the machine learning framework introduced in Ref. 4, the data is modelled as random samples from a probability distribution $P(\mathbf{x}, \mathbf{y})$. In this model, straight anomaly detection seeks points $\mathbf{x}, \mathbf{y}$ on the "tail" of the distribution; that is, where $P(\mathbf{x}, \mathbf{y})$ is small. But straight anomaly detection identifies pixels where $\mathbf{x}$ and $\mathbf{y}$ are individually unusual (*e.g.,* they might correspond to particularly dark or particularly bright pixels), as well as pixels where the *relationship* between $\mathbf{x}$ and $\mathbf{y}$ is unusual. If we write $P(\mathbf{x})$ as the distribution just of the pixels in the x-image, then this $P(\mathbf{x})$ will be the marginal distribution of $P(\mathbf{x}, \mathbf{y})$. We can similarly write $P(\mathbf{y})$ as the distribution of pixels in the y-image. Then the product $P(\mathbf{x})P(\mathbf{y})$ describes a distribution of $\mathbf{x}$ and $\mathbf{y}$ values that are independent of each other. When $P(\mathbf{x})P(\mathbf{y})$ is small, then we have that either $\mathbf{x}$ or $\mathbf{y}$ (or both) are individually unusual, without saying anything about whether the relationship between them is unusual. In seeking unusual relationships, then, it was proposed[4] to consider the ratio

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

When this ratio is small, then we say that the x-to-y *change* is unusual, regardless of whether $\mathbf{x}$ or $\mathbf{y}$ are individually unusual. Thus, we isolate the notion of anomalous change from straight anomaly.

2.1. Gaussian Model

The ratio in Eq. (1) takes a simple form when the distribution is modelled as a multivariate Gaussian. In general, a d-dimensional Gaussian depends on a d-dimensional mean $\boldsymbol{\mu} \in \mathcal{R}^d$ and a covariance matrix $K \in \mathcal{R}^{d \times d}$. We can write

$$\boldsymbol{\mu} = \langle \mathbf{z} \rangle \text{, and} \tag{2}$$

$$K = \langle (\mathbf{z} - \boldsymbol{\mu})(\mathbf{z} - \boldsymbol{\mu})^T \rangle \tag{3}$$

where the angle brackets denote a mean over the distribution (in practice, these quantities are estimated by taking a sample average over the data), and the superscript T denotes a matrix transpose. The density of the distribution at a point $\mathbf{z} \in \mathcal{R}^d$ is given by (*e.g.,* see Eq. (2.5) in Kay[5])

$$P(\mathbf{z}) = (2\pi)^{-d/2} \, |K|^{-1/2} \exp\left[-\frac{1}{2} (\mathbf{z} - \boldsymbol{\mu})^T K^{-1} (\mathbf{z} - \boldsymbol{\mu}) \right]. \tag{4}$$

It is useful to more specifically identify a stacked vector

$$\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \in \mathcal{R}^{d = d_x + d_y} \tag{5}$$

which denotes a corresponding pixel pair in both images. This leads to

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix} \tag{6}$$

*When we say "pixel" we are usually referring to the observed radiance spectrum observed at that pixel.

where $\boldsymbol{\mu}_x = \langle \mathbf{x} \rangle$ and $\boldsymbol{\mu}_y = \langle \mathbf{y} \rangle$. Further, we can write

$$K = \begin{bmatrix} X & C^T \\ C & Y \end{bmatrix} \tag{7}$$

where

$$\begin{aligned} X &= \left\langle (\mathbf{x} - \boldsymbol{\mu}_x)(\mathbf{x} - \boldsymbol{\mu}_x)^T \right\rangle, & (8) \\ Y &= \left\langle (\mathbf{y} - \boldsymbol{\mu}_y)(\mathbf{y} - \boldsymbol{\mu}_y)^T \right\rangle, \text{ and} & (9) \\ C &= \left\langle (\mathbf{y} - \boldsymbol{\mu}_y)(\mathbf{x} - \boldsymbol{\mu}_x)^T \right\rangle. & (10) \end{aligned}$$

The marginal distributions $P(\mathbf{x})$ and $P(\mathbf{y})$ are also Gaussian, and are given by

$$\begin{aligned} P(\mathbf{x}) &= (2\pi)^{-d_x/2} \, |X|^{-1/2} \exp\left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_x)^T X^{-1} (\mathbf{x} - \boldsymbol{\mu}_x) \right], \text{ and} & (11) \\ P(\mathbf{y}) &= (2\pi)^{-d_y/2} \, |Y|^{-1/2} \exp\left[-\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu}_y)^T Y^{-1} (\mathbf{y} - \boldsymbol{\mu}_y) \right]. & (12) \end{aligned}$$

Finally, we can combine Eq. (4) with Eq. (11) and Eq. (12) to express the ratio in Eq. (1)

$$\begin{aligned} \frac{P(\mathbf{x}, \mathbf{y})}{P(\mathbf{x})P(\mathbf{y})} &= \frac{(2\pi)^{-d/2} \, |K|^{-1/2} \exp\left[-\frac{1}{2} (\mathbf{z} - \boldsymbol{\mu})^T K^{-1} (\mathbf{z} - \boldsymbol{\mu}) \right]}{(2\pi)^{-(d_x + d_y)/2} \, |X|^{-1/2} \, |Y|^{-1/2} \exp\left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_x)^T X^{-1} (\mathbf{x} - \boldsymbol{\mu}_x) - \frac{1}{2} (\mathbf{y} - \boldsymbol{\mu}_y)^T Y^{-1} (\mathbf{y} - \boldsymbol{\mu}_y) \right]} \\ &= \left[\frac{|K|}{|X| \, |Y|} \right]^{-1/2} \exp\left[-\frac{1}{2} (\xi_z - \xi_x - \xi_y) \right] & (13) \end{aligned}$$

where the three scalars ξ_x, ξ_y and ξ_z are given by

$$\begin{aligned} \xi_x &= (\mathbf{x} - \boldsymbol{\mu}_x)^T X^{-1} (\mathbf{x} - \boldsymbol{\mu}_x), & (14) \\ \xi_y &= (\mathbf{y} - \boldsymbol{\mu}_y)^T Y^{-1} (\mathbf{y} - \boldsymbol{\mu}_y), \text{ and} & (15) \\ \xi_z &= (\mathbf{z} - \boldsymbol{\mu})^T K^{-1} (\mathbf{z} - \boldsymbol{\mu}). & (16) \end{aligned}$$

Since the most anomalous changes occur when the ratio $P(\mathbf{x}, \mathbf{y}) / [P(\mathbf{x})P(\mathbf{y})]$ is small, we can write an expression for anomalousness

$$\mathcal{A}'(\mathbf{x}, \mathbf{y}) = -2 \log \left[\frac{P(\mathbf{x}, \mathbf{y})}{P(\mathbf{x})P(\mathbf{y})} \right] = \log \left[\frac{|K|}{|X| \, |Y|} \right] + \xi_z - \xi_x - \xi_y. \tag{17}$$

Since $|K|$, $|X|$, and $|Y|$ are constants that do not depend on $\mathbf{x}$ or $\mathbf{y}$, we can also write

$$\mathcal{A}(\mathbf{x}, \mathbf{y}) = \mathcal{A}'(\mathbf{x}, \mathbf{y}) - \log \left[\frac{|K|}{|X| \, |Y|} \right] = \xi_z - \xi_x - \xi_y \tag{18}$$

as a measure of anomalousness. Equivalently, $\mathcal{A}(\mathbf{x}, \mathbf{y}) = (\mathbf{z} - \boldsymbol{\mu})^T Q (\mathbf{z} - \boldsymbol{\mu})$, where the quadratic coefficient matrix is given by

$$Q = \begin{bmatrix} X & C^T \\ C & Y \end{bmatrix}^{-1} - \begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix}^{-1}. \tag{19}$$

2.2. Elliptically-Contoured Distributions

The class of elliptically-contoured (EC) distributions has found utility both for radar[6] and hyperspectral imagery.[7,8] These distributions depend on the covariance matrix K and are of the form

$$P(K; \mathbf{z}) = |K|^{-1/2} H(d, \xi_z) \tag{20}$$

where $|K|$ is the determinant of the covariance matrix K, and H is a function which depends on the dimension d of the vector $\mathbf{z}$, and on $\mathbf{z}$ via the scalar $\xi_z = (\mathbf{z} - \boldsymbol{\mu})^T K^{-1} (\mathbf{z} - \boldsymbol{\mu})$, which is the squared Mahalanobis distance to the centroid of the data. The function H is a monotonically decreasing scalar function of ξ_z. Note that for the Gaussian distribution, $H(d, \xi) = (2\pi)^{-d/2} e^{-\xi/2}$. It bears remarking that this is not a single distribution, but a family of distributions indexed by the dimension d.

2.2.1. Consistent families

A consistent family $H(d, \xi_z)$ has the property that: if $P(\mathbf{z}) = |K|^{-1/2} H(d, \xi_z)$, where $\mathbf{z}$ is the stacked vector in Eq. (5), then $P(\mathbf{x}) = |X|^{-1/2} H(d_x, \xi_x)$ is the marginal distribution assocatited with the projection of $\mathbf{z}$ onto the $d_x < d$ dimensional subspace corresponding to $\mathbf{x}$. Given $H(d, \xi)$ for a given d, there exists a consistent family of lower dimensional distributions,[9] given by

$$H(d', \xi) = c(d', d) \int_0^\infty w^{(n-p)/2-1} H(d, w + \xi)\, dw, \tag{21}$$

where $c(d', d)$ is a scalar constant that ensures that the distribution is normalized.

For a consistent family, we can write an explicit expression for the ratio in Eq. (1):

$$\frac{P(\mathbf{x}, \mathbf{y})}{P(\mathbf{x})P(\mathbf{y})} = \left[\frac{|K|}{|X|\,|Y|}\right]^{-1/2} \frac{H(d, \xi_z)}{H(d_x, \xi_x) H(d_y, \xi_y)} \tag{22}$$

and from that derive a closed-form expression for anomalousness $\mathcal{A}(\mathbf{x}, \mathbf{y})$.

The Gaussian is an example of a consistent family, and as already seen in Eq. (18), provides a simple anomalous change detector. Another example of a consistent family is the multivariate-t statistic.[7,9] Here,

$$H(d, \xi) = \frac{\Gamma\left(\frac{d+\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \pi^{d/2} (\nu - 2)^{d/2}} \left(1 + \frac{\xi}{\nu - 2}\right)^{-(d+\nu)/2}. \tag{23}$$

This is a fatter tailed distribution than the Gaussian, and it gets fatter as ν gets smaller. In fact, as $\nu \to 2$, the variance diverges. The limit $\nu \to \infty$ recovers the Gaussian distribution.

The multivariate-t leads to an anomalousness measure

$$\mathcal{A}(\nu; \mathbf{x}, \mathbf{y}) = (d_x + d_y + \nu) \log(\xi_z + \nu - 2) - (d_x + \nu) \log(\xi_x + \nu - 2) - (d_y + \nu) \log(\xi_y + \nu - 2). \tag{24}$$

that may be more effective when the data is fatter tailed than Gaussian.

One simplifying limit takes place for $d_x = d_y \gg \nu$ and $\nu \to 2$:

$$\mathcal{A}(\mathbf{x}, \mathbf{y}) = \frac{\xi_z}{\sqrt{\xi_x \xi_y}}. \tag{25}$$

Although the expression in Eq. (21) ensures the *existence* of a family of distributions, it does not say that the family can be expressed in a tractable closed-form. For instance, a popular choice of EC distribution is given by the generalized gaussian:

$$H(d, \xi) = c(d, \beta, \gamma) \exp(-\gamma \xi^\beta) \tag{26}$$

with $c(d, \beta, \gamma)$ a scalar constant. Here $\beta = 1$ produces the Gaussian distribution, and $\beta < 1$ is a fatter tailed distribution. However the generalized gaussian does not satisfy the condition in Eq. (21), and it is *not* a consistent family.

One can, however, take the expression in Eq. (26) for a specific value of d and derive a consistent family of distributions for other values of d, but the corresponding expressions for these other values of d will not have the nice form in Eq. (26).[†]

[†]We can show that it *is* possible to create a family of distributions which for even d take the form $H(d, \xi) \sim G(d, \beta, \gamma; \xi) \exp(-\gamma \xi^\beta)$, where $G(d, \beta; \xi)$ is a polynomial in ξ^β. But to find the members of this family for odd d, one must fall back on the less tractable expression in Eq. (21).

3. UNCORRELATION AS AN APPROXIMATION TO INDEPENDENCE

It is for these inconsistent families that we have introduced the concept of uncorrelated versus independent distributions as a denominator in the ratio in Eq. (1). Here, if

$$P(K;\mathbf{x}) = P\left(\left[\begin{array}{cc} X & C^T \\ C & Y \end{array}\right];\mathbf{x},\mathbf{y}\right) \tag{27}$$

is an EC distribution for the stacked vector $\mathbf{z} = \left[\begin{array}{c} \mathbf{x} \\ \mathbf{y} \end{array}\right]$, then we will approximate the product of the marginals $P(\mathbf{x})P(\mathbf{y})$ which corresponds to independence of $\mathbf{x}$ and $\mathbf{y}$ with an EC distribution in which $\mathbf{x}$ and $\mathbf{y}$ are merely uncorrelated:

$$P(\mathbf{x})P(\mathbf{y}) \approx P_u(\mathbf{x},\mathbf{y}) = P\left(\left[\begin{array}{cc} X & 0 \\ 0 & Y \end{array}\right];\mathbf{x},\mathbf{y}\right). \tag{28}$$

That is,

$$P_u(\mathbf{x},\mathbf{y}) = \left|\begin{array}{cc} X & 0 \\ 0 & Y \end{array}\right|^{-1/2} H(d,\xi'), \tag{29}$$

where

$$\xi' = (\mathbf{z}-\boldsymbol{\mu})^T \left[\begin{array}{cc} X & 0 \\ 0 & Y \end{array}\right]^{-1} (\mathbf{z}-\boldsymbol{\mu}) = \xi_x + \xi_y. \tag{30}$$

The anomalousness measure then varies inversely with the modified ratio

$$\frac{P(\mathbf{x},\mathbf{y})}{P_u(\mathbf{x},\mathbf{y})} = \left[\frac{|K|}{|X||Y|}\right]^{-1/2} \frac{H(d,\xi_z)}{H(d,\xi_x+\xi_y)}. \tag{31}$$

This can be applied to any H. For the Gaussian case, we have that $P_u(\mathbf{x},\mathbf{y}) = P(\mathbf{x})P(\mathbf{y})$ exactly, and we get the same anomalousness measure this gives the same result as the independent case in Eq. (18). For the multivariate-t in Eq. (23), we obtain a result that is different from Eq. (24); we have the simpler result that

$$\mathcal{A}(\nu;\mathbf{x},\mathbf{y}) = \frac{\xi_z+\nu-2}{\xi_x+\xi_y+\nu-2}. \tag{32}$$

Here, it is interesting that there is no dependence on d_x or d_y, and that the $\nu \to 2$ limit produces

$$\mathcal{A}(\mathbf{x},\mathbf{y}) = \frac{\xi_z}{\xi_x+\xi_y}, \tag{33}$$

which recalls Eq. (25), except that the harmonic mean is replaced by an arithmetic mean. We also remark that the $\nu \to \infty$ limit of Eq. (32) leads to the Gaussian anomolousness measure defined in Eq. (18).

Even though the family of generalized Gaussian distributions in Eq. (26) in inconsistent, we can still write a simple expression for the uncorrelation-based anomalous change detector:

$$\mathcal{A}(\beta;\mathbf{x},\mathbf{y}) = \xi_z^\beta - (\xi_x+\xi_y)^\beta. \tag{34}$$

Here $\beta = 1$ corresponds to a Gaussian distribution. In the fat-tailed limit, $\beta \to 0$, and the anomalousness expression becomes $\mathcal{A}(\mathbf{x},\mathbf{y}) = \xi_z/(\xi_x+\xi_y)$ which agrees with the fat-tailed limit of the multivariate-t in Eq. (33).

4. NUMERICAL EXPERIMENTS

To illustrate the utility of EC distributions for anomalous change detection, and in particular to evaluate the effect of approximating independence with uncorrelation in this context, we provide three sets of numerical experiments: pure simulation, a hybrid simulation in which pervasive differences and anomalous changes are artificially generated in a real image, and a real pair of images which exhibits both pervasive differences in the images and some actual changes in the scene.

4.1. Pure simulation

In this experiment we generate two image pairs, one Gaussian and one elliptically contoured with very fat tails. The pervasive differences between the x and y images are encapsulated in the cross-covariance term C. For these single-channel Gaussian images, we take $X = 2$, $Y = 1$, and $C = 1.3$. For the EC distribution, we use a multivariate-t distribution with $\nu = 2.1$; we again take $X = 2$ and $Y = 1$, but for the EC data, we use $C = 1.41$. Note that the closer C is to $\sqrt{2}$, the more nearly deterministic is the relationship between the pixels in the two images.

Following the simulation framework described in Ref. [10], we generate anomalous changes by randomly shuffling the pixels in one of the images.

Fig. 1(a) shows that of the four detectors, the "Hyper" detector, which is optimized for Gaussian distributions, performs best. Also shown are the generalized gaussian (EC-beta) with $\beta = 0.5$, the independence-based multivariate t statistic (EC-indep) with $\nu = 2.1$, and the uncorrelation-based multivariate t (EC-uncorr) again with $\nu = 2.1$. Choosing ν so close to 2 produces an extreme variant of the change detector that is optimized for very fat tails.

When applied to non-Gaussian fat-tailed data, however, the EC detectors outperform Hyper. As shown in Fig. 1(b), the best detector is EC-indep, the detector that is exactly matched to the statistics of the data. But even in this extreme case, we see that the approximation provided by EC-uncorr performs nearly as well as EC-indep. The generalized Gaussian with $\beta = 0.5$ and the Gaussian itself (Hyper) perform less well.

In Fig. 2, we revert to a less highly optimized variant of the EC detector, and use a moderate $\nu = 10$. This figure also includes the performance of some previously described detectors; these detectors are surveyed in Ref. [10], and include the simple difference (SD), the chronochrome (CC) of Schaum and Stocker,[11,12] and a straight anomaly detector (RX) based on Mahalanobis distance from the centroid of the stacked $\mathbf{z}$ space.[13] For the Gaussian data in Fig. 2(a), the performance of all three of the EC detectors is virtually identical to that of Hyper, which is known to be optimal. For the EC data in Fig. 2(b), even though it is extremely fat tailed, the more moderate EC detectors (EC-indep, EC-uncorr) still perform the best, and are virtually identical in their performance. Comparing the moderate EC detectors in Fig. 2(b) with the extreme EC detectors in Fig. 1(b), we see that the moderate detectors perform nearly as well.

4.2. Simulating pervasive differences and anomalous changes

In this hybrid simulation, we begin with real 224-channel hyperspectral data from the AVIRIS sensor.[14,15] See Fig. 3. Using the simulation framework outlined in Ref. 10, we generate pervasive differences by applying some operation to the whole scene. For the results shown in Fig. 4, four operations were considered: multiplicative noise, splitting the image into two 112-channels images, smoothing the image with a 3×3 kernel, and (after smoothing) misregistering the image by one pixel. Anomalous changes are produced by randomly moving a pixel from one part of the image to another. After this simulation, the first ten principal components are used in place of the full image as a dimension reduction measure. In all four cases, the EC detectors outperformed the Gaussian-based detectors, and in particular EC-uncorr and EC-indep performed essentially identically.

4.3. Real anomalies in real imagery

In a long running experiment, Eismann *et al.*[16] made a series of hyperspectral images of a grassy field with trees in the background (see Fig. 5). As well as the grass and trees, four panels were placed in the scene, exhibiting spectra that were unlike that of most of the background, and which might be identified as anomalies in the image. But those panels were kept in place throughout the experiment, so they did not represent anomalous *changes*. Periodically, a pair of tarps would be placed on the grass, and these were the anomalous changes that algorithms were challenged to find. In particular, two images were taken on October 14th, one without and one with the emplaced tarps. The images are 800×1024 pixels, and have 124 spectral channels. Following the approach taken by Meola and Eismann,[2] we reduce that to ten channels each, using the principal components computed from a third image (taken August 25).

Fig. 6 compares the performance of the different detectors applied to this data. Although one should be cautious about drawing conclusions from single examples, we see in this case that the EC detectors outperform

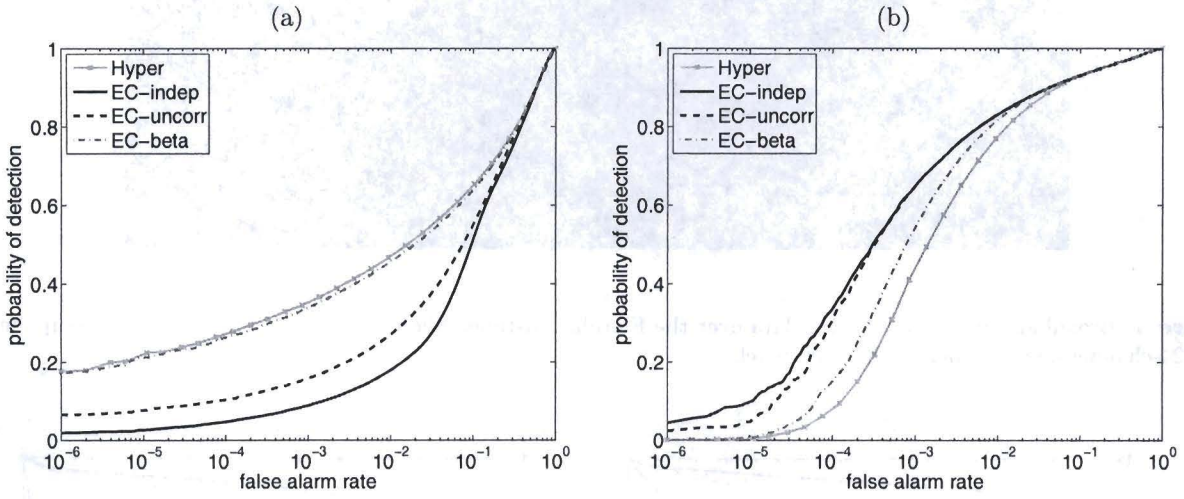

Figure 1. ROC curves for Gaussian and EC detectors applied to simulated data. In both cases, $n = 10^6$ samples are drawn from a single-channel ($d_x = d_y = 1$) distribution specified by covariances X and Y, and cross-covariance C. (a) Gaussian data is generated with $X = 2$, $Y = 1$, and $C = 1.3$. For this data, we see that the Gauss-based anomalous change detector achieves the highest performance. (b) Elliptically-contoured multivariate-t data is generated using $X = 2$, $Y = 1$, and $C = 1.41$, and $\nu = 2.1$; see Eq. (23). Here, we see that the EC detectors outperform the Gaussian detector. We also see that the detector based on the independence formula, in Eq. (24), outperforms (though only slightly) the detector based on the approximate, but simpler, formula in Eq. (32). Both plots also show the performanceof the detector in Eq. (34) with $\beta = 0.5$.

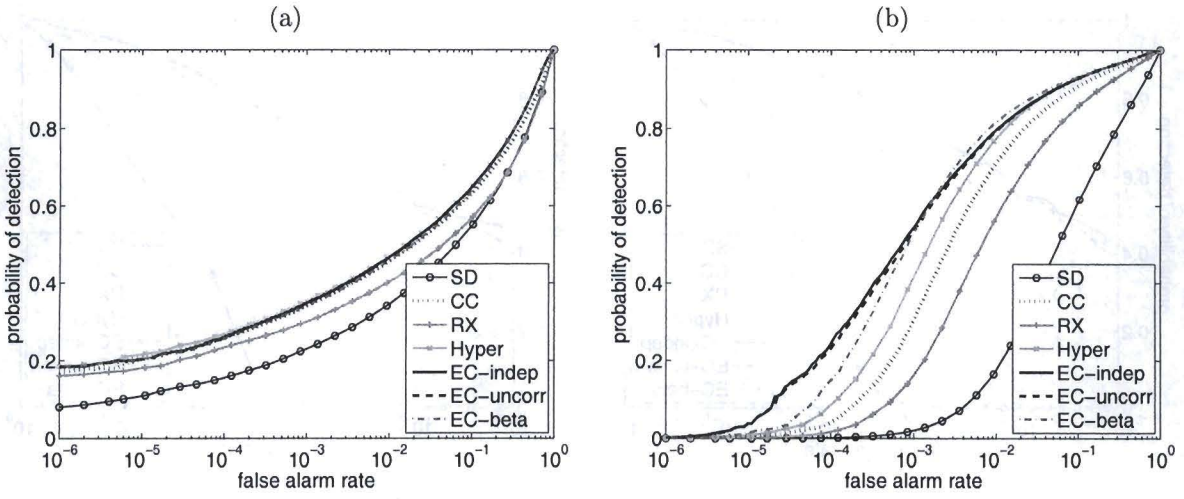

Figure 2. Same data as previous figure, but including other change detection algorithms, and using more moderate value of $\nu = 10$ instead of $\nu = 2.1$ in the EC-indep and EC-uncorr detectors.

Figure 3. Broadband image of AVIRIS data over the Florida coastline. Shown here is the first principal component of the 224-channel image, which is 150×500 pixels.

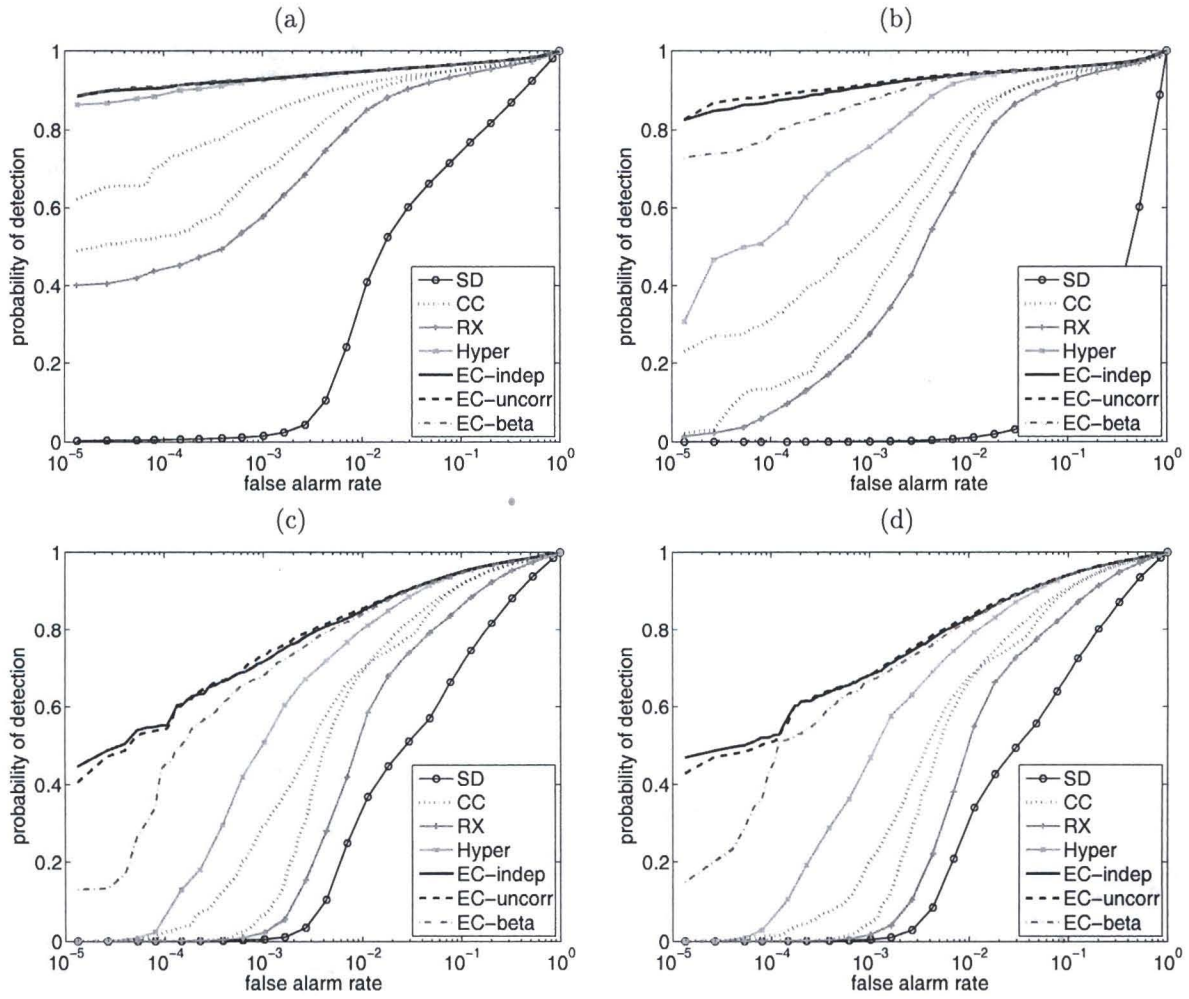

Figure 4. ROC curves produced by various anomalous change detectors, applied to the AVIRIS data shown in Fig. 3 in which pervasive differences have been simulated in four ways: (a) multiplicative noise (each pixel is multiplied by an value randomly chosen from the interval [1,2]; (b) the image bands are split into two groups, one with the first 112 bands of the image, and one with the last 112 bands; (c) the image is smoothed with a 3 × 3 kernel; and (d) the image (after smoothing) is misregistered by one pixel.

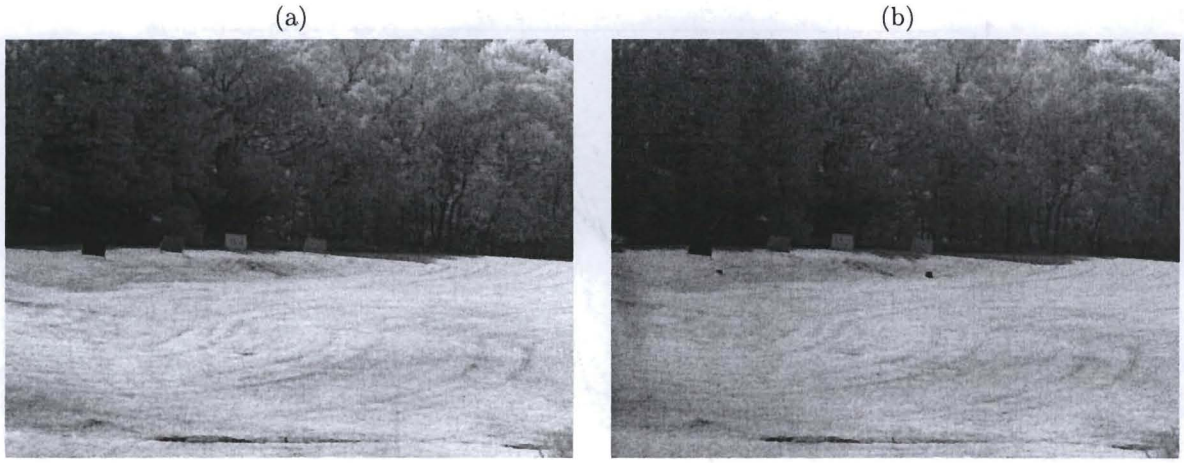

Figure 5. First principal component of two hyperspectral images taken of a grassy field with trees in the background, a set of four panels on the horizon line, and an actual change, evident in the second image as two darker spots near the center of the image.

the Gaussian (Hyper) detector in the very low false-alarm rate regime, but the multivariate-t based detectors (EC-indep and EC-uncorr) perform more poorly in the intermediate false-alarm-rate regime. The detector based on a generalized Gaussian (EC-beta) outperforms Hyper over the whole range. We recall that EC-beta is an uncorrelation-based detector and that we found that a formally optimal generalized Gaussian detector (*i.e.*, one that used independence of $P(\mathbf{x})P(\mathbf{y})$) was intractable, but that the uncorrelation approximation that is so successful on this data is given by the very simple formula in Eq. (34).

5. VISUALIZATION

The simple form given in Eq. (31) ensures that any EC-uncorr anomalous change detector will be a function only of the two scalars ξ_z and $\xi_x + \xi_y$. This permits a two-dimensional visualization of the data, as shown in Fig. 7.

Acknowledgements

We are grateful to Joe Meola and Mike Eismann for providing hyperspectral data with ground-truth changes.

This work was supported by the Laboratory Directed Research and Development (LDRD) program at Los Alamos National Laboratory.

REFERENCES

1. R. Mayer, F. Bucholtz, and D. Scribner, "Object detection by using "whitening/dewhitening" to transform target signatures in multitemporal hyperspectral and multispectral imagery," *IEEE Trans. Geosci. Remote Sens.* **41**, pp. 1136–1142, 2003.
2. J. Meola and M. T. Eismann, "Image misregistration effects on hyperspectral change detection," *Proc. SPIE* **6966**, p. 69660Y, 2008.
3. J. Theiler, "Sensitivity of anomalous change detection to small misregistration errors," *Proc. SPIE* **6966**, p. 69660X, 2008.
4. 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.
5. S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, vol. II, Prentice Hall, New Jersey, 1998.
6. E. Conte and M. Longo, "Characterisation of radar clutter as a spherically invariant random process," *IEE Proc. F (Communications, Radar and Signal Processing)* **134**, pp. 191–197, 1987.

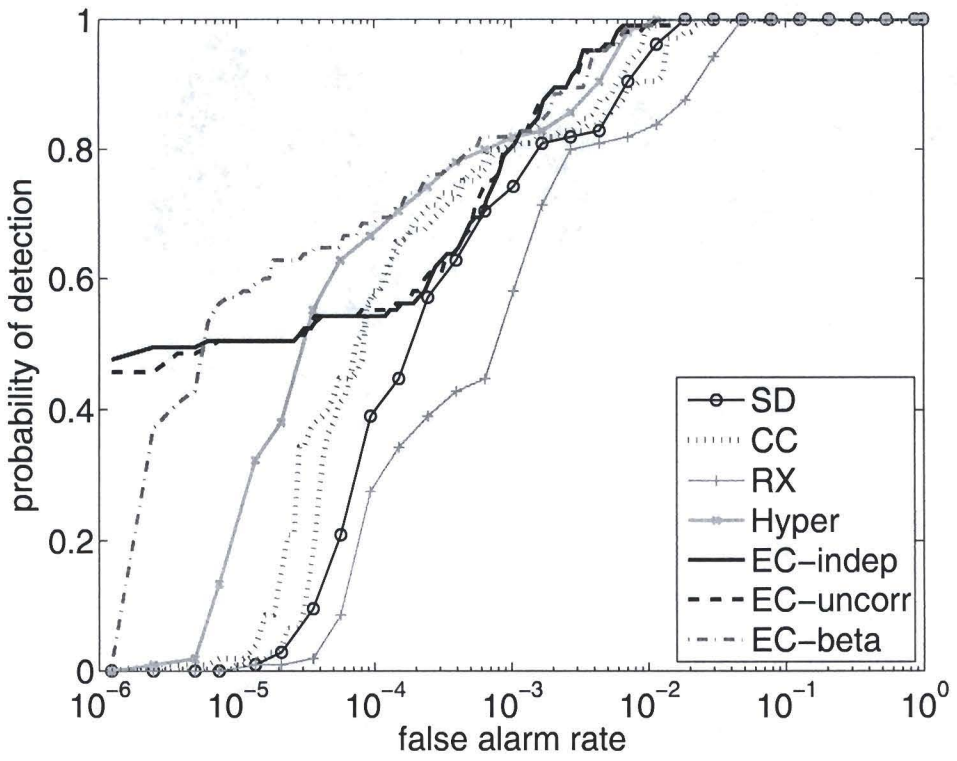

Figure 6. ROC curves for the detection of the actual changes seen in Fig. 5.

7. D. B. Marden and D. Manolakis, "Modeling hyperspectral imaging data," *Proc. SPIE* **5093**, pp. 253–262, 2003.
8. A. Schaum, E. Allman, J. Kershenstein, and D. Alexa, "Hyperspectral change detection in high clutter using elliptically contoured distributions," *Proc. SPIE* **6565**, p. 656515, 2007.
9. E. Gomez-Sanchez-Manzano, M. A. Gomez-Villegas, and J. M. Marin, "Sequences of elliptical distributions and mixtures of normal distributions," *J. Multivariate Analysis* **97**, pp. 295–310, 2006.
10. J. Theiler, "Quantitative comparison of quadratic covariance-based anomalous change detectors," *Applied Optics* **47**, pp. F12–F26, 2008.
11. A. Schaum and A. Stocker, "Spectrally selective target detection," *Proc. International Symposium on Spectral Sensing Research* , 1997.
12. A. Schaum and A. Stocker, "Long-interval chronochrome target detection," *Proc. 1997 International Symposium on Spectral Sensing Research* , 1998.
13. 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* **38**, pp. 1760–1770, 1990.
14. Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), Jet Propulsion Laboratory (JPL), National Aeronautics and Space Administration (NASA) `http://aviris.jpl.nasa.gov/`.
15. AVIRIS Free Standard Data Products, Jet Propulsion Laboratory (JPL), National Aeronautics and Space Administration (NASA). `http://aviris.jpl.nasa.gov/html/aviris.freedata.html`.
16. 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* **46**, pp. 237–249, 2008.
17. S. Cambanis, S. Huang, and G. Simons, "On the theory of elliptically contoured distributions," *J. Multivariate Analysis* **11**, pp. 368–385, 1981.
18. K.-T. Fang and Y.-T. Zhang, *Generalized multivariate analysis*, Springer-Verlag, Science Press, Berlin, Beijing, 1980.

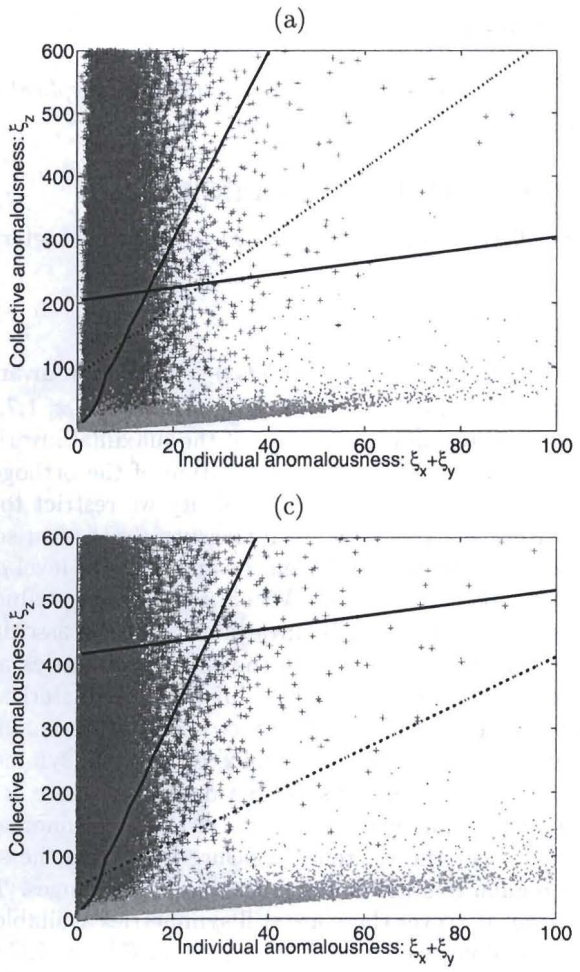

Figure 7. Scatterplots of $\xi_x + \xi_y$ versus ξ_z. The lines correspond to anomalous change detectors, and each has been calibrated to give a false alarm rate of 10^{-4}. The solid lines correspond to the Hyper detector in Eq. (18), the dash-dotted lines correspond to the multivariate-t detector in Eq. (32) with $\nu = 22$, and the dotted line is the fat-tailed limit given in Eq. (33). The data shown is (a) AVIRIS data using band splitting to model different images of the same scene, (b) a Gaussian simulation with the same covariances and cross-covariances as the AVIRIS data, and (c) a simulation with the same covariances and cross-covariances but using a very fat-tailed EC distribution corresponding to $\nu = 3$ in Eq. (23).

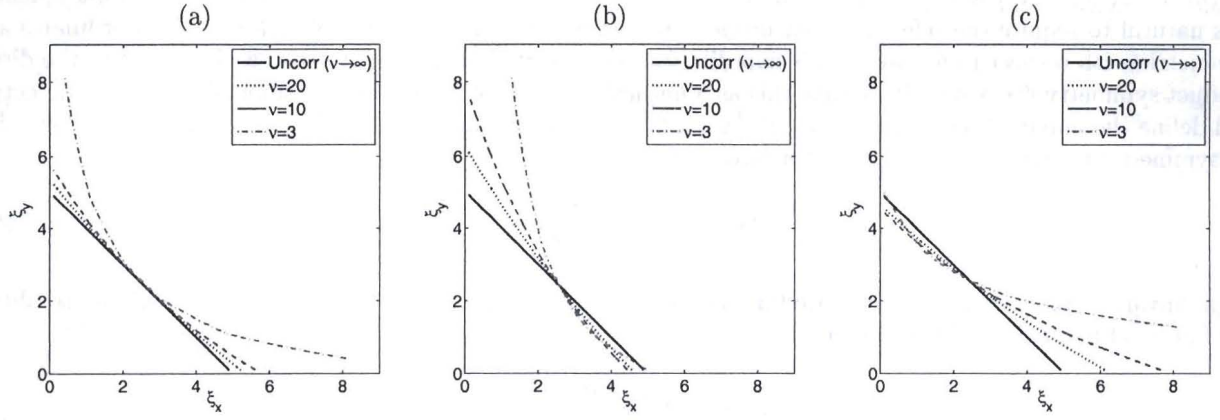

Figure 8. Contours in ξ_x-ξ_y space (for fixed ξ_z) for EC-based ACD algorithms based on the multivariate-t statistic. The three cases shown here correspond to: (a) $d_x = d_y = 10$, (b) $d_x = 10$, $d_y = 3$, and (c) $d_x = 3$, $d_y = 10$. The uncorrelation approximation corresponds to contours in which $\xi_x + \xi_y =$ constant, which agrees with the independence results in the $\nu \to \infty$ limit. Only a single contour is shown, and in all cases it is the the contour that includes the point $\xi_x = \xi_y = 2.5$. The actual levels of interest will depend on the data, but these contours show how the (correct) curved contours are approximated by the straight curve, and that that approximation appears to work best when $d_x = d_y$.

19. I. Steinwart, D. Hush, and C. Scovel, "A classification framework for anomaly detection," *J. Machine Learning Research* **6**, pp. 211–232, 2005.
20. C. Scovel, D. Hush, and I. Steinwart, "Learning rates for density level detection," *Analysis and Applications-Special Issue on Machine Learning* **3**(4), pp. 357–371, 2005.

APPENDIX A. GROUP INVARIANT ANOMALY DETECTION

According to Cambanis et. al.[17] a measure μ on $\mathcal{R}^d$ is an EC distribution with parameter Σ if the characteristic function

$$\phi(t) := \int_{\mathcal{R}^d} e^{it\cdot x} d\mu(x) \tag{35}$$

is a function of $\langle \Sigma t, t\rangle$. When Σ is invertible it is easy to show that the function $t \mapsto \langle \Sigma t, t\rangle$ is a maximal invariant for the representation $g \mapsto \Sigma^{\frac{1}{2}} g \Sigma^{-\frac{1}{2}}, g \in O(d)$ of the orthogonal group $O(d)$ in the sense (see, *e.g.,* Ex. 1.7.1 in Ref. [18]) that any function which is invariant under this representation is a function of the maximal invariant. Consequently, *EC* distributions are simply those which are invariant under a representation of the orthogonal group. Similar statements can be made when the matrix Σ is degenerate but for simplicity we restrict to the nondegenerate case. In the anomaly detection framework of Refs. [19, 20], for a given measure μ one must select a reference measure ν such that μ is absolutely continuous with respect to ν and then the anomolies at level ρ are defined to be the set $\{x : \frac{d\mu}{d\nu}(x) \leq \rho\}$ where $\frac{d\mu}{d\nu}$ is the Radon-Nykodym derivative. When μ is an EC distribution it appears reasonable to assume that the sets of anomlies at any level should be invariant under the associated orthogonal symmetry group. Consequently, the reference measure ν should have the same symmetries as μ. This leads to the conclusion that different choices of reference measure lead only to different parameterization of the level function. Namely that all symmetric anomaly detectors are reparameterizations of Mahalanobis distance. Or said differently, if we look over the set of level parameters ρ there really is only one symmetric anomaly detector- the Mahalanobis distance. Now let us consider anomalous change detection where μ is a measure on $\mathcal{R}^{d_x} \times \mathcal{R}^{d_y}$. If μ is an EC distribution and therefore symmetric, and we wish to detect anomalous changes, then the above argument implies that we must *not* select the reference measure ν to have the same symmetry as μ, for otherwise we would simply be obtaining regular anomalies and not anomalous changes. That is, symmetric anomalous change detection is *symmetry breaking.* However there are still symmetries available. In particular, Thm. 2.6.3 in Ref. [18] implies that the marginal distributions μ_X on $\mathcal{R}^{d_x}$ and μ_Y on $\mathcal{R}^{d_y}$ are EC with parameters Σ_{XX} and Σ_{YY} respectively and therefore they are orthogonally symmetric with the representions $\Phi_X(g) := \Sigma_{XX}^{\frac{1}{2}} g_X \Sigma_{XX}^{-\frac{1}{2}}$ with $g_X \in O(d_x)$ and $\Phi_X(g) := \Sigma_{YY}^{\frac{1}{2}} g_Y \Sigma_{YY}^{-\frac{1}{2}}$ with $g_Y \in O(d_y)$ respectively. Consequently, it is natural to require the reference measure ν to be symmetric with respect to Φ_X in its $\mathcal{R}^{d_x}$ coordinates and symmetric with respect to Φ_Y in its $\mathcal{R}^{d_y}$ coordinates. That is, it should be symmetric with respect to the direct product symmetry $\Phi_X \times \Phi_Y$. If we take this as a natural assumption for symmetric anomalous change detection and define the maximal invariants $\xi_z := \langle \Sigma^{-1} z, z\rangle$, $\xi_x = \langle \Sigma_{XX}^{-1} x, x\rangle$ and $\xi_y = \langle \Sigma_{YY}^{-1} y, y\rangle$, it therefore follows that all symmetric anomalous change detector have the following form

$$\{z = (x, y) | \frac{h(\xi_z)}{g(\xi_x, \xi_y)} \leq \rho\} \tag{36}$$

for a univariate function h and bivariate function g. To simplify one can choose $g(\xi, \eta) := h(\xi + \eta)$ and so obtain the family of anomalous change detectors

$$\{z = (x, y) | \frac{h(\xi_z)}{h(\xi_x + \xi_y)} \leq \rho\} \tag{37}$$

determined by a single univariate function h.