LA-UR- *10-00660*

Title:	Parametric Probability Distributions for Anomalous Change Detection
Author(s):	James Theiler, ISR-2 Bernard R. Foy, C-PCS Brendt Wohlberg, T-5 Clint Scovel, CCS-3 Los Alamos National Laboratory Los Alamos, NM 87545
Intended for:	

Los Alamos
NATIONAL LABORATORY
———— EST.1943 ————

Form 836 (7/06)

Parametric Probability Distributions for Anomalous Change Detection

February 2010

James Theiler, Bernard R. Foy, Brendt Wohlberg, and Clint Scovel

Los Alamos National Laboratory
Los Alamos, NM 87544

ABSTRACT

The problem of anomalous change detection arises when two (or possibly more) images are taken of the same scene, but at different times. The aim is to discount the "pervasive differences" that occur thoughout the imagery, due to the inevitably different conditions under which the images were taken (caused, for instance, by differences in illumination, atmospheric conditions, sensor calibration, or misregistration), and to focus instead on the "anomalous changes" that actually take place in the scene. In general, anomalous change detection algorithms attempt to model these normal or pervasive differences, based on data taken directly from the imagery, and then identify as anomalous those pixels for which the model does not hold. For many algorithms, these models are expressed in terms of probability distributions, and there is a class of such algorithms that assume the distributions are Gaussian. By considering a broader class of distributions, however, a new class of anomalous change detection algorithms can be developed. We consider several parametric families of such distributions, derive the associated change detection algorithms, and compare the performance with standard algorithms that are based on Gaussian distributions. We find that it is often possible to significantly outperform these standard algorithms, even using relatively simple non-Gaussian models.

*"Just because everything is *different* doesn't mean anything has *changed."*
–Irene Peter

1.0   Introduction

Given two or more images of the same scene, taken at different times and under different conditions, what anomalous change detection (ACD) seeks is the "interesting" changes that occurred in the scene. Unfortunately, a mathematics of "interesting" has not been developed[1] so our approach is to identify the rare, or *anomalous changes*. The idea is to distinguish them from the *pervasive differences* that occur throughout the scene (*e.g.*, see Fig. 1) due to disparities in illumination, calibration, registration, look angle, or even the choice of remote sensing platform. They can also be due to diurnal and seasonal variations [2] in the scene.

Part of the motivation for this is the intuition that interesting changes are anomalous. But even when that intuition fails – after all, "anomalous" is not synonymous with "interesting" – then: 1/ since anomalous changes are rare, one will not at least be overwhelmed by uninteresting anomalous changes; and 2/ if pervasive differences are in fact interesting, they will be large enough or plentiful enough that the the analyst can readily find them without the aid of the change detection algorithm.

Anomalous changes are assumed to be relatively rare, and occur in only a small part of the image or image archive. Because the nature of the change is not known beforehand, algorithms for *anomalous* change detection are unsupervised. If the nature of specific changes of interest were already known (and if an adequate and representative sample of those changes were available in the data), then supervised classification might be employed to identify and delineate those changes.

2.0   Probability distributions

In this section, we will derive ACD algorithms in terms of probability distributions that characterize both pervasive differences and anomalous changes. An explicit model for *anomalous* changes sems to defy the meaning of "anomaly" – it is what Rumsfeld would call an unknown unknown [3] – but a number of existing algorithms for anomly detection and anomalous change detection have effectively employed such a model, even if it was not explicitly stated as part of the model.

The use of probability distributions opens up a number of options. The most pragmatic option is pretend these distributions are Gaussian. This leads to simple close-form solutions (and in some cases to well established algorithms), and requires only that covariance matrices be estimated.

The "purist" option is to make no assumptions about the distribution at all. Following Vapnik's dictum [4], we would never model the distribution directly, but instead model the boundary that optimally separates the distribution of anomalous changes from the distribution of pervasive differences, and base this model only on the data that are available. As described in Ref. [5] and illustrated in Fig. 2, samples from the pervasive-differences class are given by the data, while samples from the anomalous-changes class are given by resampling either from the data or from a uniform distribution. This approach does have some theoretical advantages, but can also be expensive and sometimes is problematic on the tails.

We will take a middle ground, and model the data with a non-Gaussian distribution that can be described by a relatively modest number of parameters. Once we fit these parameters to the data, it is straightforward to plug them into our expressions that involve arbitrary distributions and produce anomalous change detectors. To the extent that these parametric distributions are better descriptors of the observed data, we expect that the resulting algorithms will better detect anomalous changes. In particular, since (detectable) anomalous behavior occurs on the tails of distributions, it will be useful to model data with distributions that better describe the tails.

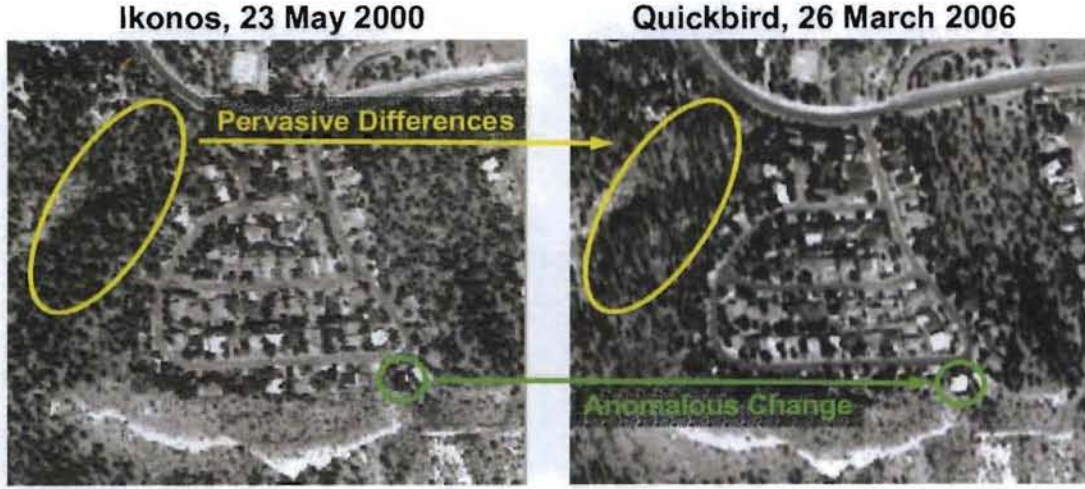[1] Indeed, some would say that "interesting mathematics" is an oxymoron.

Figure 1. Two satellite images of the Camino Redondo neighborhood in Los Alamos, New Mexico, taken roughly six years apart. Illustrated are "pervasive differences" (such as brightness, contrast, shadows and focus) which occur throughout the image, and "anomalous change" (such as the roof that has evidently been replaced) which occurs only once, or in only a small subset of the pixels.

2.1 Arbitrary distributions

Begin with two (approximately) co-registered images which we will call the χ-image and the γ-image. Let $\mathbf{x} \in \mathbb{R}^{d_x}$ be the spectrum of a pixel in the χ-image, and $\mathbf{y} \in \mathbb{R}^{d_y}$ the spectrum of the corresponding pixel in the γ-image. Here, d_x and d_y correspond to the number of spectral channels in the χ-image and γ-image, respective, and that they need not be equal. 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 the "irregular" differences (or *anomalous* changes): these are the pixels $(\mathbf{x}, \mathbf{y})$ for which $P(\mathbf{x}, \mathbf{y})$ is smallest.

Following Refs. [6, 7, 8], 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 $P(\mathbf{x}, \mathbf{y})$. The motivation for a likelihood ratio approach was discussed in Refs. [5, 9].

Since our aim is anomalous change detection, versus straight anomaly detection, we find *conditional* anomalousness a useful concept. Instead of looking for mall values of the *joint* distribution $P(\mathbf{x}, \mathbf{y})$, we can instead use the *conditional* distribution $P(\mathbf{y}|\mathbf{x}) = P(\mathbf{x}, \mathbf{y})/P(\mathbf{x})$. When the pixel value $\mathbf{y}$ is unusual given the value of $\mathbf{x}$, then the conditional distribution will be small. For the multivariate Gaussian case, it can be shown [9] that this formalism leads to the chronochrome detector [10]. 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. There are in fact *two* chronochrome detectors.

A framework for anomalous change detection proposed in Ref. [5]d, leads to the symmetric likelihood ratio

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

Anomalous changes are associated with small values of this ratio. When $P(\mathbf{x}, \mathbf{y})$ is Gaussian, this ratio produces the Hyperbolic Anomalous Change Detector (HACD), so named for the hyperbolic boundary
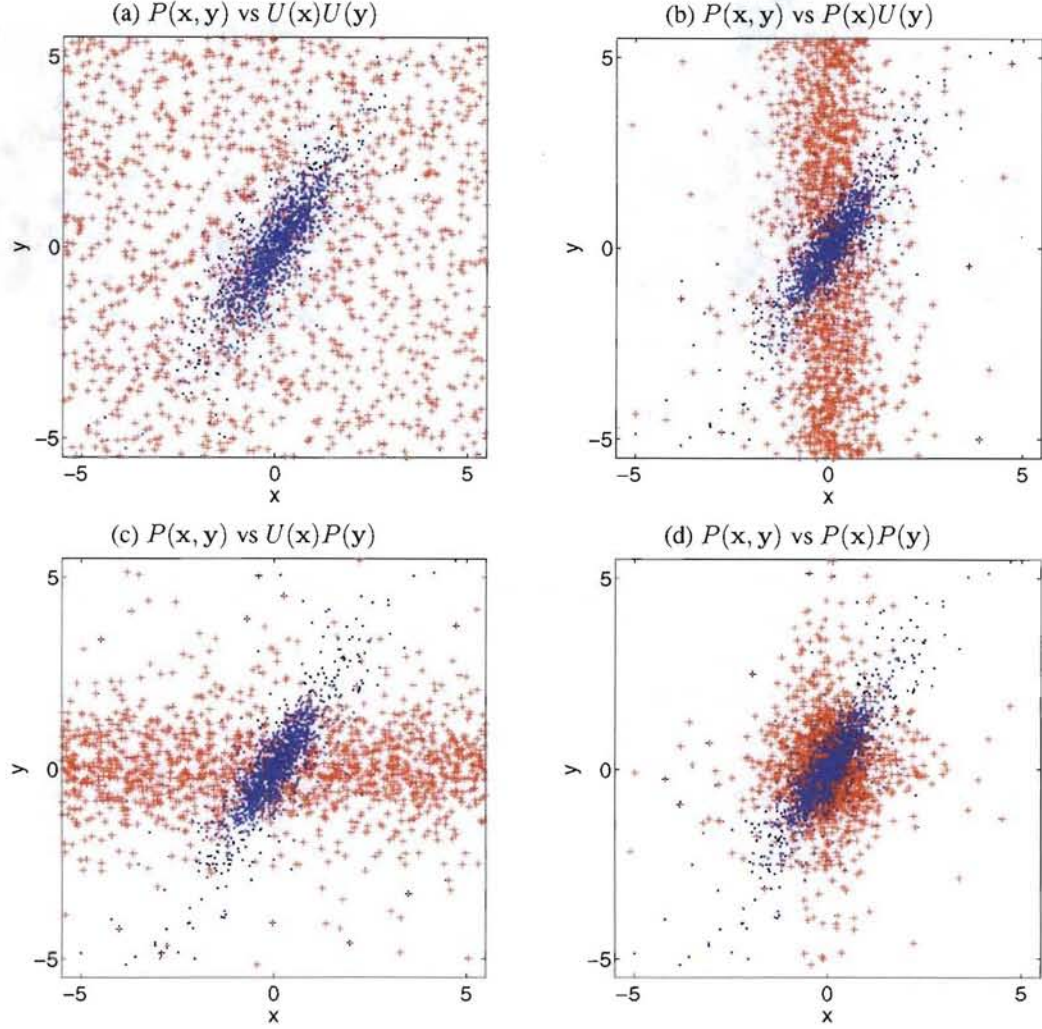
3

Figure 2. Illustrates the resampling approach for generating a background distribution of anomalous changes. Here, the x and y correspond to the two images, and the diagonal swath of blue dots blue to the pervasive changes. The simulated anomalous changes are shown as red pluses and are obtained by resampling from the normal data. (a) $P(\mathbf{x}, \mathbf{y})$ vs $U(\mathbf{x})U(\mathbf{y})$ gives level curves of $P(\mathbf{x}, \mathbf{y})$ and produces the RX-style straight anomaly detector; here the background data is generated by drawing from a uniform distribution. (b) $P(\mathbf{x}, \mathbf{y})$ vs $P(\mathbf{x})U(\mathbf{y})$ produces a generalized chronochrome. Here the x component is randomly sampled from the χ-image, and the y component of the anomalous background is drawn from a uniform distribution. (c) $P(\mathbf{x}, \mathbf{y})$ vs $U(\mathbf{x})P(\mathbf{y})$ produces the "other" generalized chronochrome. (d) $P(\mathbf{x}, \mathbf{y})$ vs $P(\mathbf{x})P(\mathbf{y})$ employs the machine learning framework. Here, x is sampled from the χ-image, and y is independently drawn from the γ-image.
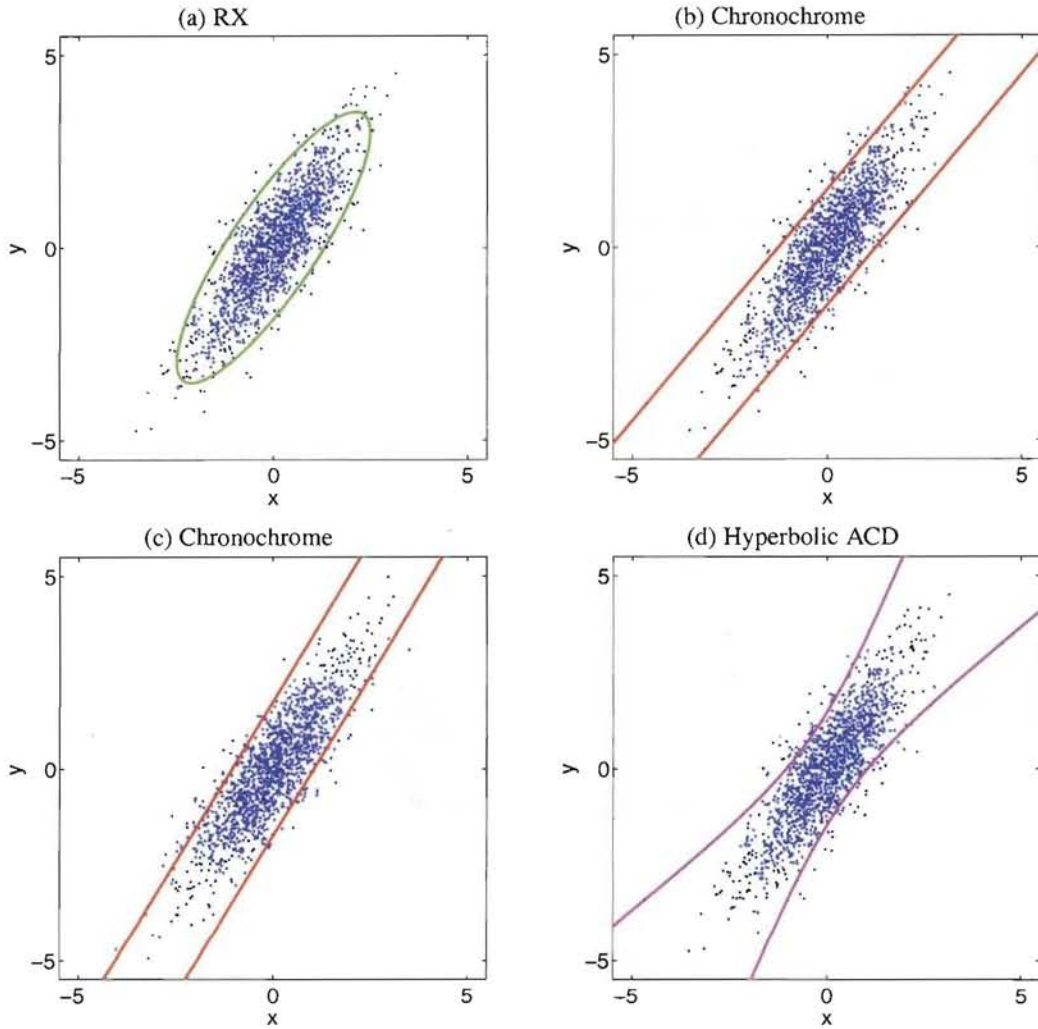
4

Figure 3. The four cases shown in Fig. 2 in the case where $P(\mathrm{x}, \mathrm{y})$ is Gaussian. Since $\mathcal{A}(\mathrm{x}, \mathrm{y})$ is quadratic, the contours will be quadratic surfaces: (a) ellipsoid, with the eigenvalues of the covariance matrix all positive; (b,c) ellipsoidal "tube" with some of the eigenvalues of the covariance matrix strictly equal to zero; (d) hyperboloid, with some of the eigenvalues negative.

5

separating regular from anomalous.

These change detection algorithms have different origins, but each of them can be treated as a ratio of probability densities (or "likelihoods"), where we write $U(\mathbf{x})$ and $U(\mathbf{y})$ to represent uniform density in the $\mathbf{x}$ and $\mathbf{y}$ space. The negative logarithm of the likelihood ratio is large when the likelihood ratio is small, and provides anomalousness measures:

$$
\begin{aligned}
\text{HACD:} \quad & \mathcal{A}(\mathbf{x}, \mathbf{y}) = -\log P(\mathbf{x}, \mathbf{y}) + \log P(\mathbf{x}) + \log P(\mathbf{y}) \\
\text{CC:} \quad & \mathcal{A}(\mathbf{x}, \mathbf{y}) = -\log P(\mathbf{x}, \mathbf{y}) + \log P(\mathbf{x}) \\
\text{CC:} \quad & \mathcal{A}(\mathbf{x}, \mathbf{y}) = -\log P(\mathbf{x}, \mathbf{y}) + \log P(\mathbf{y}) \\
\text{RX:} \quad & \mathcal{A}(\mathbf{x}, \mathbf{y}) = -\log P(\mathbf{x}, \mathbf{y})
\end{aligned}
\tag{2}
$$

2.2 Gaussian distributions

In this subsection, we consider the case that $P(\mathbf{x}, \mathbf{y})$ is a multivariate Gaussian. We can subtract the mean so from here on out, we assume that the distribution is centered at the origin; thus: $\langle \mathbf{x} \rangle = 0$ and $\langle \mathbf{y} \rangle = 0$. It is convenient to introduce

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

with $d = d_x + d_y$, as the pixel in the "stacked" image. Then, we write the covariance matrix for the stacked pixel $\mathbf{z}$ defined in Eq. (3).

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

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

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

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 is the RX anomaly detector [11].

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

$$
\text{HACD:} \quad Q = \begin{bmatrix} X & C^T \\ C & Y \end{bmatrix}^{-1} - \begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix}^{-1}
\tag{6}
$$

$$
\text{CC:} \quad Q = \begin{bmatrix} X & C^T \\ C & Y \end{bmatrix}^{-1} - \begin{bmatrix} X^{-1} & 0 \\ 0 & 0 \end{bmatrix}
\tag{7}
$$

$$
\text{CC:} \quad Q = \begin{bmatrix} X & C^T \\ C & Y \end{bmatrix}^{-1} - \begin{bmatrix} 0 & 0 \\ 0 & Y^{-1} \end{bmatrix}
\tag{8}
$$

$$
\text{RX:} \quad Q = \begin{bmatrix} X & C^T \\ C & Y \end{bmatrix}^{-1}.
\tag{9}
$$

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

$$
\begin{aligned}
\xi_x &= \mathbf{x}^T X^{-1} \mathbf{x} \\
\xi_y &= \mathbf{y}^T Y^{-1} \mathbf{y} \\
\xi_z &= \mathbf{z}^T Z^{-1} \mathbf{z}.
\end{aligned}
\tag{10}
$$

6

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

$$
\begin{aligned}
\text{HACD:} \quad & \mathcal{A}(\mathbf{x}, \mathbf{y}) = \xi_z + \xi_x + \xi_y \\
\text{CC:} \quad & \mathcal{A}(\mathbf{x}, \mathbf{y}) = \xi_z + \xi_x \\
\text{CC:} \quad & \mathcal{A}(\mathbf{x}, \mathbf{y}) = \xi_z + \xi_y \\
\text{RX:} \quad & \mathcal{A}(\mathbf{x}, \mathbf{y}) = \xi_z
\end{aligned}
\tag{11}
$$

These four detectors are illustrated for the case $d_x = d_y = 1$ in Fig. 3.

2.3 Parametric non-Gaussian distributions

We will in particular consider elliptically contoured distributions [12], where $P(\mathbf{z})$ depends on the covariance matrix Z and can be written

$$
P(\mathbf{z}) = |Z|^{-1/2} H(d, \xi_z)
\tag{12}
$$

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. As an 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 ξ_x, ξ_y, and ξ_z defined in Eq. (10). In particular, we can write

$$
\begin{aligned}
\text{HACD:} \quad & \mathcal{A}(\mathbf{x}, \mathbf{y}) = h(d, \xi_z) - h(d_x, \xi_x) - h(d_y, \xi_y) \\
\text{CC:} \quad & \mathcal{A}(\mathbf{x}, \mathbf{y}) = h(d, \xi_z) - h(d_x, \xi_x) \\
\text{CC:} \quad & \mathcal{A}(\mathbf{x}, \mathbf{y}) = h(d, \xi_z) - h(d_y, \xi_y) \\
\text{RX:} \quad & \mathcal{A}(\mathbf{x}, \mathbf{y}) = h(d, \xi_z)
\end{aligned}
\tag{13}
$$

where $h(d, \xi) = -\log H(d, \xi)$. Note that the RX depends only on ξ_z and threfore it is equivalent to the Gaussian RX.

Kano [13] defines 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 ξ_z is the scalar defined in Eq. (10); 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 $\mathbf{x}$.

The Gaussian is an example of a consistent family, and as already seen in Eq. (11), 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:

$$
H(d, \alpha, \gamma, \xi) = c(d, \alpha, \gamma) \exp(-\gamma \xi^\alpha)
\tag{14}
$$

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 and it is *not* a consistent family [13]. It is, in principle, possible to take the expression in Eq. (14) for a specific value of d and derive a consistent family of distributions for smaller values $d' < d$, but the corresponding expressions for these other values of d will not have the nice form in Eq. (14).

A generalization of the Gaussian which *is* a consistent family is the multivariate t distribution [14, 13, 15]:

$$
H(d, \nu, \xi) = \frac{\Gamma\left(\frac{d+\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \pi^{d/2} (\nu - 2)^{d/2}} \left(1 + \frac{\xi}{\nu - 2}\right)^{-(d+\nu)/2}.
\tag{15}
$$

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. Not only is Eq. (15) 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. (13), and dropping unimportant additive constants, we obtain the following expressions for anomalousness of change

$$\text{EC-HACD:} \quad \mathcal{A}(\mathbf{x}, \mathbf{y}) = (d_x + d_y + \nu) \log (\xi_z + \nu - 2) \; -(d_x + \nu) \log (\xi_x + \nu - 2)$$
$$-(d_y + \nu) \log (\xi_y + \nu - 2)$$
$$\text{EC-CC:} \quad \mathcal{A}(\mathbf{x}, \mathbf{y}) = (d_x + d_y + \nu) \log (\xi_z + \nu - 2) \; -(d_x + \nu) \log (\xi_x + \nu - 2) \quad (16)$$
$$\text{EC-CC:} \quad \mathcal{A}(\mathbf{x}, \mathbf{y}) = (d_x + d_y + \nu) \log (\xi_z + \nu - 2) \; -(d_y + \nu) \log (\xi_y + \nu - 2)$$

Note that as $\nu \to \infty$ (and in particular for $\nu \gg d_x + d_y$), and dividing out an irrelevant factor of ν, this expression reduces to the Gaussian limit in Eq. (11).

Another limit of interest is $\nu \to 2$. This simplifies the above expression considerably, in in particular, for $d_x = d_y \gg 2$, and again dropping unimportant constants, we can write:

$$\text{HACD:} \quad \mathcal{A}(\mathbf{x}, \mathbf{y}) = \xi_z^2 / (\xi_x \xi_y)$$
$$\text{CC:} \quad \mathcal{A}(\mathbf{x}, \mathbf{y}) = \xi_z^2 / \xi_x \quad (17)$$
$$\text{CC:} \quad \mathcal{A}(\mathbf{x}, \mathbf{y}) = \xi_z^2 / \xi_y$$

3.0 Validation of Anomaly Detection Algorithms

One problem with validating anomaly detcection algorithms is that anomalies are rare. While there is indeed value to anecdotal examples of real anomalies detected in real images, it is difficult to do statistical analysis. And when using these examples for algorithm development, the dangers of overfitting are considerable.

On the other hand, it is difficult to trust pure simulation; imagery is notoriously difficult to simulate. Even apart from all the physical issues of radiation transfer, atmospheric distortions, sensor noise, *etc.*, there is also the problem of simulating "clutter" in remote sensing imagery. It is asking a lot for a simulation to include the plethora of junk that people leave lying around on the ground.

Our hybrid approach [16] is to start with real data – which will naturally include whatever noise, distortions, and clutter are in that imagery – and introduce our own pervasive differences and anomalous changes. The pervasive differences are produced by applying some operator to *all* of the pixels; the anomalous changes are produced by applying some other operator just to *one* pixel (or in some cases, to a small patch of pixels [17]), as illustrated in Fig. 4. The "trick" is to employ appropriate operators.

3.1 Simulating pervasive differences

One of the best ways to simulate pervasive differences is to use two actual images of the same scene, taken at different times. One drawback to this approach is that there may be anomalous changes in the scene that are not known. Our opinion is that this is a minor issue, since those anomalous changes will be small and/or subtle, and when used for comparing different algorithms, all of the algorithms will be up against the same artifacts. A second drawback is that only one kind of pervasive differences can be examined this way – the pervasive differences that are exhibited in that particular image pair. Both of these drawbacks can be addressed by simulating the pervasive differences. For the simulation, only a single image of a scene is used, and from that single image, a pair of images are produced. One can, for instance, take that image and add noise to it, apply some spatial operator (such as smoothing) to it, modify the brightness or contrast of the image, or spatially translate the image to produced a misregistered pair. For multispectral, and especially hyperspectral, imagery, once can also "split" the spectral bands. For instance, AVIRIS data has 224 spectral bands; one can take the first 112 bands and consider that the first image, and the second 112 bands and call that the second image. This simulates the situation where the two images are taken with different cameras. Finally, one can also take combinations of these.

8

3.2 Simulating anomalous changes

When anomalous changes are simulated, the idea is to make the change only at a single pixel. Having chose which pixel that will be, the anomalous change could be something like a brightening or a darkening of the pixel, but our approach has been to simulate the anomalous changes with another pixel chosen randomly in the image.

The idea is to distinguish anomalous changes from outright anomalies; so at the location where the anomalous change will be simulated, one replaces that pixel with another pixel taken from elsewere in the image. Along these lines, subpixel anomalous changes can be generated by taking a linear combination of the current pixel with the other randomly chosen pixel.

As described, this scheme simulates only a single anomalous change at a time. For computational efficiency, once can produce an entire image of anomalous changes by scrambling the locations of all the pixels. In this scenario, one uses a pair of images produced by the pervasive difference operator to compute covariance matrices and to produce a curve of false alarm rate as a function of ACD threshold. Then, one applies the ACD algorithm to a pair of images in which the same pervasive differences are present but for which one of the image pixels have been scrambled, and one uses this to produce a curve of detection rate versus the same threshold values. Combining these two curves, one produces a ROC curve of detection rate versus false alarm rate.

One further twist: in our experiments, the pixels are randomly partitioned into separate "training" and "testing" sets. The ACD algorithm is trained (*i.e.*, the covariance matrix and other parameters – such as the parameter ν in the multivariate-t distribution – are fit) on the training set, and the curves detailing detection rate and false alarm rate versus thershold are computed on the testing set. We can re-do this with different partitions and this provides an ensemble of ROC curves that provide a sense of how variable the ROC curve estimates are.

3.3 Results

We applied the simulation framework shown in Fig. 4 to an AVIRIS image with 224 channels [18], using two different pervasive change cases, and two different dimension reduction schemes. We found in Fig. 5 that HACD outperformed both chronochromes, both of which in turn outperformed RX. But using the non-Gaussian parametric distribution given by the multivariate t, we found that EC-HACD outperformed HACD and EC-CC outperformed CC. The value of ν was fit using the moment method described in Ref. [19].

In an exercise that ran over many months, Eismann *et al.* [2] took a series of hyperspectral images of the same scene (see Fig. 6). In addition to a grassy field with trees in the background, four 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 the top two panels of Fig. 6 provides an example with real pervasive differences due to seasonal changes from August to October. Following Refs. [2, 1], the data were reduced to $d = 10$ bands by taking principal components from the August dataset. We used the simulation framework to introduce anomalous changes and computed ROC curves, as seen in Fig. 7(a). The results agree with those seen in the full simulation in Fig. 5: HACD beat CC beat RX, EC-HACD beat HACD, and EC-CC beat CC.

As a further check, we considered an example with actual anaomlous changes, as seen in the left two panels of Fig. 6. These two images are taken two months apart, and there are two folded tarps in the second image, which provides anomalous changes. Although results with real anomalies, shown in Fig. 7(b), are necessarily anecdotal, they still confirm what was observed for the simulated anomalies: in the low false alarm rate regime, the EC-based change detectors outperformed their Gaussian counterparts.
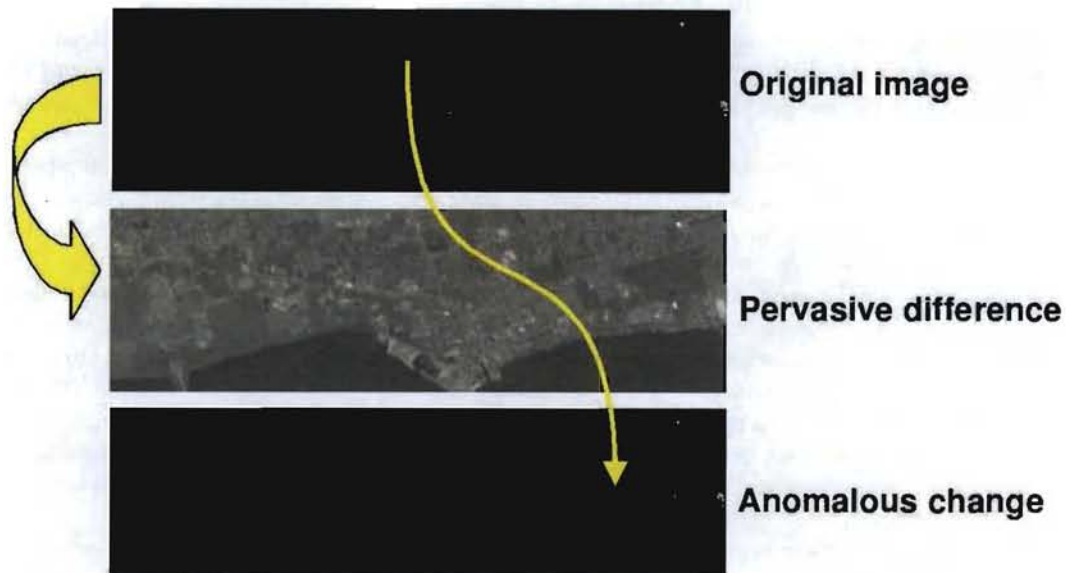
Original image

Pervasive difference

Anomalous change

Figure 4. Simulation framework: pervasive differences are simulated with an operator applied to every pixel in the scene; anomalous changes are simulated with only a single pixel. Our favorite way to simulate anomalous changes is to move a pixel from one part of the scene to another. That way the pixel itself is not anomalous, only the change is.

4.0 Summary

The distinction is made between anomalous changes and pervasive differences, and each of these two classes is modelled by a probability distribution. One *learns* the distribution for the pervasive difference – this is provided by the observed data – and one derives from this a distribution for the anomalous changes. Finally, from these two distributions, one can produce detectors of anomalous changes: these find the pixels in an image pair where the changes are most unusual.

When these distributions are Gaussian, then familiar ACD methods are recovered. But by using a broader (*i.e.*, non-Gaussian) class of distributions, new ACD algorithms can be produced. In particular, observed data generally exhibits much fatter tails than those given by Gaussian distributions. Since it is the tails of the data distribution where the distinction between normal data and outliers is most difficult, better models of the tails have the potential to produce better anomaly detection.
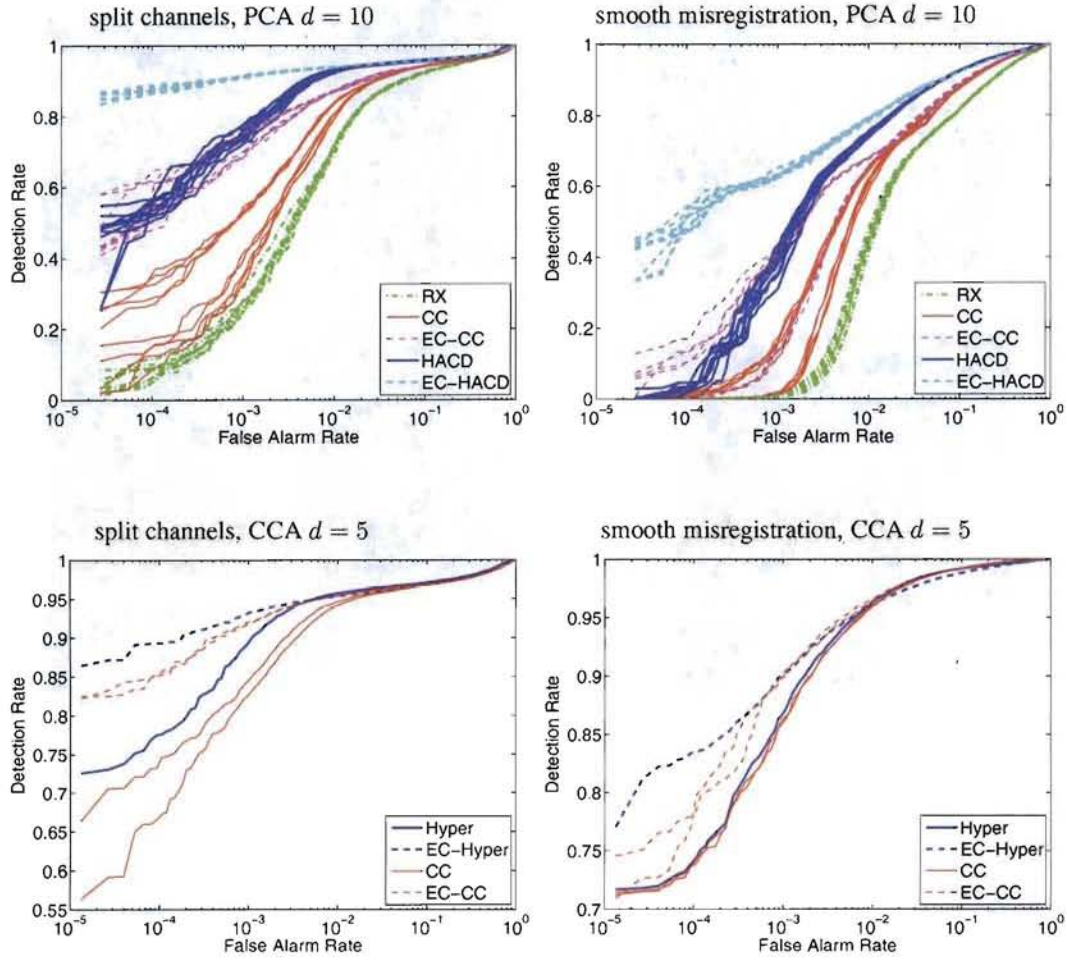
5.0 Acknowledgements

Figure 5. ROC curves for simulated anomalous changes. In the top two panels, these are based on ten trials, each one a different in-sample/out-of-sample partition. The bottom panels are based on a single trial, but use canonoical components analysis (CCA) instead of principal components analysis (PCA) for dimension reduction. In all cases for this experiment, HACD outperformed CC, which in turn outperformed RX. But the main point is that EC-HACD outperforms HACD and EC-CC outperforms CC. We recall that there are two EC-CCs and two CCs, and this is reflected in the bunching of curves in the figures above

taken Aug 25, 2005

taken Oct 14, 2005

taken Oct 14, 2005, after placing two dark tarps on the grass

Figure 6. Images corresponding to the hyperspectral data taken by Eismann *et al.* [2].

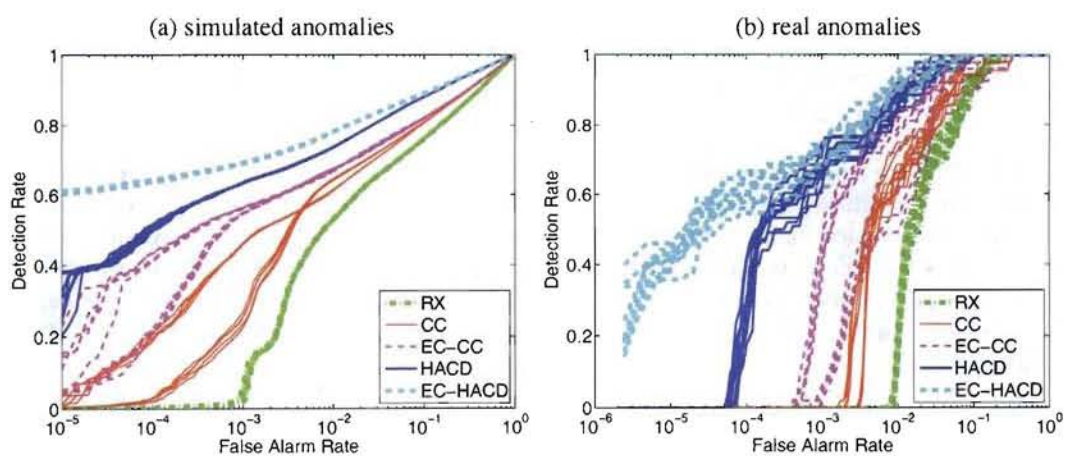(a) simulated anomalies

(b) real anomalies

Figure 7. ROC curves for the data shown in Fig. 6.

6.0 References

1. J. Meola and M. T. Eismann, "Image misregistration effects on hyperspectral change detection," *Proc. SPIE*, vol. 6966, p. 69660Y, 2008.

2. 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.

3. D. Rumsfeld. "As we know, there are known knowns. There are things we know we know. We also know there are known unknowns. That is to say we know there are some things we do not know. But there are also unknown unknowns, the ones we don't know we don't know." Department of Defense news briefing, 2002.

4. V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer, 2nd ed., 1999.

5. 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.

6. 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.

7. J. Theiler and D. M. Cai, "Resampling approach for anomaly detection in multispectral images," *Proc. SPIE*, vol. 5093, pp. 230–240, 2003.

8. I. Steinwart, D. Hush, and C. Scovel, "A classification framework for anomaly detection," *J. Machine Learning Research*, vol. 6, pp. 211–232, 2005.

9. J. Theiler and S. Perkins, "Resampling approach for anomalous change detection," *Proc. SPIE*, vol. 6565, p. 65651U, 2007.

10. A. Schaum and A. Stocker, "Long-interval chronochrome target detection," *Proc. 1997 International Symposium on Spectral Sensing Research*, 1998.

11. 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.

12. S. Cambanis, S. Huang, and G. Simons, "On the theory of elliptically contoured distributions," *J. Multivariate Analysis*, vol. 11, pp. 368–385, 1981.

13. Y. Kano, "Consistency property of elliptical probability density functions," *J. Multivariate Analysis*, vol. 51, pp. 139–147, 1994.

14. D. B. Marden and D. Manolakis, "Modeling hyperspectral imaging data," *Proc. SPIE*, vol. 5093, pp. 253–262, 2003.

15. 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.

16. J. Theiler, "Quantitative comparison of quadratic covariance-based anomalous change detectors," *Applied Optics*, vol. 47, pp. F12–F26, 2008.

17. J. Theiler, N. R. Harvey, R. Porter, and B. Wohlberg, "Simulation framework for spatio-spectral anomalous change detection," *Proc. SPIE*, vol. 7334, 2009.

18. 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.

19. J. Theiler, C. Scovel, B. Wohlberg, and B. R. Foy, "Elliptically-contoured distributions for anomalous change detection in hyperspectral imagery," To appear in: *IEEE Geoscience and Remote Sensing Letters*, 2010. doi: 10.1109/LGRS.2009.2032565.