# Right spectrum in the wrong place: a framework for local hyperspectral anomaly detection

*James Theiler and Amanda K. Ziemann; Los Alamos National Laboratory; Los Alamos, NM 87545*

## Abstract

*This paper formulates the hyperspectral anomaly detection problem in terms of a local context by modeling the relationship of individual pixels with the annuli of pixels that surround them. In this formulation, a locally anomalous pixel is one that might even be quite typical in the context of the whole image, but is "out of place" with respect to its local neighborhood. The problem of anomaly detection is cast as a supervised learning problem, in which samples from one class (normal) are provided by pixel/annulus pairs that occur in the scene, and samples from the second class (anomalous) can be created by making pixel/annulus pairs in which the pixels and annuli are effectively scrambled with respect to each other.*

*Although the formulation is in terms of machine learning, the experiments performed here use a simplified approach in which parametric (multivariate Gaussian and fatter-tailed multivariate t) distributions are fit to the data. This leads to a suite of local anomaly detectors that we compare to standard local RX and global RX detectors.*

> I been in the right place
> But it must have been the wrong time
> I'd of said the right thing
> But I must have used the wrong line
> I been in the right trip
> But I must have used the wrong car
> My head was in a bad place
> And I'm wondering what it's good for
> — Dr. John

## Introduction

The rich variety of anomaly detection algorithms that have been proposed for hyperspectral imagery [1] speaks to their appeal. Anomaly detection offers a way to deal with the burgeoning glut of image data by providing a principled and scalable approach to cue more expensive (*e.g.*, human) analysis on an adjustably small fraction of the available imagery. Hyperspectral imagery is especially attractive for remote sensing applications because each pixel contains information about the material properties of whatever is in the scene. A hyperspectral pixel can contain hundreds of channels of spectral information (far more than the three – red, green, and blue – channels of traditional imagery). In fact, because there is so much information per pixel, many hyperspectral analysis algorithms treat the image as a "bag of pixels," processing each multichannel pixel independently, and a pixel is considered anomalous if it is unlike all of the other pixels in the bag. For the *local* anomaly detection problem, however, spatial context is taken into account as well.

The main contribution in this paper is the formulation of local anomaly detection as a binary classification problem, for which the tools of machine learning (both traditional and modern) can be applied. In this formulation, a pixel is characterized by two (usually vector-valued) quantities, $y$ and $\mathbf{x}$. Here, $y$ corresponds to the pixel itself and $\mathbf{x} = [x_1^\mathsf{T}, x_2^\mathsf{T}, \ldots, x_K^\mathsf{T}]^\mathsf{T}$ corresponds to the $K$ pixels in the annulus surrounding the pixel of interest. For multispectral images with $d$ channels, $y$ is a $d$-dimensional vector, while $\mathbf{x}$ has $Kd$ components. See Fig. 1.

As a practical matter, we can reduce the number of vectors $K$ that we use to characterize the annulus, by selecting linear combinations of the pixel values. A symmetry-inspired feature selection approach, described in [2], can reduce $K = 40$ to $K = 7$.

The key to any target (or anomaly) detection problem is the characterization of the background [3]; the background, or non-target, or "normal" class, corresponds to the bulk of the image, and samples drawn from this class are provided by the image itself. In this case, the background model includes the coupled pixel-annulus system, denoted $(\mathbf{x}, y)$. We write $p(\mathbf{x}, y)$ as the probability distribution function associated with the normal class.

Anomalies are deviations from the background model, but in order to optimize the detection of anomalies, we need a model for those anomalies. If we write $p_a(\mathbf{x}, y)$ as a model for local anomalies, then we can use the likelihood ratio to measure anomalousness:

$$\mathscr{A}(\mathbf{x}, y) \sim \frac{p_a(\mathbf{x}, y)}{p(\mathbf{x}, y)}. \tag{1}$$

We use '∼' to indicate a kind of informal equality. We treat "anomalousness" as a relative term, so any positive monotonic function (*e.g.*, a logarithm) of the likelihood ratio will serve equally well as an anomalousness measure.

We argue that the proposed formulation enables optimized detection (in the Neyman-Pearson sense) of local anomalies in hyperspectral imagery. This will be demonstrated with numerical experiments on real data with artificial anomalies (obtained by taking pixels from one part of an image and moving them to another) showing ROC curves that compare the performance of detectors based on the proposed formulation versus traditional and regression-based anomaly detectors.

In what follows, we write $p(\mathbf{x})$ and $p(y)$ as the marginal distributions associated with $p(\mathbf{x}, y)$.

### *Global and Local RX*

As a point of comparison, we also consider anomaly detectors that *do not* use the local annulus information in $\mathbf{x}$. That is:

$$\mathscr{A}(y) \sim \frac{p_a(y)}{p(y)}. \tag{2}$$
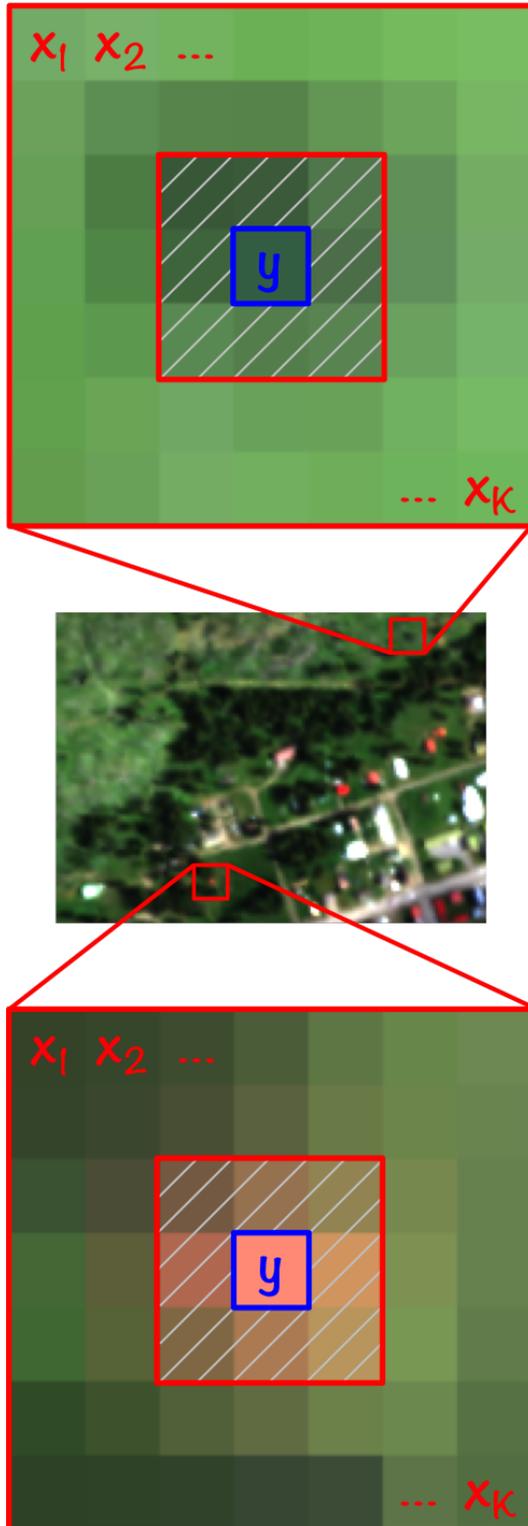
**Figure 1.** *A pixel $y$ is surrounded by an annulus of pixels $x_1, x_2, \ldots, x_K$. The question of interest is whether $y$ is anomalous in the local context of the $x$-pixels. The annuli shown here are $7 \times 7$ with a $3 \times 3$ hole, so $K = 40$. Exploiting the eight-fold symmetry of the square annulus, we can work with a reduced dimensionality of $K = 7$ features per band.*

If we take $p_a(y)$ to be uniform and $p(y)$ to be multivariate Gaussian, then we obtain one of the simplest anomaly detectors, given by the Mahalanobis distance [4] to the center of the data. This gives

$$\mathscr{A}(y) = (y - \hat{y})^{\mathsf{T}} R_{yy}^{-1} (y - \hat{y})^{\mathsf{T}}, \qquad (3)$$

where $R_{yy} = \langle (y - \hat{y})(y - \hat{y})^{\mathsf{T}} \rangle$. Here, $\hat{y}$ is a mean spectrum. For global RX, $\hat{y}$ is the average pixel spectrum over the whole image. A global anomaly is something like a pink spaceship, which would be an unusual sighting anywhere in the image.

For local RX,

$$\hat{y} = \langle \mathbf{x} \rangle = (1/K) \sum_{k=1}^{K} x_k \qquad (4)$$

is the average pixel over the annulus surrounding the pixel of interest. The average that defines $R_{yy}$ can be over the same annulus that defines $\hat{y}$ (as originally suggested by Reed, Yu, and Stocker [5, 6]), over a larger annulus (as suggested by Matteoli *et al.* [1]), or over the entire image. For computational efficiency and to avoid issues of covariance regularization [7–16], we use a global average for $R_{yy}$ even as we are using a local average for $\hat{y}$. The pink spaceship will likely still be anomalous, but so will more ordinary objects that are in unusual locations (*e.g.*, a blue Prius in the middle of the dense forest).

### Wrong (i.e., conditionally anomalous) spectrum

The anomalous $y$ is modeled as a broad flat distribution, *e.g.*, $p_a(y) = \text{const}$. Thus, $p_a(\mathbf{x}, y) = \text{const} \times p(\mathbf{x})$, which leads to an anomaly detector of the form

$$\mathscr{A}(\mathbf{x}, y) \sim \frac{p(\mathbf{x})}{p(\mathbf{x}, y)}. \qquad (5)$$

This expression recalls the definition of conditional probability: $p(y|\mathbf{x}) = p(\mathbf{x}, y)/p(\mathbf{x})$. Thus, Eq. (5) can be interpreted as a kind of conditional anomalousness: $\mathscr{A}(\mathbf{x}, y)$ is large when $p(y|\mathbf{x})$ is small. In other words, $y$ is unusual *in the context* of the annulus pixels $\mathbf{x}$. This anomaly detector is similar to Local RX, in that it is sensitive not only to pink spaceships but to hybrid automobiles in rural settings. Where Local RX makes an *ad hoc* estimate of anomalousness, however, the conditional anomalousness in Eq. (5) aims for optimality by employing a likelihood ratio.

### Right spectrum in the wrong place

Another kind of anomalous pixel is anomalous in the context of the annulus pixels without being anomalous in its own right. That is, the spectrum $y$ is ordinary compared to the other pixels in the image (*i.e.*, it has the "right spectrum"), but it is locally anomalous in the context of the pixels in its surrounding annulus (*i.e.*, it is "in the wrong place"). For example, a tree in a forest might not be at all unusual, but a lone tree surrounded by buildings (assuming there are no other lone trees in the city) would qualify as anomalous in this framework. Here, instead of taking $p_a(y) = \text{const}$, we take $p_a(y) = p(y)$, where $p(y)$ is the distribution associated with image pixel spectra $y$.

Our local anomaly detector is again expressed as a likelihood ratio:

$$\mathscr{A}(\mathbf{x},y) \sim \frac{p(\mathbf{x})p(y)}{p(\mathbf{x},y)}. \tag{6}$$

When this quantity is large for a specific $(\mathbf{x},y)$ pair, then we take $y$ as anomalous with respect to its local context $\mathbf{x}$. Although it is possible to estimate $p(\mathbf{x},y)$ directly – *e.g.*, as a Gaussian or heavy-tailed elliptically-contoured distribution – we observe that the expression $p(\mathbf{x})p(y)/p(\mathbf{x},y)$ can be interpreted as a likelihood ratio, which means that it can also be estimated indirectly using binary classification algorithms from machine learning.

This formulation treats location as paramount. Our aim is not to determine whether the spectrum of a pixel $y$ is by itself anomalous, but whether it is anomalous in the context of the pixels surrounding it. And, in contrast to the conditional anomalousness defined in Eq. (5), a pixel that is unusual in its own right (*e.g.*, that pink spaceship) will be less anomalous in this RSWP (Right Spectrum in the Wrong Place) formulation.

## Gaussian and Elliptically-contoured models

Without loss of generality, assume that the (global) mean values of $\mathbf{x}$ and $y$ have been subtracted, and are now zero. For a zero-centered Gaussian model of $\mathbf{x}$ and $y$, we specify a covariance $R_{zz}$, which is estimated by the second moments of the data:

$$R_{zz} = \left[ \begin{array}{cc} R_{xx} & R_{yx}^{\mathsf{T}} \\ R_{yx} & R_{yy} \end{array} \right] = \left[ \begin{array}{cc} \langle \mathbf{x}\mathbf{x}^{\mathsf{T}} \rangle & \langle \mathbf{x}y^{\mathsf{T}} \rangle \\ \langle y\mathbf{x}^{\mathsf{T}} \rangle & \langle yy^{\mathsf{T}} \rangle \end{array} \right], \tag{7}$$

where we have written $R_{xx} = \langle \mathbf{x}\mathbf{x}^{\mathsf{T}} \rangle$ and $R_{yy} = \langle yy^{\mathsf{T}} \rangle$ and $R_{yx} = \langle y\mathbf{x}^{T} \rangle$. Let us also write the Mahalanobis distances: $\xi_x(\mathbf{x}) = \mathbf{x}^{\mathsf{T}} R_{xx}^{-1} \mathbf{x}$, and $\xi_y(y) = y^{\mathsf{T}} R_{yy}^{-1} y$, and $\xi_z(\mathbf{x},y) = \left[ \begin{array}{cc} \mathbf{x}^{\mathsf{T}} & y^{\mathsf{T}} \end{array} \right] R_{zz}^{-1} \left[ \begin{array}{c} \mathbf{x} \\ y \end{array} \right]$.

In this context, we can write the wrong-spectrum detector as

$$\mathscr{A}(\mathbf{x},y) = \xi_z(\mathbf{x},y) - \xi_x(\mathbf{x}) \tag{8}$$

and the right-spectrum-wrong-place anomaly detector as

$$\mathscr{A}(\mathbf{x},y) = \xi_z(\mathbf{x},y) - \xi_x(\mathbf{x}) - \xi_y(y). \tag{9}$$

In fact, we can also write the Global RX detector in Eq. (3) as

$$\mathscr{A}(y) = \xi_y(y). \tag{10}$$

This expression can be extended from multivariate Gaussian to more general elliptically-contoured distributions. In particular, for the multivariate $t$ distribution suggested by Manolakis *et al.* [17] for hyperspectral data in general, we can again follow the lead of anomalous change detection to write [18]:

$$\xi' = H(d,\nu,\xi) = (d+\nu)\log\left(1+\frac{\xi}{\nu-2}\right), \tag{11}$$

where $\nu$ is the degrees of freedom parameter in the $t$ distribution. Observe that in the $\nu \to \infty$ limit, $\xi'$ becomes $\xi$. Fig. 2 shows a plot of $\xi'$ vs $\xi$; we see that $\xi'$ increases monotonically with $\xi$, but
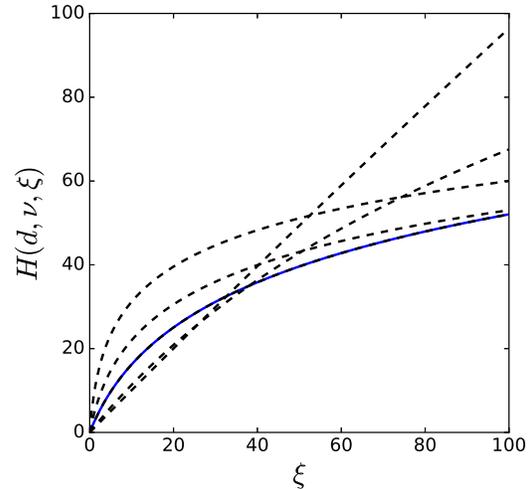
**Figure 2.** *Function $H(d,\nu,\xi)$ defined in Eq. (11) for elliptically-contoured multivariate t distribution. Here $d=10$ and $\nu$ takes on values 3, 5, 10 (solid), 50, and 1000 (very nearly linear).*

with decreasing slope as $\xi$ gets larger. Informally, we can think about this in terms of the effect of outlier values of $\xi$ being less significant.

This leads to an algorithmically trivial modification (replace $\xi$ with $\xi'$) of Eq. (8)

$$\mathscr{A}(\mathbf{x},y) = \xi_z'(\mathbf{x},y) - \xi_x'(\mathbf{x}) \tag{12}$$

$$= H(d_x+d_y,\nu,\xi_z(\mathbf{x},y)) - H(d_x,\nu,\xi_x(\mathbf{x})) \tag{13}$$

$$= (d_x+d_y+\nu)\log\left(1+\frac{\xi_z(\mathbf{x},y)}{\nu-2}\right)$$

$$- (d_x+\nu)\log\left(1+\frac{\xi_x(\mathbf{x})}{\nu-2}\right) \tag{14}$$

and of Eq. (9):

$$\mathscr{A}(\mathbf{x},y) = \xi_z'(\mathbf{x},y) - \xi_x'(\mathbf{x}) - \xi_y'(y) \tag{15}$$

$$= (d_x+d_y+\nu)\log\left(1+\frac{\xi_z(\mathbf{x},y)}{\nu-2}\right)$$

$$- (d_x+\nu)\log\left(1+\frac{\xi_x(\mathbf{x})}{\nu-2}\right)$$

$$- (d_y+\nu)\log\left(1+\frac{\xi_y(y)}{\nu-2}\right). \tag{16}$$

Here, $d_x$ and $d_y$ are the dimensions, respectively, of the $\mathbf{x}$ and $y$ vectors. In particular, $d_y = d$ is the number of channels in the hyperspectral image, and $d_x = Kd$, where $K$ is the (effective) number of pixels in the annulus.

Here, the covariance matrix is estimated from second moments of the data. But for anomaly detection, where it is the periphery of the distribution that is important, an alternative approach might be to estimate the covariance associated with an ellipsoid that encloses most of the data [19].

## Analogy with Anomalous Change Detection

Since we have an $\mathbf{x}$ and a $y$ for each pixel, we can think of them as two separate images (the $\mathbf{x}$ image will have many more

**Figure 3.** *Cooke City, MT; HyMap sensor with 126 bands [25].*

bands than the *y* image), and can imagine for the purposes of this thought experiment that the **x** image was taken before the *y* image. For most of the pixel pairs $(\mathbf{x}, y)$, the "change" from **x** to *y* is normal. But saying that a given pixel *y* is anomalous in the context of an associated annulus **x** would be interpreted as saying that *y* indicates an anomalous change at that pixel compared to the **x** that was observed at that pixel in the previous image.

The analogy may be imperfect, but the mathematics is very similar to that developed for the anomalous change detection problem [20–22], including the use of elliptically-contoured distributions [18].

## Relation to Regression Framework

The formulation here extends what was previously developed as a "regression framework" [2, 23]. In that work, one uses regression to fit a function $f$ that provides an estimate $\hat{y} = f(\mathbf{x})$ of a center pixel *y* as a function of annulus pixels **x**. From this estimator, one can compute a covariance of the estimation error, $R_{ee} = (y - \hat{y})^\mathsf{T}(y - \hat{y})$, and from this an anomaly measure:

$$\mathscr{A}(\mathbf{x}, y) = (y - f(\mathbf{x}))^\mathsf{T} R_{ee}^{-1}(y - f(\mathbf{x})). \tag{17}$$

If $f(\mathbf{x})$ is a linear least-squares estimator, then this is equivalent to the formulation in Eq. (5) under the assumption that $p(y|\mathbf{x})$ is Gaussian. One thing that the new formulation provides is the ability to generalize the regression framework by treating $p(y|\mathbf{x})$ as non-Gaussian, such as an elliptically-contoured distribution.

Using a classification-based approach instead of a regression-based approach follows Vapnik's dictum, in that it avoids solving a more general problem as an intermediate step [24]. This frees us from the need to make assumptions about underlying distributions, and permits us to make use of well-developed off-the-shelf machine learning algorithms. It also provides a more direct solution to the detection of anomalies corresponding to pixels that exhibit the "right spectrum in the wrong place."

## Experiments

To illustrate and compare these anomaly detectors, we perform experiments using artificial anomalies in real data.

### Datasets

As background imagery, we selected three hyperspectral datasets and one multispectral dataset. All of these images were atmospherically compensated to approximate ground surface reflectance.

The Cooke City dataset [25] is a 280×800 pixel image with $d = 126$ spectral channels that span the visible, near infrared, and shortwave infrared electromagnetic spectrum (VNIR-SWIR,



**Figure 4.** *SHARE 2012, near Rochester, NY; SpecTIR VS sensor with 229 bands [26].*



**Figure 5.** *Reno, NV; SpecTIR sensor with 72 bands [27].*



**Figure 6.** *Omaha, NB; WorldView-2 sensor with 8 bands [28, 29].*

wavelengths 0.45-2.5$\mu$m). This dataset was collected by the Rochester Institute of Technology (RIT) using the HyMap sensor as part of the 2006 CHARM Collection; it was released in 2008 as a "blind test" package [30] and has been widely used for target detection tasks. The imagery is of Cooke City, MT, and has an approximate ground sample distance (GSD, *i.e.*, spatial resolution) of 2m. See Fig. 3.

The SHARE 2012 experimental campaign [26] was conducted by RIT to create standardized datasets available to the entire remote sensing community. In this image (Fig. 4) there are $280 \times 170$ pixels, and 360 bands ranging from 0.4-2.45$\mu$m, but we used a hand-selected subset of $d = 229$ of the highest-quality bands: indices 7-114,128,140-143,153-177,211-252,292-340. The image was collected by the ProSpecTIR VS sensor with an approximate GSD of 1m.

The Reno dataset [27] was taken with a SpecTIR sensor; there are $d = 72$ bands in the $600 \times 320$ pixel image shown in Fig. 5.

In contrast with the first three hyperspectral datasets, a fourth dataset was collected by a multispectral sensor. The WorldView-2 [28, 29] image of Omaha, Nebraska was provided by Digital Globe. The sensor has 8 spectral bands, and this $512 \times 512$ pixel subset is cropped from a much larger image of the region. See Fig. 6.

### Implanting targets

Asserting that a given pixel in a natural image is in fact anomalous is inevitably a judgement call [31]. To assess the quality of different anomaly detectors, without falling into anecdotal comparisons, we will implant anomalous targets into natural backgrounds. By asserting that all the implanted targets are anomalous and that the rest of the image is non-anomalous, we have a basis for computing detection rates, false alarm rates, and receiver-operator characteristic (ROC) curves.

Two schemes were employed for implanting targets, corresponding to different concepts of what a local anomaly is. In both schemes, a small number of pixels (small compared to the total number of pixels in the image) are chosen at random to be anomalous.

In the first ("misplaced pixel") scheme, the pixels in those randomly chosen locations are replaced with pixels from some other random location in the same image. By construction, then, these new pixels are not by themselves anomalous. But they *are* anomalous with respect to the neighborhoods in which they now find themselves.

In the second ("uniform") scheme, those pixels are replaced with pixels whose spectra are generated from a uniform random distribution. That is, each component of the vector-valued pixel spectrum is chosen from a uniform distribution over the range from the minimum to the maximum that that component takes on over all the pixels in the image.

A variant is to consider sub-pixel anomalies. Here, based on a parameter $0 \le \alpha \le 1$, the new pixel has the form $y'' = (1-\alpha)y + \alpha y'$, where $y$ is the original spectrum at a given pixel location, and $y'$ is the random pixel value. We employ subpixel anomalies in this experiment because full-pixel uniform anomalies are *so* anomalous that all of our detectors are able to detect virtually all of our implanted targets.

**Table 1. Anomaly detectors used in this exposition. Here, RX refers to straightforward Mahalanobis distance based detectors. G- and EC- refer to the Gaussian and Elliptically-Contoured (based on multivariate t distribution) variants of the algorithms. WS is the "wrong spectrum" anomaly detector and RSWP is the "right spectrum in the wrong place."**

| Name | Equations | ROC Curve legend |
|------|-----------|------------------|
| Global RX | Eq. (3) | Green |
| Local RX | Eq. (3), Eq. (4) | Black |
| G-WS | Eq. (5), Eq. (8) | Red solid |
| EC-WS | Eq. (5), Eq. (14) | Red dashed |
| G-RSWP | Eq. (6), Eq. (9) | Blue solid |
| EC-RSWP | Eq. (6), Eq. (16) | Blue dashed |

### Results

Fig. 7 illustrates several points. One is that global RX anomaly detection is useless for detecting the local anomalies produced by an implanting method that recycles pixels already in the image. This is demonstrated by the green curve, which shows worse performance than the other detectors here. Two is that the use of multiple features, instead of simply the local mean, leads to better performance. Here we use the "dihedral" feature set, which provides seven annulus features for each band. A third illustration is that, when the anomalies are "purely local" – *i.e.*, they are pixels taken randomly from other parts of the image – the blue RSWP ("right spectrum in the wrong place") detector outperforms the red WS ("wrong spectrum") detectors. Finally, considerable improvement is observed using the transformation in Eq. (11), shown here as dashed lines. This transformation corresponds to treating $p(\mathbf{x}, y)$ as a fat-tailed elliptically-contoured distribution instead of a Gaussian. Here we used $\nu = d = 126$. (Although $\nu$ can be estimated from the data, for instance using the scheme described in the appendix of [18], the simpler suggestion of using $\nu = d$ was found to be useful in a target detection context [32].)

Error bars on ROC curves can be tricky [33], and all the more so given our approach of implanting targets at random, so we have simply run multiple (five) trials and plotted the ROC curves for each trial.

It is important to recognize that the relative performance of the different anomaly detectors depends on the nature of the anomalies. Fig. 11 shows what happens when the anomalies are obtained from a flat distribution and included as subpixel ($\alpha = 0.005$) anomalies. Here the (green) global anomaly detector is not that much different from the (black) simple local anomaly detector. Also, opposite to the case in Fig. 7, the (solid red) Gaussian WS detectors outperform the (solid blue) RSWP detectors. Interestingly, the non-Gaussian detectors (dashed lines) very substantially outperform their Gaussian counterparts.

## Conclusions and future work

Inspired by the regression framework, we developed a more direct anomaly detection approach, and this enabled us to employ a more general formulation. Although the Gaussian conditional anomaly ("wrong spectrum") detector in Eq. (8) is equivalent to a variant of the regression model, the use of distributions allows
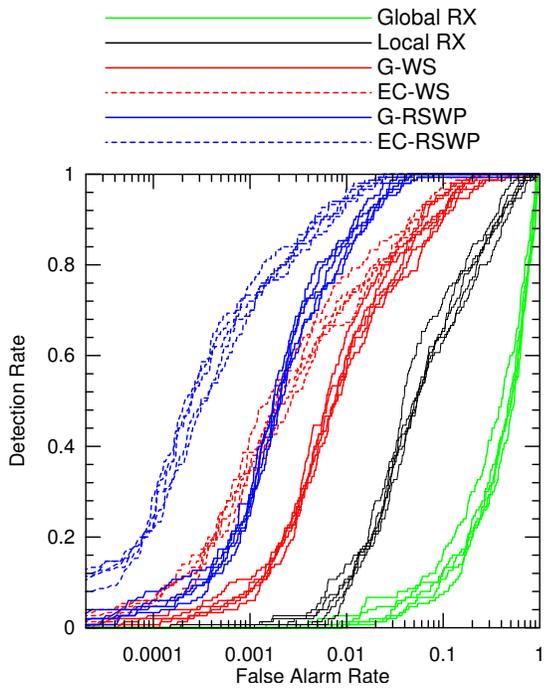
**Figure 7.** *Using the 126-channel Cooke City hyperspectral dataset [25] as "background," we implanted 150 targets at random, using the "misplaced pixel" scenario, and computed ROC curves for various anomaly detectors. The experiment was performed five times (i.e., five different sets of 150 targets), and the ROC curves for each trial are plotted. The annulus was of size $7 \times 7$ with a $3 \times 3$ central hole. Here, green corresponds to the global RX anomaly detector, black to local RX (but with global covariance), and red and blue to feature-based annulus descriptors ($D_4 \Sigma$) with seven values per annulus. Red uses the "wrong spectrum" detectors, solid assumes Gaussian, and dashed assumes multivariate t with $\nu = 126$ chosen to agree with the dimension of the data. Blue uses the "right spectrum in the wrong place" detectors; i.e., Eq. (9) for solid and Eq. (16) for dashed.*
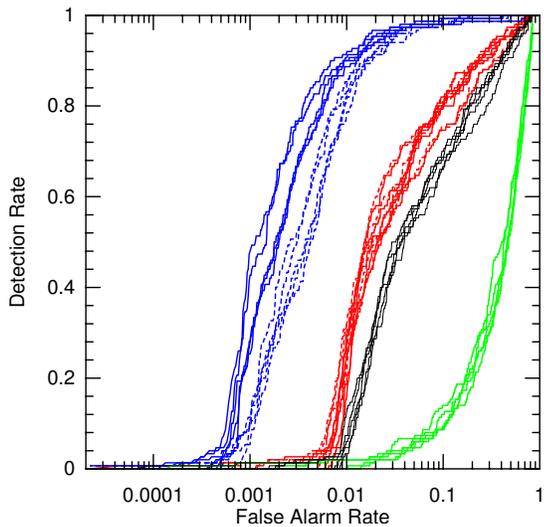


**Figure 8.** *Same as Fig. 7 but using 229 bands of the 360-band SHARE 2012 dataset.*
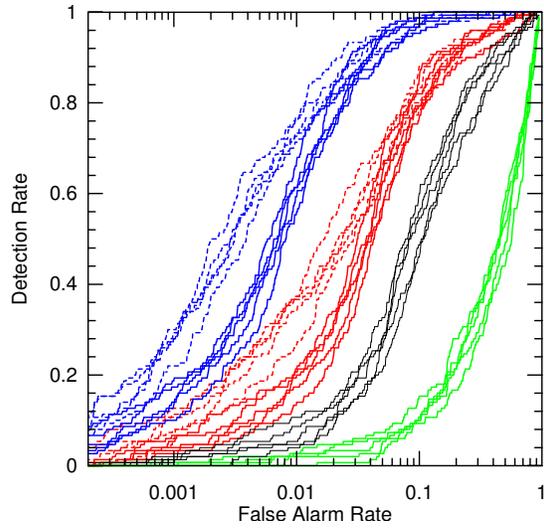


**Figure 9.** *Same as Fig. 7 but using the 72-band Reno dataset.*
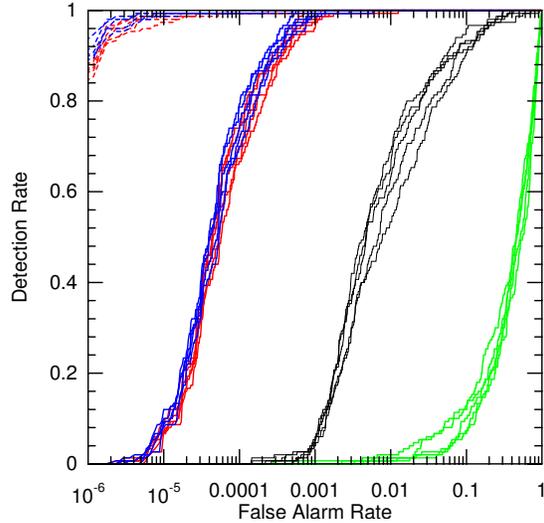


**Figure 10.** *Same as Fig. 7 but using the eight-band WorldView-2 dataset.*
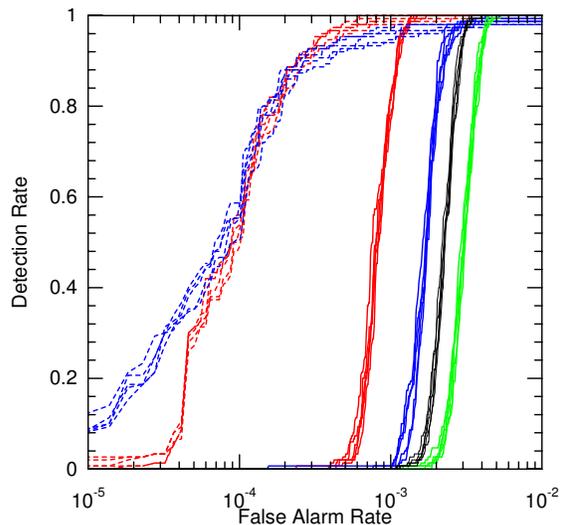


**Figure 11.** *Experiment is similar to that shown in Fig. 7, but the implanted anomalies are subpixel uniform ($\alpha = 0.005$).*
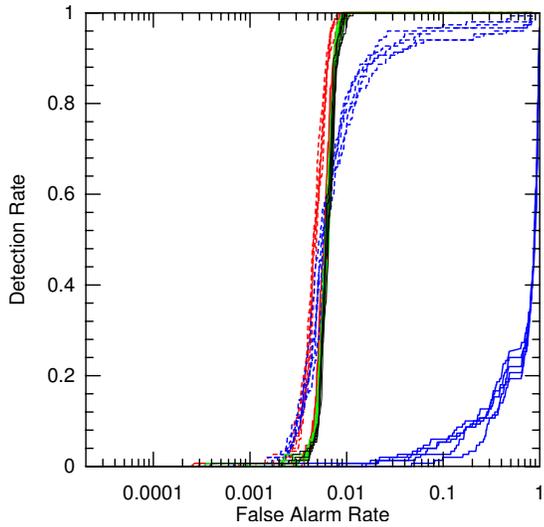
**Figure 12.** *Same as Fig. 11 but using 229 bands of the 360-band SHARE 2012 dataset – also, with a much smaller $\alpha = 0.0005$ (!). Similar to Fig. 13, the Gaussian RSWP detector exhibits abysmal performance; the non-Gaussian variant, however, is competitive (except in the very high detection rate regime) with the other detectors.*
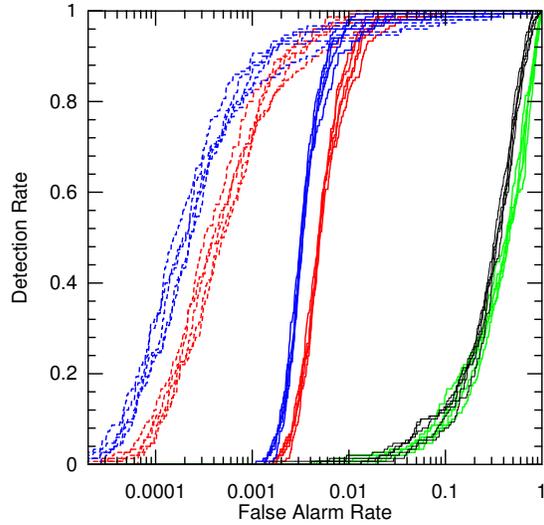


**Figure 14.** *Same as Fig. 11 but using the eight-band WorldView-2 dataset. Unlike Fig. 11, and contrary to our expectations, we see that the (blue) RSWP detector slightly outperforms the (red) WS detector.*



**Figure 15.** *Urban pixels in the Cooke City image, indicated in red.*
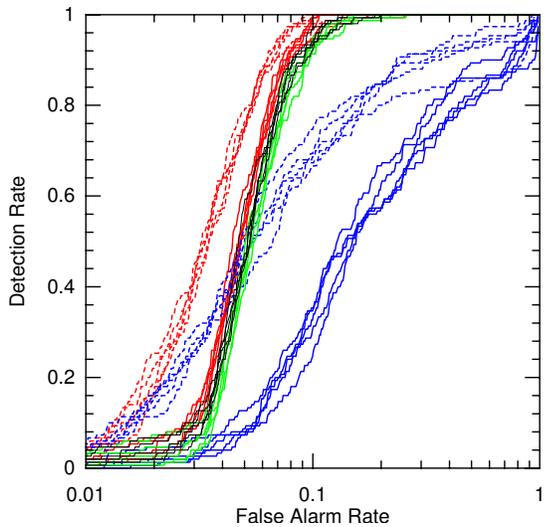


**Figure 13.** *Same as Fig. 11 but using the 72-band Reno dataset.*
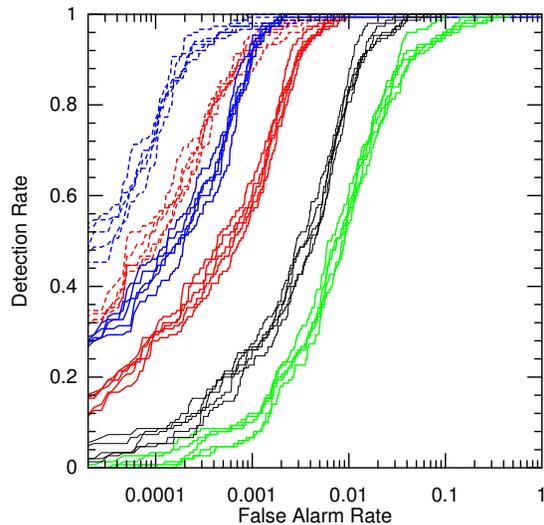


**Figure 16.** *Using the same 126-channel Cooke City hyperspectral image as in Fig. 7, but using only the "urban" pixels as anomalies. In each of the five trials, 150 pixels are chosen at random from the urban area (from the pixels indicated in red in Fig. 15), and implanted at random locations throughout the image. Comparing to Fig. 7, we see that urban pixels are more readily identified as anomalies, but that the relative performance of the different algorithms is preserved.*

us to make two improvements. One of them is to more directly address the "right spectrum in the wrong place" scenario and to produce an anomaly detector that is optimized for that case. The second improvement (which appears, empirically, to have more benefit) enables us to use non-Gaussian distributions to characterize our data, and in particular we can use a fat-tailed elliptically-contoured distribution instead of a Gaussian, and get better performance with only a minor change in the operation of the algorithm (using $\xi' = H(d, \nu, \xi)$ in place of $\xi$). This minor change improves the performance of both the "wrong spectrum" and the "right spectrum in the wrong place" anomaly detectors.

The open-ended nature of the new formalism provides many opportunities for further practical improvements. Given that this is a machine learning approach, we have a wide variety of machine learning tools – from Fisher discriminants and support vector machines to random forests and deep neural nets – that we can potentially use to find the boundary in $(\mathbf{x}, y)$ space that separates the normal pixels from the anomalies.

The moving window approach used here provides one way to achieve a local context; another approach is use segmentation of the image [34, 35]. In fact, we might be able to use segmentation and moving windows together, defining a different $p(\mathbf{x}, y)$ in each different segment.

The Mahalanobis distances based on Gaussian (or elliptically-contoured) distributions can be replaced with distances to subspaces, and several authors have considered local subspace models [36, 37] for background estimation and anomaly detection. We were particularly intrigued with an innovation employed by Ranney and Soumekh [36] (compare Fig. 2 of that reference with Fig. 1 here). In this case, four distances are computed – a separate distance for each of four cardinal directions – and a minimum is taken of the four. Thus, the dimension of the subspace is kept small, but nonstationarity is ameliorated. We believe that this four-distance innovation might be applicable to our approach as well.

In our implantation of subpixel uniform targets, the parameter $\alpha$ corresponds to the fraction of the pixel occupied by the anomaly. But the values that we used ($\alpha = 0.005$ and $\alpha = 0.0005$) are unrealistically small. We used such small values because larger $\alpha$ would lead to anomalies that are "too easy" to detect, and as a result we would be unable to distinguish the relative performance of different algorithms. But given that we are interested in optimizing subpixel performance, we could follow the analogy to anomalous change detection and employ expressions that were developed for subpixel anomalous changes [38].

Although hyperspectral data is not Gaussian, it has been observed that it can be more Gaussian in some directions, particularly the directions with lower variance. [39–41] Thus, we might see improvement with a Gaussian/Non-Gaussian (G/NG) scheme [42, 43]. In general, this works by using principal components analysis to decompose data into two components, a high-variance non-Gaussian component (first few principal components) and a lower-variance near-Gaussian component (remaining principal components). The non-Gaussian component is low-dimensional, and so can be modeled in more sophisticated ways (the simplest perhaps is using the scheme in Eq. (11)), while the Gaussian component is treated essentially by using a local version of subspace RX [44].

## Acknowledgment

## References

[1] S. Matteoli, M. Diani, and G. Corsini, "A tutorial overview of anomaly detection in hyperspectral images," *IEEE A&E Systems Magazine*, vol. 25, pp. 5–27, 2010.

[2] J. Theiler, "Symmetrized regression for hyperspectral background estimation," *Proc. SPIE*, vol. 9472, pp. 94721G, 2015.

[3] S. Matteoli, M. Diani, and J. Theiler, "An overview background modeling for detection of targets and anomalies in hyperspectral remotely sensed imagery," *IEEE J. Sel. Topics in Applied Earth Observations and Remote Sensing*, vol. 7, pp. 2317–2336, 2014.

[4] P. C. Mahalanobis, "On the generalised distance in statistics," *Proc. National Institute of Sciences of India*, vol. 2, pp. 49–55, 1936.

[5] 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.

[6] A. D. Stocker, I. S. Reed, and X. Yu, "Multi-dimensional signal processing for electro-optical target detection," *Proc. SPIE*, vol. 1305, pp. 218–231, 1990.

[7] J. P. Hoffbeck and D. A. Landgrebe, "Covariance matrix estimation and classification with limited training data," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, pp. 763–767, 1996.

[8] G. Cao and C. A. Bouman, "Covariance estimation for high dimensional data vectors using the sparse matrix transform," in *Advances in Neural Information Processing Systems 21*. 2009, pp. 225–232, MIT Press.

[9] N. M. Nasrabadi, "Regularization for spectral matched filter and RX anomaly detector," *Proc. SPIE*, vol. 6966, pp. 696604, 2008.

[10] C. E. Caefer, J. Silverman, O. Orthal, D. Antonelli, Y. Sharoni, and S. R. Rotman, "Improved covariance matrices for point target detection in hyperspectral data," *Optical Engineering*, vol. 7, pp. 076402, 2008.

[11] G. Cao, C. A. Bouman, and J. Theiler, "Weak signal detection in hyperspectral imagery using sparse matrix transform (SMT) covariance estimation," in *Proc. WHISPERS (Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing)*. 2009, IEEE.

[12] Stefania Matteoli, Marco Diani, and Giovanni Corsini, "Improved estimation of local background covariance matrix for anomaly detection in hyperspectral images," *Optical Engineering*, vol. 49, pp. 046201, 2010.

[13] Avishai Ben-David and Charles E. Davidson, "Estimation of hyperspectral covariance matrices," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2011, pp. 4324–4327.

[14] J. Theiler, G. Cao, L. R. Bachega, and C. A. Bouman, "Sparse matrix transform for hyperspectral image processing," *IEEE J. Sel. Topics in Signal Processing*, vol. 5, pp. 424–437, 2011.

[15] L. Bachega, J. Theiler, and C. A. Bouman, "Evaluating and improving local hyperspectral anomaly detectors," *Proc. IEEE Applied Imagery Pattern Recognition Workshop*, 2011.

[16] J. Theiler, "The incredible shrinking covariance estimator," *Proc. SPIE*, vol. 8391, pp. 83910P, 2012.

[17] D. Manolakis, D. Marden, J. Kerekes, and G. Shaw, "On the statistics of hyperspectral imaging data," *Proc. SPIE*, vol. 4381, pp. 308–316, 2001.

[18] J. Theiler, C. Scovel, B. Wohlberg, and B. R. Foy, "Elliptically-contoured distributions for anomalous change detection in hyperspectral imagery," *IEEE Geoscience and Remote Sensing Lett.*, vol. 7, pp. 271–275, 2010.

[19] G. Grosklos and J. Theiler, "Ellipsoids for anomaly detection in remote sensing imagery," *Proc. SPIE*, vol. 9472, pp. 94720P, 2015.

[20] A. Schaum and A. Stocker, "Long-interval chronochrome target detection," *Proc. ISSSR (International Symposium on Spectral Sensing Research)*, 1998.

[21] 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.

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

[23] J. Theiler and B. Wohlberg, "Regression framework for background estimation in remote sensing imagery," *Proc. 5th IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2013.

[24] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, 2nd edition, 1999.

[25] D. Snyder, J. Kerekes, I. Fairweather, R. Crabtree, J. Shive, and S. Hager, "Development of a web-based application to evaluate target finding algorithms," *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 2, pp. 915–918, 2008.

[26] A. Giannandrea, N. Raqueno, D. W. Messinger, J. Faulring, J. P. Kerekes, J. van Aardt, K. Canham, S. Hagstrom, E. Ontiveros, A. Gerace, J. Kaufman, K. M. Vongsy, H. Griffith, B. D. Bartlett, E. Ientilucci, J. Meola, L. Scarff, and B. Daniel, "The SHARE 2012 data campaign," *Proc. SPIE*, vol. 8743, pp. 87430F, 2013.

[27] SpecTIR Advanced Hyperspectral and Geospatial Solutions, "Free data samples," Online: `http://www.spectir.com/free-data-samples/`.

[28] G. Marchisio, F. Pacifici, and C. Padwick, "On the relative predictive value of the new spectral bands in the WorldView-2 sensor," *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 2723–2726, 2010.

[29] N. T. Anderson and G. B. Marchisio, "WorldView-2 and the evolution of the DigitalGlobe remote sensing satellite constellation: introductory paper for the special session on WorldView-2," *Proc. SPIE*, vol. 8390, 2012.

[30] Rochester Institute of Technology (RIT) Digital Imaging and Remote Sensing Laboratory, "Target detection blind test," Online: `http://dirsapps.cis.rit.edu/blindtest/`.

[31] J. Theiler, "By definition undefined: adventures in anomaly (and anomalous change) detection," *Proc. 6th IEEE Workshop on Hyperspectral Signal and Image Processing: Evolution in Remote Sensing (WHISPERS)*, 2014.

[32] 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. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, p. I:221, 2008.

[33] J. Kerekes, "Receiver operating characteristic curve confidence intervals and regions," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, pp. 251–255, 2008.

[34] E. A. Ashton, "Detection of subpixel anomalies in multispectral infrared imagery using an adaptive bayesian classifier," *IEEE Transactions On Geoscience and Remote Sensing*, vol. 36, 1998.

[35] J. Theiler and L. Prasad, "Overlapping image segmentation for context-dependent anomaly detection," *Proc. SPIE*, vol. 8048, pp. 804807, 2011.

[36] K. I. Ranney and M. Soumekh, "Hyperspectral anomaly detection within the signal subspace," *IEEE Geoscience and Remote Sensing Lett.*, vol. 3, pp. 312–316, 2006.

[37] S. Matteoli, N. Acito, M. Diani, and G. Corsini, "An automatic approach to adaptive local background estimation and suppression in hyperspectral target detection," *IEEE Trans. Geoscience and Remote Sensing*, vol. 49, pp. 790–800, 2011.

[38] J. Theiler, "Subpixel anomalous change detection in remote sensing imagery," *Proc. IEEE Southwest Symposium on Image Analysis and Interpretation*, pp. 165–168, 2008.

[39] J. Theiler, B. R. Foy, and A. M. Fraser, "Characterizing non-Gaussian clutter and detecting weak gaseous plumes in hyperspectral imagery," *Proc. SPIE*, vol. 5806, pp. 182–193, 2005.

[40] P. Bajorski, "Maximum Gaussianity models for hyperspectral images," *Proc. SPIE*, vol. 6966, pp. 69661M, 2008.

[41] S. M. Adler-Golden, "Improved hyperspectral anomaly detection in heavy-tailed backgrounds," in *Proc. 1st Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*. 2009, IEEE.

[42] J. Theiler, "Ellipsoid-simplex hybrid for hyperspectral anomaly detection," in *Proc. 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*. 2011, IEEE.

[43] G. A. Tidhar and S. R. Rotman, "Target detection in inhomogeneous non-Gaussian hyperspectral data based on nonparametric density estimation," *Proc. SPIE*, vol. 8743, pp. 87431A, 2013.

[44] A. Schaum, "Spectral subspace matched filtering," *Proc. SPIE*, vol. 4381, pp. 1–17, 2001.

## Author Biographies

*JT received a Ph.D. in Physics from Caltech in 1987, and subsequently held appointments at UCSD, MIT Lincoln Laboratory, Los Alamos National Laboratory, and the Santa Fe Institute. He joined the technical staff at Los Alamos in 1994, and was named a Laboratory Fellow in 2005. His professional interests include statistical modeling, machine learning, image processing, and remote sensing.*

*AKZ received a Ph.D. in Imaging Science from Rochester Institute of Technology in 2015, and currently holds a postdoctoral appointment at Los Alamos National Laboratory. Her interests include remote sensing, graph-based modeling, manifold learning, and rescue dogs.*