BY DEFINITION UNDEFINED: ADVENTURES IN ANOMALY (AND ANOMALOUS CHANGE) DETECTION

James Theiler

Los Alamos National Laboratory, Los Alamos, NM 87545 USA

ABSTRACT

This paper is a personal and purposefully idiosyncratic survey of issues, some technical and some philosophical, that arise in developing algorithms for the detection of anomalies and anomalous changes in hyperspectral imagery.

The technical emphasis in anomaly detection is on modeling background clutter. For hyperspectral imagery this is a challenge because there are so many channels (the hyperspectral part) and because there is so much spatial structure (the imagery part). A wide range of models has been proposed for characterizing hyperspectral clutter: global and local models, Gaussian and non-Gaussian models, full-rank and subspace models, parametric and nonparametric models. And hybrid combinations, thereof. In discussing how these models relate to each other, an important theme will be characterizing the quality of a model in the absence of ground truth.

The anomaly itself of more of a philosophical creature. It is a deviation from what is typical or expected. In general, the detection of anomalies is complicated by the fact that anomalies are rare and that anomalies tend to defy any kind of precise specification. One might even say of anomalies that they are, by definition, undefined.

Index Terms— Detection, Target, Anomaly, Background, Hyperspectral, Imagery

1. INTRODUCTION

The detection of anomalies is rarely a goal in its own right. In many operational scenarios, the anomaly is something that is passed along for human inspection, or for further processing. It is, as Fig. 1 illustrates, a cueing process, and it has two purposes: the first, and arguably most important, is to reduce the quantity of incoming data to a level that can be handled by the more expensive downstream analysis; the second is to pass along the potentially interesting items. The distinction between "interesting" and "potentially interesting" (like the distinction between "dead" and "mostly dead" [1]) can be very important, but if the anomaly detector has succeeded at its first goal (of reducing data quantity), then the downstream analysis can make this judgment. This judgment can be very complicated and domain-specific, but what makes anomaly detection useful as a concept is that the anomaly detection module has more generic goals.

Broadly speaking, anomalies are rare and unusual. One way (but not the only way) to make this notion more formal is to invoke the mathematics of probability distributions. Write $\mathbf{x} \in \mathbb{R}^d$ to indicate a data sample (often a single pixel in a hyperspectral image), and $p(\mathbf{x})$ to indicate the probability density associated with that sample. We can take the view that anomaly detection is a binary classification problem, with the two classes being anomalous and background. If $p_a(\mathbf{x})$ is the probability density for anomalous items, and

$p_b(\mathbf{x})$ is the density for the non-anomalous background class, then the likelihood ratio provides the optimal detector of anomalousness:

$$\mathcal{A}(\mathbf{x}) = \frac{p_a(\mathbf{x})}{p_b(\mathbf{x})}. \tag{1}$$

One then compares to a threshold ($\mathcal{A}(\mathbf{x}) \gtreqless \eta$) to decide whether a given pixel is anomalous or not.

Although (1) is unambiguously optimal, one can fairly object that neither $p_a(\mathbf{x})$ nor $p_b(\mathbf{x})$ are really known. And in fact, both are problematic, though in different ways.

1.1. $p_a(\mathbf{x})$ is not really known – a *philosophical* problem

If individual anomalies resist definition, how can we expect to write down a probability distribution for *all* anomalies? There are two ways to address this objection:

One way to respond is by acknowledging (or at least feigning) ignorance of individual anomalies, and expressing that ignorance by saying that $p_a(\mathbf{x})$ should be a broad flat distribution that treats all possible anomalies equally. (This is akin to the Bayesian notion of an "uninformative prior.") For instance, if we say that $p_a(\mathbf{x})$ is uniform over some range that is much larger than the support of $p_b(\mathbf{x})$, then
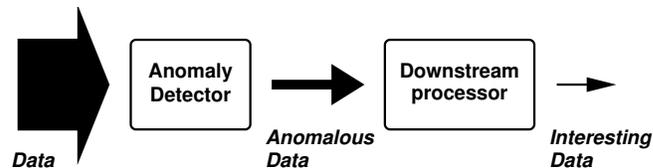
$$\mathcal{A}(\mathbf{x}) = \frac{\text{constant}}{p_b(\mathbf{x})}, \tag{2}$$

so $\mathcal{A}(\mathbf{x}) \gtreqless \eta$ is equivalent to saying $p_b(\mathbf{x}) \lesseqgtr \eta'$ with the threshold $\eta' = \text{constant}/\eta$. This shows how anomaly detection is equivalent to "density level detection" [2].

A second way to respond is by invoking the generalized likelihood ratio (GLR). The GLR is widely used, and has proven useful over a broad range of disciplines, particularly including remote sensing data analysis. Unlike the simple likelihood ratio, however, claims of optimality no longer apply. For the anomaly detection problem, one can set up a hypothesis testing scenario

$$H_o : \mathbf{x} = \mathbf{z} \tag{3}$$
$$H_1 : \mathbf{x} = \mathbf{z} + \mathbf{w} \tag{4}$$

Anomaly Detector → **Downstream processor** →

Data *Anomalous Data* *Interesting Data*

Fig. 1. Anomaly detection as a cueing process. The difference between anomalous (or "potentially interesting") data and truly interesting data is a judgment that is made downstream of the anomaly detector. But that judgment is feasible because it is only performed on a small subset of the original data.

where H_o is the null hypothesis of no anomaly, and H_1 is the alternative hypothesis that an anomalous quantity ($\mathbf{w}$) has been added to a non-anomalous value $\mathbf{z}$. Here, $\mathbf{z}$ is the non-anomalous pixel value whose probability density function is given by $p_b(\mathbf{z})$, and $\mathbf{w}$ is the unknown anomaly. The GLR takes the form

$$\mathcal{A}(\mathbf{x}) = \frac{\max_{\mathbf{w}} p_b(\mathbf{x} - \mathbf{w})}{p_b(\mathbf{x})} = \frac{\text{constant}}{p_b(\mathbf{x})} \quad (5)$$

which gives the same result as (2).

One interpretation of the GLR result is that it "avoids" making any assumptions about $p_a(\mathbf{x})$. A second interpretation is that it "effectively" assumes that $p_a(\mathbf{x})$ is uniform, but because the assumption is not explicit, the analyst does not recognize it, and therefore cannot question it. Or alter it.

The GLR is itself problematic. As has been noted by several authors [3, 4, 5], it is a non-unique and non-optimal solution to the composite hypothesis testing problem. Also, in order to even use the GLR for anomaly detection, one has to invoke an additive model in (3), which is a fairly prescriptive view of what constitutes an anomaly.

And it bears remarking that some kinds of anomaly detection prefer choices of $p_a(\mathbf{x})$ that are not uniform. A detector of spectrally sparse anomalies (arising, *e.g.*, from gas-phase chemical plumes) was developed in [6], put more weight on anomalies $\mathbf{w}$ whose spectrum was concentrated in a small number of channels. A restriction to anomalous color used a different $p_a(\mathbf{x})$ in [7]. As we'll see later in Section 5, anomalous change detection can be derived in terms of yet another $p_a(\mathbf{x})$ which depends on background distributions from the individual images [8]. But even here, the choice of $p_a(\mathbf{x})$ is broad and relatively flat, compared to the background distribution $p_b(\mathbf{x})$.

1.2. $p_b(\mathbf{x})$ is not really known – a *technical* problem

Regardless of whether a GLR is employed, or a $p_a(\mathbf{x})$ is posited, or what choice is made for that $p_a(\mathbf{x})$, the key *technical* challenge for good anomaly detection is the modeling of the background distribution $p_b(\mathbf{x})$ [9]. A natural candidate is the multivariate Gaussian, which leads to a Mahalanobis distance: $\mathcal{A}(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu})^T R^{-1} (\mathbf{x} - \boldsymbol{\mu})$ where $\boldsymbol{\mu}$ and R are the mean and covariance of the Gaussian distribution. In practice, these quantities are estimated from the data: $\widehat{\boldsymbol{\mu}} = (1/N) \sum_{i=1}^{N} \mathbf{x}_i$, and $\widehat{R} = (1/N) \sum_{i=1}^{N} (\mathbf{x}_i - \widehat{\boldsymbol{\mu}})(\mathbf{x}_i - \widehat{\boldsymbol{\mu}})^T$, and then

$$\mathcal{A}(\mathbf{x}) = (\mathbf{x} - \widehat{\boldsymbol{\mu}})^T \widehat{R^{-1}} (\mathbf{x} - \widehat{\boldsymbol{\mu}}). \quad (6)$$

The usual choice is $\widehat{R^{-1}} = \widehat{R}^{-1}$, that is: estimate the inverse with the inverse of the estimate. But regularized approximations are often better [10, 11, 12].

Equation (6) is often accompanied by the name RX, although the original RX detector [13, 14] referred specifically to the local variant. As written, (6) corresponds to a global RX, in which a single $\boldsymbol{\mu}$ and a single R are computed for the whole image. In the local variant, a separate $\boldsymbol{\mu}$ is computed for every pixel in the image, based on the mean of pixels in an annulus around that pixel. The covariance R is again computed as an average of $(\mathbf{x} - \widehat{\boldsymbol{\mu}})(\mathbf{x} - \widehat{\boldsymbol{\mu}})^T$, with the average computed over the same annulus used to compute $\boldsymbol{\mu}$, or over a larger annulus [15], or over the whole image (local mean, global covariance). In [16], $\widehat{\boldsymbol{\mu}}$ is estimated not as a local mean of pixel values in the annulus, but as a more general function of those pixel values, where that function is learned from the rest of the image.

An interesting extension of (6) is to map $\mathbf{x}$ into a feature space $\Phi(\mathbf{x})$ and compute Mahalanobis distance in that feature space; this was originally suggested in [17], and applied to hyperspectral anomaly detection in [18] (and corrected in [19]).

1.2.1. Generative and Discriminative models

"When solving a given problem," Vapnik advises [20, p. 30], "try to avoid solving a more general problem as an intermediate step." Methods that seek to estimate $p_b(\mathbf{x})$ as a first step, and then compare $p_b(\mathbf{x}) \lessgtr \eta$ as a second step are doing more work in that first step than is actually necessary. One doesn't need to know *everything* about $p_b(\mathbf{x})$; one only needs the contour associated with $p_b(\mathbf{x}) \gtrless \eta$ for some η.

The emphasis on estimating $p_b(\mathbf{x})$ with some $\widehat{p}_b(\mathbf{x})$ leads to "generative" models. If you assume $p_b(\mathbf{x})$ is Gaussian, as in RX, or that it is a finite mixture model of Gaussians, or if you use a kernel density estimator, then you are building a generative model.

The emphasis on identifying the values of $\mathbf{x}$ that are on either side of the $\widehat{p}_b(\mathbf{x}) \gtrless \eta$ boundary leads to "discriminative" models. The support vector machine is a classic discriminative model [21], but there are simpler examples. For instance, one might model the data by a single ellipsoid that encloses most of the data. Such a model is robust (indeed, oblivious) to variations in $p_b(\mathbf{x})$ that occur in the "core" of the distribution, and instead models the distribution on the "periphery" where the contour that discriminates background from anomalous will be drawn. The MINVOL algorithm [22] provides one solution to this problem; a support vector machine with a quadratic kernel also achieves this goal [23].

2. PERFORMANCE CRITERIA

Given that they elude definition, it's no surprise that anomalies are hard to measure. And anomaly detectors are hard to compare. The literature on anomaly detection tends to lean on anecdotal demonstrations (think: Forest Radiance). But because anomalies are rare, it is difficult to find enough anecdotes to achieve any kind of statistical precision. And it is even more difficult to have any confidence that the known anomalies in a given dataset are at all representative of the anomalies that you'd like your anomaly detector to find.

One very sensible approach is to implant targets into an image [24, 25]. But an alternative (and essentially equivalent) approach is to compute the volume inside an anomaly detection surface; this is proportional to the number of implanted anomalous targets (assuming that the implanted targets are drawn from a uniform $p_a(\mathbf{x})$) that are incorrectly identified as background. Thus, a good anomaly detector will have a small volume. Since the anomaly detection surface will be calibrated to a given false alarm rate, one can make a coverage plot of log-volume versus log-false alarm rate [23, 26, 27].

This solution is not a panacea. It for instance doesn't work for subspace methods such as SSRX [28, 29], whose anomaly surface is open and therefore has infinite volume. Yet SSRX is often observed to be an operationally effective anomaly detector. A second potential issue, described below in Section 6, is that coverage curves are not strictly invariant to nonlinear coordinate changes.

3. IN-PLANE VS CROSS-PLANE ANOMALIES

When anomaly detection occurs in a high-dimensional space, as it does with hyperspectral imagery, then a qualitative difference emerges between the high-variance dimensions and the low-variance dimensions. One difference that has been empirically observed is that the high-variance dimensions tend to be less Gaussian than the low-variance dimensions [30, 31, 32]. This difference has led to anomaly detectors that treat those directions differently. For instance, the SSRX (subspace RX) algorithm [28, 29] uses principal

components and projects out the largest variance directions, performing RX detection on the remaining directions. In [26], the high-variance directions are modeled by a simplex and the low-variance directions by Gaussians. The Gaussian/Non-Gaussian (G/NG) decomposition is also explored in [33], using a non-parametric histogram-based model for the high-variance directions.

One can qualitatively characterize anomalies as being "in-plane" (anomalous because they are far from the mean in the high-variance directions) or "cross-plane" (far from the mean in the low-variance directions). In-plane anomalies tend to be "qualitatively similar but quantitatively different" from most of the data, whereas cross-plane anomalies tend to be more qualitatively different. These are subjective descriptions, but in-plane and cross-plane anomalies tend to have subjectively different characters.

When SSRX projects out the large-variance directions before performing traditional RX, it preferentially seeks the cross-plane anomalies. By contrast, the kernel-RX detector described in [18, 19] effectively projects out the directions (in the kernel-induced space) that are not in the subspace spanned by the data samples used to train the detector; thus, it seeks in-plane anomalies.

4. TRANSDUCTIVE MODELING

Traditional machine learning is careful to distinguish the labeled data from which the model is trained from the unlabeled data to which the model is applied. This makes operational sense: if the purpose of the model is to supply labels, then there is no point in learning a model to supply labels to data that is already labeled. One is often tempted to evaluate the performance of a classifier on the same data from which the classifier is trained, but this is properly regarded as a sin.

In the transductive variant of traditional machine learning [20, p. 293], one is allowed to include the unlabeled data samples that the model will be applied to as part of the model building process. The reason this is not sinful is that those data samples do not have labels.

One interpretation of the anomaly detection problem is that there are two classes (anomalous and background), and labeled training data is available only for one class (background). Under this interpretation, one trains an anomaly detector on the non-anomalous data, and then applies it to wholly distinct set of data.

But what happens more commonly in practice is that one has a given dataset and one presumes that the vast majority of samples in that dataset are not anomalous. For training purposes, one treats *all* of the data as if it were non-anomalous, learns a model, and then applies this model back to the very same data that was used to train it. The samples that are identified as most anomalous are the candidates for anomalies. Again, because the data were not labeled, sin is avoided, grace is attained, and with any luck, some interesting (or at least potentially interesting) data samples are identified.

5. ANOMALOUS CHANGE DETECTION

For the change detection problem, one has two images that are co-registered to each other, and the anomalousness at a pixel depends on the values, call them $\mathbf{x}$ and $\mathbf{y}$, of the corresponding pixels in the two images. Perhaps even more so than for anomaly detection, anomalous change detection is an excellent way to reduce data and cue analysts [34]. Although a (surprisingly large) number of change detection algorithms begin with the simple subtraction $\mathbf{x} - \mathbf{y}$, these algorithms tend to be overly-sensitive to pervasive differences (*e.g.*, in lighting or calibration) between the two images. In [35], the "chronochrome" was introduced; this also involves subtracting the

two images, but only after doing a least-squares fit of one to the other; the anomalousness of change is measured using an RX detector on the difference image.

Distribution-based (instead of subtraction-based) anomalous change detection was introduced in [8]. Here, one employs a likelihood ratio like the one in (1), but in terms of the pixels in both of the images $(\mathbf{x}, \mathbf{y})$:

$$\mathcal{A}(\mathbf{x}, \mathbf{y}) = \frac{p_a(\mathbf{x}, \mathbf{y})}{p_b(\mathbf{x}, \mathbf{y})}. \tag{7}$$

This still leaves open the question of what to use for $p_a(\mathbf{x}, \mathbf{y})$. A uniform distribution is one option, but this says more about whether the pair $(\mathbf{x}, \mathbf{y})$ is anomalous than whether the change $\mathbf{x} \leftrightarrow \mathbf{y}$ is unusual. If we focus on one-directional change $\mathbf{x} \rightarrow \mathbf{y}$, and consider only anomalousness in $\mathbf{y}$, then we can write $p_a(\mathbf{x}, \mathbf{y}) = p_b(\mathbf{x})p_a(\mathbf{y})$ and if we let $p_a(\mathbf{y})$ be uniform, then

$$\mathcal{A}(\mathbf{x}, \mathbf{y}) = \frac{p_b(\mathbf{x})}{p_b(\mathbf{x}, \mathbf{y})}. \tag{8}$$

If we fit Gaussian models to $p_b(\mathbf{x})$ and $p_b(\mathbf{x}, \mathbf{y})$, then it can be shown [36] that the chronochrome detector emerges. Note that there is a fundamental asymmetry in (8); using $p_b(\mathbf{y})$ instead of $p_b(\mathbf{x})$ for the numerator in (8) produces a different anomalous change detector, one that is optimized for $\mathbf{y} \rightarrow \mathbf{x}$ changes. If we care about both $\mathbf{x} \rightarrow \mathbf{y}$ and $\mathbf{y} \rightarrow \mathbf{x}$, we can write

$$\mathcal{A}(\mathbf{x}, \mathbf{y}) = \frac{p_b(\mathbf{x})p_b(\mathbf{y})}{p_b(\mathbf{x}, \mathbf{y})}. \tag{9}$$

When $p_b(\mathbf{x}, \mathbf{y})$ is Gaussian, this leads to the hyperbolic anomalous change detector (HACD) [36]. But more generally, (9) allows both generative modeling with elliptically-contoured distributions [37] and discriminative modeling with support vector machines [38].

Note, by the way, that (9), unlike (8), does not suffer from the ambiguity of coordinate choice described below in Section 6.

6. CHANGE OF COORDINATES: A CAUTIONARY TALE

One should choose $p_a(\mathbf{x})$ to be generally broad and flat. The uniform distribution is a choice that is intuitive, natural, and convenient. But one should not assume that it is somehow "optimal." To see this, consider expressing $\mathbf{x}$ in some different coordinates; for instance $\widetilde{\mathbf{x}} = \log \mathbf{x}$. A distribution that is uniform in $\widetilde{\mathbf{x}}$ will be equivalent to a non-uniform distribution in $\mathbf{x}$. In general, if $\widetilde{\mathbf{x}} = f(\mathbf{x})$, then

$$p_a(\widetilde{\mathbf{x}}) = \left| \frac{\partial f}{\partial \mathbf{x}} \right|^{-1} p_a(\mathbf{x}). \tag{10}$$

If $f(\mathbf{x})$ is a linear function, then the coefficient is a constant, and the distribution will be uniform in $\mathbf{x}$ *and* in $\widetilde{\mathbf{x}}$, but if $f(\mathbf{x})$ is nonlinear, then $p_a(\mathbf{x})$ and $p_a(\widetilde{\mathbf{x}})$ cannot *both* be uniform.

For hyperspectral imagery, two popular choices for expressing $\mathbf{x}$ are radiance and reflectance. For linear transformations between the two, there is no effect. But one cannot expect precise equivalence if there are nonlinearities in the conversion.

All of this suggests that one should not put too much sweat into the question of exactly what $p_a(\mathbf{x})$ should be. The philosophy tells us that it should be broad and flat, but not what its exact functional form should be. Instead, we should concentrate our technical efforts on getting a good model for $p_b(\mathbf{x})$. As is often the case in life, the technical trumps the philosophical.

And vice versa.

7. REFERENCES

[1] William Goldman, "The Princess Bride," 1987.

[2] S. Ben-David and M. Lindenbaum, "Learning distributions by their density levels: A paradigm for learning without a teacher," *J. Computer and System Sciences*, vol. 55, pp. 171–182, 1997.

[3] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*, Springer, New York, 2005.

[4] A. Schaum, "Continuum fusion: a theory of inference, with applications to hyperspectral detection," *Optics Express*, vol. 18, pp. 8171–8181, 2010.

[5] J. Theiler, "Confusion and clairvoyance: some remarks on the composite hypothesis testing problem," *Proc. SPIE*, vol. 8390, pp. 839003, 2012.

[6] J. Theiler and B. Wohlberg, "Detection of spectrally sparse anomalies in hyperspectral imagery," *Proc. IEEE Southwest Symposium on Image Analysis and Interpretation*, pp. 117–120, 2012.

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

[8] 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.

[9] 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*, 2014, To appear (doi:10.1109/JSTARS.2014.2315772).

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

[11] 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.

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

[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*, vol. 38, pp. 1760–1770, 1990.

[14] 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.

[15] 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.

[16] J. Theiler and B. Wohlberg, "Regression framework for background estimation in remote sensing imagery," *Proc. WHISPERS*, 2013.

[17] D. Cremers, T. Kohlberger, and C. Schnörr, "Shape statistics in kernel space for variational image segmentation," *Pattern Recognition*, vol. 36, pp. 1929–1943, 2003.

[18] H. Kwon and N. M. Nasrabadi, "Kernel RX: a new nonlinear anomaly detector," *Proc. SPIE*, vol. 5806, pp. 35–46, 2005.

[19] N. M. Nasrabadi, "Hyperspectral target detection: an overview of current and future challenges," *IEEE Signal Processing Magazine*, vol. 34, pp. 34–44, Jan 2014.

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

[21] D. Tax and R. Duin, "Data domain description by support vectors," in *Proc. ESANN99*, M Verleysen, Ed., Brussels, 1999, pp. 251–256, D. Facto Press.

[22] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*, Wiley-Interscience, New York, 1987.

[23] J. Theiler and D. Hush, "Statistics for characterizing data on the periphery," *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 4764–4767, 2010.

[24] W. F. Basener, E. Nance, and J. Kerekes, "The target implant method for predicting target difficulty and detector performance in hyperspectral imagery," *Proc. SPIE*, vol. 8048, pp. 80481H, 2011.

[25] Y. Cohen, Y. August, D. G. Blumberg, and S. R. Rotman, "Evaluating sub-pixel target detection algorithms in hyperspectral imagery," *J. Electrical and Computer Engineering*, vol. 2012, pp. 103286, 2012.

[26] J. Theiler, "Ellipsoid-simplex hybrid for hyperspectral anomaly detection," *Proc. WHISPERS*, 2011.

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

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

[29] A. Schaum, "Hyperspectral anomaly detection: Beyond RX," *Proc. SPIE*, vol. 6565, pp. 656502, 2007.

[30] S. M. Adler-Golden, "Improved hyperspectral anomaly detection in heavy-tailed backgrounds," *Proc. WHISPERS*, 2009.

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

[32] 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.

[33] 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.

[34] M. T. Eismann, J. Meola, A. D. Stocker, S. G. Beaven, and A. P. Schaum, "Airborne hyperspectral detection of small changes," *Applied Optics*, vol. 47, pp. F27–F45, 2008.

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

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

[37] 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 Letters*, vol. 7, pp. 271–275, 2010.

[38] I. Steinwart, J. Theiler, and D. Llamocca, "Using support vector machines for anomalous change detection," *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 3732–3735, 2010.