Anisotropic background models for spectral target detection

James Theiler

Space Remote Sensing and Data Science Group,
Los Alamos National Laboratory,
Los Alamos, NM 87545, USA

ABSTRACT

Algorithms are derived for detecting targets in cluttered backgrounds, where the background is modeled as a product of univariate distributions independently fit to each of the principal component projections. Thus, fatter-than-Gaussian tails are fit to the data, with a different fatness parameter for each principal component. Comparisons are made to elliptically-contoured distributions (which, unlike these product distributions, are isotropic in the whitened space), including the multivariate t and the Gaussian. Numerical experiments are performed on hyperspectral data from the SHARE 2012 exercise, with target detection performance evaluated on both actual and simulated targets. Both direct and residual data are considered, with the residual data obtained from local background subtraction – these residual data are found to exhibit not only lower variance, but qualitatively different tail statistics. More direct target-agnostic measures are also employed to asses how well these models fit the different kinds of background clutter.

Keywords: Clutter, Clairvoyant fusion, Composite hypothesis testing, Elliptically-contoured distribution, Anisotropic distribution, Hyperspectral imagery, Target detection

1. INTRODUCTION

To detect targets in spectral imagery, one needs a model for the target, a model for the background, and a model for how the target interacts with the background.

The traditional background model is the multivariate Gaussian, characterized by a mean and a covariance matrix. Although subtle issues remain in optimally estimating a covariance matrix from data, the Gaussian is straightforward and leads to simple closed-form expressions for target and anomaly detection. A somewhat broader class of background distribution, first developed for signal detection in radar clutter,[1,2] but adapted for hyperspectral target detection by Manolakis and colleagues,[3,4] and since then developed more widely,[5–9] are the elliptically-contoured (EC) distributions. These distributions are similar to the Gaussian (indeed, the Gaussian is a special case) in that they are characterized by a mean vector and a covariance matrix. But instead of the e^{-r^2} scaling of the density in the radial direction, usually fatter tails are employed. The multivariate t distribution is perhaps the most popular of the EC distributions, and it has the advantage that projecting a d_1-dimensional t distribution to $d_2 < d_1$ dimensions leads to a d_2-dimensional t distribution.*

Much more general non-parametric (and non-isotropic) distributions have also been proposed and employed in hyperspectral image analysis,[10] including manifolds,[11,12] endmembers,[13] kernel-based models[14–16], variable-bandwidth kernels,[17,18] rotation-based iterative Gaussianization,[19–21] and even a highly anisotropic "urchin" model.[22]

The approach taken here is a kind of middle-ground between the isotropic EC models and generic non-parametric models. Following an approach introduced by Adler-Golden[23] (and discussed more recently by Schaum[24]), we will consider distributions that are products of marginal (*i.e.*, one-dimensional) distributions, and in particular, are products of marginal distributions in the principal component directions.

Email: jt@lanl.gov

*Note that the converse is *not* true. If the marginals are t-distributed, that does not guarantee that the higher-dimensional distribution is t-distributed; in particular, the product of lower-dimensional EC distributions is not in general an EC distribution.

This approach is motivated, but not constrained, by the observation[5,23] that higher-numbered (lower variance) principal components tend to be more Gauss-like in their tails, while lower-numbered (higher variance) components tend to have fatter tails.

Unlike the fully non-parametric approaches, this approach also begins with a mean and covariance matrix, and from that a whitening of the the data that preserves the principal component directions. This transformed (mean subtracted and whitened) data has zero mean and unit variance in each of the principal component directions. But, unlike EC models, this whitened data is not isotropic, and in particular the nature (*i.e.*, the "fatness") of the tails will be separately modeled for each of the principal components.

2. ANISOTROPIC TAILS

In a seminal paper, published in the first WHISPERS proceedings, Adler-Golden introduced a model for heavy-tailed background clutter, and used it to improve anomaly detection.[23] His model was an explicit distribution, given by the product

$$P_{\text{bkg}}(\mathbf{x}) = P_1(x_1)P_2(x_2)\cdots P_d(x_d) \tag{1}$$

where x_k corresponds to the kth principal component of $\mathbf{x}$.[†] This is an approximation because principal components are not, in general, truly *independent*; but they are uncorrelated; and in the special case that the data are Gaussian, they *are* independent. So although it is "only" an approximation, it is an approximation which includes the Gaussian as a special case (if the background distribution *were* Gaussian, then it would be exact).

In Adler-Golden's formulation,[23] each principal component is modeled by a different heavy tailed distribution, taken to be of the form

$$P_k(x) = c_k \, \exp(-|a_k x|^{-p_k}). \tag{2}$$

and a different exponent p_k is fit to each principal component. Here a_k and c_k are functions of p_k (see Appendix A).

An alternative form employs the t distribution. Here,

$$P_k(x) = c_k \left[(\nu_k - 2) + x^2\right]^{-(1+\nu_k)/2} \tag{3}$$

where ν_k characterizes how fat the tail of the distribution is.

In these anisotropic models, the parameters p_k (or ν_k) are generally different for each k. If it were to turn out that all the p_k's (or ν_k's) were equal to the same value, we would *still* have an anisotropic distribution. (An exception is if $p_k = 2$ for all k, or $\nu_k = \infty$ for all k; in that case, the distributions would be Gaussian.)

3. FORMULATION FOR TARGET DETECTION IN CLUTTER

Let $\mathbf{x} \in \mathbb{R}^d$ represent the reflectance spectrum that is observed for a pixel in an image with d spectral channels. If $\mathbf{z}$ is the background spectrum (*i.e.*, what the spectrum would be for a pixel if no target were present), write $\mathbf{x} = \boldsymbol{\xi}(\mathbf{z})$ to indicate the effect of the target on the background spectrum. For the replacement target model that will be used in the experiments here, $\boldsymbol{\xi}(\mathbf{z}) = (1-a)\mathbf{z} + a\mathbf{t}$, where a is the target abundance and $\mathbf{t}$ is the target spectrum. As noted in [25], we can write

$$\mathcal{L}(\mathbf{x}) = \frac{P_{\text{target}}(\mathbf{x})}{P_{\text{bkg}}(\mathbf{x})} = \frac{P_{\text{bkg}}(\boldsymbol{\xi}^{-1}(\mathbf{x}))}{P_{\text{bkg}}(\mathbf{x})} \left|\frac{d\boldsymbol{\xi}}{d\mathbf{x}}\right|^{-1} \tag{4}$$

as the likelihood ratio of target to non-target. Here $|d\boldsymbol{\xi}/d\mathbf{x}|$ is the absolute value of the determinant of the Jacobian of $\boldsymbol{\xi}$. For the replacement target model, this becomes

$$\left|\frac{d\boldsymbol{\xi}}{d\mathbf{x}}\right| = (1-a)^d \tag{5}$$

[†]Note on notation: Scalar x_k (with x un-bolded) is the kth principal component of the vector $\mathbf{x}$. But vector $\mathbf{x}_n$ (with $\mathbf{x}$ in boldface) is the nth data sample (*e.g.*, the nth pixel in a spectral image).

so that

$$\mathcal{L}(a, \mathbf{x}) = (1-a)^{-d} \frac{P_{\text{bkg}}\left(\dfrac{\mathbf{x} - a\mathbf{t}}{1-a}\right)}{P_{\text{bkg}}(\mathbf{x})}. \tag{6}$$

This expression provides an optimal detector when a is known (the so-called clairvoyant case[26]), or can be used to produce a GLRT detector by maximizing over a (*e.g.*, see [27] for Gaussian P_{bkg} or [8] for multivariate t), or a Bayesian detector by integrating over a (*e.g.*, see [28,29]). Here, the Veritas[25] method is employed, which uses Eq. (6) with a fixed at three-sigmas relative to the background.

3.1 Flow loss

Although the ultimate measure of model quality is how well it detects targets of interest, it is useful to have a direct, and target-agnostic, measure of how well a given approximation matches the true distribution. A natural choice is the Kullback-Liebler divergence, which naturally splits into two terms:

$$\begin{aligned} D_{KL}(P_{\text{true}}, P_{\text{model}}) &= \int P_{\text{true}}(\mathbf{x}) \log\left(\frac{P_{\text{true}}(\mathbf{x})}{P_{\text{model}}(\mathbf{x})}\right) d\mathbf{x} \\ &= -\underbrace{\int -P_{\text{true}}(\mathbf{x}) \log P_{\text{true}}(\mathbf{x})\, d\mathbf{x}}_{\text{true entropy}} + \underbrace{\int -P_{\text{true}}(\mathbf{x}) \log P_{\text{model}}(\mathbf{x})\, d\mathbf{x}}_{\text{flow loss}}. \end{aligned} \tag{7}$$

The first term is the entropy of the true distribution, and although it is not known, it does not depend on the choice of model; it is a fixed constant. The second term we call the *flow loss*,[‡] though it is essentially the average negative log likelihood of the data samples with respect to the model. In particular, if we have data $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ that is presumed to be drawn from $P_{\text{true}}(\mathbf{x})$, then:

$$L = \int -P_{\text{true}}(\mathbf{x}) \log P_{\text{model}}(\mathbf{x})\, d\mathbf{x} \approx \frac{1}{N} \sum_{n=1}^{N} -\log P_{\text{model}}(\mathbf{x}_n) \tag{8}$$

The flow loss provides a measure of how well the model distribution fits the data.

4. EXPERIMENTS

Fig. 1 shows ROC curves for detection of painted panels (green and yellow) in imagery collected as part of the SHARE 2012 campaign.[30,31] There are only a few targets, and they are spatially clumped, so they do not give a robust statistical characterization of algorithm performance. They do show, however, that these algorithms can detect real targets in real backgrounds.

A less realistic, but statistically more informative, approach is to implant targets into the scene. This enables more targets to be in the scene, and enables the targets to be observed against the full range of background variation that is available in the image. This also avoids the target mismatch problem, which arises when the "known" spectrum of the target (often obtained in a laboratory setting) doesn't match the spectrum exhibited in the image.[§]

We use the two-histogram/matched-pair approach,[32–34] and make two copies of the image, one with targets implanted in every pixel, and one without any implanted targets. We train on the data without implanted targets, but apply the detector to both images: detections in the image with targets contribute to the detection rate while detections in the image without targets contribute to the false alarm rate. ROC curves (from five resamplings of in-sample/out-of-sample) are shown in Figures 2 and 3, with summarizing statistics (based on 15

[‡]The term arises from the normalizing flows literature (*e.g.*, see [19]), we are using it here more generally. A smaller flow loss indicates a better fit of the model to the true distribution.

[§]This is both a feature and a bug; the target mismatch problem is real, and by avoiding it here, we are not "fully" testing the detection algorithm. In this case, however, our main interest is not in absolute characterization of ultimate performance, but in the relative performance differences that arise from different background models.

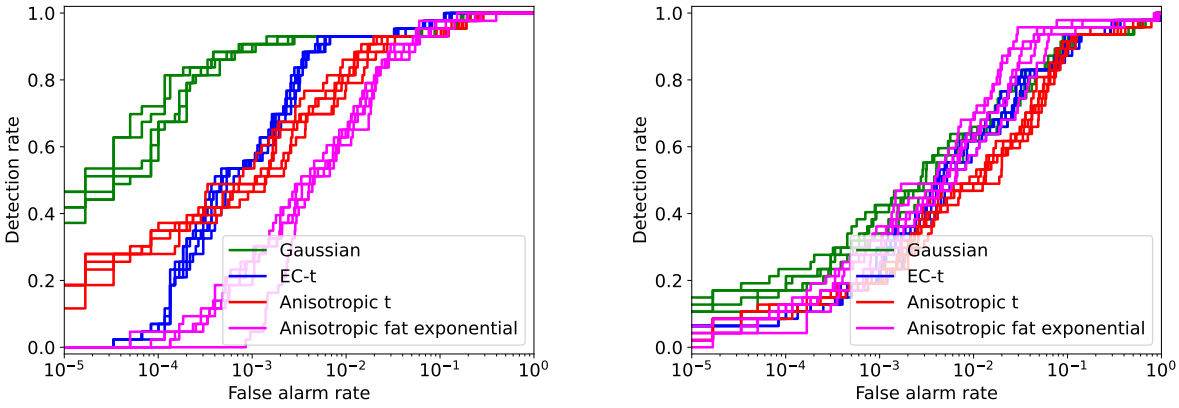

Figure 1. ROC curves for RIT SHARE 2012 imagery with yellow targets (left) and green targets (right).

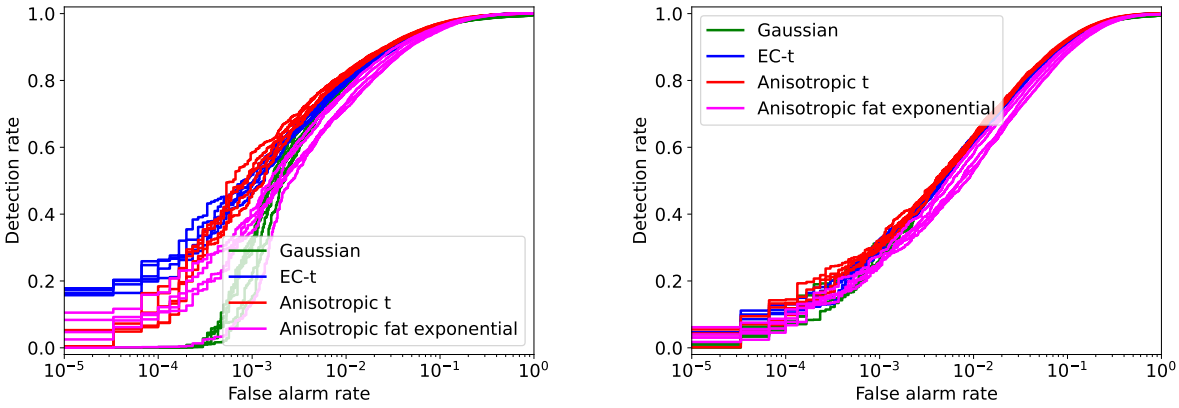

Figure 2. ROC curves for RIT SHARE 2012 imagery with *simulated* yellow targets (left) and green targets (right). The targets are added with an abundance factor of $a = 0.015$.

resamplings) shown in Tables 1 and 2. As with the real data, we see that the yellow targets are more detectable than the green targets at the same abundance. Unlike the real data, however, we see the fat-tailed distributions achieving the best detections.

A more careful comparison of algorithm performance is shown in Tables 1 and 2. Three different statistics are used for comparison: a false alarm rate (at the threshold for which detection rate is one half), area under the ROC curve (the reported value is 1-AUC so that smaller values are better), and the flow loss described in Section 3.1. For both the yellow and the green targets, we see the best performance given by the t-distributed background models. The anisotropic t is slightly better than the EC-t, though the difference is smaller than the run-to-run variation characterized by the error bars. The anisotropic fat exponential model outperformed the Gaussian model for the yellow targets, but not for the green targets. Anecdotally, it was found that fitting the fat exponential model parameters was sensitive to details of the fitting process. The moment-based method described in Appendix A was used for fitting the parameter, but better performance was observed when p was limited to a range $0.1 \leq p \leq 2.0$.

5. CONCLUSIONS

An appealing feature of the marginal product distributions that were investigated here is that they exhibit "adaptably fat" tails, with different fatness in different directions. For target detection applications, the tails are particularly important, because it is only in the tails of the background that targets can be detected. Direct fitting of the distribution at the tails is challenging because that is where the sample data is most sparse. The

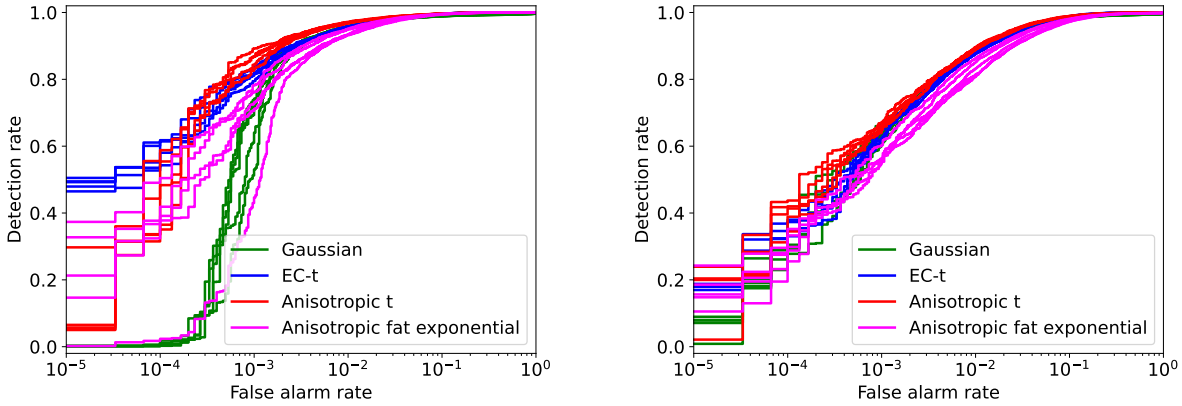

Figure 3. ROC curves for RIT SHARE 2012 imagery with *simulated* yellow targets (left) and green targets (right). The targets are added with an abundance factor of $a = 0.02$.

Table 1. Performance measures using different model distributions, applied to the detection of yellow panels in RIT SHARE 2012 imagery. Scores and error bars are based on 25 runs with different in-sample/out-of-sample splits; the reported values are for out of sample performance. We observe that the Gaussian model is outperformed by all of the alternatives. The anisotropic-t had the lowest false alarm rate (FAR@DR=0.5) and the highest area under the ROC curve (1-AUC is lowest), but the EC-t has the lowest flow loss.

Model Distribution	FAR@DR=0.5	1-AUC	Flow loss
Gaussian	0.001862 ± 0.000198	0.021708 ± 0.000718	1.424240 ± 0.016900
EC-t	0.001056 ± 0.000164	0.014936 ± 0.000858	1.321160 ± 0.003094
Anisotropic-t	0.000859 ± 0.000152	0.013172 ± 0.000679	1.371280 ± 0.002676
Anisotropic fat exponential	0.001179 ± 0.000232	0.018172 ± 0.001498	1.328920 ± 0.004757

Table 2. Same as Table 1, but with the green panels as targets. Since the background is the same, the flow loss numbers are identical with to those in Table 2. For the same strength implanted target, we see that the green panels are harder to detect (higher false alarm rates and lower AUC scores), but we again see the anisotropic-t as the best detector. For this target, however, the anisotropic fat exponential was even worse than the Gaussian.

Model Distribution	FAR@DR=0.5	1-AUC	Flow loss
Gaussian	0.004969 ± 0.000572	0.037216 ± 0.001953	1.424240 ± 0.016900
EC-t	0.004756 ± 0.000516	0.033540 ± 0.002394	1.321160 ± 0.003094
Anisotropic-t	0.004668 ± 0.000561	0.030024 ± 0.001709	1.371280 ± 0.002676
Anisotropic fat exponential	0.007864 ± 0.000917	0.041644 ± 0.002163	1.328920 ± 0.004757

advantage of fitting parametric models is that the density at the tails can in part be "extrapolated" from the density nearer to the core of the distribution. The emphasis of the models considered here (the Student-t and the fat exponential) is to characterize those tails analytically, without putting much effort into characterizing the details of the core of the distribution even though the core is "where all the data are".

The use of these simple distributional models pushes back against the over-fitting that is virtually inevitable when high-dimensional anisotropy is considered. These distributions are products of d one-dimensional distributions, and these d separate factors are independently fit, and each factor is fit based on all of the data samples.

As good as these marginal product distributions look "on paper," however, what this preliminary study found is that they did not substantially outperform their isotropic counterparts. The anisotropic models (especially those based on the Student-t distribution) were certainly competitive, and even slightly better, and this author's opinion is that they merit further study. What follows are some thoughts on what issues those further studies might consider.

Since the background distribution is, above all, a characterization of the data in the hyperspectral imagery, a natural next step would be to consider other datasets – data from other locations and/or from other sensors could well yield different background distributions.

Another limitation of the current study is that it considers only the replacement target model. As Eq. (4) and Eq. (19) show, any target model can be used, including Beer's Law gas detection.[35–38] Adler-Golden's original application was anomaly detection,[23] and it may be that this kind of explicit background model is most well adapted to anomaly, and possibly also anomalous change,[39–43] detection.

The fitting of both the Student-t and the fat exponential models was achieved by computing higher moments of the data, and then identifying the parameter associated with the higher moment value. One disadvantage of this approach is that extremely fat-tailed distributions have unbounded higher moments. For example, the moment-based approach for the multivariate t distribution described in [9] only works for $\nu \geq 3$. An alternative approach would be a direct estimate of maximum likelihood value by literally maximizing the likelihood (or, equivalently, minimizing the flow loss). Whether that is computationally practical, or will yield more stable estimates, are questions that a future study could resolve.

One final item on this "to-do" list would be to employ local, instead of global, mean subtraction. Here, the mean $\mathbf{m}$ at a given pixel $\mathbf{x}$ is estimated from a local annulus surrounding that pixel; this idea been employed already in a variety of studies,[44–55] and Appendix B shows how the likelihood function is altered for generic target detection. One of the earliest advocates of local mean subtraction was the original RX anomaly detection paper by Reed and Yu,[44] which also remarked that the distribution of the residual $\mathbf{x} - \mathbf{m}$ is often more nearly Gaussian than the unsubtracted $\mathbf{x}$ values. It remains for future studies to clarify whether this remark applies as well to the *tails* of the background distributions.

ACKNOWLEDGMENTS

This work was supported in part by the United States Department of Energy (DOE) through the Laboratory Directed Research and Development (LDRD) program at Los Alamos National Laboratory.

REFERENCES

1. E. Conte and M. Longo, "Characterisation of radar clutter as a spherically invariant random process," *IEE Proc. F (Communications, Radar and Signal Processing)* **134**, pp. 191–197, 1987.
2. M. Rangaswamy, D. Weiner, and A. Ozturk, "Non-Gaussian random vector identification using spherically invariant random processes," *IEEE Trans. Aerospace and Electronic Systems* **29**, pp. 111–124, 1993.
3. D. Manolakis, D. Marden, J. Kerekes, and G. Shaw, "On the statistics of hyperspectral imaging data," *Proc. SPIE* **4381**, pp. 308–316, 2001.
4. D. B. Marden and D. Manolakis, "Using elliptically contoured distributions to model hyperspectral imaging data and generate statistically similar synthetic data," *Proc. SPIE* **5425**, pp. 558–572, 2004.

5. J. Theiler, B. R. Foy, and A. M. Fraser, "Characterizing non-Gaussian clutter and detecting weak gaseous plumes in hyperspectral imagery," *Proc. SPIE* **5806**, pp. 182–193, 2005.
6. A. Schaum and E. Allman, "Detection in hyperspectral imagery based on non-Gaussian elliptically contoured distributions," *Proc. MSS (Military Sensing Symposium) on Camouflage, Concealment, and Deception* , 2006.
7. A. Schaum, E. Allman, J. Kershenstein, and D. Alexa, "Hyperspectral change detection in high clutter using elliptically contoured distributions," *Proc. SPIE* **6565**, p. 656515, 2007.
8. 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.
9. 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* **7**(2), pp. 271–275, 2010.
10. S. Matteoli, M. Diani, and J. Theiler, "An overview of background modeling for detection of targets and anomalies in hyperspectral remotely sensed imagery," *J. Selected Topics in Applied Earth Observations and Remote Sensing (JSTARS)* **7**, pp. 2317–2336, 2014.
11. C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina, "Exploiting manifold geometry in hyperspectral imagery," *IEEE Trans. Geoscience and Remote Sensing* **43**, pp. 441–454, 2005.
12. A. K. Ziemann and D. W. Messinger, "An adaptive locally linear embedding manifold learning approach for hyperspectral target detection," *Proc. SPIE* **9472**, pp. 254–268, 2015.
13. N. B. Gallagher, D. M. Sheen, J. M. Shaver, B. M. Wise, and J. F. Shultz, "Estimation of trace vapor concentration-pathlength in plumes for remote sensing applications from hyperspectral images," *Proc. SPIE* **5093**, pp. 184–194, 2004.
14. G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geoscience and Remote Sensing* **43**, pp. 1351–1362, 2005.
15. H. Kwon and N. M. Nasrabadi, "Kernel RX-algorithm: a nonlinear anomaly detector for hyperspectral imagery," *IEEE Trans. Geoscience and Remote Sensing* **43**, p. 388–397, 2005.
16. J. Theiler and G. Grosklos, "Problematic projection to the in-sample subspace for a kernelized anomaly detector," *IEEE Geoscience and Remote Sensing Letters* **13**, pp. 485–489, 2016.
17. S. Matteoli, T. Veracini, M. Diani, and G. Corsini, "Background density nonparametric estimation with data-adaptive bandwidths for the detection of anomalies in multi-hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters* **11**, p. 163–167, 2014.
18. S. Matteoli, T. Veracini, M. Diani, and G. Corsini, "A locally adaptive background density estimator: An evolution for RX-based anomaly detectors," *IEEE Geoscience and Remote Sensing Letters* **11**, p. 323–327, 2014.
19. C. Meng, Y. Song, J. Song, and S. Ermon, "Gaussianization flows," *Proc. 23rd International Conference on Artificial Intelligence and Statistics* **108**, pp. 4336–4345, 2020.
20. J. A. Padrón-Hidalgo, V. Laparra, and G. Camps-Valls, "Unsupervised anomaly and change detection with multivariate Gaussianization," *IEEE Trans. Geoscience and Remote Sensing* **60**, 2021.
21. J. Theiler and C. X. Ren, "Iterative R&R (Rotation and Remarginalization) for detecting targets in spectral imagery," *Proc. SPIE* **12335**, p. 123350C, 2022.
22. B. Daniel and A. Schaum, "Urchin: an RX-derivative accounting for anisotropies in whitened clutter," *Proc. SPIE* **7695**, p. 769504, 2010.
23. 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)*, IEEE, 2009.
24. A. Schaum, "Sub-pixel detection for disparate target and background probability models," *Proc. SPIE* **12094**, p. 120940P, 2022.
25. J. Theiler, "Veritas: an admissible detector for targets of unknown strength," *Proc. SPIE* **11727**, 2021.
26. S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, vol. II, Prentice Hall, New Jersey, 1998.

27. A. Schaum and A. Stocker, "Spectrally selective target detection," in *Proc. ISSSR (International Symposium on Spectral Sensing Research)*, p. 23, 1997.
28. J. Theiler, S. Matteoli, and A. Ziemann, "Bayesian detection of solid subpixel targets," in *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 3213–3216, 2021.
29. S. Matteoli and J. Theiler, "Bayesian target detection algorithms for solid subpixel targets in hyperspectral images." Under review.
30. 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* **8743**, p. 87430F, 2013.
31. J. P. Kerekes, K. Ludgate, A. Giannandrea, N. G. Raqueno, and D. S. Goldberg, "SHARE 2012: Subpixel detection and unmixing experiments," *Proc. SPIE* **8473**, p. 84730H, 2013.
32. M. Bar-Tal and S. R. Rotman, "Performance measurement in point target detection," *Infrared Physics & Technology* **37**, pp. 231–238, 1996.
33. J. Theiler, "Matched-pair machine learning," *Technometrics* **55**, pp. 536–547, 2013.
34. A. Ziemann, M. Kucer, and J. Theiler, "A machine learning approach to hyperspectral detection of solid targets," *Proc. SPIE* **10644**, p. 1064404, 2018.
35. A. Beer, "Bestimmung der absorption des rothen lichts in farbigen flussigketiten," *Ann. Physik* **86**, pp. 78–88, 1852.
36. D. R. Thompson, I. Leifer, H. Bovensmann, M. Eastwood, M. Fladeland, C. Frankenberg, K. Gerilowski, R. O. Green, S. Kratwurst, T. Krings, B. Luna, and A. K. Thorpe, "Real-time remote detection and measurement for airborne imaging spectroscopy: a case study with methane," *Atmospheric Measurement Techniques* **8**, pp. 4383–4397, 2015.
37. J. Theiler, B. R. Foy, C. Safi, and S. P. Love, "Onboard cubesat data processing for hyperspectral detection of chemical plumes," *Proc. SPIE* **10644**, p. 1064405, 2018.
38. S. P. Love, L. A. Ott, J. Theiler, B. R. Foy, C. L. Safi, M. E. Dale, C. G. Peterson, A. A. Guthrie, N. A. Dallman, K. G. Boyd, P. S. Stein, J. A. Wren, M. C. Proicou, and M. K. Dubey, "High-resolution hyperspectral imaging of dilute gases from cubesat platforms," *American Geophysical Union, Fall Meeting* , pp. abstract #A41K–3107, 2018.
39. A. Schaum and A. Stocker, "Long-interval chronochrome target detection," *Proc. ISSSR (International Symposium on Spectral Sensing Research)* , 1998.
40. A. A. Nielsen, K. Conradsen, and J. J. Simpson, "Multivariate alteration detection (MAD) and MAF post-processing in multispectral bi-temporal image data: new approaches to change detection studies," *Remote Sensing of the Environment* **64**, pp. 1–19, 1998.
41. A. Schaum and A. Stocker, "Linear chromodynamics models for hyperspectral target detection," *Proc. IEEE Aerospace Conference* , pp. 1879–1885, 2003.
42. 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.
43. A. Ziemann and S. Matteoli, "Detection of large-scale and anomalous changes," in *Hyperspectral Image Analysis: Advances in Machine Learning and Signal Processing*, S. Prasad and J. Chanussot, eds., pp. 351–375, Springer, 2020.
44. I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoustics, Speech, and Signal Processing* **38**, pp. 1760–1770, 1990.
45. Y. Cohen and S. R. Rotman, "Spatial-spectral filtering for the detection of point targets in multi- and hyperspectral data," *Proc. SPIE* **5806**, pp. 47–55, 2005.
46. C. E. Caefer, M. S. Stefanou, E. D. Nelson, A. P. Rizzuto, O. Raviv, and S. R. Rotman, "Analysis of false alarm distributions in the development and evaluation of hyperspectral point target detection algorithms," *Optical Engineering* **46**, p. 076402, 2007.
47. 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* **47**, p. 076402, 2008.
48. C. C. Borel and R. F. Tuttle, "Improving the detectability of small spectral targets through spatial filtering," *Proc. SPIE* **7812**, p. 78120K, 2010.

49. C. C. Borel, "Methods to find sub-pixel targets in hyperspectral data," in *Proc. 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, IEEE, 2011.
50. Y. Cohen, D. G. Blumberg, and S. R. Rotman, "Subpixel hyperspectral target detection using local spectral and spatial information," *J. Applied Remote Sensing* **6**, p. 063508, 2012.
51. Y. Cohen, Y. August, D. G. Blumberg, and S. R. Rotman, "Evaluating sub-pixel target detection algorithms in hyper-spectral imagery," *J. Electrical and Computer Engineering* **2012**, p. 103286, 2012.
52. J. Theiler and B. Wohlberg, "Regression framework for background estimation in remote sensing imagery," in *Proc. 5th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, IEEE, 2013.
53. N. Hasson, S. Asulin, S. R. Rotman, and D. Blumberg, "Evaluating backgrounds for subpixel target detection: when closer isn't better," *Proc. SPIE* **9472**, p. 94720R, 2015.
54. J. Theiler, "Symmetrized regression for hyperspectral background estimation," *Proc. SPIE* **9472**, p. 94721G, 2015.
55. J. Theiler and A. Ziemann, "Local background estimation and the replacement target model," *Proc. SPIE* **10198**, p. 101980V, 2017.

APPENDIX A. FITTING FAT EXPONENTIALS TO DATA

The fat exponential is of the form $P(x) = c\,\exp(-|ax|^{-p})$. In the case $p = 2$, this is a Gaussian distribution, but for $p < 2$, the tails of the distribution are fatter than that of a Gaussian.

A.1 Dependence on p

Although there are nominally three parameters – c, a, and p – we also have two constraints on the distribution and that leads to c and a being expressible in terms of p. The first constraint is that the distribution integrates to one; that is:

$$1 = \int_{-\infty}^{\infty} c\,\exp(-|ax|^p)\,dx \tag{9}$$

The condition that the mean is zero is built into the parameterization of the fat exponential; since it is symmetric about zero, it has mean zero. Finally, since the model is to be applied to whitened data, there is a condition that the second moment is unity. That is:

$$1 = \int_{-\infty}^{\infty} x^2\,c\,\exp(-|ax|^p)\,dx \tag{10}$$

To begin, we derive an expression for arbitrary moments. Write

$$\begin{aligned}\kappa_m &= \int_{-\infty}^{\infty} |x|^m\,c\,\exp(-|ax|^p)\,dx = 2c\int_0^{\infty} x^m\,\exp(-(ax)^p)\,dx \\ &= 2c\,a^{-m}\int_0^{\infty} (ax)^m\,\exp(-(ax)^p)\,dx\end{aligned} \tag{11}$$

Let $u = (ax)^p$, so that $du = ap(ax)^{p-1}dx$. Since $ax = u^{-p}$, we have that $dx = \frac{1}{ap}u^{(1-p)/p}du$, and

$$\begin{aligned}\kappa_m &= 2c\,a^{-m}\int_0^{\infty} u^{m/p}\,\exp(-u)\times\frac{1}{ap}u^{(1-p)/p}du \\ &= 2c\,a^{-m-1}p^{-1}\int_0^{\infty} u^{(m+1-p)/p}\,\exp(-u)\,du\end{aligned} \tag{12}$$

At this point, we can use the definition of the gamma function: $\Gamma(1+z) = \int_0^{\infty} u^z\,e^{-u}\,du$ to write

$$\begin{aligned}\kappa_m &= 2c\,a^{-m-1}p^{-1}\Gamma\left(1+\frac{m+1-p}{p}\right) \\ &= \frac{2c}{pa^{m+1}}\Gamma\left(\frac{m+1}{p}\right)\end{aligned} \tag{13}$$

Now from the condition that the integral is one, we have that $1 = \kappa_0 = \frac{2c}{pa}\Gamma(1/p)$, which leads to

$$c = \frac{ap}{2\Gamma(1/p)}. \tag{14}$$

We also have the condition that the variance is one, so that

$$1 = \kappa_2 = \frac{2c}{pa^3}\Gamma(3/p) = 2 \times \underbrace{\left[\frac{ap}{2\Gamma(1/p)}\right]}_{c} \times \frac{1}{pa^3}\Gamma(3/p) = a^{-2}\frac{\Gamma(3/p)}{\Gamma(1/p)} \tag{15}$$

which gives

$$a = \left[\frac{\Gamma(3/p)}{\Gamma(1/p)}\right]^{1/2} \tag{16}$$

Finally, we can use these expressions for c and a to write a general formula for the m'th moment:

$$\begin{aligned}\kappa_m &= \frac{2c}{pa^{m+1}}\Gamma\left(\frac{m+1}{p}\right) = 2 \times \left[\frac{ap}{2\Gamma(1/p)}\right] \times \frac{1}{pa^{m+1}}\Gamma\left(\frac{m+1}{p}\right) = a^{-m}\frac{\Gamma((m+1)/p)}{\Gamma(1/p)} \\ &= \left[\frac{\Gamma(3/p)}{\Gamma(1/p)}\right]^{-m/2} \frac{\Gamma((m+1)/p)}{\Gamma(1/p)}\end{aligned} \tag{17}$$

is an explicit closed-from expression for κ_m as a function of p.

A.2 Estimating p from data

Given a set of scalar points, based on a projection to one of the principal component axes, and scaled to have zero mean and unit variance, we can estimate p for that axis by first estimating one of the moments

$$\widehat{\kappa}_m = \frac{1}{N}\sum_{n=1}^{N} |x_n|^m \tag{18}$$

and then inverting Eq. (17) to find the value of p that leads to this value of κ_m. Since higher moments are more vulnerable to outliers, a smaller value of m is preferred, and in the work here, $m = 1$ was used.

Rather than attempt a direct closed-form inversion of Eq. (17), a numerical approximation was obtained by taking a range of p values, computing the associated κ_m values, and using interpolation to obtain a function that yields p as a function of κ_m.

APPENDIX B. LOCAL MEAN SUBTRACTION

In estimating the background clutter distribution P_{bkg}, each pixel is treated essentially independently. One simple approach (though it can become more complicated) for exploiting the spatial structure in a typical image is to use neighboring pixels to estimate the "local" mean of the distribution. In this case, there is a different P_{bkg} function for each pixel; we can write this as $P_{\text{bkg}}(\mathbf{x}, \mathbf{m})$ where $\mathbf{m}$ is the local estimate of the mean. In practice a global distribution, based on the "residual" background, can be estimated from the values of $\mathbf{x} - \mathbf{m}$ in the residual image. Here, $P_{\text{residual}}(\mathbf{x} - \mathbf{m})$ describes this distribution function, which depends only on the difference $\mathbf{x} - \mathbf{m}$. In this formulation, the likelihood ratio in Eq. (4) becomes

$$\mathcal{L}(\mathbf{x}, \mathbf{m}) = \frac{P_{\text{residual}}(\boldsymbol{\xi}^{-1}(\mathbf{x}) - \mathbf{m})}{P_{\text{residual}}(\mathbf{x} - \mathbf{m})} \left|\frac{d\boldsymbol{\xi}}{d\mathbf{x}}\right|^{-1}. \tag{19}$$

Note that this is *not* (except in special cases[¶]) the same as replacing the image with the residual image and doing standard target detection on the residual image.[55] To use this formula, one must keep track of both $\mathbf{x}$ and $\mathbf{m}$ for every pixel.

[¶]The additive target model is one such special case, because $\mathbf{x} = \boldsymbol{\xi}(\mathbf{z}) = \mathbf{z} + a\mathbf{t}$ implies $\boldsymbol{\xi}^{-1}(\mathbf{x}) - \mathbf{m} = \boldsymbol{\xi}^{-1}(\mathbf{x} - \mathbf{m})$, which implies that $\mathcal{L}(\mathbf{x}, \mathbf{m})$ depends only on $\mathbf{x} - \mathbf{m}$.