

# On the Statistics of Hyperspectral Imaging Data

Dimitris Manolakis, David Marden, John Kerekes and Gary Shaw

MIT Lincoln Laboratory  
244 Wood Street  
Lexington, MA 02420-9185

## ABSTRACT

Characterization of the joint (among wavebands) probability density function (pdf) of hyperspectral imaging (HSI) data is crucial for several applications, including the design of constant false alarm rate (CFAR) detectors and statistical classifiers. HSI data are vector (or equivalently multivariate) data in a vector space with dimension equal to the number of spectral bands. As a result, the scalar statistics utilized by many detection and classification algorithms depend upon the joint pdf of the data and the vector-to-scalar mapping defining the specific algorithm. For reasons of analytical tractability, the multivariate Gaussian assumption has dominated the development and evaluation of algorithms for detection and classification in HSI data, although it is widely recognized that it does not always provide an accurate model for the data. The purpose of this paper is to provide a detailed investigation of the joint and marginal distributional properties of HSI data. To this end, we assess how well the multivariate Gaussian pdf describes HSI data using univariate techniques for evaluating marginal normality, and techniques that use unidimensional views (projections) of multivariate data. We show that the class of elliptically contoured distributions, which includes the multivariate normal distribution as a special case, provides a better characterization of the data. Finally, it is demonstrated that the class of univariate stable random variables provides a better model for the heavy-tailed output distribution of the well known matched filter target detection algorithm.

**Keywords:** Target detection, anomaly detection, hyperspectral, elliptically contoured distributions, spherically invariant random vectors, HYDICE

## 1. INTRODUCTION

The most successful algorithms for target detection and background classification using HSI data employ information about the potential targets and backgrounds. Therefore, their performance depends upon the accuracy and proper use of this information. The spectral signature of a given target or background type can significantly vary as a result of differences in illumination conditions, material composition, and spatial shape. In addition to this inherent variability of targets and backgrounds is the variability caused by the presence of atmospheric effects and sensor noise.

Accurately modeling target variability in target detection algorithms<sup>1,2</sup> may lead to improvement in detection performance. Accurate modeling of background variability leads to improved detection performance and facilitates the development of CFAR detectors. The design of target detectors for hyperspectral imaging applications has been dominated by the use of the multivariate normal distribution for modeling background and noise. However, the performance of Gaussian detectors in non-Gaussian background and noise is inferior to that of detectors specifically optimized for such environments. As expected, optimization for non-Gaussian background requires the development of non-Gaussian statistical models that accurately describe the background and noise variability.

This paper describes an effort to understand and model the single band and multiple (joint) band statistics of HSI data and its effects on the output statistics of anomaly detectors and detectors for targets with known spectral signature. The class of elliptically contoured (EC) distributions, or spherically invariant random vectors (SIRV), provide a powerful model for many non-Gaussian multivariate data, and are examined as a potential model for HSI

---

Address correspondence to D. Manolakis: Tel. 781-981-0524, fax 781-981-7271, e-mail [dmanolakis@ll.mit.edu](mailto:dmanolakis@ll.mit.edu)



**Figure 1.** Run 07 data and the three regions used for statistical analysis.

backgrounds. In Section 2, we discuss the specific HSI data sets used for the analysis and evaluation of the proposed statistical models. The statistical characterization of individual single bands is discussed in Section 3. In Section 4, we provide a brief summary of the key aspects of elliptically contoured distributions as they apply to the characterization of spherically invariant random vectors. These results are used in Section 5 to analyze the joint distribution of HSI data. Section 6 discusses the application of stable distributions to characterize the univariate distribution of certain detection statistics. Finally, we present some conclusions in Section 7.

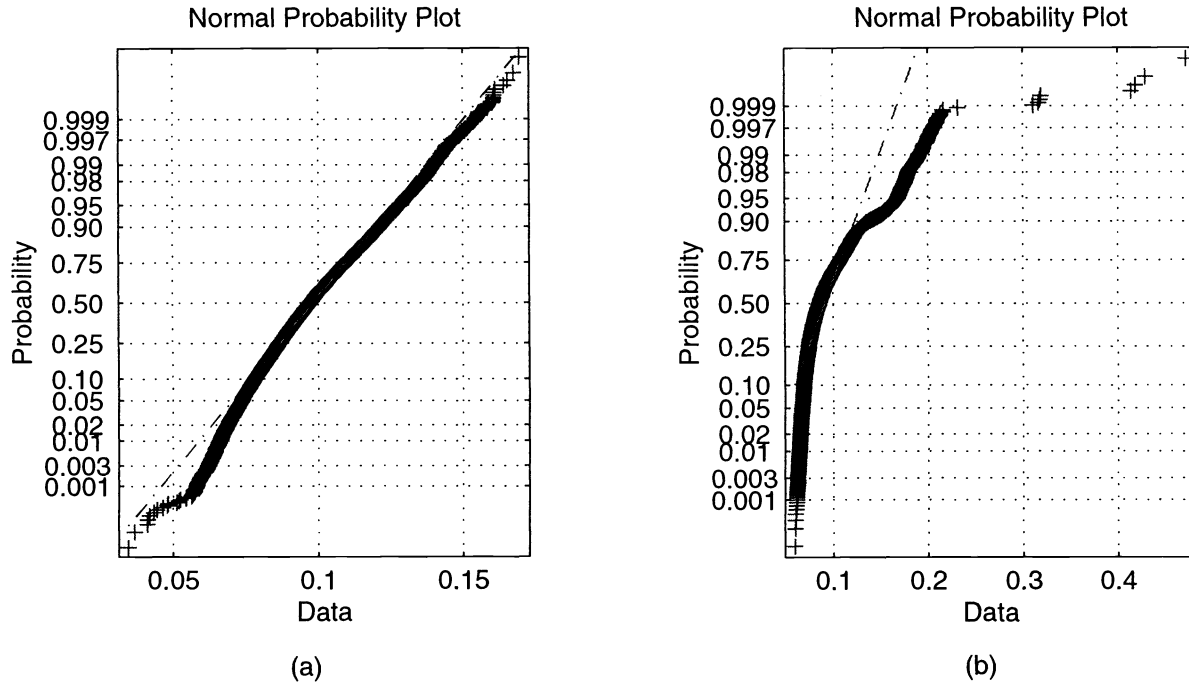
## 2. HSI DATA DESCRIPTION

Airborne hyperspectral imagery data collected by the HYDICE<sup>3</sup> sensor at the U.S. Army Aberdeen Proving Grounds on 24 August 1995 were used in this analysis. HYDICE collects calibrated (after processing) spectral radiance data in 210 wavelengths spanning 0.4 to 2.5  $\mu\text{m}$  in nominally 10 nm wide bands. Figure 1 shows a single band ( $\lambda = 0.565 \mu\text{m}$ ) image of the Run 07 data collected at 9:27 AM local time under clear conditions from an altitude of 3 km. The spatial resolution of the imagery is approximately 1.5 meters.

In examining the statistical properties of the data, several groupings, or classes, were considered. Three regions are identified in the white boxes in Figure 1 describing three classes that were selected by their spatial proximity. In the lower right is a "Grass" region, the middle top is a "Tree" region, and on the left is a "mixed" region. These regions define the pixels selected for three of the classes considered. Also considered were two classes resulting from a supervised classification process performed to isolate spectrally similar (not necessarily spatially adjacent) pixels. Data from two classes selected from this analysis were labeled "Class 2 Grass" and "Class 9 Tree". Table 1 summarizes the classes analyzed and their sample sizes (number of pixels).

**Table 1.** Classes selected for statistical analysis.

Class name	Selection Technique	Sample Size
Grass	Spatially adjacent	7,760
Tree	Spatially adjacent	8,232
Mixed	Spatially adjacent	7,590
Class 2 Grass	Supervised classification	27,351
Class 9 Tree	Supervised classification	25,872



**Figure 2.** Quantile-Quantile (Q-Q) plots of HYDICE data from band 39 ( $\lambda = 0.565\mu\text{m}$ ) for (a) Class 2 Grass and (b) Mixed background data. Also shown are straight lines corresponding to the normal distributions.

The data were analyzed in units of calibrated spectral radiance for the characterization of joint statistics (Section 5) and surface reflectance for the modeling of target detection statistics (Section 6). While hyperspectral data is often transformed to apparent surface reflectance through an atmospheric compensation algorithm, such transformations are usually linear, and as such, do not affect the statistical distributions. Also, for the multivariate analyses examples, only 144 of the 210 channels were used to avoid regions of low signal-to-noise ratio.

One note of caution. While it is known that data artifacts exist in the calibrated spectral radiance due to an interpolation process used to replace "dead" pixels, only the data selected by the supervised classification technique had any type of screening applied for anomalous pixels. The other classes represent typical data as it would be processed by an unsupervised automated algorithm.

### 3. SINGLE WAVEBAND STATISTICS

Before discussing the joint, or multivariate, statistics, it is worthwhile to examine some of the scalar, or marginal, statistics. Figure 2 presents quantile-quantile (Q-Q) plots for two of the classes and one spectral channel, showing the cumulative probability distribution as a function of the data value, overlaid with straight lines representing Gaussian data. As one can see, neither class would be well represented by the Gaussian assumption, although the class selected through the supervised classification process is slightly better matched to the Gaussian assumption. A recent paper<sup>4</sup> provides some additional insight into single waveband statistics. Since, detection and classification applications mainly depend upon the joint distribution of the data, we are not going to further investigate single waveband statistics in this paper.

### 4. RANDOM VECTORS WITH ELLIPTICALLY CONTOURED (EC) DISTRIBUTIONS

The class of EC distributions, or SIRV, contains distributions that have similar features to the multivariate normal distribution, but which exhibit either longer or shorter tails than the normal. In this section, we consider the class of

mixtures of normal distributions, which is a subclass<sup>5</sup> of EC distributions.

A random real vector  $\mathbf{x} = [x_1 x_2 \dots x_p]^T$  with mean  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Gamma}$  is a SIRV if and only if its PDF has the form<sup>5</sup>

$$f_p(\mathbf{x}) = (2\pi)^{-p/2} |\boldsymbol{\Gamma}|^{-1/2} h_p(d) \quad (1)$$

where  $d$  is a quadratic form (Mahalanobis distance) defined by

$$d = (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Gamma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \quad (2)$$

and  $h_p(d)$  is a positive, monotonically decreasing function for all  $p$ . We shall denote such a distribution using the shorthand notation  $EC(\boldsymbol{\mu}, \boldsymbol{\Gamma}, h)$ .

The multivariate normal distribution (MVN) is a special case with

$$h_p(d) = \exp\left(-\frac{1}{2}d\right) \quad (3)$$

In addition, the EC class includes the multivariate  $t$ , the multivariate Cauchy, and the double exponential distributions. The discrete normal (Gaussian) mixture, which is widely used in supervised or unsupervised pattern recognition algorithms, is a special case of SIRV. The normal mixture PDF is a simple finite weighted sum of normal PDFs and can be used for the approximation of many other SIRVs. The common feature of all EC distributions is that they all have elliptical contours of equiprobability. Many of the properties of the MVN distribution extend also to the class of EC distributions. Thus, all the marginals of an EC distribution, and the conditional distributions of some variables, given the values of the others, are also EC.

Any SIRV,  $\mathbf{x}$ , can be represented by

$$\mathbf{x} = \boldsymbol{\Gamma}^{1/2}(v\mathbf{z}) + \boldsymbol{\mu} \quad (4)$$

where

$$\mathbf{z} \sim N(\mathbf{0}, \mathbf{I}) \quad (5)$$

and  $v$  is a non-negative random variable independent of  $\mathbf{z}$ . The type of a SIRV is uniquely determined by the PDF  $f_v(v)$  of  $v$ , which is known as the *characteristic PDF* of  $\mathbf{x}$ . The random variable  $v$  is normalized so the  $E\{v^2\} = 1$ , that is, to unit mean squared error. Clearly,  $f_v(v)$  and  $\boldsymbol{\Gamma}$  can be specified independently.

If we have an analytical expression for  $f_v(v)$ , the class of admissible  $h_p(d)$  functions is obtained by

$$h_p(d) = \int_0^\infty v^{-p} \exp\left(-\frac{1}{2} \frac{d}{v^2}\right) f_v(v) dv \quad (6)$$

The PDF of Mahalanobis distance  $d$  is given by

$$f_p(d) = \frac{1}{2^{p/2} \Gamma(p/2)} d^{p/2-1} h_p(d) \quad (7)$$

and can be used to *uniquely* reduce the multidimensional PDF modeling problem to an equivalent one-dimensional one. If  $\mathbf{x}(n), \dots, \mathbf{x}(N)$  form a random sample from a  $N(\boldsymbol{\mu}, \boldsymbol{\Gamma})$  population, then

$$d(n) = [\mathbf{x}(n) - \boldsymbol{\mu}]^T \boldsymbol{\Gamma}^{-1} [\mathbf{x}(n) - \boldsymbol{\mu}] \quad (8)$$

for  $n = 1, \dots, N$  correspondingly form a random sample from a  $\chi_p^2$  distribution with  $p$  degrees of freedom.

If the PDF  $f_v(v)$  is not available in closed form, we can generate SIRV's and determine  $h_p(d)$  by expressing  $\mathbf{x}$  in spherical coordinates  $r, \theta$ , and  $\phi_1, \phi_2, \dots, \phi_{p-2}$ . The PDF of  $r$  is given by

$$f_p(r) = \frac{r^{p-1}}{2^{p/2-1}\Gamma(p/2)} h_p(r^2) u(r) \quad (9)$$

The angles  $\theta$  and  $\phi_k$  are independent of the radius  $r$  and they do not affect the type of the SIRV.

One of the most important implications of (7) is that the multivariate PDF for any SIRV is uniquely determined by the univariate PDF of the Mahalanobis distance (2). As a result, the multivariate PDF identification problem is reduced to a simpler univariate one. This result provides the cornerstone for our investigations in the statistical characterization of HSI data. Several researchers<sup>6,7</sup> have developed libraries and generation techniques for SIRVs specified by their characteristic PDF or  $h_p(d)$  that can be applied to our objective of modeling unknown hyperspectral backgrounds.

## 5. MULTIPLE (JOINT) WAVEBAND STATISTICS

Since the distribution of multivariate data in the observation space is inherently "sparse", it is impractical to use goodness-of-fit tests to evaluate the capacity of the MVN PDF to characterize the data. In practice, we can assess multivariate normality using techniques from the following categories<sup>8</sup>: (a) Univariate techniques for evaluating marginal normality, (b) Multivariate techniques for evaluating joint normality, and (c) Techniques that use unidimensional views (projections) of multivariate data. These techniques can be extended for the characterization of elliptical<sup>5-7</sup> distributions.

If we know the mean and covariance of a multivariate random sample, we can check for normality by comparing the distribution of the Mahalanobis distance (8) against a chi-squared distribution. However, in practice, we have to estimate<sup>9,10</sup> the mean and covariance from the available data. To this end, we use the maximum likelihood estimates of the mean

$$\hat{\boldsymbol{\mu}} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}(n) \quad (10)$$

and the covariance matrix

$$\hat{\boldsymbol{\Gamma}} = \frac{1}{N} \sum_{n=1}^N [\mathbf{x}(n) - \hat{\boldsymbol{\mu}}][\mathbf{x}(n) - \hat{\boldsymbol{\mu}}]^T \quad (11)$$

to compute the Mahalanobis distance

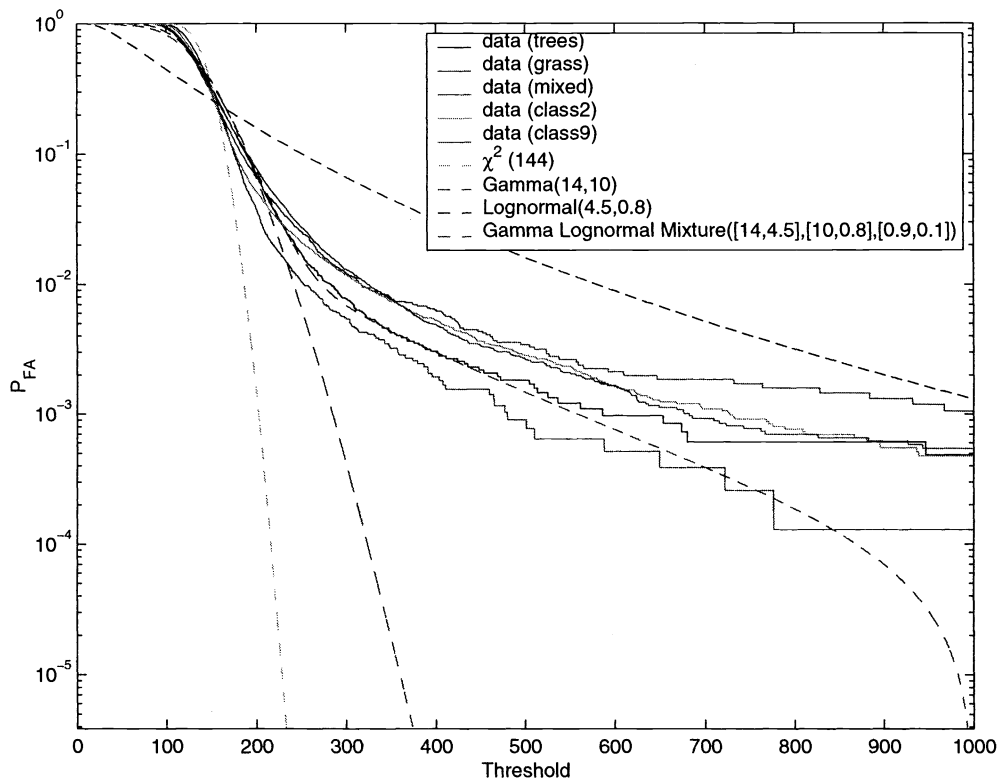
$$d(n) = \frac{1}{N-1} [\mathbf{x}(n) - \hat{\boldsymbol{\mu}}]^T \hat{\boldsymbol{\Gamma}}^{-1} [\mathbf{x}(n) - \hat{\boldsymbol{\mu}}] \quad (12)$$

for all pixels  $1 \leq n \leq N$ . If  $\mathbf{x}(n) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma})$ , it can be shown that  $d(n)$  is distributed according to the beta distribution

$$p_d(d) = \frac{1}{B(a, b)} d^{a-1} (1-d)^{b-1}, \quad 0 \leq d \leq 1 \quad (13)$$

where the parameters

$$a = \frac{p}{2} \quad \text{and} \quad b = \frac{N-p-1}{2} \quad (14)$$



**Figure 3.** Modeling the distribution of the Mahalanobis distance of HSI data.

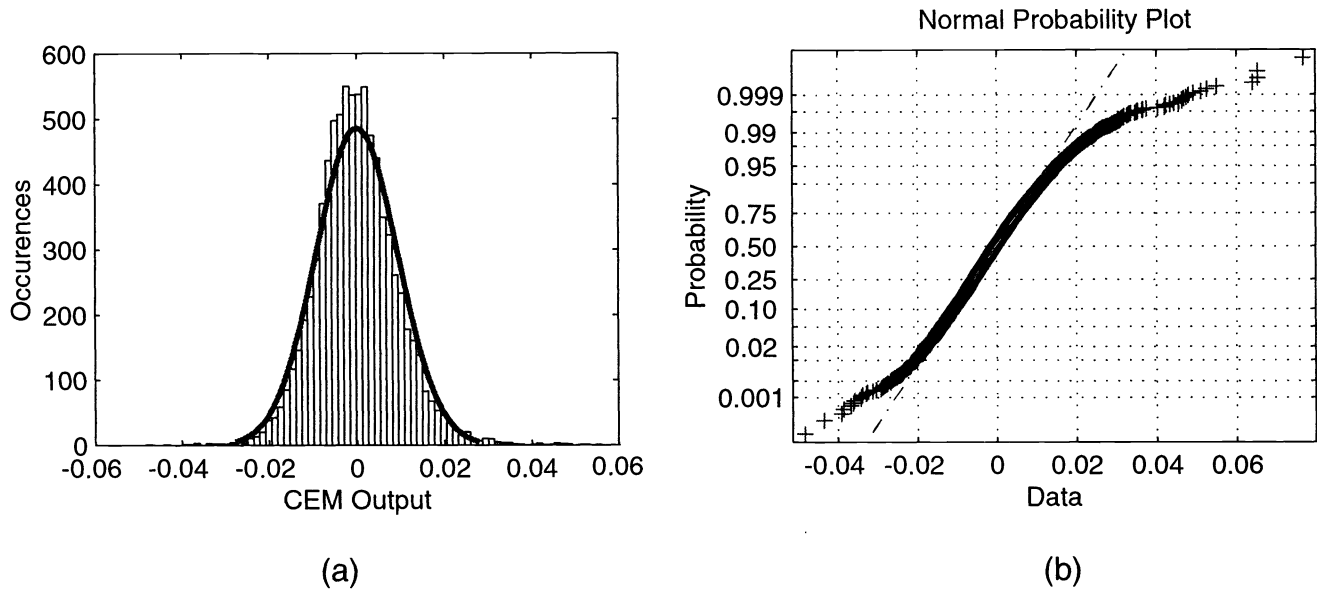
When the sample size,  $N$ , is greater than about  $10p$ , the difference between the true beta distribution and the chi-squared approximation,  $\chi_p^2$ , is negligible for our purposes. Therefore,  $\chi^2$  probability plots provides a good graphical method for testing multivariate normality.

For SIRV processes, the distribution of the Mahalanobis distance completely characterizes their joint PDF. Therefore, to test whether an  $EC(\mu, \Gamma, h)$  fits the data, we follow three steps: (a) compute the Mahalanobis distances (12), (b) compute the expected PDF (7), and (c) compare the two distributions using a probability plot. However, for detection applications, where the key goal is accurate prediction of the probability of false alarm, it is more interesting to compare the empirical and theoretically predicted probabilities of false alarms. It is interesting to note that the Mahalanobis distance (12) is the statistic used by the popular RX algorithm<sup>11,12</sup> for anomaly detection.

Figure 3 shows the probability of false alarm for the data sets discussed in Section 2 and theoretical predictions based on a chi-squared, Gamma, lognormal, and a Gamma-lognormal mixture. These distributions can be uniquely identified by the distribution of their Mahalanobis distance. The corresponding elliptical distributions are determined by uniquely defined marginal distributions. The chi-squared distribution results when the data are obtained from a multivariate normal distribution. Evidently, the Gamma-lognormal mixture provides a good description for the body and the tails of the underlying distribution.

## 6. DISTRIBUTION OF UNIVARIATE MATCHED FILTER DETECTION STATISTICS

The heavy tails in the univariate distribution of the Mahalanobis distance imply heavy tails in the multivariate distribution of the data. Therefore, heavy tails may appear not only in the quadratic Mahalanobis distance, but in



**Figure 4.** CEM output statistic (a) histogram and fitted normal density curve and (b) normal probability plot for the tree data set.

other linear and quadratic statistics employed in several widely used<sup>13,14</sup> target detection techniques. In this section, for illustration purposes, we shall investigate the statistic of the popular matched filter<sup>14</sup> algorithm termed Constrained Energy Minimization<sup>15</sup> (CEM). The algorithm is given by

$$T(\mathbf{x}) = \frac{(\mathbf{s} - \hat{\boldsymbol{\mu}})^T \hat{\boldsymbol{\Gamma}}^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}})}{(\mathbf{s} - \hat{\boldsymbol{\mu}})^T \hat{\boldsymbol{\Gamma}}^{-1} (\mathbf{s} - \hat{\boldsymbol{\mu}})} \quad (15)$$

where  $\mathbf{s}$  is the desired target spectral signature and  $\mathbf{x}$  is the test pixel. If  $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma})$ , and  $N$  is large compared to the number of bands  $L$ ,  $T(\mathbf{x})$  should be normally distributed. Figure 4 shows the histogram of the CEM output for the tree data set and the corresponding normal probability plot. Clearly, the normal probability plot indicates the existence of heavy tails. Similar plots, with smaller or larger deviations from normality, have been produced using the other data sets. Hence, using a normal distribution to predict the probability of false alarm will provide optimistic estimates.

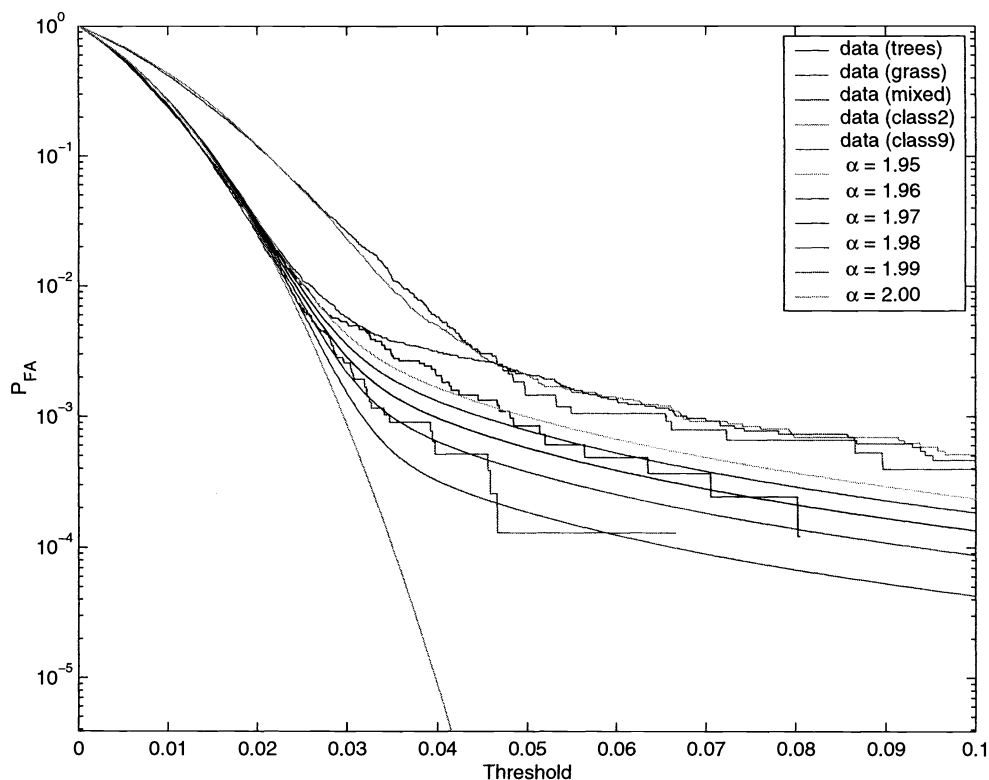
The family of symmetric  $\alpha$ -stable (S $\alpha$ S) distributions provides a good model for data with impulsive behavior. They are characterized by a parameter  $\alpha$  (characteristic exponent) that takes values in the range  $0 < \alpha \leq 2$ . The value  $\alpha = 1$  leads to the Cauchy distribution and the value  $\alpha = 2$  to the Gaussian. The stable distributions result from the central limit theorem if we remove the finite variance constraint. The only stable distribution with finite second-order moments is the normal distribution. Since  $\alpha$ -stable distributions follow from the central limit theorem they are invariant under linear transformations. Since there is no closed-form expression for their probability density function, S $\alpha$ S random variables are specified<sup>16-18</sup> by their characteristic function (that is, the Fourier transform of the PDF)

$$\Phi(\xi) = \exp(j\mu\xi - |\sigma\xi|^\alpha) \quad (16)$$

where  $\alpha$  is the characteristic exponent,  $\sigma$  is a scale parameter, and  $\mu$  is a location parameter. The heaviness of the tails increases as  $\alpha$  increases from 1 (Cauchy) to 2 (Gaussian). The estimation of the parameters of a stable distribution from data is a challenging due to the presence of “spikes”. The published compilation<sup>18</sup> provides a comprehensive

review of statistical techniques for stable distributions from a practical perspective. The estimation method used in this paper<sup>19</sup> is based on the use of the characteristic function.

Figure 5 shows the probability of false alarm when the CEM detector is used for different scenes as well as superimposed theoretical curves obtained using the family of  $S\alpha S$  distributions for various values of  $\alpha$ . It can be seen that the tails of the empirical  $P_{FA}$  curves can be modeled by the heavier tails of the stable distribution.



**Figure 5.** CEM output statistics and their modeling using stable distributions.

## 7. CONCLUSIONS

The preliminary results shown in this paper suggest that the family of EC distributions, which is a generalization of the multivariate normal distribution, can provide a more accurate characterization of HSI data. Further work is in progress to identify the specific members of EC family that capture accurately the the joint statistics of the available HSI data from the HYDICE sensor. Also, an effort to model the occasionally heavy tails of the CEM detector statistics using the family of stable distributions was presented. Again, preliminary results demonstrated the effectiveness of the proposed model. More extensive investigations, which are currently under way, will be presented in future publications.

## REFERENCES

1. B. Thai and G. Healey, "Invariant subpixel target identification in hyperspectral imagery," in *Algorithms for Multispectral and Hyperspectral Imagery IV*, S. S. Shen and M. R. Descour, eds., vol. 3717, pp. 14–24, SPIE, (Orlando, Florida), 1999.
2. D. Manolakis, C. Siracusa, and G. Shaw, "Adaptive matched subspace detectors for hyperspectral imaging applications," IEEE International Conference on ASSP, (Salt Lake City, Utah), May 2001.

3. L. J. Rickard, R. Basedow, E. Zalewski, P. Silvergate, and M. Landers, "HYDICE: An airborne system for hyperspectral imaging," *Proceedings of SPIE, Imaging Spectrometry of the Terrestrial Environment* **1937**, 1993.
4. J. S. Tyo, J. Robertson, J. Wollenbecker, and R. C. Olsen, "Statistics of target spectra in HSI scenes," *SPIE Imaging Spectrometry VI*, August 2000.
5. K. T. Fang, S. Kotz, and K. W. Ng, *Symmetric Multivariate and Related Distributions*, Chapman and Hall, New York, 1990.
6. M. Rangaswamy, D. Weiner, and A. Ozturk, "Non-gaussian random vector identification using spherically invariant random processes," *IEEE Transactions on Aerospace and Electronic Systems* **29**, pp. 111–124, January 1993.
7. M. Rangaswamy, D. Weiner, and A. Ozturk, "Computer generation of correlated non-gaussian radar clutter," *IEEE Transactions on Aerospace and Electronic Systems* **31**, pp. 106–116, January 1995.
8. R. Gnanadesikan, *Methods for Statistical Data Analysis of Multivariate Observations*, Wiley, New York, second ed., 1997.
9. M. Healy, "Multivariate normal plotting," *Appl. Stat.* **17**, pp. 157–161, 1968.
10. D. L. Kessel and K. Fukunaga, "A test for multivariate normality with unspecified parameters," tech. rep., Purdue University School of Electrical Engineering, 1972.
11. I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. on Acoustics, Speech, and Signal Processing* **38**, pp. 1760–1770, October 1990.
12. D. Stein and S. Beaven, "The fusion of quadratic detection statistics applied to hyperspectral imagery," *IRIA-IRIS proceedings of the 2000 Meeting of the MSS Specialty Group on Camouflage, Concealment, and Deception*, pp. 271–280, March 2000.
13. D. Manolakis, G. Shaw, and N. Keshava, "Comparative analysis of hyperspectral adaptive matched filter detectors," in *Algorithms for Multispectral and Hyperspectral Imagery VI*, S. S. Chen and M. R. Descour, eds., vol. 4049, pp. 2–17, SPIE, (Orlando, FL), April 2000.
14. D. Manolakis, C. Siracusa, D. Marden, and G. Shaw, "Hyperspectral adaptive matched filter detectors: Practical performance comparison," in *Algorithms for Multispectral and Hyperspectral Imagery*, SPIE, (Orlando, FL), April 2001.
15. W. H. Farrand and J. C. Harsanyi, "Mapping the distribution of mine tailings in the Coeur D' Alene River Valley, Idaho, through the use of a constrained energy minimization technique," *Remote Sens. Environ.* **59**, pp. 64–76, 1997.
16. D. G. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and Adaptive Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering and Array Processing*, McGraw-Hill Companies, Inc., Boston, 2000.
17. G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Processes: Stochastic Models with Infinite Variance*, Chapman and Hall, New York, 1994.
18. R. J. Adler, R. E. Feldman, and M. S. Taqqu, *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, Birkhauser, Boston, 1998.
19. S. Kogon and D. Williams, "Characteristic function based estimation of stable distribution parameters," in *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, R. J. Adler, R. E. Feldman, and M. S. Taqqu, eds., pp. 311–335, Birkhauser, Boston, 1998.

## ACKNOWLEDGMENTS

We wish to express our gratitude to CAPT Frank Garcia, DUSD (S&T), Program Manager HTAP, for his enthusiastic support. We also express our gratitude to SITAC for providing the calibrated and ground-truthed HYDICE data used in this work.

This work was sponsored by the Department of the Defense Air Force contract F19628-95-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the author and not necessarily endorsed by the United States Air Force.