EC-GLRT: DETECTING WEAK PLUMES IN NON-GAUSSIAN HYPERSPECTRAL CLUTTER USING AN ELLIPTICALLY-CONTOURED GENERALIZED LIKELIHOOD RATIO TEST

James Theiler and Bernard R. Foy

Los Alamos National Laboratory
Los Alamos, NM 87545

ABSTRACT

We investigate the behavior of a detector for weak gaseous plumes in hyperspectral imagery that can be derived in terms of a generalized likelihood ratio test (GLRT) applied to an elliptically-contoured (EC) model for the distribution of background clutter. Two limiting cases of this EC-GLRT detector are the adaptive matched filter (AMF) and the adaptive coherence estimator (ACE). While the general EC-GLRT detector does not share the specific optimality or invariance properties exhibited by these limiting cases, it provides an in-between model that can be competitive with both of them over a broad range of scenarios.

Index Terms— hyperspectral imagery, chemical plume, matched filter, generalized likelihood ratio test, non-Gaussian distribution

1. INTRODUCTION

The optimal detector for a gaseous plume (of a specific chemical) in a hyperspectral image depends on the statistical distribution of the background clutter and on the strength of the plume, if present; when both of these are known, then the optimal detector can be written in terms of a likelihood ratio test. Specifically, let $\mathbf{r} \in \mathcal{R}^d$ correspond to the d-channel spectrum at a single hyperspectral pixel; for a weak plume, it can be expressed as the sum of the plume-free background given by $\mathbf{x} \in \mathcal{R}^d$ and a plume whose effect is the product of the scalar plume strength ε, and the chemical signature of the target gas $\mathbf{t}$. That is:

$$\mathbf{r} = \mathbf{x} + \varepsilon\mathbf{t}. \qquad (1)$$

We remark that this formulation neglects a lot of relevant nonlinear physical phenomena [1, 2, 3, 4], but for a weak plume, this linearized expression provides a useful approximation [1, 4]. In Eq. (1), $\mathbf{t}$ is known and $\mathbf{r}$ is measured; the background $\mathbf{x}$ and plume strength ε are both unknown.

While $\mathbf{x}$ itself is unknown, we will assume that we know (or, at least, can adequately estimate) a probability distribution function $P(\mathbf{x})$. The *simple* hypothesis testing formula-

tion considers the two hypotheses:

$$H_o : \qquad \mathbf{r} = \mathbf{x} \qquad (2)$$
$$H_1 : \qquad \mathbf{r} = \mathbf{x} + \varepsilon\mathbf{t}, \text{ with } \varepsilon = \varepsilon_o. \qquad (3)$$

The null hypothesis (H_o) is that there is no plume present; the alternative (H_1) is that there is a plume of strength ε_o. In this case, the optimal detector can be directly expressed as the ratio of likelihoods.

$$\mathcal{L}(\varepsilon_o; \mathbf{r}) = \frac{P(\mathbf{r} - \varepsilon_o\mathbf{t})}{P(\mathbf{r})}. \qquad (4)$$

The likelihood ratio test (LRT) compares $\mathcal{L}(\varepsilon_o; \mathbf{r})$ with a threshold η; observations $\mathbf{r}$ are labeled on-plume or off-plume, depending on whether $\mathcal{L}(\varepsilon_o; \mathbf{r})$ is greater or less than η. By adjusting η, the trade-off can be made between missed detections and false alarms.

In fact, this optimal detector is sometimes referred to as the "clairvoyant" detector [5] because it assumes that the signal strength ε_o for the alternative hypothesis is known.

In practice, however, a *composite* hypothesis test is more appropriate; here, the alternative hypothesis

$$H_1 : \qquad \mathbf{r} = \mathbf{x} + \varepsilon\mathbf{t}, \text{ with } \varepsilon > 0 \qquad (5)$$

covers a multitude of cases. In general, optimal detectors for composite hypothesis tests are not available, but a popular heuristic is the generalized likelihood ratio test (GLRT) [6, 7, 5], which addresses the unknown ε by using the maximum likelihood estimate as if it were the known value.

$$\mathcal{G}(\mathbf{r}) = \max_{\varepsilon > 0} \mathcal{L}(\varepsilon; \mathbf{r}) = \frac{\max_{\varepsilon > 0} P(\mathbf{r} - \varepsilon\mathbf{t})}{P(\mathbf{r})}. \qquad (6)$$

It turns out that the GLRT is optimal for Gaussian $P(\mathbf{x})$ [7], but that is a special case.

In what follows, we can assume without loss of generality that mean values have been subtracted, so $\langle \mathbf{r} \rangle = \langle \mathbf{x} \rangle = \langle \varepsilon \rangle = 0$; we will also assume that ε is small enough (and the sample size is large enough) that we can approximate the covariance K of the clutter $\mathbf{x}$ with the sample covariance of the measurements $\mathbf{r}$; that is: $K \approx \langle \mathbf{rr}^T \rangle$. The validity of this approximation is discussed in Ref. [8].
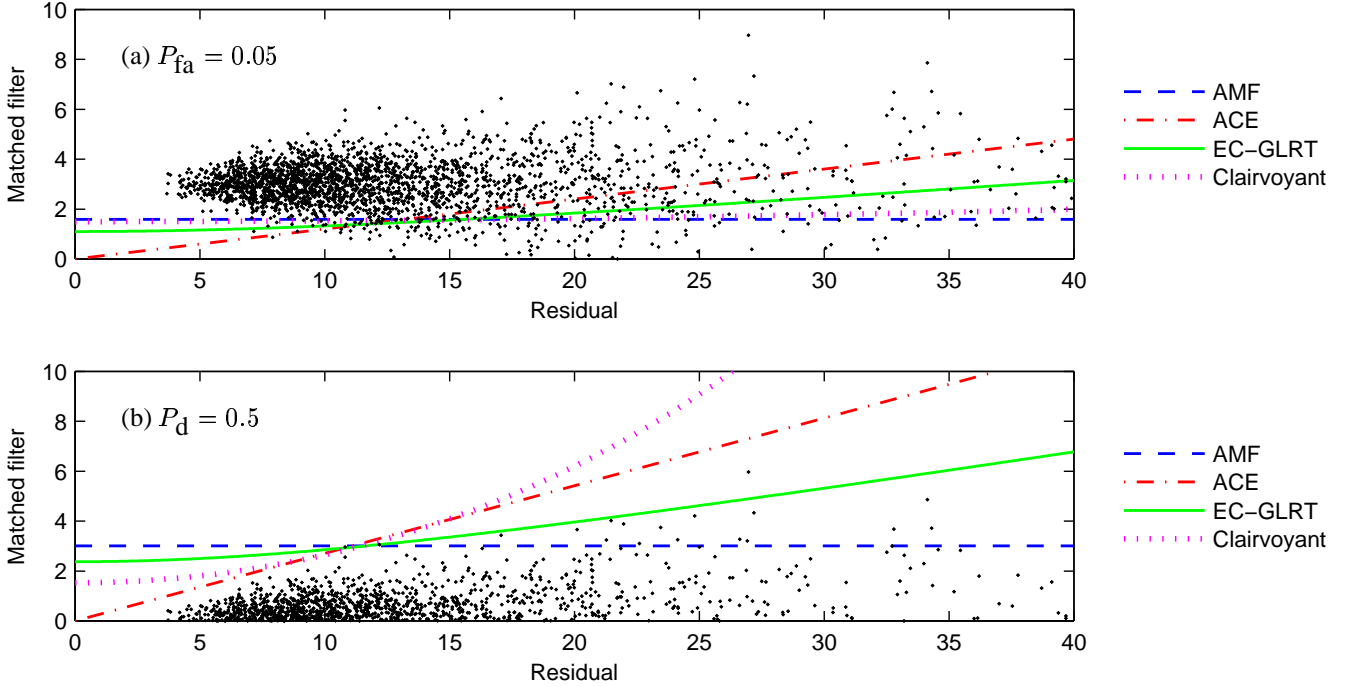
Fig. 1. Detector boundaries for the AMF, ACE, and EC-GLRT (with $\nu = d + 2$) detectors are shown in matched-filter-residual coordinates. Also shown is the clairvoyant detector, assuming signal strength of 3 sigmas, and an EC-distribution with $\nu = 2.5$. (a) In this plot, the detectors are calibrated to have the same $P_{\mathrm{fa}} = 0.05$ false alarm rate; data are shown for a three-sigma signal, and those points above the curves will be detections. Here, AMF exhibits the highest detection rate (fewest missed detections). (b) In this plot, the detectors are calibrated to have the same $P_{\mathrm{d}} = 0.5$ detection rate for three-sigma signals; the data shown have no signal at all, so all the points above the curves are false alarms. Here, ACE has the lowest false alarm rate. These results correspond to the three-sigma points on the curves shown in Fig. 2(c,d).

The class of elliptically-contoured (EC) distributions has found utility both for radar [9] and hyperspectral imagery [10, 11]. These distributions are of the form $P(\mathbf{x}) = H(\mathbf{x}^T K^{-1}\mathbf{x})$, where $\mathbf{x}^T K^{-1}\mathbf{x}$ is the squared Mahalanobis distance to the centroid of the data (which is the origin, since the data have been mean-subtracted), and $H : \mathcal{R}^+ \to \mathcal{R}^+$ is a monotonically decreasing scalar function. Note that for the Gaussian distribution, $H(z) = (2\pi)^{-d/2}|K|^{-1/2}e^{-z/2}$.

Following Ref. [10], we consider the multivariate t-distribution, given by

$$P(\mathbf{x}) = \text{constant} \times \left(1 + \frac{\mathbf{x}^T K^{-1}\mathbf{x}}{\nu - 2}\right)^{-(d+\nu)/2}. \qquad (7)$$

where the constant depends on ν, d, and $|K|$, but not on $\mathbf{x}$. In general, we require $\nu > 2$, and note that the $\nu \to \infty$ limit leads to the Gaussian distribution.

With this EC model, we can write an expression for the clairvoyant detector (which requires knowledge of signal strength ε), in terms of a discriminator $\mathcal{D}$ which varies monotonically with the likelihood ratio $\mathcal{L}(\varepsilon; \mathbf{r})$ given in Eq. (4),

using the $P(\mathbf{x})$ from Eq. (7):

$$\mathcal{D}(\varepsilon; \mathbf{r}) = \mathcal{S}\left((\nu - 1)\left[1 - \mathcal{L}(\varepsilon; \mathbf{r})^{-2/(d+\nu)}\right]\right) \qquad (8)$$

$$= \mathcal{S}\left(\frac{(\nu - 1)\left(2\varepsilon \mathbf{t}^T K^{-1}\mathbf{r} - \varepsilon^2 \mathbf{t}^T K^{-1}\mathbf{t}\right)}{(\nu - 2) + \mathbf{r}^T K^{-1}\mathbf{r}}\right) \qquad (9)$$

where $\mathcal{S}(z) = \text{sign}(z)\sqrt{|z|}$ is the signed square root. To obtain the EC-GLRT detector, note the ε that maximizes this expression is given by $\varepsilon = \mathbf{t}^T K^{-1}\mathbf{r}/\mathbf{t}^T K^{-1}\mathbf{t}$. Thus:

$$\text{EC-GLRT} = \sqrt{\frac{(\nu - 1)}{(\nu - 2) + \mathbf{r}^T K^{-1}\mathbf{r}}}\,\frac{\mathbf{t}^T K^{-1}\mathbf{r}}{\sqrt{\mathbf{t}^T K^{-1}\mathbf{t}}}. \qquad (10)$$

In the limit $\nu \to \infty$, the distribution in Eq. (7) becomes Gaussian, and the detector in Eq. (10) becomes the adaptive matched filter:

$$\text{AMF} = \frac{\mathbf{t}^T K^{-1}\mathbf{r}}{\sqrt{\mathbf{t}^T K^{-1}\mathbf{t}}}. \qquad (11)$$

In the limit $\nu \to 2$, the detector in Eq. (10) becomes the adaptive coherence estimator [12]:

$$\text{ACE} = \frac{\mathbf{t}^T K^{-1}\mathbf{r}}{\sqrt{(\mathbf{r}^T K^{-1}\mathbf{r})(\mathbf{t}^T K^{-1}\mathbf{t})}}. \qquad (12)$$
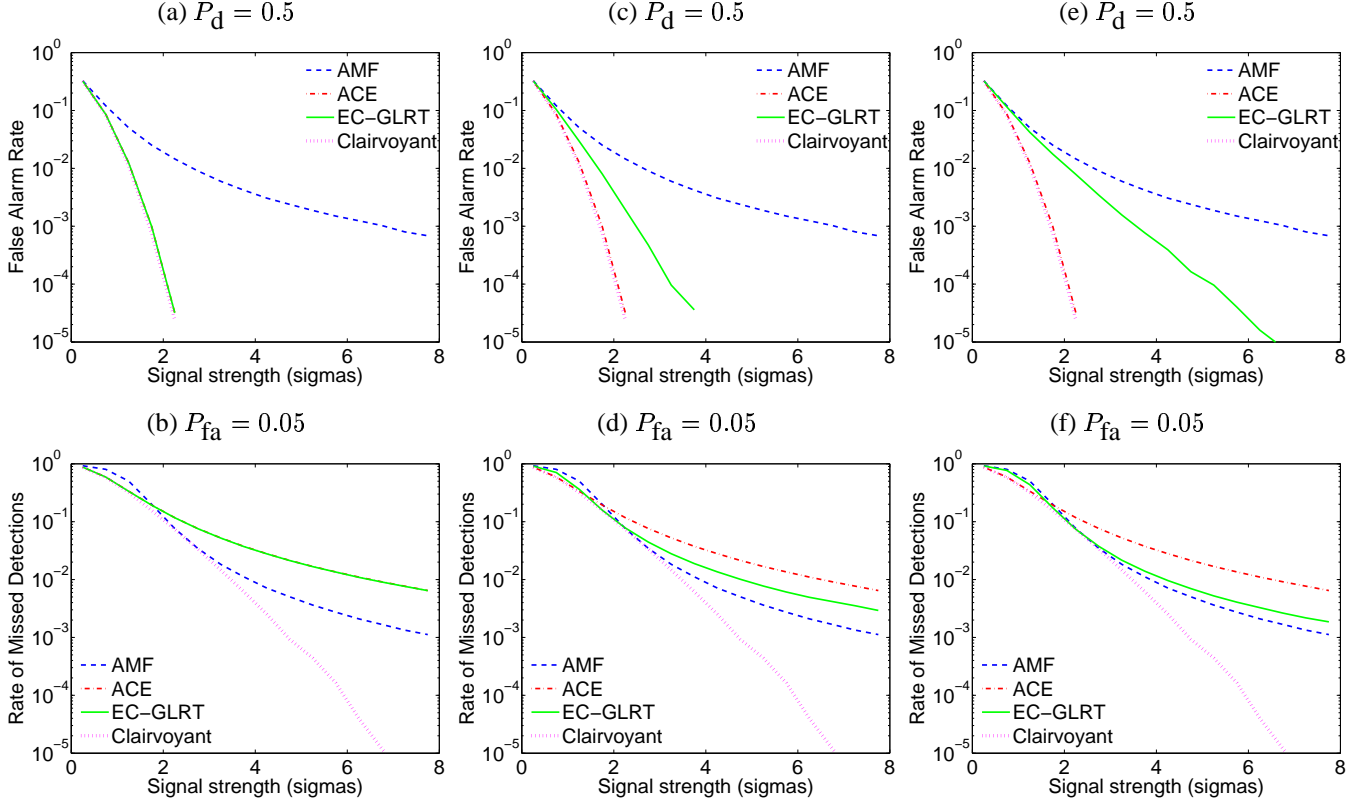
Fig. 2. Performance of AMF, ACE, and EC-GLRT detectors on a simulated elliptically contoured dataset (simulated with $\nu = 2.5$) with a simulated plume. In panels (a,b), the EC-GLRT detector uses $\nu = 2.5$, matching that of the data; in panels (c,d), the EC-GLRT detector uses $\nu = d + 2 = 226$. In panels (e,f), the EC-GLRT detector uses $\nu = 1000$.

It is useful to plot data in matched-filter-residual (MFR) coordinates (introduced in Ref. [13]); here a pixel with value $\mathbf{r}$ is plotted with MF $= \mathbf{t}^T K^{-1} \mathbf{r} / \sqrt{\mathbf{t}^T K^{-1} \mathbf{t}}$ and residual $R = \sqrt{\mathbf{r}^T K^{-1} \mathbf{r} - (\text{MF})^2}$. The AMF, ACE, EC-GLRT, and the optimal "clairvoyant" detectors all depend only on MF and R, and they are shown in these coordinates in Fig. 1.

We have compared the performance of the AMF, ACE, and EC-GLRT detectors over a range of parameters, using simulated Gaussian, EC, and actual hyperspectral data. In the experiment with Gaussian data (not shown), using $d = 224$ channels, we confirmed that the AMF was optimal, but we saw nearly identical (and therefore nearly optimal) performance for the ACE and EC-GLRT algorithms as well.

The results of an experiment with EC data are shown in Fig. 2. Here simulated data is a multivariate t-distribution with $\nu = 2.5$ (extremely fat-tailed). The simulation is for a 500×500 pixel image with $d = 224$ spectral channels. If we use $\nu = 2.5$ for our EC-GLRT detector [Fig. 2(a,b)], then its performance is virtually identical to that of the ACE detector. Using an EC-GLRT detector with $\nu = d + 2 = 226$, we find a detector whose performance is mid-way between that of AMF and ACE. With $\nu = 1000$; the EC-GLRT detector is much better than AMF in the low false-alarm rate regime, and better than ACE in the high detection rate regime.

2. CONCLUSIONS

For Gaussian data, the AMF is known to be the optimal detector; but it is a fragile optimality because, for high-dimensional Gaussian data, all of the algorithms give nearly identical performance. For deviations from Gaussianity that can be characterized with an EC distribution, as has been suggested for hyperspectral data [10], substantial differences between the algorithms can be observed. Which algorithm is better, however, depends not so much on the background distribution as the regime of interest.

In the very low false alarm rate region (with fixed $P_{\mathrm{d}} = 0.5$), the EC-GLRT (with parameter ν chosen to match the clutter distribution) appears to be nearly optimal. For high dimensional data ($d \gg \nu$), the EC-GLRT detector (with ν matched to that of the distribution) is nearly identical to the ACE detector. When d is large, but ν is also large, then the distribution is nearly Gaussian and the performance of ACE and EC-GLRT is nearly identical. Thus, a practical recommendation for high dimensional data in the very low false alarm rate regime is to use the ACE detector. On the other hand, in the regime of fixed false alarm ($P_{\mathrm{fa}} = 0.05$) and very high detection rate, the AMF is seen to outperform the ACE and EC-GLRT detectors by a wide margin.

(a) Broadband image   (b) $P_{\mathrm{fa}} = 0.05$ (c) $P_{\mathrm{d}} = 0.5$
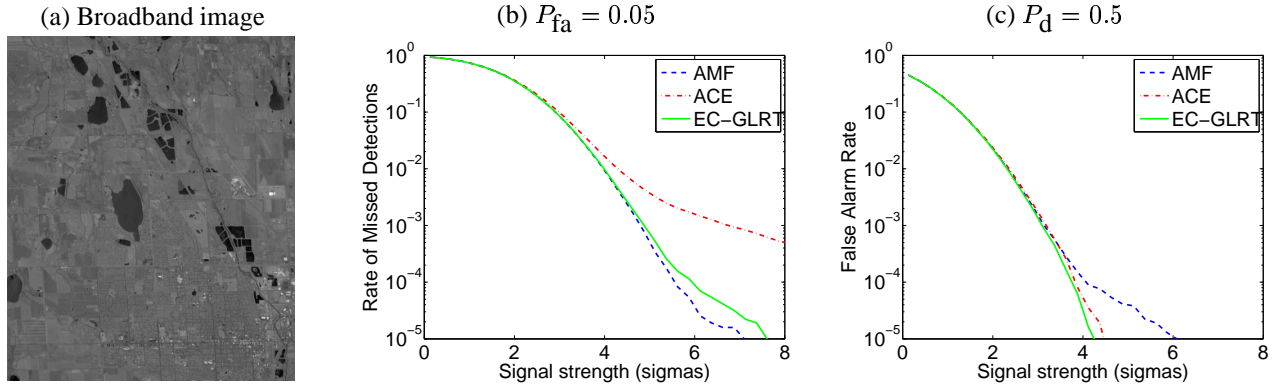
Fig. 3. Performance of AMF, ACE, and EC-GLRT detectors on a dataset with a simulated plume on an AVIRIS hyperspectral image, shown in panel (a). For this image, the AMF is seen in panel (b) to outperform the ACE in a fixed false alarm rate regime, but those roles are reversed in panel (c) which corresponds to a fixed detection rate regime. It will depend on the operational scenario which regime is more important. An EC-GLRT detector (using $\nu = 1000$) is seen to perform well in both regimes.

We have seen that although the EC-GLRT can be derived in terms of a specific background distribution, the optimal EC-GLRT detector is not necessarily the one whose parameter ν is matched to that distribution. On average, the magnitude of $\mathbf{r}^T K^{-1} \mathbf{r}$ is roughly equal to d, the number of spectral channels. By taking the EC-GLRT parameter ν to be of order d, even if the ν of corresponding to the distribution $P(\mathbf{x})$ is much smaller than that, it is possible to use the EC-GLRT to bridge the performance of the ACE and AMF detectors. This is seen in Fig. 3. Applied to real data, the EC-GLRT with $\nu = O(d)$ is competitive with the better of the ACE and AMF detectors in *both* the high detection *and* low false alarm rate regimes. In this usage of the EC-GLRT, there is no need to estimate the actual ν for the underlying data.

3. REFERENCES

[1] A. Hayden, E. Niple, and B. Boyce, "Determination of trace-gas amounts in plumes by the use of orthogonal digital filtering of thermal-emission spectra," *Applied Optics*, vol. 35, pp. 2803–2809, 1996.

[2] S. J. Young, "Detection and quantification of gases in industrial-stack plumes using thermal-infrared hyperspectral imaging," Tech. Rep. ATR-2002(8407)-1, The Aerospace Corporation, 2002.

[3] B. R. Foy, R. R. Petrin, C. R. Quick, T. Shimada, and J. J. Tiee, "Comparisons between hyperspectral passive and multispectral active sensor measurements.," *Proc. SPIE*, vol. 4722, pp. 98–109, 2002.

[4] 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.

[5] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, vol. II, Prentice Hall, New Jersey, 1998.

[6] E. J. Kelly, "An adaptive detection algorithm," *IEEE Trans. Aerospace and Electronic Systems*, vol. 22, pp. 115–127, 1986.

[7] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Processing*, vol. 42, pp. 2146–2156, 1994.

[8] J. Theiler and B. R. Foy, "Effect of signal contamination in matched-filter detection of the signal on a cluttered background," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, pp. 98–102, 2006.

[9] E. Conte and M. Longo, "Characterisation of radar clutter as a spherically invariant random process," *IEE Proc. F (Communications, Radar and Signal Processing)*, vol. 134, pp. 191–197, 1987.

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

[11] A. Schaum, Eric Allman, John Kershenstein, and Drew Alexa, "Hyperspectral change detection in high clutter using elliptically contoured distributions," *Proc. SPIE*, vol. 6565, pp. 656515, 2007.

[12] L. L. Scharf and L. T. McWhorter, "Adaptive matched subspace detectors and adaptive coherence estimators," *Proc. Asilomar Conference on Signals, Systems, and Computers*, 1996.

[13] J. Theiler, B. R. Foy, and A. M. Fraser, "Beyond the adaptive matched filter: nonlinear detectors for weak signals in high-dimensional clutter," *Proc. SPIE*, vol. 6565, pp. 656503, 2007.