

Full Paper

Characterizing Clutter in the Context of Detecting Weak Gaseous Plumes in Hyperspectral Imagery

Tom Burr *, Bernard R. Foy, Herb Fry, and Brian McVey

Mail Stop F600, Los Alamos National Laboratory, Los Alamos NM 87545, USA

E-mails: tburr@lanl.gov, bfoy@lanl.gov, ha_fry@lanl.gov, bdm@lanl.gov

*Author to whom correspondence should be addressed. E-mail: tburr@lanl.gov

Received: 29 September 2006 / Accepted: 17 November 2006 / Published: 23 November 2006

Abstract: Weak gaseous plume detection in hyperspectral imagery requires that background clutter consisting of a mixture of components such as water, grass, and asphalt be well characterized. The appropriate characterization depends on analysis goals. Although we almost never see clutter as a single-component multivariate Gaussian (SCMG), alternatives such as various mixture distributions that have been proposed might not be necessary for modeling clutter in the context of plume detection when the chemical targets that could be present are known at least approximately. Our goal is to show to what extent the generalized least squares (GLS) approach applied to real data to look for evidence of known chemical targets leads to chemical concentration estimates and to chemical probability estimates (arising from repeated application of the GLS approach) that are similar to corresponding estimates arising from simulated SCMG data. In some cases, approximations to decision thresholds or confidence estimates based on assuming the clutter has a SCMG distribution will not be sufficiently accurate. Therefore, we also describe a strategy that uses a scene-specific reference distribution to estimate decision thresholds for plume detection and associated confidence measures.

Keywords: clutter, single-component multivariate Gaussian, mixture distribution, generalized least squares, near infrared, central limit theorem

1. Introduction

Remote detection and identification of weak chemical plumes using passive hyperspectral sensors is a challenging problem. Here we consider an airborne visible and near infrared (NIR) sensor (AVIRIS [1,2]) having 224 spectral channels spanning a wavelength range of approximately 0.5 μm to 2.5 μm

and wavelength resolution of approximately $10^{-8}m$. One challenge in weak gaseous plume detection is that background clutter consisting of a mixture of components such as water, grass, and asphalt must be well characterized.

There are many ways to characterize background clutter; the appropriate characterization depends on analysis goals. We focus on ways that are most relevant for a particular method (generalized least squares, GLS) of plume detection. Plume-like pixels are those thought to have a gas plume influencing the signal; background pixels are those thought not to have a plume influencing the signal. One common set of simplifying assumptions leads to a GLS problem [3] in which chemical concentration estimates and associated decision thresholds are used to infer which if any pixels are plume-like. The GLS solution is commonly also referred to as an adaptive matched filter. This matched filter is “adaptive” because it uses the scene to estimate the covariance matrix (described below) of the background pixels.

When we look at real scenes, we tend to see them as a mixture of components, or “clutter,” rather than as, for example, a single-component multivariate Gaussian (SCMG). Whether formal mixture models are effective for analysis depends on the goals, and although modeling clutter as SCMG may seem simplistic, it has proven to be surprisingly effective [3-11]. A SCMG refers to a Gaussian distribution having a single mean vector and covariance matrix. Because real scenes typically have several physical components such as asphalt, vegetation, and water, etc., one would not expect a SCMG to be an effective model for clutter. Not surprisingly, other distributions have been proposed (such as mixtures of single-component multivariate Gaussians) for hyperspectral IR data, but either there is no corresponding plume detection strategy, or the corresponding plume detection methods have not yet consistently outperformed the GLS approach presented here [6,11,12-15]. Also, in evaluations of machine learning approaches [16], in which synthetic plumes are added to real scenes, again the GLS approach remains highly competitive.

It is not our purpose here to recommend a particular detection method, but because GLS has been shown to be effective, we will compare GLS-based results from real scenes to those from simulated SCMG scenes, in order to assess the adequacy of the SCMG approximation in this context. Therefore, our main goal is to show to what extent the popular GLS approach applied to real data to look for evidence of known chemical targets leads to chemical concentration estimates and to chemical probability estimates (arising from repeated application of the GLS approach) that are similar to corresponding estimates arising from simulated SCMG data.

In order to perform gas plume detection using GLS it is necessary to characterize the distribution of GLS values in real scenes that do not contain plumes so that thresholds corresponding to false positive rates can be estimated. Recent results suggest that scalar-valued GLS estimates, which arise from a linear transformation applied to scenes consisting of a mixture of background emissivities (but no or negligible plumes) are surprisingly close to Gaussian in distribution [3]. The evaluations in [3] considered how nearly Gaussian the scalar GLS estimates were from a particular scene, using random target directions (not from a chemical library, but randomly generated). There were somewhat limited checks for Gaussian behavior; for example, false positive rates based on GLS values from the real scene were not considered nor compared to rates based on a simulated reference SCMG distribution. Therefore, we extend the evaluation in [3] of the Gaussian approximation for GLS estimates, and also

evaluate related Bayesian model averaging (BMA, see section 3) values that rely on repeated application of the GLS approach to each of many candidate subsets of possible plume chemicals.

In some cases, approximations to decision thresholds or confidence estimates based on assuming the clutter has a SCMG will not be sufficiently accurate. Therefore, we also describe a strategy that uses a scene-specific reference distribution to estimate decision thresholds for plume detection and associated confidence measures. To simplify the discussion, we will use the informal expressions “more Gaussian” and “less Gaussian” to describe the extent to which a particular estimate based on a cluttered scene is behaving like an estimate from a simulated SCMG scene.

Section 2 gives additional background. Section 3 describes the radiance data model and simplifications leading to the GLS approach and/or to BMA approaches that rely on GLS. Section 4 describes tests to compare GLS-based results from real data to those from simulated SCMG data. Because in many cases the GLS-based result applied to SCMG data will be Gaussian, the tests are checks for Gaussian behavior, such as skewness, kurtosis, and quantiles. Example results are also presented in Section 4. Section 5 explores why the GLS transform improves the Gaussian approximation. Section 6 describes the notion of a scene-specific reference distribution in cases where approximation based on the SCMG assumption is not adequate. Section 7 is a summary.

2. Background

Our goal is to characterize clutter in the context of gas plume detection in cluttered scenes that contain a mixture of background emissivities, such as water, vegetation, asphalt, concrete, buildings, etc., for which various models and approaches have been proposed. A typical scene consists of approximately 128 to 256 spectral channels in each of approximately 500×500 spatial pixels, where the detected radiance at each pixel depends on the ground radiance, atmospheric transmission, instrument noise, and whether a chemical plume lies between the ground and the detector.

Figure 1 is the broadband image (the average radiance over all 224 spectral channels in this case) of an example scene (614 vertical by 512 horizontal pixels) which is cluttered because it contains mountains, buildings, and water, but contains (to our knowledge) no plumes.

We will use data from this same scene throughout and refer to it as scene A. We assume that scene A contains no plumes and will evaluate to what extent the collection of 224-dimensional radiance vectors at each pixel behaves in the context of GLS-based plume detection as a collection of 224-dimensional Gaussians having a single 224-dimensional mean and 224-by-224 dimensional covariance matrix (described in Section 3). In practice, it is not known whether a real scene contains any plumes, so there is typically an iterative procedure that first assumes there are no plumes, then looks for plumes using the covariance estimate from the “off-plume” pixels, then removes “plume-like” pixels and re-estimates the “off-plume” covariance matrix and mean vector. Because we focus entirely on characterizing the background in cases having weak and rare plumes, we will not consider such an iterative procedure. However, plume-detection performance is expected to degrade when on-plume pixels are used to estimate the background covariance matrix. Performance also degrades if the chemical target directions are similar to the clutter directions (as defined, for example, by the eigenvector directions in the spectral decomposition of the background covariance matrix [5, 6, 14]).

In the scenes we consider, small (up to a few or a few tens of pixels) and weak (both in terms of temperature difference between the ground and plume and in terms of chemical strength, see section 3) plumes from a library of possible chemicals might be present. These chemicals have effects on the measured at-sensor radiance that we refer to as “chemical signatures,” which are based on “known” spectra (the spectral library values are not known perfectly) that must be transformed (introducing an error source). Therefore, the “chemical signatures” are also not known perfectly. However, our focus is on characterizing the background clutter in the context of this situation, so we will ignore errors in the chemical signatures and deal strictly with real and simulated scenes that have (or are assumed to have) no plume.

We focus on the GLS values or BMA probability estimates and associated thresholds for deciding whether a plume is present in a pixel in scenes having no plumes. This is an effective way to characterize background clutter in the context of plume detection because decision thresholds impact performance as defined by false positive and negative rates. Therefore, we compare thresholds estimated from one real cluttered scene and from several simulated cluttered scenes to thresholds from corresponding uncluttered, SCMG scenes. For completeness, we also compute other aspects of the distribution of GLS and BMA values, such as kurtosis and skewness, so that the Gaussian approximation can be more fully assessed.

One caveat is that approaches other than GLS are necessary if we cannot assume we have an exhaustive list of all possible chemical targets in the scene (“chemical target” is described Section 3); in such cases, there might be more relevant ways to characterize the background clutter. In addition, other approaches become available if the ground pixels are viewed twice or if other regions of the spectrum are used. For example, mid-wave IR data might allow the opportunity to observe plume effects as a shadow effect on ground pixels and also as we do here for NIR data, as a signal effect when the plume lies in the line of sight between the sensor and the corresponding ground pixels.

There are several implications of concluding that an SCMG model is effective in our context. First, it is a simple model-based summary of complicated clutter. Second, decision thresholds could be based on SCMG data, and therefore could be computed analytically (analogous, for example, to claiming that a decision threshold of ± 2 standard deviations corresponds to a 5% false alarm probability by appealing to a Gaussian approximation). Third, it would suggest that it might be difficult to find robust methods that improve the GLS plume-detection performance (defined as the false negative rate for a given small fixed positive rate).

3. Data Models and Associated GLS-based Plume Detection Methods

3.1. AVIRIS signals

The photons detected by a visible/NIR hyperspectral detector associated with background pixel i at frequency ν_j (all terms depend on frequency ν_j) can be modeled as

$$S_i^b = (\varepsilon_i L_i^g + (1 - \varepsilon_i) L_i^{down}) \tau + L_i^{up} + N_i, \quad (1)$$

where ε_i is the emissivity, L_i^g is the Planck function at ground temperature, τ is the atmospheric transmission, L_i^{up} is upwelling radiance, $(1-\varepsilon_i)L_i^{down}$ is downwelling radiance reflected off the ground, and N_i includes unmodeled effects and instrument noise [3, 15]. Throughout we assume that context can be used to distinguish scalars, vectors, and matrices.

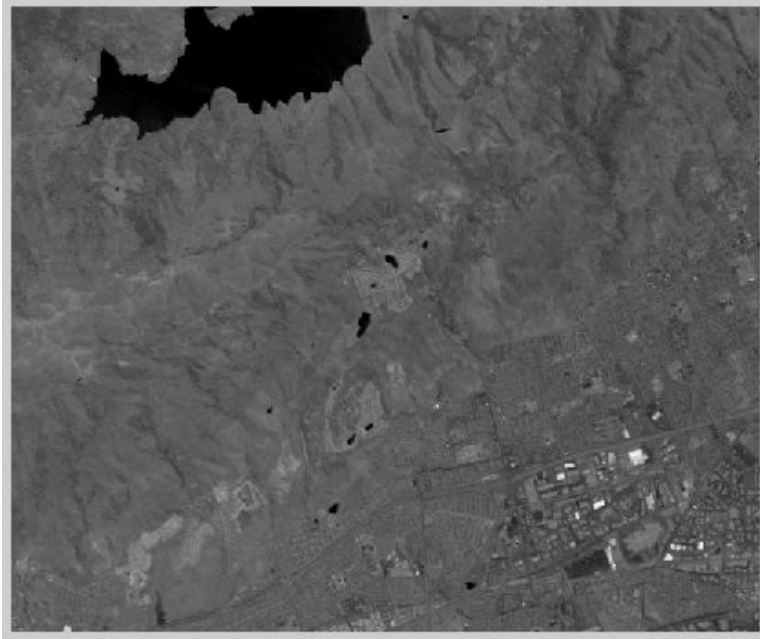


Figure 1. (top) Broadband image of a scene A, which is cluttered because it includes mountains, buildings, and water, but to our knowledge contains no plumes.

The signal from plume pixel i is S_i^b plus terms to model the plume effect,

$$S_i^p = S_i^b + \alpha_p [L_i^p - (\varepsilon_i L_i^g + (1-\varepsilon_i)L_i^{down})]\tau, \quad (2)$$

where $\alpha_p = 1 - \tau_p$ is the plume absorption and L_i^p is the Planck function at plume temperature. The ε_i terms (emissivities) depend on the properties of the background. Concrete, asphalt, buildings, grass, dirt, water, and other common background features each have their characteristic emissivity.

Using the approximation $\alpha_p = 1 - \tau_p \approx \sum_{k=1}^{n_c} \sigma_k C_k = \sigma\beta$, (assuming Beer's law [3] for chemical absorption, and a weak, optically thin plume, so that $1 - e^{-x} \cong x$), where C_k is the concentration of chemical k integrated over the pathlength through the plume, $\sigma = (\sigma_1 \sigma_2 \dots \sigma_{n_c})$ is the $p \times n_c$ matrix of chemical spectra with σ_k the chemical spectra (cross section) for chemical k , $\beta = (C_1, C_2, \dots, C_{n_c})$ is the unknown gas concentration vector, n_c is the number of chemicals in the plume, and defining A_i as $A_i = \{[L_i^p - (\varepsilon_i L_i^g + (1-\varepsilon_i)L_i^{down})]\tau\}$, we can rewrite Eq. (2) as

$$S_i^p = A_i \beta + S_i^b. \quad (3)$$

If any of the estimated components in the β parameter is large, this is evidence of a plume at pixel i . Typically the units for β are parts per million per meter per degree Kelvin temperature difference between the plume and ground, and the radiance units are Watts/(cm² sr). The temperature difference term arises from evaluating Planck's function at ground and plume temperature. Note that for ε_i values near 1, A_i will be nearly zero if there is no temperature difference between the ground and plume, making plume detection nearly impossible. Therefore, it is important to realize that a plume has a stronger effect and is therefore more easily detected if its temperature difference (positive or negative) with the ground is large.

To summarize, we can write equation (3) generically as

$$r = A_i \beta + z, \quad (4)$$

where r is the measured, calibrated radiation at pixel i , β is the amount of the chemical "signature" A_i at pixel i which we want to estimate, and z is the background effect ($z = S_i^b$) [16]. We will refer to $A_i = [L_i^p - (\varepsilon_i L_i^g + (1 - \varepsilon_i) L_i^{down})] \tau \sigma$, as the "chemical signature" or "target" in the task of distinguishing background from background plus target. In Eq. (4), we typically assume that r has been mean-centered to have zero sample mean in the scene.

A common alternative to Eq. (4) decomposes r as a linear superposition of background albedo effects and signal effects plus white noise, which leads to fitting each pixel to the best combination of background and target signatures. Again, the GLS approach has performed well compared to this and all other alternatives attempted to date [14], although performance comparisons depend on specific goals and contexts, including spectral library sizes.

In Eq. (4), A_i is the matrix of chemical signatures, defined as $A_i = [L_i^p - (\varepsilon_i L_i^g + (1 - \varepsilon_i) L_i^{down})] \tau \sigma$, and the factor $S = [L_i^p - (\varepsilon_i L_i^g + (1 - \varepsilon_i) L_i^{down})] \tau$, which multiplies σ in A_i , can be estimated using an in-scene method [17] and/or atmospheric models such as FASCODE [18] to estimate τ . In our examples, we assumed a standard United States atmosphere and FASCODE to estimate τ . We also assumed that the factor $S = [L_i^p - (\varepsilon_i L_i^g + (1 - \varepsilon_i) L_i^{down})] \tau$ is constant among pixels for the spectral channels of interest [19], so we assume and write $A_i = A$, which is probably more correct in the long wave IR than in the visible or NIR range. Future work will assess the impact of this assumption on estimates of β . Because the factor S is estimated (by estimating τ and typically by assuming $[L_i^p - (\varepsilon_i L_i^g + (1 - \varepsilon_i) L_i^{down})]$ is constant across spectral channels although in reality it depends on pixel temperature via Planck's function) and because the spectral cross sections σ in the chemical library are measured with error, we have an "errors in predictors" issue [20] in addition to the clutter issue. Strategies to deal with these predictor errors is the subject of ongoing work. However, a tentative finding is that more elaborate methods that account for the error in A in Eq. (3) are difficult to implement (too many pixels for routine fitting) and do not substantially outperform the GLS method that ignores errors in predictors [21-23]. In addition, because we want to compare clutter-related AVIRIS results to corresponding results from long-wave IR sensors, we used long-wave IR chemical signatures. Therefore, the signatures are realistic, but not

real, and the major challenge we focus on is background clutter, rather than challenges related to mispecifying the target signatures.

The approach considered here uses all of the image's pixels to characterize the background by computing $\hat{\Sigma}$ (recall that there is no iteration to remove plume-like pixels). For each frequency ν_j , the mean response \bar{r} (over all pixels) is subtracted from the responses so that the centered response r has mean 0. We then assume that $r \sim N(0, \hat{\Sigma})$ for the background pixels, where $\hat{\Sigma}$ is the p -by- p sample covariance matrix (with variances on the diagonal and covariances on the off-diagonal), and p is the number of frequency channels. The symbol $N(0, \hat{\Sigma})$ denotes a multivariate ($p = 224$ for this paper) Gaussian distribution having mean 0 and covariance matrix $\hat{\Sigma}$, which is the SCMG assumption. Combining all these assumptions, the simplistic SCMG model for plume pixels is

$$r = A\beta + z, \text{ with } z \sim N(0, \hat{\Sigma}) \quad (5)$$

and $\hat{\Sigma}$ and \bar{r} are estimated using all of the scene's pixels.

3.2. GLS

The GLS solution to Eq. (5) for a given pixel is

$$\hat{\beta} = (A^T \hat{\Sigma}^{-1} A)^{-1} A^T \hat{\Sigma}^{-1} r, \quad (6)$$

which is a vector consisting of k concentration estimates [24, 25]. We assume that real plumes contain at most three chemicals, so $k \leq 3$. However, in some contexts, we must evaluate all possible subsets of 1, 2, or 3 chemicals from a chemical library. Because the number of spectral channels is typically at least $p = 128$, this approach restricts attention to the overdetermined case ($p > k$). As an aside, we always have enough background pixels to estimate the p -by- p covariance matrix $\hat{\Sigma}$. Issues concerning the number of pixels required for $\hat{\Sigma}$ to be a high-quality estimate will not concern us here because we force all scenes being compared to have exactly the same sample covariance (Section 4). Our synthetic background scenes will be generated from various mixtures, each compared to synthetic SCMG data.

In Eq. (6), note three features of the multiplication by $\hat{\Sigma}^{-1}$ prior to the multiplication by A^T : (1) in contrast to the often-used principal components, the transformation to $\hat{\Sigma}^{-1} r$ does not result in uncorrelated coordinates, but instead the covariance matrix of $\hat{\Sigma}^{-1} r$ is $\hat{\Sigma}^{-1}$; (2) in the case that $\hat{\Sigma}^{-1}$ has nonzero off-diagonal entries, note that $\hat{\Sigma}^{-1} r$ includes an axis rotation, which again differs from the often-used principal component rotation, and (3) in the case that $\hat{\Sigma}^{-1}$ is diagonal (all off-diagonal entries are zero), note that $\hat{\Sigma}^{-1} r$ can be expressed as $(\frac{r_1}{\sigma_1^2}, \frac{r_2}{\sigma_2^2}, \dots, \frac{r_p}{\sigma_p^2})$, which very strongly emphasizes

the low variance directions, which contrasts to the more typical scaling which would divide each variable r_i by σ_i .

BMA involves repetitive application of the GLS for an exhaustive list of possible chemical subsets, as described below.

3.3. BMA

In order to decide which chemicals are present in a candidate plume, we apply the GLS to all subsets of 1, 2, or 3 chemicals [3], and then Bayesian Model Averaging (BMA) [26] or a stepwise least squares procedure is used to estimate the probability that each chemical from the library was present. These probability estimates are impacted by nonGaussian behavior, as we show below. Here we briefly describe BMA for subset selection.

For a given data set D and probability model for the data, it would be ideal if we could calculate the exact probability of each subset. By Bayes theorem, $P(M_1 | D) \propto P(D | M_1)P(M_1)$, where $P(M_1)$ is the prior probability for model (subset) M_1 . This calculation requires calculation of the expression $P(D | M_1) = \int P(D | M_1, \beta_1)\pi(\beta_1 | M_1)d\beta_1$, where β_1 is the coefficient vector for the chemicals in model M_1 . Such integrals are notoriously difficult in most real problems requiring either numerical integration, analytical approximation, or Markov Chain Monte Carlo methods [27, 28]. Even if the integral could be computed accurately, we would rarely know the exact subset probabilities because real data never follows any probability model exactly. Therefore, various approximations are in common usage, with the BIC (Bayesian Information Criterion) being perhaps the most common [26].

Following [26] and [29] we approximate $P(M_j | D)$ using the approximate result that $P(M_j) \propto e^{-BIC_j/2}$ where $BIC_j = p \ln(RSS_j / p) + (n_j + 1)\log(p)$, n_j is the number of predictor variables in subset (model) j , $RSS_j = \sum_{i=1}^p (\hat{\Sigma}^{-1/2}r_i - \hat{\Sigma}^{-1/2}A\hat{\beta}_j)^2$ is the residual sum of squares for subset j , p is the number of spectral channels per pixel, and $\hat{\beta}_j$ is the GLS-based estimate of β_j . The BIC expression is derived using the Laplace method for approximating the integral required to calculate $P(M_j | D)$, and assuming a flat prior (over the region where the integrand is nonnegligible) for the value of β_j .

The probability that chemical C is present is $P(C|D) = \sum_{j=1}^M I(C \in M_j)P(M_j | D)$, where $I(\cdot)$ equals 1 if

its argument is true. That is, to estimate $P(C|D)$, we simply sum the model probabilities for each subset that includes chemical C . Although these probabilities have varying accuracy, depending on the data (see section 4), BMA is one of the most effective strategies for chemical subset selection, particularly when prior information such as some chemical combinations being highly likely or unlikely is available [21].

4. Evaluation of the multivariate Gaussian assumption

Kurtosis and skewness are two of many measures used to gauge how close a distribution is to Gaussian [30]. Testing for univariate or multivariate normality is somewhat of an art because: (1) there are many possible tests; (2) any test applied to real data is practically guaranteed to reject the Gaussian hypothesis provided the sample size is large enough, and (3) some types of nonnormal behavior can be present, yet the Gaussian approximation is still adequate for some goals. Regarding (2), the probability of detecting nonGaussian behavior approaches one as the sample size increases, even if the departure from normality is very small. Therefore, in view of (3), it is often prudent, as we

do below, to compare conclusions made using the real data to corresponding conclusions made from simulated Gaussian data.

Here we will include the following comparisons of real data to corresponding Gaussian data: (a) kurtosis; (b) skewness; (c) the quantiles of each element of $\hat{\beta}$ (because these quantiles are used to select decision thresholds to decide whether the corresponding chemical is present), and (d) BMA-based estimates of the probability that a given chemical from a chemical library is present. We will do so in the cases of both scalar-valued and vector-valued $\hat{\beta}$ for each pixel.

To simplify notation, we will continue to use r as the response and A as the predictor matrix, and ignore measurement error in A . We will use a simulated Gaussian reference distribution (the SCMG) that exactly follows Eq. (5) for comparison.

The next subsection describes mixture distributions and gives example quantiles from mixtures of scalar-valued Gaussian distributions. The following two subsections define skewness and kurtosis, and give examples with scalar mixtures. Following subsections describe simulated scenes and give results of our comparisons to SCMG data for a few cases.

4.1. Kurtosis and Skewness

Kurtosis is defined as the ratio of the fourth moment to the square of the second moment, $\frac{\sum_{i=1}^n z_i^4}{(\sum_{i=1}^n z_i^2)^2}$,

for a mean-centered variable z . The expected sample kurtosis is 3 for a Gaussian distribution. Often, as we do throughout, the expected value of 3 is subtracted so that the expected value of the “kurtosis excess” (which we will still refer to as kurtosis) is 0 for a Gaussian distribution. The scalar ($k = 1$ case) GLS solution was evaluated in [3] and found to be much closer to Gaussian, as measured by kurtosis, than $A^T r$.

We will also evaluate skewness, a common measure of symmetry, defined as $\frac{\sum_{i=1}^n z_i^3}{(\sum_{i=1}^n z_i^2)^{3/2}}$, which also

has an expected value of 0 for Gaussian data.

Throughout, we will use $\hat{\beta} = (A^T \hat{\Sigma}^{-1} A)^{-1} A^T \hat{\Sigma}^{-1} r$ to denote the GLS solution and $\hat{\beta} = (A^T A)^{-1} A^T r$ to denote the classical least squares (CLS) solution [24]. The CLS solution is also called a standard matched filter or a “white-noise filter.” The CLS solution is almost guaranteed to perform worse (as defined by the false negative rate in plume detection for a given small false positive rate) than the GLS solution because the CLS solution does not use the variances and covariances to properly weight the solution vector. Here, we focus on how the GLS (and CLS) estimates on real data compare to corresponding estimates on simulated SCMG data. However, in Section 4, we mention a few detection probability comparisons that confirm that the GLS approach has higher plume detection probability than the CLS when they have the same false positive probability. Perhaps surprisingly, the GLS estimates do not always lead to “more Gaussian” behavior than the CLS estimates.

4.2. Mixtures of Scalar-valued Gaussian Variables

Because real data r is a mixture of many physical components, in the important special case of scalar-valued $\hat{\beta} = (A^T \hat{\Sigma}^{-1} A)^{-1} A^T \hat{\Sigma}^{-1} r$ at each pixel, we might expect $\hat{\beta}$ values to fall in clusters corresponding to the various physical components. Such clustering arises from mixtures of distributions [11]. As we show below, this clustering is strongly mitigated (but not eliminated) by the linear combinations of the components of r that are used in the calculation of $\hat{\beta}$. This subsection defines mixtures of scalar-valued Gaussian variables and then briefly evaluates to what extent Gaussian mixtures can have Gaussian-like quantiles.

A random variable X arising from mixing N_{comp} scalar-valued Gaussian random variables can be expressed as $X \sim \sum_{i=1}^{N_{\text{comp}}} \pi_i z_i$, where π_i is the probability that component i is chosen, and component i is Gaussian with mean μ_i and standard deviation σ_i , denoted as $z_i \sim N(\mu_i, \sigma_i)$. Mixtures of vector-valued Gaussian random variables are expressed similarly, and have been proposed to describe real IR hyperspectral imagery [3, 11]. Mixtures of other vector-valued random variables have also been proposed [7].

It is sufficient here to consider the case where the means μ_i differ among groups, but the standard deviations σ_i are the same in each group, denoted σ . The mean and standard deviation of the mixture are then $\mu_{\text{mix}} = \sum_{i=1}^{N_{\text{comp}}} \pi_i \mu_i$ and $\sigma_{\text{mix}}^2 = \sigma^2 + \sum_{i=1}^{N_{\text{comp}}} \pi_i \mu_i^2 - \mu_{\text{mix}}^2$. Our main interest in the plume-detection context is in the probabilities of exceeding specified thresholds. It can be shown by straightforward calculation that $P(|X - \mu_{\text{mix}}| > k\sigma_{\text{mix}}) = \sum_{i=1}^{N_{\text{comp}}} \pi_i (1 - \varphi(\frac{\mu_i - \mu_{\text{mix}}}{\sigma} + k \frac{\sigma_{\text{mix}}}{\sigma_i}))$.

Perhaps surprisingly, for many mixtures, these probabilities are smaller than those of the corresponding reference distribution, which is a single-component univariate Gaussian having the same standard deviation (σ_{mix}) as the mixture. Therefore, the commonly-observed tendency for GLS values in hyperspectral IR image analysis to have fatter tails (higher probabilities of mean-centered values exceeding $k\sigma$) is not necessarily expected. However, for many other mixtures, particularly those having very unequal π_i , the tails are fatter than the reference Gaussian. For example, suppose the random variable X arises a mixture consisting of 3 components, with rather unequal component fractions $\pi_1 = 0.083$, $\pi_2 = 0.83$, and $\pi_3 = 0.083$ and component means $\mu_1 = -3$, $\mu_2 = 0$, and $\mu_3 = 5$, and with $\sigma = 1$. Then it is easily shown that $P(|X - \mu_{\text{mix}}| > k\sigma_{\text{mix}}) = 0.079$, 0.022 , and 0.0006 for $k = 2$, 3 , and 4 , respectively, and the kurtosis equals 2.76 . The corresponding probabilities for the single-component reference Gaussian are 0.046 , 0.003 , and 0.00006 , which are significantly smaller, indicating that this particular mixture has fatter-than-Gaussian tails. On the other hand, if the random variable X arises a symmetric mixture consisting of 3 components having $\pi_1 = 0.25$, $\pi_2 = 0.5$, and $\pi_3 = 0.25$ and $\mu_1 = -3$, $\mu_2 = 0$, and $\mu_3 = 3$ with $\sigma = 1$, then $P(|X - \mu_{\text{mix}}| > k\sigma_{\text{mix}}) = 0.023$, 0.000014 , and 4.4×10^{-11} for $k = 2$, 3 , and 4 , respectively, indicating that this mixture has thinner-than-Gaussian tails. And, its kurtosis equals -0.67 .

Although kurtosis measures tail behavior fairly well, it is possible for a distribution to have less-than-Gaussian kurtosis, yet have higher probability of exceeding some of these thresholds, and vice

versa. For example, one example 10-component Gaussian having very different component fractions has kurtosis -0.333 but $P(|X - \mu_{\text{mix}}| > k\sigma_{\text{mix}}) = 0.0002$. This is significantly larger than the corresponding probability for a single-component Gaussian, which is 0.00006 .

4.3. Simulated scenes to compare to SCMG scenes

The simulated mixture scenes will be motivated by real data examples such as shown in Figure 2 from scene A. The top plot in Figure 2 is example spectra from selected pixels from Scene A. The bottom plot in Figure 2 is the same, except the spectra have been mean-centered to zero mean by subtracting the scene mean from each pixel's spectrum. Note that the response varies considerably.

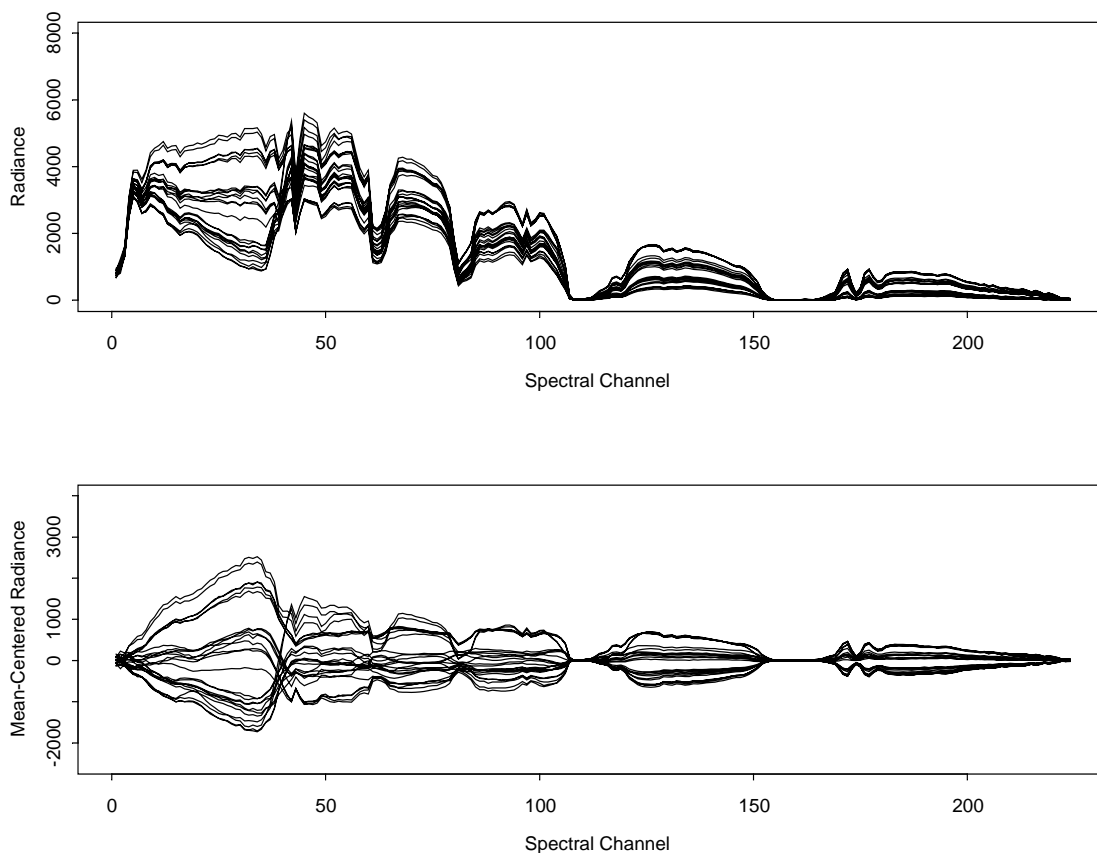


Figure 2. (top) Example radiance from selected pixels from Scene A.; (bottom) Example mean-centered radiance from selected pixels from Scene A.

Figure 3 is the coefficient of variation (COV) for each spectral channel for a random selection of 16,384 pixels from Scene A. The COV is the standard deviation (across pixels in this case) divided by the mean. Except for the lowest-response channels, most COVs are approximately 20% to 50%. In addition, simple cluster analyses of scene A and similar scenes suggest that the mixture components are represented with widely varying percentages. For example, there is relatively little water in Scene A, compared to dry land. These observations suggest two features to include in simulated mixture models for Scene A. First, the simulated data sets will each have the same COV as Scene A (this turns out not to be important for our data, see below). Second, we randomly draw the mixture fractions π_i

from a lognormal distribution, which results in quite unequal fractions, such as those in Scene A (this is important for our data). For example, there is a large fraction of mountainous vegetation in Scene A, a moderate fraction of water, and a small amount of asphalt.

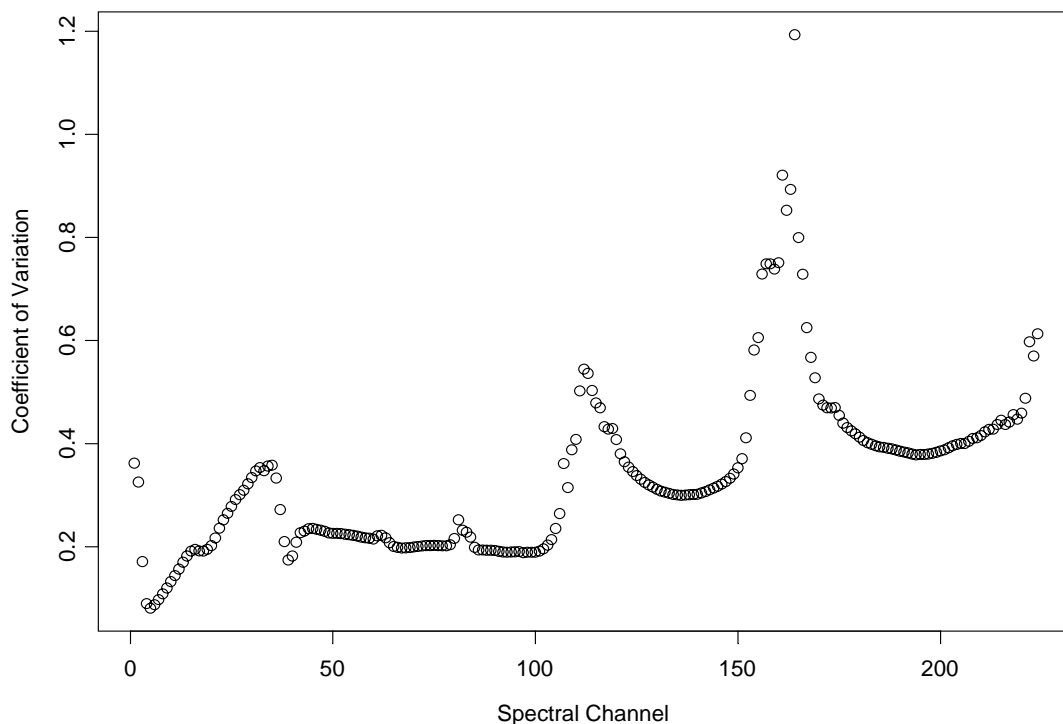


Figure 3. Coefficient of variation of mean-centered spectra from selected pixels from Scene A.

Figure 4 is a normal probability plot of CLS and GLS values for each of 2 example chemical directions (the example chemical directions were randomly chosen from the library of 296 chemical directions) from a subset consisting of 16,384 randomly-selected pixels from Scene A. A normal probability plot plots the sorted data versus the expected values of sorted data from a simulated reference Gaussian distribution. It is an effective qualitative tool for detecting nonnormality. Mixtures typically exhibit sharp bends or jumps in normal probability plots. In large samples from a Gaussian distribution (16,384 is large in this context, but small enough to carry out the many simulations presented here), the normal probability plot will be very nearly a straight line. We see that the GLS values are “more Gaussian” (more linear) than the CLS values, as noted by [3] using kurtosis evaluations. And, the GLS and CLS values are both much closer to Gaussian on the basis of any measure than are any of the individual 224 spectral channel values.

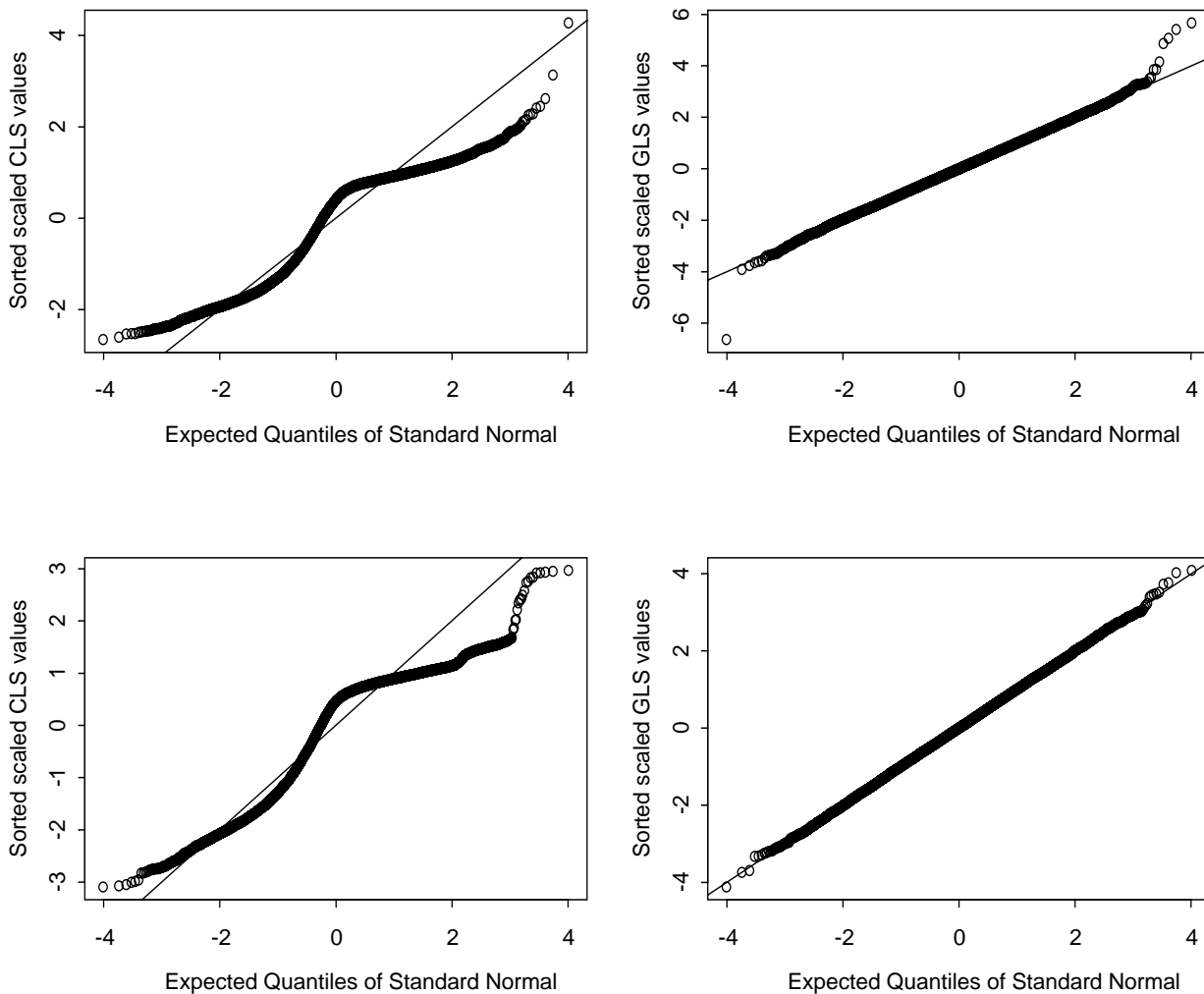


Figure 4. Normal probability plots of F and GLS values in two different chemical directions (the top row is one direction; the bottom row is another direction) from 16,384 randomly selected pixels from scene A.

We will describe CLS and GLS results for several simulated scenes and for subsets of pixels from scene A. The simulated scenes include a 7-component Gaussian mixture example with randomly generated mean vectors; another 7-component Gaussian mixture example, but with the component mean vectors mimicking those in Figure 2; a Gaussian mixture example with means given by randomly chosen background albedos from an albedo spectral library; a multivariate t with 5 degrees of freedom [12], and 3 sets of 16,384 randomly-selected pixels from scene A. All simulations were performed in Splus [31].

In all cases, we can arrange for the simulated reference SCMG to have either exactly or approximately the same sample $\hat{\Sigma}$ as the scene of interest. Because the ranked eigenvalues of $\hat{\Sigma}$ from real scenes typically decay to nearly zero after the largest 2 to 3 eigenvalues, we anticipated that simulated reference scenes should have exactly the same $\hat{\Sigma}$ (and eigenvalues) as the corresponding scene of interest (real or simulated). Therefore, results reported here have forced the simulated

reference Gaussian to have exactly the same (to within machine accuracy) $\hat{\Sigma}$ as the corresponding scene of interest. Also, in computing $\hat{\Sigma}$ and/or $\hat{\Sigma}^{-1}$ it is sometimes advisable to discard very small eigenvalues or use some other type of regularization. We do not report results here arising from any type of regularization. In our context of characterizing clutter by comparing results to a corresponding simulated reference SCMG, it is important to do the same calculations on both data sources, but less important to experiment with options to improve some type of performance.

A secondary consideration is whether the 224 channel variances should be equal. Recall from Figure 3 that the COV is not constant across spectral channels; however, we have observed that forcing a constant COV across channels gives similar results to choosing a COV similar to that in Figure 3. Therefore, for brevity, all results shown here using simulated mixtures assume the same COV across spectral channels as in a random selection of 16,384 pixels from real scene A, and also assume the same lognormal distribution of component fractions.

Gaussian mixture 1 was generated by randomly selecting 7 components of 224 means to represent 7 mixture components, each having a randomly-generated mean for each of the 224 spectral channels. The component fractions (π_i as used above) were randomly generated from a lognormal distribution and therefore varied considerably. For example, to generate the first 2 of 224 radiance values, we first randomly choose (with probability given by the component fractions π_i , for $i = 1, 2, \dots, 7$) a component mean (component 1 might represent water, component 2 might represent asphalt, etc.) Suppose the randomly-generated mean vector is (5.2, -3.3). We then add Gaussian noise to simulate within-component variation, resulting, for example in a simulated r value of (5.3, -3.2). This is repeated 16,384 times, each time randomly choosing which component to generate and then adding Gaussian noise to simulate within-component variation. The relative sizes of between-component and within-component variance could be chosen from real scenes if the goal were to select a model that is most characteristic of observed r values. Alternatively, as we did here, the relative sizes of between-component and within-component variance can simply be varied empirically. Recall that in all cases, we forced the simulated reference Gaussian to have exactly the same (to within machine accuracy) $\hat{\Sigma}$ as the corresponding scene of interest (simulated Gaussian mixture 1). Probably largely because of that constraint, there was not much variation in results among multiple realizations of this procedure.

Recall from Figure 2 (bottom plot) that the real data is better described by a mixture in which the 224 channel means are either all positive or negative. Therefore, Gaussian mixture 2 was generated in the same manner as Gaussian mixture 1, except the 7 component means were chosen by simulating one random direction (positive or negative representing above or below the mean) for all 224 channels and then choosing 224 random magnitudes. This resulted in simulated data that is more similar to that arising from simply mixing together varying fractions (using the log-normal distribution to produce quite different component fractions) of randomly-chosen r values from Scene A. Note that in this case we would not have a mean vector such as in the Gaussian Mixture 1 case having first 2 components of (5.2, -3.3), because we forced all components to have the same sign.

Gaussian mixture 3 was generated in the same manner as mixture 1, except the 7 component means were randomly chosen from a library of 40 common background albedos.

The multivariate- t has been described [12] as a possible model for IR hyperspectral data, primarily because of the heavier-than-Gaussian tail behavior. The degrees of freedom can be chosen on the basis of fits to the various aspects of the data [12]; we chose $df = 5$ ([12] also found that $df = 5$ gave a

relatively good fit) and followed [12] to generate data having the multivariate- t distribution with a covariance equal to the covariance from 16,384 pixels randomly chosen from scene A. Note that the multivariate- t is not a mixture distribution, but [12] also evaluated mixtures of the multivariate- t .

We also randomly-selected subsets of 16,384 pixels from real scene A are used for our real data examples.

4.4. Scalar-valued $\hat{\beta}$ for each pixel

The case that $\hat{\beta}$ is scalar-valued for each pixel corresponds either to searching for a plume containing one particular chemical in a scene, or to searching for a plume that contains only one chemical, but the chemical could be any chemical in a library containing, say, L chemicals. In the later case, it is necessary to consider the distribution of the maximum of L GLS values. Various types of non-Gaussian behavior could manifest themselves differently when such an overall maximum is evaluated. We consider these two sub-cases separately and use a library of $L = 296$ chemicals, which is one of several libraries we typically use in practice. Contact the first author for more detail about this chemical library.

4.4.1 Searching for a plume containing one particular chemical

We will compute the scalar GLS values for one chemical at a time (and report average results over all 296 chemicals) for simulated scenes, the real scene, and the corresponding simulated SCMG scenes.

For each of these cases, Table 1 gives the fraction of pixels (16,384 pixels were used for each case) for which the scaled CLS or GLS value (scaled to unit variance) exceeds 2, 3, 4, or 5, the kurtosis, and the correlation in the normal probability plot. The correlation in the normal probability for any distribution plot will be quite high because it is a plot of sorted values versus expected sorted values computed as if the data were Gaussian. Evaluating the correlation in the normal probability plot is very similar to the Shapiro-Wilks test for normality [30]. Here we rely on direct comparison to a value of 1.0 which is obtained in such large samples from a Gaussian distribution. Multiple repeats of each simulated case allowed us to determine that any correlations of 0.99 or less are statistically significantly less than 1.0, indicating a larger departure from the Gaussian distribution case (having a correlation of 1.0 in such large samples) than can be explained by chance alone.

As an important aside, evaluation of detection probabilities for injected signals is beyond our scope here because our focus is on the Gaussian approximation; however in all cases in Table 1, the GLS-based DP values were significantly higher than the CLS-based DP values.

From Table 1 we note that Gaussian Mixture 1 is indistinguishable from its corresponding simulated reference SCMG. This is somewhat surprising, but indicates the power of the central limit effect when each of the 224 mean values is generated randomly for each of the 7 mixture components.

The Gaussian Mixture 2 results are qualitatively more typical of those from real scenes: there is a significantly larger fraction of large GLS values (exceeding 3, 4, or 5) than in the corresponding reference scene. However, this particular mixture distribution appears to be “worse” than our real data in terms of the corresponding GLS’s distance from normality (see the low correlations in the normal

probability plot for example). Interestingly, the CLS values often have large kurtosis, but mostly because of the fraction of CLS values exceeding modest values such as 2 than because of CLS values exceeding 3, 4, or 5. Recall that kurtosis values should be near 0 for the Gaussian distribution. There is also a tendency for the GLS values to have larger-than-Gaussian kurtosis, mostly due to the fraction of values exceeding the larger thresholds such as 3, 4, and 5. Also, note that the GLS values have larger kurtosis than do the CLS values for Gaussian Mixture 2; in this respect, the CLS values are therefore more Gaussian than the GLS values in this case. Although not shown here, there was a tendency for the kurtosis of the GLS values to be more Gaussian for random directions than in actual chemical directions randomly chosen from the library of 296 chemical spectra.

The multivariate- t with 5 degrees of freedom is somewhat like the real data, as was found in [10] using different data than here. Other values for the degrees of freedom were examined; for example, results for 3 degrees of freedom were considerably different than results for real data. Notice however, that contrary to the general tendency, the GLS values are not closer to Gaussian than are the CLS values.

In the three randomly-chosen subsets of pixels from real scene A, note that again the GLS values are closer to Gaussian than are the CLS values. Also note that the fraction of GLS values exceeding 3, 4, or 5 are distinctly larger than the corresponding SCMG values. Therefore, while the GLS transform is generally more Gaussian than is the CLS transform, it is not necessarily adequate to rely on Gaussian approximations (see Section 6).

Finally, note that in all cases, the normal probability plots of the GLS values are more Gaussian than those of the CLS values, as indicated by their significantly higher correlations (last column in Table 1).

4.4.2 Searching for a plume containing one chemical from a library of L chemicals

In this subsection, we still consider a scalar value for each pixel, but the scalar is the value of the maximum GLS value from the selected library of 296 chemicals. Table 2 gives results based on choosing the maximum for the same cases as in Table 1. We have experimented with both random and actual chemical directions; results shown here are for actual chemical directions. There is a tendency for the GLS values in random directions to show somewhat lower kurtosis than the GLS values in chemical directions. We believe that random directions are more typical of so-called “narrow-line absorber” chemicals, but have not yet attempted a systematic evaluation by classes of target signatures.

Table 1. Cell entries in columns 1 to 4 are 100 times the fraction of (CLS, GLS) values (averaged over 296 chemicals, each scaled-to-unit-variance), denoted Z , that exceed 2, 3, 4, or 5, and the kurtosis, skewness, and correlation in the normal probability plot. These (CLS, GLS) entries are given twice per cell: once for the scene and once for a corresponding simulated Gaussian reference scene. Entries for the 100 times the fractions exceeding thresholds are within approximately ± 0.05 or less of results in hypothetical repeats of the same procedure. Entries for skewness and kurtosis are within approximately ± 0.01 of results in hypothetical repeats of the same procedure. Zero entries are less than 10^{-5} .

Case	P($Z>2$)	P($Z>3$)	P($Z>4$)	P($Z>5$)	kurtosis	skewness	Correlation in the normal probability plot
Gauss. Mix 1	2,2	0.08,0.1	0.01,3e-3	0,6e-5	-0.13,-04e-3	-0.07,-0.01	0.999,1
	2,2	0.1,0.1	3e-3,3e-3	4e-5,0	0.004,-2e-3	0.1e-3,-2e-3	1,1
Gauss. Mix 2	4,3	4,1	1,1	0.03,0.3	6.7,18.2	2.48,-0.12	0.78,0.86
	2,2	0.09,0.01	8e-5,3e-3	0,2e-5	-0.07,-3e-3	2e-3,3e-4	1,1
Gauss.Mix 3	3,2	0.2,0,3	0.02,0.4	0,0.01	-0.25,1.05	0.33,-0.16	0.984,0.991
	2,2	0.2,0.1	3e-3,3e-3	2e-5,4e-5	-0.017,0.01	-5e-35,7e-3	1,1
Multi- t with 5 df.	2,2	0.1,0.4	4e-3,0.1	2e-4,2e-4	0.02,2.25	-0.02,0.02	1, 0.990
	2,2	0.1,0.1	2e-3,2e-3	0,4e-5	0.02,2e-3	-0.01,-0.01	1,1
Scene A:	3,2	0.4,0.2	0.05,0.02	0.02,6e-3	1.81,0.47	0.41,4e-3	0.987,0.989
Pixelsubset 1	2,2	0.2,0.1	0.01,4e-3	0,6e-5	0.01,2e-3	0.02,-0.002	1,1
Scene A:	2,2	0.3,0.2	0.2,0.02	0.1,4e-3	16.1,0.43	1.26,7e-3	0.95,0.999
Pixelsubset 2	2,2	0.1,0.1	1e-3,e3-3	0,0	-0.04,2e-3	9e-4,-4e-3	1,1
Scene A:	0.5,2	0.1,0.2	0.03,0.02	0.01,5e-3	4.4,0.49	-0.14,7e-3	0.91,0.99
Pixelsubset 3	2,2	0.2,0.1	4e-3,3e-3	0,2e-5	-6e-3,-4e-3	0.03,2e-3	1,1

Again, Gaussian Mixture 1 is almost indistinguishable from its corresponding simulated reference SCMG. There is a hint of the typical pattern in the extreme tails for $P(Z>5)$: the mixture has slightly heavier than Gaussian tails. This pattern is more pronounced for the other four cases (three simulated and one real scene A with several subsets examined separately). And, that pattern is the key observation from Table 2. That is, in all cases except Gaussian Mixture 1, the extreme tail behavior, $P(Z>3)$, $P(Z>4)$, and $P(Z>5)$, is quite different than its corresponding simulated Gaussian reference distribution. As an aside, the maximum of 296 Gaussian values does not have a Gaussian distribution,

so it is not obvious what the kurtosis should be. Regardless, results for each case can simply be compared to corresponding results from the simulated reference SCMG distribution.

Table 2. Cell entries in columns 1 to 4 are 100 times the fraction of the maximum over 296 (CLS, GLS) values that exceed 2, 3, 4, or 5, and the kurtosis. These (CLS, GLS) entries are given twice per cell: once for the scene and once for a corresponding simulated Gaussian reference scene. Entries for the fractions exceeding thresholds are within approximately ± 0.05 or less of results in hypothetical repeats of the same procedure. Entries for kurtosis are within approximately ± 0.01 of results in hypothetical repeats of the same procedure. The boldface entries are the only two examples in which the kurtosis of the GLS values was less Gaussian than the kurtosis of the CLS values.

Case	P(Z>2)	P(Z>3)	P(Z>4)	P(Z>5)	kurtosis
Gauss. Mixture 1	78,75	11,14	0.3,0.5	0,0.01	-0.16,-4e-3
	74,76	14,15	0.6,0.6	0.01,0	0.01,0.03
Gauss. Mixture 2	12,24	5,22	4,16	3,6	-0.76,-0.57
	7,88	0.5,17	0.02,0.6	0,6e-3	-0.14,-1e-3
Gauss. Mixture 3	89,89	40,36	4,10	0,4	-0.46,1.33
	72,69	12,11	0.4,0.5	6e-3,0.01	-0.02,0.02
Multi- <i>t</i> with 5 df.	18,66	3,27	0.4,10	0.05,4	0.19,2.14
	15,81	1,4	0.04,0.5	0,0.01	0.18,0.01
Scene A: pixel subset 1	14,92	3,22	0.5,2	0,3,0.5	1.59,0.24
	14,94	1,21	0.05,0.7	0,6e-3	-6e-3,-1e-3
Scene A: pixel subset 2	11,91	1,22	0.5,2	0.4,0.5	21.6,0.54
	14,94	1,20	0.04,0.8	0,0	-0.05,2e-3
Scene A: pixel subset 3	4,89	1,23	0.4,2	0.3,0.6	4.46,0.37
	10,94	0.9,21	0.03,0.7	0,6e-3	-5e-3,7e-3

4.5. Vector-valued $\hat{\beta}$ for each pixel

In this section we consider the vector-valued GLS estimate. We consider the case where $\hat{\beta} = (A^T \hat{\Sigma}^{-1} A)^{-1} A^T \hat{\Sigma}^{-1} r$ is 2-dimensional. Results for the 3-dimensional case are very similar, and 3 is the largest number of dimensions considered in any of our fitted models. To summarize the distribution of each entry in the $\hat{\beta}$ vector, in Table 3 we report the fraction (times 100) of scaled-to-unit variance values exceeding thresholds of $T = 2, 3,$ and $4,$ and the kurtosis and skewness for three cases, depending on whether the directions are randomly selected from the 296-chemical library, selected from distinct chemical groups in the 296-chemical library (on the basis of simple clustering of the 296 spectra), or purely random. Table 3 is based on 45 pairs of chemical directions for each of 10 random selections of 16,384 pixels from scene A.

It appears from Table 3 that the target direction (random or from the chemical library) does impact the quantiles and the kurtosis. A graphical summary of Table 3 is given by the hierarchical clustering [31] result in Figure 5. Figure 5 is a typical hierarchical clustering that displays the Euclidean distance between each pair of vector-valued results. Nearby results cluster into groups. It is clear that the Gaussian (G) results are more similar to each other than to the CLS or GLS results; it is also clear that the GLS results are closer than the CLS results to the Gaussian results. Distances in the left plot are based on the probabilities of scaled values exceeding 2, 2.5, 3, 3.5, and 4. Distances in the right plot are based on these same probabilities plus the skewness and kurtosis. Clustering results are typically very sensitive to whether the variables being cluster are scaled to have the same variance. We examined hierarchical clustering results for both scaled and unscaled variables, and in all cases, the same pattern was clear. That is, Gaussian values are more similar to each other than to the CLS or GLS values, and GLS values are closer than the CLS values to the Gaussian values. In addition, we estimated the number of modes in the distribution (using smooth density estimates rather than histograms to identify modes) and the GLS values were very much more likely than the CLS values to have one mode, although both CLS and GLS showed multimodal behavior compared to the unimodal Gaussian distribution. Finally, note that the random directions subcase (subcase 3) tends to be different from the chemical directions subcases (subcases 1 and 2) for the GLS values and for the CLS values, and that for the GLS, the random directions subcase is closer to Gaussian.

4.6. BMA-based estimates of the probability that a given chemical from the library is present

Although exact calculations can be made for the probability that a given chemical from the library is present, these calculations rely strongly on all the assumptions made, including all the relevant probability distributions. Therefore, it is common to use approximate calculations that are typically in practice as effective as the “exact” calculations, as described in the BMA subsection. By “effective,” it is meant that estimated chemical probabilities are well calibrated. For example, if one records the estimated probability of chemical C in repeated experiments, then among all those instances in which $P(C)$ lies in, say, the interval (0.65,0.74), the fraction of instances in which chemical C is present is approximately 0.70.

Table 3. Results summary (averages over the results for 2 chemicals for multiple cases) for the distribution of each entry in the $\hat{\beta}$ vector. Cell entries are CLS, GLS, and G values, where G refers to GLS results from the reference Gaussian. The Chemical* directions case is chemicals chosen from distinct chemical groups rather than completely at random from the 296-chemical library.

Direction	$P(Z>2) \times 100$	$P(Z>3) \times 100$	$P(Z>4) \times 100$	kurtosis	skewness
Chemical*	1.9,2.3,2.2	0.3,0.2,0.1	0.09,0.03,2e-3	1.82,0.77,-5e-3	0.46,4e-3,5e-3
Chemical	2.2,2.3,2.3	0.3,0.2,0.2	0.1,0.03,0.03	14.7,0.47,0.47	1.0,0.02,0.02
Random	2.8,2.3,2.3	0.5,0.2,0.1	0.1,5e-3,3e-3	17.3,4e-3,-1e-3	1.46,0.02,2e-3

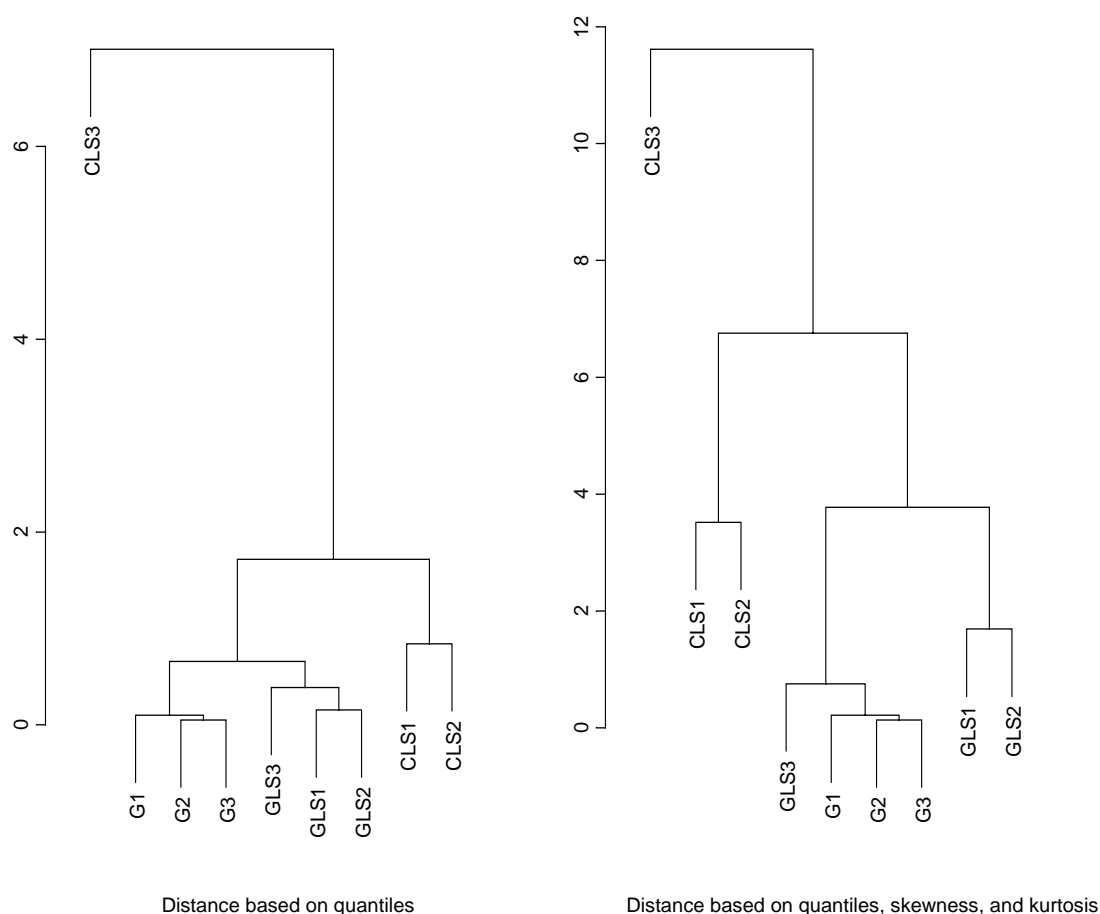


Figure 5. Hierarchical clustering result showing that the GLS values are closer to Gaussian (G) than are the CLS values. The subcases are denoted 1, 2, and 3. Subcase 1 is in chemical directions chosen from the library of 296 chemicals to be mutually distinct. Subcase 2 is randomly chosen chemical directions from the library of 296 chemicals. Subcase 3 is in random directions. The plot depicts the

average result of 45 pairs of chemical directions for each of 10 randomly chosen sets of 16,384 pixels per set.

It has been shown [25] that BMA applied to simulated SCMG data is well calibrated, although the quality of the agreement between estimated and observed chemical probabilities depends on the dimension and condition (how close to singular) of the X matrix. If instead we apply BMA to a real scene such as scene A, then we expect some performance differences due to the departure from the SCMG modeling assumption. For example, we randomly selected 16,384 pixels and applied BMA to each pixel. After using the 16,384 pixels to estimate Σ , random amounts of 0, 1, 2, or 3 chemical were injected on a per-pixel basis following Equation (3). Because [3] considered only random chemical directions (rather than using chemical cross sections from a library), we again consider both random directions and actual chemical directions randomly chosen from the same library of 296 chemicals. We repeated this experiment 10 times, each time choosing 16,384 random pixels. Results are summarized in Table 4.

Table 4. The average absolute difference (and the R^2 describing the fit) between the estimated and observed probability of a chemical being present, for each of four chemicals, denoted as chemical 1, 2, 3, and 4. The first entry in each cell is for the case of four randomly-selected chemicals from the chemical library. The second entry is for the case of four random directions simulated from a Gaussian distribution. There are two simulated reference distributions. The first has the same mean and Σ as the real data, but is a SCMG; the second has a diagonal Σ with equal variances on the diagonal. Simulated data entries are within approximately 0.01 of a repeat of the 10 sets of 16,384 pixels. Real data entries would not vary if the 10 sets of 16,384 were analyzed again.

Data Source	Agreement Measure	Chemical			
		1	2	3	4
Real Data: 10 random subsets of 16,384 pixels from scene A	Avg Abs Diff	0.50, 0.07	0.20, 0.08	0.19, 0.08	0.43, 0.07
	R^2	0.09, 0.80	0.60, 0.78	0.46, 0.78	0.54, 0.79
Reference Gaussian Data having same Σ as Scene A	Avg Abs Diff	0.46, 0.07	0.20, 0.07	0.20, 0.07	0.43, 0.07
	R^2	0.06, 0.78	0.57, 0.78	0.45, 0.78	0.57, 0.77
Reference Gaussian Data having Diagonal Σ	Avg Abs Diff	0.11, 0.04	0.04, 0.04	0.04, 0.04	0.06, 0.04
	R^2	0.91, 0.93	0.96, 0.94	0.92, 0.94	0.79, 0.94

Because all entries are repeatable to within approximated 0.01, Table 4 demonstrates a slight tendency for lower agreement (higher average absolute difference) between the estimated and

observed chemical probabilities in the real compared to in the corresponding simulated one-component Gaussian data. Similarly, if the agreement between estimated and observed chemical probabilities is defined on the basis of the quality of a fitted line using R^2 (which measures the percent of variance in the response that is explained by a linear fit to the predictor) as in Table 4 (as in [25]), then again, there is a slight tendency for worse agreement in the real data than in the corresponding simulated data.

To summarize Table 4, the three paired-average (over the four chemicals) absolute differences are (0.33, 0.08), (0.32, 0.07), and (0.06, 0.04) for the real scene, for the simulated SCMG having the same Σ as the real scene, and for the simulated SCMG with Σ being diagonal with equal variances (the off-diagonal values are 0, in strong contrast to Σ s from real scenes, which have strongly correlated spectral channels, as can be anticipated from Figure 2), respectively. The first entry in each pair is for four randomly-selected target directions from the chemical library. The second entry is for four random target directions. The corresponding pairs for the R^2 measure are (0.42, 0.79), (0.41, 0.78), and (0.90, 0.94). We empirically verified that these average results are repeatable to within approximately ± 0.01 or less. Thus, when averaged over the four chemicals, there is a slight tendency toward lower agreement in the real data both for the average absolute difference and for the R^2 measure.

This is consistent with earlier findings using a different sensor (SEBASS, which has 128 spectral channels) [3,32] that led to a scene-specific reference distribution approach; however, in this AVIRIS case with 224 channels, the difference between the real data and the Gaussian data having the same Σ as the real scene is considerably smaller. Notice also that Gaussian data having a diagonal Σ with equal variances exhibits much better agreement between observed and estimated probabilities. This is also consistent with [25] which showed that the BIC-based BMA approximation tends to be better when Σ is more nearly diagonal.

5. Exploring why the GLS values are closer than CLS values to Gaussian

We have demonstrated that GLS values are generally but not always closer to Gaussian than are CLS values using real and simulated data. This has also been demonstrated for a long-wave IR sensor having 128 channels [32].

Both the CLS and GLS transforms involve linear combinations of 224 variables, each of which has a mixture distribution in real scenes. The central limit theorem (CLT) suggests that, provided no terms dominate, a linear combination will be more Gaussian than any of the individual variables. There are many versions of the CLT [33], some of which allow various forms of dependence among the variables. And, there are various results for rates of convergence to the Gaussian distribution as the number of summands increases. All of our examples combine 224 correlated variables, and it is clear that both the CLS and GLS transforms, which both combine the 224 variables, are more Gaussian than any particular variable (spectral channel).

The term $\hat{\Sigma}^{-1}$ that appears in the GLS transform but not in the CLS transform typically has approximately 50% of its 224 entries greater than zero in a given row, with some terms being considerably larger than the others. These large terms have the potential to dominate the sum, mitigating the central limit effect. Therefore, we do not anticipate providing a theoretical result but will rely on the following additional evidence.

Figure 6 (top left) is a plot of a 2-component Gaussian (a mixture) characterized by 2 small circular point clouds. Although it is simulated, it is similar to scatter plots of some pairs of spectral channels from scene A. The two long thin horizontal elliptical point clouds are the transformed coordinates corresponding to multiplication by $\hat{\Sigma}^{-1}$. The top center (right) plot is a histogram of the first (second) transformed coordinate values. Notice that the low variance directions in the top left plot have become the high variance directions in new coordinates (the middle plot has a much wider range than the right plot). Also notice that the two components still appear, in the right plot, but that their separation is less than in the left plot. The three bottom plots are the same, except there is no multiplication by $\hat{\Sigma}^{-1}$. In the bottom left plot, the 2-component Gaussian is identical to that in the corresponding top plot, but because the scale is smaller, both components are long and thin bivariate Gaussians. The bottom center plot shows the first coordinate (without a transform) and the bottom right plot shows the second coordinate (without a transform). The bottom right plot is more spread out than the corresponding top right plot, by a factor of approximately 2. Multiplication by $\hat{\Sigma}^{-1}$ tends to reduce the spread among component distributions; this makes the central limit effect have less work to do to approach Gaussian behavior.

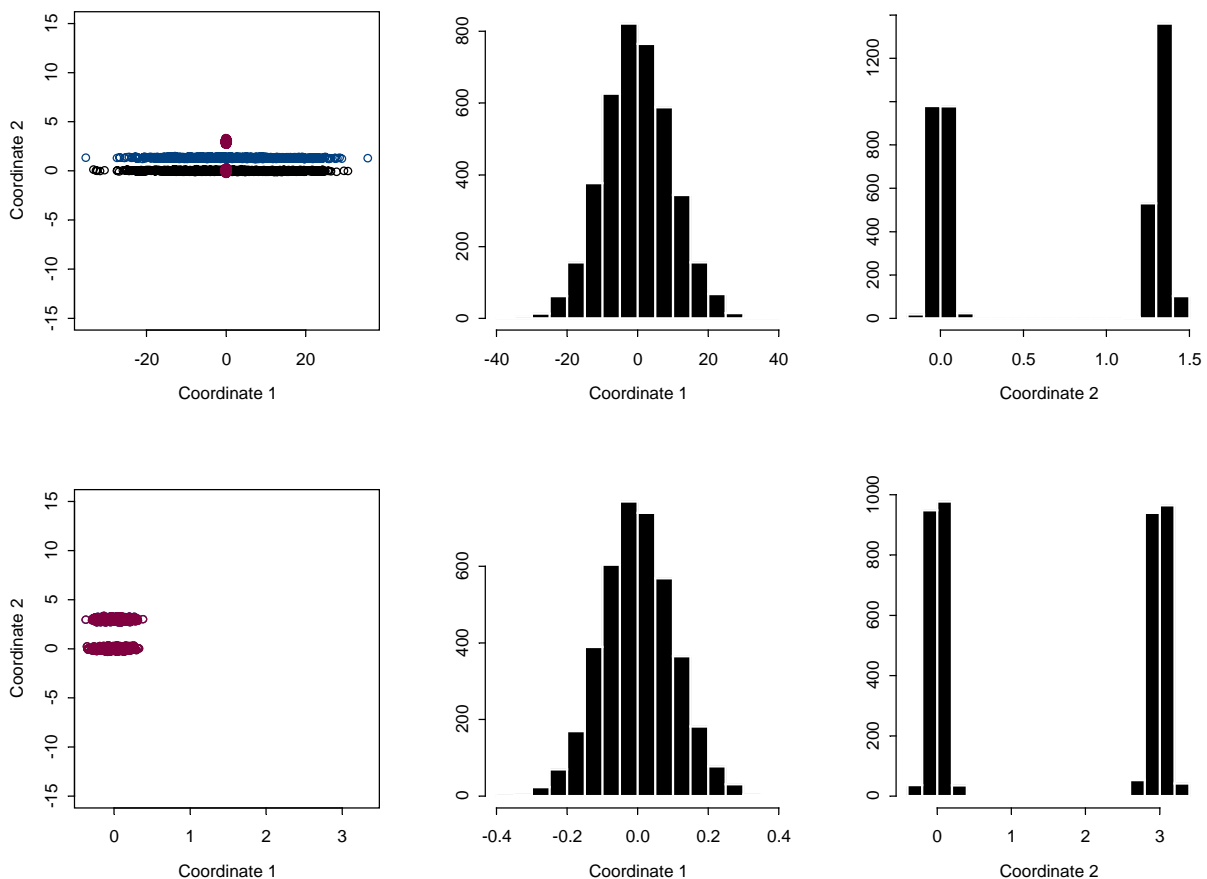


Figure 6. (top left) A 2-component Gaussian (a mixture) characterized by two small circular point clouds; (top center) the first component of the GLS transform; (top right) the second component of the GLS transform. The bottom left plot is the same as the top left plot except the horizontal axis is scaled differently. The bottom center and right plots are the same as the corresponding plots in the top row, except are for the CLS transform.

Figure 7 (top left) shows the 224-channel response from each of 4 pixels selected from scene A because they were well-separated, staying in order from lowest to highest across most channels. The 4 sets of 224 corresponding values that result from multiplication by $\hat{\Sigma}^{-1}$ are shown in the top right plot. For each of the 224 channels, multiplication by $\hat{\Sigma}^{-1}$ has resulted in bringing the mixture components much closer, to the extent that the ordering of the 4 component values varies much more across channels. This behavior is present even if we use only 2 channels, and has occurred for any number of channels that we have experimented with to date (2, 3, 4, 16, 32, 64, 128, and 224).

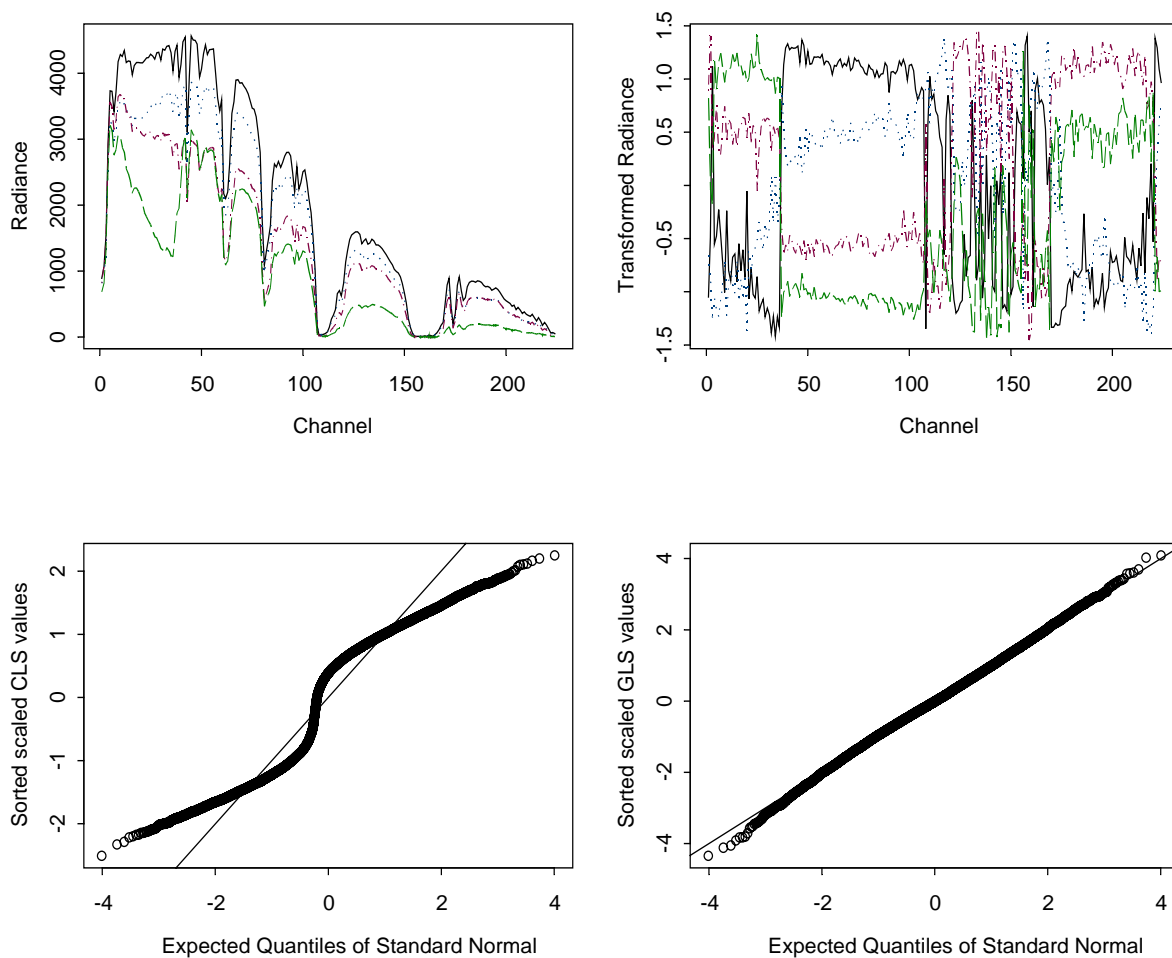


Figure 7. (top left) The well-separated 224-channel response from each of 4 pixels; (top right) The response in the top left plot, multiplied by $\hat{\Sigma}^{-1}$ and then scaled to unit variance, corresponding to the GLS operation; (bottom left) normal probability plot of corresponding CLS values in a random chemical direction using a random mixture of the 4 pixels in the top left plot; (bottom right) same as bottom left, except for GLS values.

The bottom plots are corresponding normal probability plots that should be very nearly linear if the CLS (bottom left) or the GLS (bottom right) values are Gaussian. The CLS and the GLS values are computed by first generating a synthetic scene consisting of 16,438 observations from a 4-component mixture, using the 224-channel response for each of the 4 pixels shown in the top left plot as the

component means. The lognormal-generated mixture fractions were 0.07, 0.25, 0.60, and 0.08, respectively for the four components. Random noise having 10% COV was added to each observation. Then, the CLS and GLS transforms were computed in a randomly-chosen chemical direction from the 296-chemical library. We see that the CLS values are not close to Gaussian (the correlation in the normal probability plot is 0.910), but the GLS values are very close to Gaussian (the correlation in the normal probability plot is 0.998) except for the extreme tails. The rapid rise in the normal probability plot of the CLS values near 0 indicates a mixture of two components; although the data was generated from a four-component mixture plus noise, only two components remain recognizable after the CLS transform.

6. Scene-specific reference distribution

As seen here and in [3], the GLS values in scenes consisting of mixtures is often well approximated by a SCMG. It is well known that GLS provides the best linear unbiased estimator of β (the BLUE, which is the minimum variance unbiased estimator), so regardless of whether the data is approximately Gaussian, GLS should be competitive [24], and has been shown to be competitive [14,16]. Therefore, its performance is likely to be competitive, although the accuracy of confidence statements that rely on the Gaussian approximation is unknown. In addition, the performance measure of interest here is in the context of predicting which chemicals are in a given plume, and the BLUE property is not customized for this goal, so it is an open question whether other estimators can perform better at chemical selection.

Regardless of whether other estimators will outperform the GLS, it is important to assess the confidence in each prediction. If the Gaussian approximation were sufficiently accurate, then in the case of using BMA to predict which chemicals are present in a given plume, the estimated chemical probabilities could be used directly in the natural fashion. We then select a tunable threshold, such as $T = 0.99$ and predict, for example, that chemical C is present if $P(C)$ exceeds T . Conditional on observed data D , the false positive probability for detecting chemical C is estimated to be $1 - T$. The actual false positive probability is likely to differ from $1 - T$ because the real data is nonGaussian. Even if the data were Gaussian, because of using the approximate result $P(M_j) \propto e^{-BIC_j/2}$ and because of choosing the maximum or top few chemical probabilities from the entire chemical library, the actual false positive probability will differ from $1 - T$ [25]. Therefore, we have developed a scene-specific reference distribution for selecting thresholds, as described next.

To characterize the behavior of estimated chemical probabilities, [4] reports results for the BMA strategy applied to real scenes in which there is no true plume. Custom software allows the user to: (1) select some number of connected pixels (typically 20 to 100); (2) apply the BMA algorithm to fit the null model and all single-chemical models to each of thousands of randomly chosen connected-pixel regions, and (3) record the highest probability of any chemical for each of the regions. This generates a reference distribution for comparison to BMA probabilities for a pixel-region that is thought to be a plume containing chemicals of interest. This is a scene-specific reference distribution used to select decision thresholds (such as if $P(C) > 0.9$ then predict that chemical C is present in the plume) that attempts to account for real-data features.

An example of the reference distribution using BMA for a real scene, using a library of $L = 10$ chemicals (and, to avoid other issues, using a pixel region of only one “connected” pixel) was given in [21]. The fraction of pixels that exceed 0.9 for example, is approximately 3%, which is a higher fraction than we observe in corresponding simulated data from the simulated reference SCMG distribution. Although there is the potential for random variation due to how the pixels are partitioned, we have observed essentially the same result for all random partitions analyzed. However, if we select a different set of 10 chemicals, we usually observe nonnegligible variation between the BMA probabilities for the first set of 10 and the second set of 10 chemicals. For example, a second set of 10 chemicals led to 3.5% of BMA probabilities exceeding 0.9 (versus 3% with the first set of 10). Using library sizes of $L = 4$ and $L = 35$, we find approximately 2% to 5%, or 1% to 3%, respectively, of pixel regions having BMA-based probabilities exceeding 0.9. We expected smaller probabilities in the “no chemicals present” case (the candidate plume region is in fact not a plume) as L increases, but we have not observed a clear trend as L increases, although in this example there was a trend as L increased from 4 to 10 to 35.

7. Summary

We have demonstrated that the GLS transform using either random or chemical directions applied to real and simulated data leads to surprisingly close-to-Gaussian values, in agreement with [3] which used random directions in real and synthetic scenes to generate GLS values. Reference [3] noted that fatter-than-Gaussian tails in GLS values have been reported by multiple authors analyzing IR images, and conjectured that the reasons might be that target directions are not random, perhaps related in a systematic way to the ground emissivities, or that image artifacts such as problematic spectral bands lead to fat tails. However, specific quantiles were not investigated, and the plotted kurtosis values, although close to those for Gaussian, tended to be slightly larger than those for Gaussian. Therefore, our findings agree qualitatively with [3] (although we demonstrated that in some cases the CLS transform was closer to Gaussian than the GLS transform), and indicate a tendency for the GLS transform to be more Gaussian in random than in chemical directions. We also showed that GLS-based BMA chemical probability estimates in one real and several synthetic scenes were fairly close to those in corresponding simulated SCMG data.

If GLS-related and/or BMA-related results for simulated SCMG data are not sufficiently accurate for a scene of interest, we described a computationally demanding scene-specific reference distribution approach. Using the scene-specific reference distribution avoids the SCMG-based approximation which we have shown has good, but varying accuracy.

Performance (false negative rate in plume detection for a given false positive rate) comparisons were beyond our scope; we note however, that GLS-based approaches have remained competitive among the several other options reported [3-11] in the context we considered (ignoring errors in predictors, which are the chemical signatures here, and assuming we have an exhaustive list of all possible chemical targets). In addition, the approximate Gaussian behavior of the GLS values suggests a computational advantage in some cases involving Gaussian-based analytical approximations rather than scene-specific reference distribution approximations to decision thresholds for desired false positive rates. Also, we note that an elliptically contoured model (such as the multivariate- t

distribution) does not appear to explain our real scenes as well as a mixture model. The fact that real scenes are a mixture of backgrounds is motivating some efforts to collect multiple scans of the same scene so that pre and post scans can be compared on a per-pixel basis. This approach requires some level of image registration, but offers the potential to improve performance.

Acknowledgements

The authors gratefully acknowledge support from the U.S. Department of Energy and from the three reviewers whose comments led to substantial improvements.

References and Notes

1. AVIRIS Free Standard Data Products, Jet Propulsion Laboratory (JPL), National Aeronautics and Space Administration (NASA). <http://aviris.jpl.nasa.gov/html/aviris.freedata.html>.
2. Vane, G.; Green, R.; Chrien, T.; Enmark, H.; Hansen, E.; Porter, W. The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). *Remote Sensing of the Environment* **1993**, *44*, 127-143.
3. Theiler, J.; Foy, B.; Fraser, A. Characterizing non-Gaussian Clutter and Detecting Weak Gaseous Plumes in Hyperspectral Imagery. Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XI, *Proc. SPIE*. **2005**, *5806*, 182-193.
4. McVey, B.; Burr, T.; Fry, H. Distribution of Chemical False Positives for Hyperspectral Image Data. Los Alamos National Laboratory Restricted Release Technical Report LA-CP-02-521, 2003.
5. Bernhardt, M.; Heather, J.; Smith, M. New Models for Hyperspectral Anomaly Detection and Un-Mixing. Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XII, edited by Sylvia S. Shen, Paul E. Lewis, *Proc. SPIE*. **2005**, *5806*, 720-730.
6. Villeneuve, P.; Fry, H.; Theiler, J.; Smith, B.; Stocker, A. Improved Matched-Filter Detection Techniques. *Proc. SPIE*. **1999**, *3753*, 278- 285.
7. Manolakis, D.; Marden, D.; Kerekes, J.; Shaw, G. On the Statistics of Hyperspectral Imaging Data. *Proc. SPIE*. **2001**, *4381*, 308-316.
8. Harsanyi, J.; Chang, C. Hyperspectral Image Classification and Dimensionality Reduction: an Orthogonal Subspace Projection Approach. *IEEE Trans. Geoscience and Remote Sensing* **1994**, *32*, 779-785.
9. Marden, D.; Manolakis, D. Modeling Hyperspectral Imaging Data. *Proc. SPIE*. **2003**, *5093*, 253-262.
10. Marden, D.; Manolakis, D. Using Elliptically Contoured Distributions to Model Hyperspectral Imaging Data and Generate Statistically Similar Synthetic Data. *Proc. SPIE*. **2004**, *5425*, 558-572.
11. Funk, C.; Theiler, J.; Roberts, D.; Borel, C. Clustering to Improve Matched Filter Detection of Weak Gas Plumes in Hyperspectral Thermal Imagery. *IEEE Trans. Geoscience and Remote Sensing* **2001**, *39*, 1410-1420.
12. Manolakis, D.; Siracusa, C.; Marden, D.; Shaw, G. Hyperspectral Adaptive Matched Filter Detectors: Practical Performance Comparison, *Proc. SPIE*. **2001**, *4381*, 18-33.
13. Manolakis, D.; Shaw, G. Detection Algorithms for Hyperspectral Imaging Applications. *IEEE Signal Processing Magazine* **2002**, *19*(1), 29-43.

14. Bajorski, P. Analytical Comparison of the Matched Filter and Orthogonal Subspace Projection Detectors in Structured Models for Hyperspectral Images. Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XII, edited by Sylvia S. Shen, Paul E. Lewis, *Proc. SPIE*. **2006**, 6233, 1-12.
15. Gallagher, N.; Sheen, D.; Shaver, J.; Wise, B.; Shultz, J. Estimation of Trace Vapor Concentration-Pathlength in Plumes for Remote Sensing Applications From Hyperspectral Images. *Proc. SPIE*. **2003**, 5093, 184-194.
16. Foy, B.; Theiler, J.; Fraser, A. Unreasonable Effectiveness of the Adaptive Matched Filter, to appear in: *Proc. MSS (Military Sensing Symposia) Passive Sensors Conference*, 2006, private communication.
17. Young, S.; Johnson, B.; Hackwell J. An In-scene Method for Atmospheric Compensation of Thermal Hyperspectral Data. *J. Geophys. Res. Atm.* **2002**, 107, 4774-4774.
18. FASCODE for atmospheric transmission calculations, see <http://www.vs.afrl.af.mil/ProductLines/IR-Clutter/fascode.aspx>
19. Foy, B.; Petrin, R.; Quick, R.; Shimada, T.; Tjee, J. Comparisons Between Hyperspectral Passive and Multispectral Active Sensor Measurements. *Proc. SPIE*. **2002**, 4722, 98-109.
20. Carroll, R.; Ruppert, D.; Stefanski, L. Measurement Error in Nonlinear Models, Chapman and Hall, London, 1996; pp 1-78.
21. Burr, T.; Fry, H.; McVey, B.; Sander, E. Chemical Identification using Bayesian Model Selection, **2002 Proc. of the American Statistical Association, Section on Physical and Engineering Sciences** [CD-ROM], Ann Arbor, Michigan: American Statistical Association, LA-UR 02-7281.
22. Heasler, P.; Hylden, J. Technical Letter Report for Non-Linear methods Task: Fits of Nonlinear Bayesian Regression (NBLR) to Tanasi and Polecat Images, Pacific Northwest National Laboratory Official Use Only Report, 2005.
23. McVey, B. Official use only internal presentation, cited with permission, 2004.
24. Montgomery, D.; Peck, E. Introduction to Linear Regression Analysis, 2nd edition, Wiley: New York, 1992.
25. Burr, T.; Fry, H.; McVey, B.; Sander, E.; Cavanaugh, J.; Neath, A. Performance of Variable Selection Methods in Regression using Variations of the Bayesian Information Criterion, LAUR-05-6324, submitted for publication, 2005
26. Raftery, A; Madigan, D.; Hoeting, J. Bayesian Model Averaging for Linear Regression Models. *Journal of the American Statistical Association* **1997**, 92, 179-191.
27. DiCiccio, T.; Kass, R.; Raftery, A.; Wasserman, L. Computing Bayes Factors by Combining Simulation and Asymptotic Approximations. *Journal of the American Statistical Association* **1997**, 92, 903-915.
28. Kass, R.; Raftery, A. Bayes Factors. *Journal of the American Statistical Association* **1995**, 90, 773-795.
29. Neath, A.; Cavanaugh, J. Regression and Time Series Model Selection Using Variants of the Schwarz Information Criterion. *Communications in Statistics-Theory and Methods* **1997**, 26, 559-580.
30. Mardia, K., Tests of Univariate and Multivariate Normality, Handbook of Statistics Vol. 1; North-Holland Publishing Co, 1980; pp 279-320.

31. Splus 7 for Windows, Insightful Corp., 2005.
32. Burr, T.; Foy, B.; Fry, F.; McVey, B. Characterizing Clutter in the Context of Detecting Weak Gaseous Plumes using the SEBASS Sensor. Los Alamos National Laboratory Restricted Release Report, LA-CP-06-0982, 2006, private communication.
33. Chung, K., A Course in Probability Theory, Academic Press: Orlando, 1974; pp 196-249.

© 2006 by MDPI (<http://www.mdpi.org>). Reproduction is permitted for noncommercial purposes.