

Chemical Plume Detection with an Iterative Background Estimation Technique

Eric Truslow^a, Steven Golowich^a, Dimitris Manolakis^a

^aMIT-Lincoln Laboratory, 244 Wood Street, Lexington, MA 02420, U.S.A.

ABSTRACT

The detection of chemical vapor plumes using passive hyperspectral sensors operating in the longwave infrared is a challenging problem with many applications. For adequate performance, detection algorithms require an estimate of a scene's background statistics, including the mean and covariance. Diffuse plumes with a large spatial extent are particularly difficult to detect in single-image schemes because of contamination of background statistics by the plume. To mitigate the effects of plume contamination, a first pass of the detector can be used to create a background mask. However, large diffuse plumes are typically not removed by a single pass. Instead, contamination can be reduced by using smoothed detection results as a background mask.

In the proposed procedure, a detector bank is run on the cube, and a threshold applied to produce a binary image. The binary image can be modeled as a spatial point process consisting of high density and low density regions. By applying a spatial filter to the detection image, regions with overall higher intensity are detected as containing plume and can be removed from background statistic estimates. The key intuition is that regions with a higher density of hits are more likely to contain plume since plumes are spatially contiguous. We demonstrate with real plume data that this method can drastically improve detection performance over the single-pass method, and explore tradeoffs between different filter sizes and thresholds.

Keywords: Chemical detection, hyperspectral, spatial processing

1. INTRODUCTION

Hyperspectral imagers operating in the long wave infrared (LWIR) have a variety of military and commercial remote sensing applications. The main objectives when using LWIR sensors are the detection of surface materials over a wide area, and the detection of chemical vapors in the atmosphere. Hyperspectral imagers have the high spatial and spectral resolutions necessary to completely both goals. While there are many algorithms available for the general detection problem, for adequate performance, there are many aspects of the sensing environment that must be considered. In particular, looking for rare single-pixel targets in the visible and near-infrared (VNIR) presents different challenges from locating a large spatially-contiguous vapor cloud in the LWIR. In this paper, we focus on cases where the plume is large (relative to the image), and provide a method for handling this scenario. The method we develop is inspired by spatial point process analysis techniques, and involves an iterative filtering process.

The detection of chemical vapors is accomplished by comparing a library of known absorption signatures to the observed data using a detection algorithm. The most common algorithms used in LWIR detection problems are the matched filter (MF) and the adaptive cosine/coherence estimator (ACE); we focus on the ACE detector. These algorithms are based on the assumption that the data are Gaussian, and require estimates of the mean and covariance of the background of the scene. The processing pipelines that we examine are shown in Fig. 1. The pipeline in Fig. 1b requires an estimate of the plume-free background parameters either obtained using the background mask from Fig. 1a, or from secondary data. The highlighted areas are places where spatial processing could be incorporated.

Parameter estimates usually come from secondary background data, when plume-free data is available. When the background is changing, because of platform movement, or atmospheric effects, statistics for the background may have to be estimated using data that contains the chemical plume. When the covariance matrix is estimated using plume-pixels, the covariance matrix is contaminated and detection performance may be significantly reduced. To avoid

Further author information: (Send correspondence to Eric Truslow)
Eric Truslow: E-mail: eric.truslow@ll.mit.edu

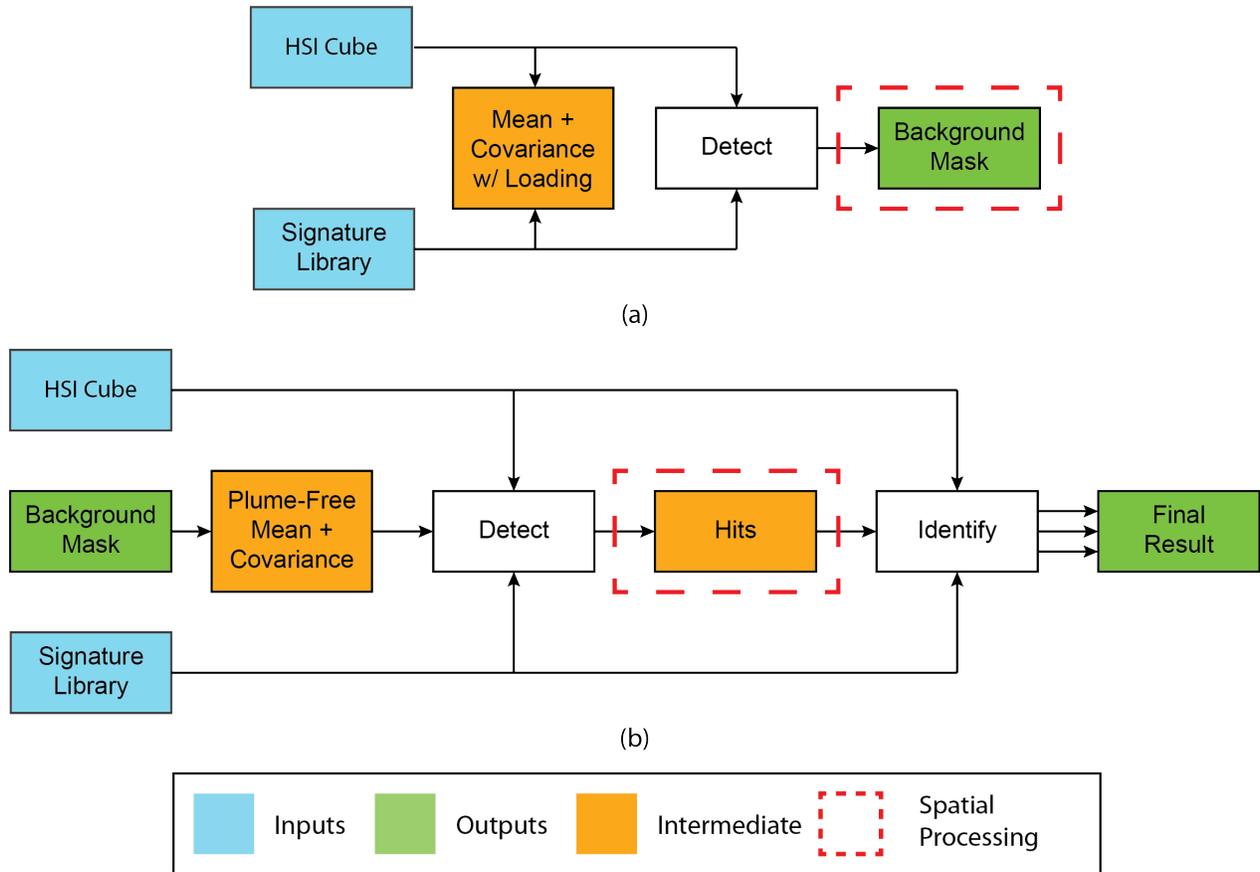


Figure 1: (a) Plume free background estimation (PFBE) pipeline. (b) Detection and identification pipeline.

contamination, the data is automatically pre-screened in a process called plume-free background estimation (PFBE), usually in a single pass. However, a single pass of the plume-free estimation procedure may not completely remove contamination. We demonstrate that using multiple passes of the PFBE with a spatial filter, can result in improved performance when the plume is relatively large.

After thresholding the output of the detection algorithm there are three types of results that are of interest:

1. unstructured background false alarms;
2. structured background false alarms; and
3. structured plume hits.

In this context, structuring refers to spatial clustering of detection pixels in the image; when the detections are unstructured, they are said to follow complete spatial randomness (CSR). Unstructured false alarms do not exhibit clustering, and can be removed through filtering. Structured false alarms cannot be removed this way, but filtering can help identify structured plume hits. It is desirable for the false alarms to be unstructured for removal with filtering. Whether the false alarms are CSR depends on several parameters of the detection system including the threshold, the PFBE procedure, the signature and the scene.

The PFBE procedure involves running a detector using a robust covariance estimate. The robust estimate involves adding a small positive value to the main diagonal of the covariance matrix. This procedure makes detection of the plume possible in the presence of some contamination, but has several negative side-effects. We show that, while it

is a useful tool for PFBE, diagonal loading can lead to structured false alarms, and reduces algorithm identification performance.

The remainder of the paper is organized as follows. Section 2 gives an overview of the chemical detection pipeline including signal models and robust detection for PFBE. Section 3 gives an review of simple spatial point processes and the concept of CSR. In Section 4, we discuss the data we used and the development of “ground-truth”. The main results are presented in Section 5, while a summary is given in Section 6.

2. CHEMICAL PLUME DETECTION OVERVIEW

There have been a variety of processing pipelines proposed and analyzed for detecting plumes in hyperspectral images. The technique we analyze consists of a PFBE step followed by a final detection or identification step, as shown in Fig. ???. The PFBE step consists of running a bank of robust detectors and then keeping a percentage of the pixels with lowest scores as the background. When only a single chemical from the library is considered, the detector bank reduces to a single detector instead.

2.1 Chemical Plume Detection Overview

To detect which pixels contain plume, a detection algorithm is run on each pixel in the image, which compares the pixel spectrum to a library signature. Starting with a simplified radiative transfer model (the three-layer model), the measured radiance can be rewritten for use in detection algorithms. A pixel with plume can be expressed in terms of the background radiance as

$$x(\lambda) \approx \Delta_T \tau_a(\lambda) \sum_i^m \alpha_i s_i(\lambda) + L_{\text{off}}(\lambda) \quad (1)$$

where

- λ is wavelength;
- x is the pixel under test;
- τ is the atmospheric transmission;
- Δ_T is the temperature contrast between the plume and background;
- m is the number of chemicals in the pixel;
- α_i is the concentration of the i th chemical;
- s_i is the i th library signature, and
- L_{off} is the radiance emitted by the background.

Sampling at a set of band centers $[\lambda_1, \dots, \lambda_p]$ yields the measurement vector $\mathbf{x} = [x_1, \dots, x_p]^T$ where p is the number of sensor channels. The library signatures $s_i(\lambda)$ are multiplied by the assumed atmospheric transmission $\tau_a(\lambda)$ and downsampled to the band centers to obtain sampled signatures s_i . Organizing the signatures as a matrix \mathbf{S} and the b_i 's as a vector \mathbf{b} , we have

$$\mathbf{x} = \sum_{i=1}^m \mathbf{s}_i b_i + \mathbf{v} = \mathbf{S}\mathbf{b} + \mathbf{v}, \quad \mathbf{v} \sim \mathcal{N}(\mathbf{m}_b, \mathbf{C}_b) \quad (2)$$

where \mathbf{m}_b is the background clutter mean and \mathbf{C}_b is the clutter covariance. The assumption in Eq. 2 is that the noise and background clutter \mathbf{v} are well modeled by a multivariate Gaussian distribution. Defining the whitening matrix $\mathbf{C}_b^{-1/2}$ and whitened vectors

$$\tilde{\mathbf{x}} = \mathbf{C}_b^{-1/2}(\mathbf{x} - \mathbf{m}_b), \quad \tilde{\mathbf{S}} = \mathbf{C}_b^{-1/2}\mathbf{S}, \quad \tilde{\mathbf{v}} = \mathbf{C}_b^{-1/2}(\mathbf{v} - \mathbf{m}_b) \quad (3)$$

yields the standard regression model

$$\tilde{\mathbf{x}} = \sum_{i=1}^m \tilde{\mathbf{s}}_i \tilde{b}_i + \tilde{\mathbf{v}} = \tilde{\mathbf{S}}\tilde{\mathbf{b}} + \tilde{\mathbf{v}}, \quad \tilde{\mathbf{v}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \quad (4)$$

where the clutter and noise is zero mean and has identity covariance.

When there is a single chemical in the library, a detection algorithm can be used to decide whether that chemical is present or not. Perhaps the most well known detection algorithm used in hyperspectral imagery is the matched filter (MF).¹ However, the adaptive cosine/coherence estimator (ACE) is considered the state-of-the-art detector. The ACE detector for the k th library signature is defined as

$$y_k = \frac{(\tilde{\mathbf{x}}^T \tilde{\mathbf{s}}_k)^2}{\|\tilde{\mathbf{x}}\|^2 \|\tilde{\mathbf{s}}_k\|^2} = \cos^2(\tilde{\theta}_k). \quad (5)$$

where $\tilde{\theta}_k$ is the angle between the whitened signature $\tilde{\mathbf{s}}_k$ and the whitened pixel $\tilde{\mathbf{x}}$. Finally, the score y_k is compared to a threshold η to make a decision about each pixels in the scene.

2.2 Robust Detection

In the detection process, the mean and covariance of the background must be estimated from the available data. Usually, these estimates are denoted $\hat{\mathbf{m}}_b$ and $\hat{\mathbf{C}}_b$, and are directly substituted into the whitening equations given earlier. For decent performance, it is crucial that any pixels that contain plume be removed from the estimates.² In the VNIR, where target materials usually consist of a few rare and often spectrally distinct pixels, a single pass of the PFBE procedure may be sufficient.³ However, in plume detection, multiple passes over the data can improve performance when the plume is large. For simplicity, a single detector with the known chemical

To mitigate the effects of contamination caused by estimating background statistics using pixels within the plume, a first pass with a robust detector can be used. Detection algorithms can be made robust to both noise and signature mismatch through covariance loading and dominant mode rejection (DMR), which both involve modifying the eigenvalues of the estimated covariance matrix.⁴ We consider the diagonal loading approach, which results in modification of the eigenvalues of $\hat{\mathbf{C}}$. Using the singular value decomposition (SVD) the modification becomes

$$\hat{\mathbf{C}} + \delta \mathbf{I} = \mathbf{U}(\mathbf{\Lambda} + \delta \mathbf{I})\mathbf{U}^T \quad (6)$$

where \mathbf{U} contains of eigenvectors of $\hat{\mathbf{C}}$, and $\mathbf{\Lambda}$ is a diagonal matrix with the singular values σ_k^2 on the main diagonal sorted from largest to smallest.

$$\hat{\mathbf{C}}_{\delta}^{-1} = (\hat{\mathbf{C}} + \delta \mathbf{I})^{-1} \quad (7)$$

where δ is the loading factor. Selection of the loading factor is critically important and depends on both the data and the application. Large eigenvalues in $\mathbf{\Lambda}$ correspond to small eigenvalues in the inverse, and vice-versa. The addition of δ affects the smallest eigenvalues of \mathbf{C} more than the largest and effectively reduces the overall spread of eigenvalues in both $\hat{\mathbf{C}}$ and its inverse.

3. SPATIAL POINT PROCESSES

After thresholding the detector output, we have a binary image or point pattern, where each one in the image is called an *event*. Usually, spatial point patterns are defined on a continuous region of the xy -plane, but the principles are very much the same.

In the continuous case, if the point process follows *complete spatial randomness* (CSR) then the number of events in any region A follow a Poisson distribution with intensity $\lambda|A|$ where λ is the *intensity* or expected number of events per unit area,⁵ and $|A|$ is the area of A . The probability of finding k events in a region of size $|A|$ becomes

$$P(k; \lambda) = \frac{e^{-\lambda|A|} (\lambda|A|)^k}{k!} \quad (8)$$

where the parameter λ is constant over A . If the process is CSR, the events E are independent and uniformly distributed on the region A . When the intensity is a function of position over the region, the process is called *inhomogeneous*. One way to estimate the intensity of an inhomogeneous process is by partitioning the region into sub regions called quadrats, and to then tally the number of events in each quadrat. This is known as a quadrat count and provides an spatially varying intensity estimate.

For a binary detection image, the region is already partitioned into cells, but at most one event can occur in each cell. In this case, the region A is the rectangular region $[0, M_S] \times [0, M_L]$ in the xy -plane. The pixel locations can be defined at equally spaced positions on the interior of the region, with the first pixel is located at $(0.5, 0.5)$. There is a subset of pixels that have a value of 1 (the hits) with locations $e_i = (x_i, y_i)$, $i = 1, \dots, N$, which form the set of events E . Since each pixel can either be a 0 or a 1, the number of events at any particular location is at most 1. If the events occur randomly and independently with some probability of occurrence p , then the image can be modeled as a series of Bernoulli trials.⁶ In this case, taking any subset of the image with M pixels, the total number of events will follow a binomial distribution. That is, the probability of having k detections in N pixels is

$$B(k; p, M) = \binom{M}{k} p^k (1-p)^{(M-k)} \quad k \in \{0, \dots, M\} \quad (9)$$

where p is the probability of a detection occurring. When the probability of occurrence p is small and the number of pixels is large, the binomial distribution can be approximated by a Poisson distribution.

The maximum likelihood estimate of p is the average number of events over the estimation area, written as

$$\hat{p}_{\text{ML}} = \frac{k}{M} \quad (10)$$

where M is the number of pixels being considered, and k is the number of events or sum of the pixels in the region. For the inhomogeneous case the spatially varying intensity can be estimated using quadrat counting, or a non-parametric estimator such as

$$z(x, y) = \sum_{i=1}^N \frac{1}{w_i} h(x - x_i, y - y_i) \quad (11)$$

where the function h is centered at each event $e_i = (x_i, y_i)$ in E and may be weighted by an edge-correction factor w_i . The edge correction factor is used when the event is near a boundary causing the estimator to be biased. The function z is a continuous function over A that, when evaluated at a grid of points results in an image. Since the image only takes values of 1 at the locations of the events, the operation in (11) is a convolution of a binary image with a filter function h .

To get an estimate of the probability of success p at each point in the image, we can create a 2-D averaging filter defined as

$$h[m, n] = \begin{cases} 1/N & (m, n) \in B \\ 0 & \text{otherwise.} \end{cases} \quad (12)$$

where B defines a square region of N points. Since plumes have smooth boundaries a disk filter or Gaussian filter; are more appropriate. Fig. 2 shows a disk filter and Gaussian filter generated using Matlab's `fspecial` function

Assuming that within the region there are at least two processes with different spatially varying intensities, it is possible to partition the region based on local intensity estimates.⁷ Of particular interest is when one of the processes has a low intensity and represents background noise and clutter while the other is a higher intensity process which contains the signal of interest.

3.1 Summary Statistics for Spatial Point Patterns

To test whether or not a particular spatial point pattern arose under CSR, one of several summary or test statistics can be developed. This type of hypothesis test does not tell us much about the actual underlying process, except that if the data are CSR the generating process does not have any underlying structure. However, it is useful for our application to test whether or not background false alarms are well modeled as CSR or not. A summary of the statistics discussed is shown in Table 1.

In analyzing a spatial point pattern, the distances between neighboring points are of primary interest. Given the set of events E , distances can be represented by an adjacency matrix D with elements $d_{ij} = \|e_i - e_j\|$, the matrix contains the pairwise distances from each event (detection) to every other event. The elements of the matrix are

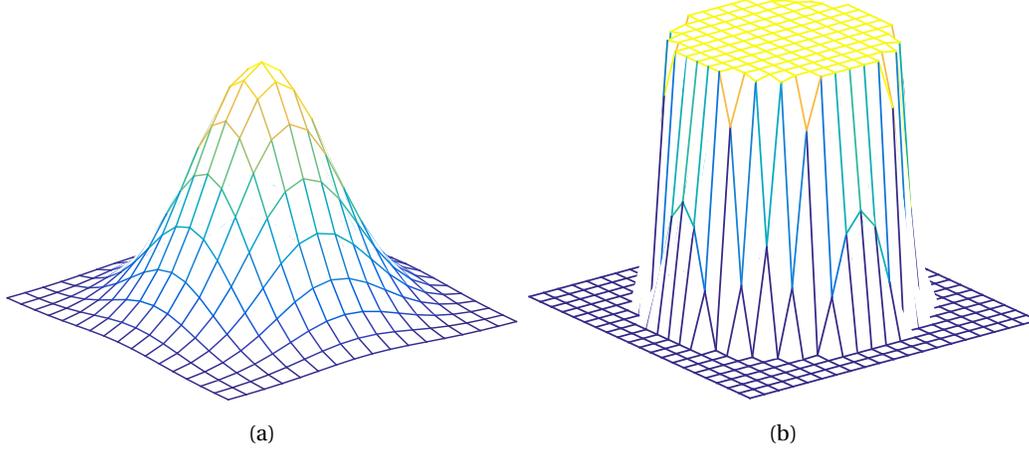


Figure 2: Filters used in spatial processing problems. (a) Gaussian filter of size 17 with $\sigma = 3$; (b) disk filter with radius 7.

Symbol	Name	Definition
$G(r)$	Nearest neighbor distance	$\frac{1}{N} \sum_{i=1}^N I(\min_{j \neq i} (d_{ij}) \leq r)$
$F(r)$	Point to nearest event distance	$\frac{1}{M} \sum_{k=1}^M I(\min_i (\ s_k - e_i\) \leq r)$
$E(r)$	Expected events in radius r	$\frac{1}{N} \sum_{i=1}^N \sum_{j \neq i} \frac{1}{w_{ij}} I(d_{ij} \leq r)$
$K(r)$	Ripley's K-function	$\frac{ A }{N(N-1)} \sum_{i=1}^N \sum_{j \neq i} \frac{1}{w_{ij}} I(d_{ij} \leq r)$

Table 1: Commonly used summary statistics for spatial processes. w_{ij} are edge corrections.

$d_{ij} = \|e_i - e_j\|$ for all. The distance matrix is size N^2 , has non-negative elements and zeros on the main diagonal. The cumulative-nearest-neighbor distribution, is defined with respect to distance r as

$$G(r) = \frac{1}{N} \sum_{i=1}^N I(\min_{j \neq i} (d_{ij}) \leq r) \quad (13)$$

where $I(\cdot)$ is the indicator function. The function G measures the cumulative fraction of events whose closest neighbor is less than a distance r away.

The point-to-nearest-event distribution $F(r)$ describes the distance from arbitrary sampling locations to the nearest event. The number of sampling points and their locations can be varied, but is typically a grid of points over the region. For a set of sampling points s_k , the function is estimated as

$$F(r) = \frac{1}{K} \sum_{k=1}^K I(\min_i (\|s_k - e_i\|) \leq r) \quad (14)$$

where M is the number of sampling locations. Both of the functions F and G follow the same distribution under CSR, namely

$$F(r) = G(r) = 1 - e^{-\lambda \pi r^2} \quad (15)$$

where πr^2 is the area of a circle of radius r . Instead of examining distributions that consider individual events, statistics that consider pairs of events may be used. The function $E(r)$ is defined as

$$E(r) = \frac{1}{N} \sum_{i=1}^N \sum_{j \neq i} \frac{1}{w_{ij}} I(d_{ij} \leq r) \quad (16)$$

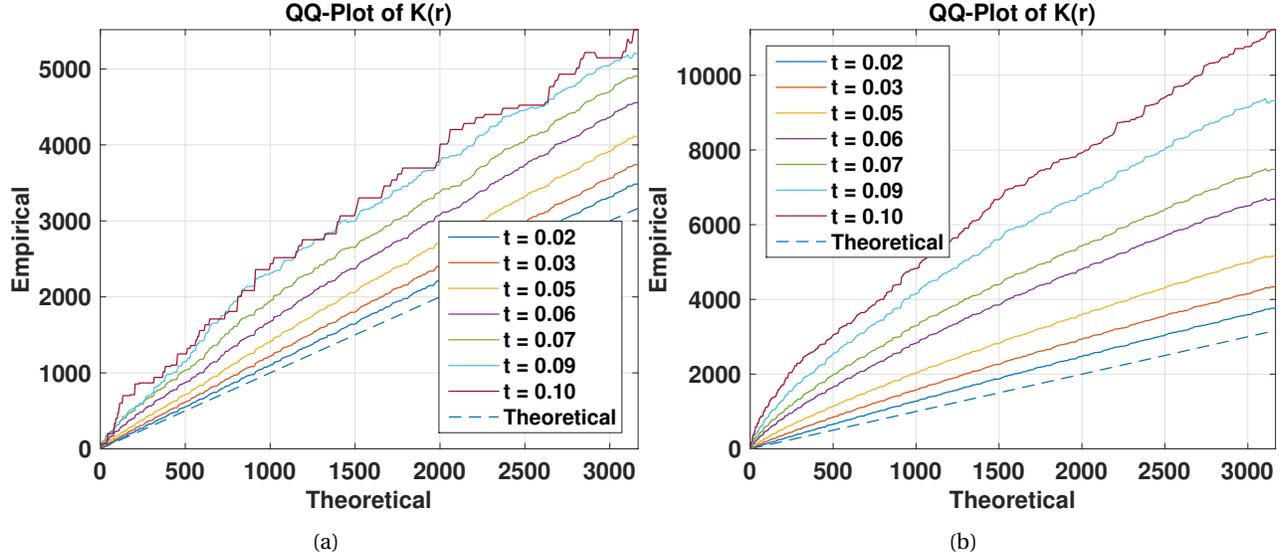


Figure 3: QQ-Plot of $K(r)$ for various single gas ACE thresholds of plume-free cube. (a) Chemical B background cube. (b) Chemical C background cube. QQ-Plot of $G(r)$ for various single gas ACE thresholds of plume-free cube.

and measures the expected number of events within a distance r from an arbitrary event. Taking the ratio $E(r)/\lambda$, we obtain a weighted inter-event distance function, also known as Ripley's K-function, defined as

$$\hat{K}(r) = \hat{E}(r) \frac{|A|}{N-1} = \frac{|A|}{N(N-1)} \sum_{i=1}^N \sum_{j \neq i} \frac{1}{w_{ij}} I(d_{ij} \leq r) \quad (17)$$

where the estimate of the intensity is approximated by $|A|/(N-1)$. The edge correction weights w_{ij} used for E and K are defined as the proportion of the circumference of a circle centered at e_i , with radius $\|e_i - e_j\|$, that intersects the region A . For a rectangular region the edge correction weights w_{ij} have a simple closed form solution.⁵ Edge-corrections for (13) and (14) are also available, but are related to the area of a circle instead of the circumference.⁸ Several edge correction weights are available for each of the estimators discussed, but for large enough sample sizes with sufficiently small choices of r , the difference in weighting should be negligible.⁹ Using Ripley's K-function $K(r)$, we can assess how CSR-like a pattern is as a function of threshold. For some choices of A , closed-form expressions for K exist, but in practice it is easier to use Monte-Carlo (MC) simulation to estimate the function under CSR, as with many of the functions used to analyze spatial point patterns. Fig. 3 shows a QQ-plot of the empirical K-function and the theoretical one for ACE scores at different thresholds. Counter-intuitively, lower thresholds lead to a closer fit between the empirical and theoretical. Since the K-function measures the distribution of pair-wise distances, as the threshold is lowered, the false alarms become more regularly spaced, leading to a better fit. This suggests that background false alarms with the highest scores exhibit spatial structure, but that the pattern of false alarms becomes more CSR-like for lower thresholds. Spatial filtering could be used to remove these false alarms, but may not remove the most problematic.

4. DATA DESCRIPTION

Per-pixel ground truth for real side-looking hyperspectral data is generally unavailable because the exact spatial extent of the plume is unknown. However, when the camera is stationary and the atmospheric conditions sufficiently similar over a viewing time period, the statistics of a plume-free cube can be used to detect the plume in later images. In practice, where the system may be running for a long time under varying weather conditions, or on a mobile system, this strategy for detecting plumes may not work.

The datasets we have that contain real chemical releases are taken over a relatively short time interval of approximately 20 minutes. During this interval, several pre-release cubes are available to estimate background statistics from. Using the estimated mean and covariance from the plume-free cube, resolves the problem of having to excise the plume

Run	Chemical	Data Characteristics
1	A	Very large plume at maximum size.
2	A	Moderately sized plume at maximum.
3	B	Weak plume throughout; structured background false alarms

Name	Notes
Library	The library contained 12 signatures.
Chemical A	Much easier to detect, and did not have many confusers.
Chemical B	More difficult to detect, and produced several large false alarm regions.

Table 2: Description of experimental data.

before estimating background statistics. Table 2 lists three different runs from the full dataset and gives a qualitative description of each. We focus on the first run because of how large the plume became.

To create masks for detection, we ran a detector using the estimated background parameters from an earlier cube in the series, as shown in Fig. 4. A disk filter with radius 25 was applied to the thresholded detection image, with low threshold of 0.1. After filtering using (11), the pixels with estimated intensity greater than 0.1 were used to construct an optimistic guard mask of where the plume might be. This mask includes pixels that initially did not pass the detection threshold, but that had many neighbors pass the threshold. The pixels that passed the threshold and had an estimated intensity greater than 0.1 were used as the ground truth. For all analysis, the pixels in the envelope but not in the true-plume, were excluded from consideration.

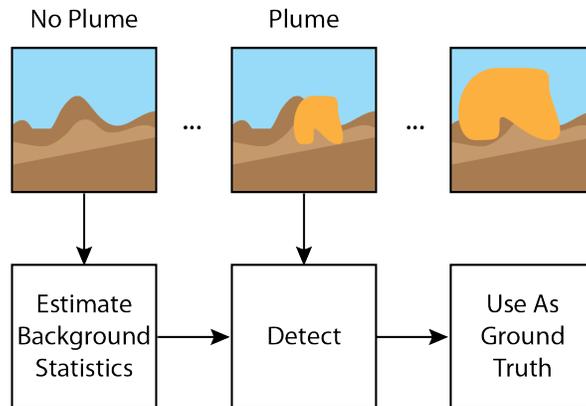


Figure 4: Illustration of ground-truth creation procedure.

Using this ground truth, the fraction of the image taken up by the plume in each cube was estimated, as shown in Fig. 5. Generally, the larger and more dispersed the plume the, harder it was to detect using background data from the same cube, due to covariance contamination, being optically thinner and potentially a lower temperature contrast. For these reasons, we considered cube 16 from run number 1. In cube number 8, the plume is still expanding, while number 16 is where the plume is at its maximum size relative to the size of the image. Chemical A was generally easier to detect and identify than chemical B, except when the plume became relatively large. Although ground truth created this way is not perfect, it provides a way to compare the best-case detection results to the single-image results.

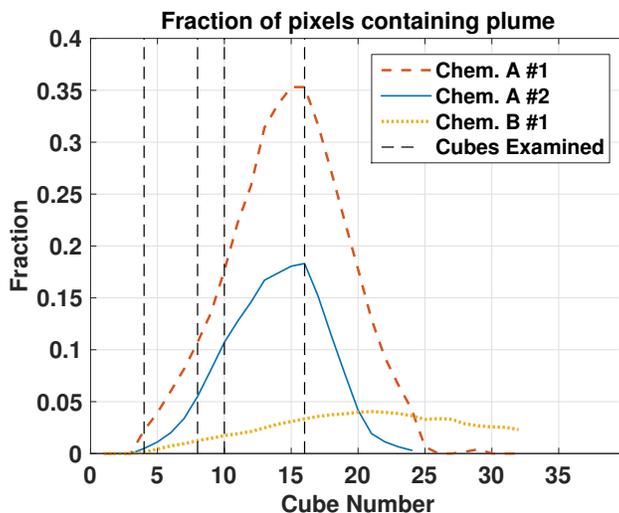


Figure 5: Fraction of pixels in each cube that were deemed to contain plume based on best case scenario used for ground truth. The vertical lines denote cube numbers 4, 8, 10, and 16.

5. RESULTS

To evaluate the effectiveness of adding an additional spatial processing step to the existing processing pipeline, a single set of cubes from the same release was used. In our analysis in this section, only a single detector with the correct library signature was used. This experiment represents the best case situation where a single chemical is in the library and is the chemical observed. In most situations, instead of having a single chemical signature, a series of signatures will need to be examined. While the library we used had 12 signatures, introducing results for each signature complicates analysis considerably.

Having a positive loading factor makes the detection and identification algorithms less sensitive to mismatches between the library signature and the measured signal.¹⁰ The effect of the loading factor on the detection scores can be substantial. In Fig. 6(a) the detection scores for a background cube match the theoretical distribution of ACE when no loading factor is used, but deviate significantly as the loading factor is increased. The theoretical distribution only applies when the whitening transformation leads to uncorrelated and identically distributed Gaussian data. Since these assumptions do not hold when a large loading factor is used, deviation from the theoretical is expected.

The assumption of perfect whitening leads to good agreement between the theoretical distribution of ACE and the empirical distribution, but also makes false alarms relatively CSR, as discussed in Section 3.1. Deviation in the ACE scores from the theoretical distribution affects some pixels more than others. If a group of affected pixels are spatially contiguous, then deviation from CSR is expected. This is useful for highlighting potential plume pixels. However, background false alarms may also be emphasized, which can lead to large structured false alarm regions. Fig. 6(b) show a QQ-plot of the theoretical K -function and the empirical K -function for false alarms at the same set of loading factors as the previous example. The same false alarm rate was set for each loading factor. Increasing the loading factor led to larger deviations from the theoretical CSR distribution, though even without a loading factor, the false alarms were not perfectly CSR. Using the preliminary detection results from a robust detector, a mask for the estimated background pixels can be created before re-estimating background statistics and re-running the detection algorithms.

Fig. 7 shows a scree plot of the eigenvalues of an estimated covariance matrix over several cubes. A scree plot shows the eigenvalues of the estimated covariance matrix sorted from largest to smallest. Since the smallest eigenvalues of the covariance matrix are on the order of 10^{-16} and the distribution of these eigenvalues is fairly flat (aside from the largest few), even a loading factor on the order of 10^{-16} is large relative to the smallest eigenvalues.

The distinct spatial structure of both background false alarms and the plume, suggests that it is possible to separate these pixels from other randomly distributed false alarms in the image. However, it is difficult to differentiate structured false alarm regions from plume regions, without additional processing. Thus, using a relatively large loading factor for a

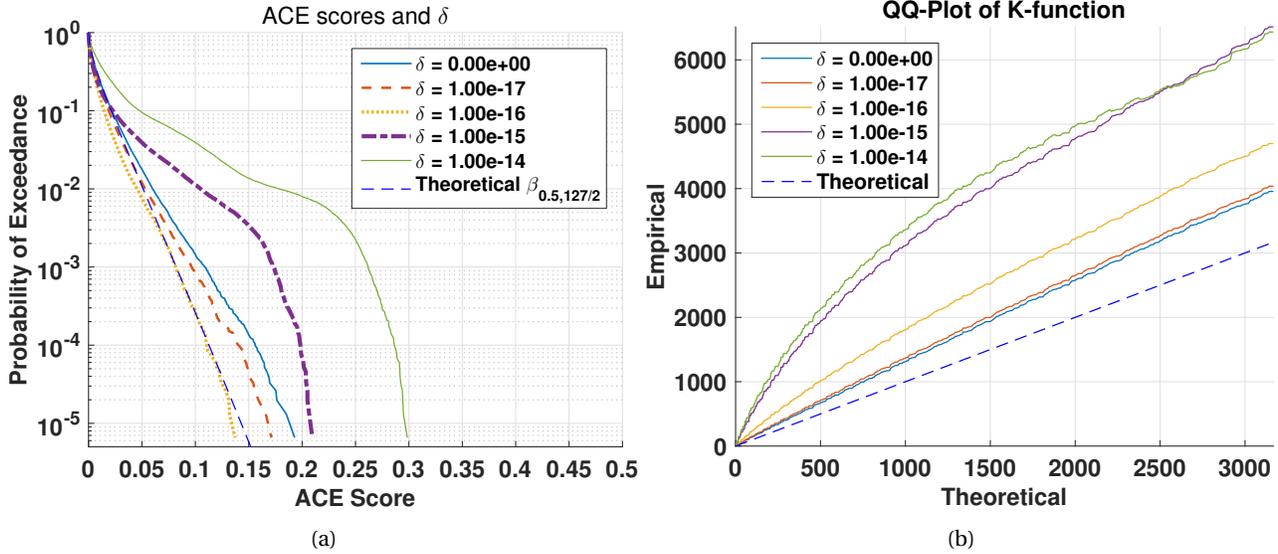


Figure 6: (a) Probability of exceedance plots of ACE scores for a background-only cube, with various loading factors. Exceedance is one minus the empirical CDF. The distribution of scores became more multi-modal as the loading factor increased. For reference, the exceedance function for a β distribution (the theoretical distribution of ACE) is plotted. (b) QQ-Plot of empirical K-function versus theoretical values (under CSR) for different values of the loading factor δ . In each case, a threshold for the top 10% of scores was set. As the loading factor increases, the detections show more spatial structure, and the empirical estimates move further from the theoretical ones.

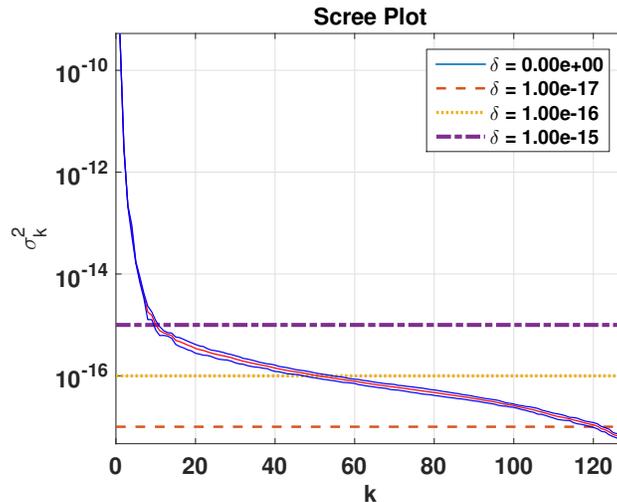


Figure 7: Scree plot of the singular values σ_k of the estimated covariance matrix for a set of cubes. The mean (red) and three standard deviations (blue) of the calculated singular values are shown. The horizontal lines show different loading factors from 10^{-17} to 10^{-15} .

first detection pass, leads to a good estimate of regions that may contain plume. Though this is a conservative estimate that may contain background pixels, contamination should be avoided.

5.1 Iterative Background Estimation

Instead of setting a static threshold on the detection scores to do PFBE, through extensive testing a quantile-based approach has been found to be both easier to implement and less sensitive to threshold. In a quantile-based approach, assume that the plume takes up at most a fraction of the image, denoted $1 - q$. Then, a plume-free background estimate can be formed by taking the pixels with detection scores up to the q th quantile. Using such a technique, the same

fraction of every cube is excluded from background estimates as plume.

When the cube is a background-only cube, background pixels will be incorrectly labeled as plume pixels and will be excluded from estimates of the background statistics. Just as excluding the plume from background estimates improves detection of the plume, excluding background pixels from background estimates, can increase the false alarm rate for a given threshold. This is especially important for structured background false alarms, because they may be spatially contiguous.

If the initial background estimates do not remove all of the plume, but a portion of the plume is detected, it is possible to iteratively improve the background estimates and thereby improve performance. If at each step, the plume pixels fall in a higher quantile, more of the plume will be excluded from the background estimates, and consequently fewer background pixels will be labeled as plume, making overall detection improve.

The iterative algorithm is described as follows:

1. Initialize background mask to whole image.
2. Run a robust ACE detector to get y_0 .
3. Form a detection mask with $y > \eta$.
4. Apply disk filter with radius r to the detection image to get intensity estimates \hat{p} .
5. Calculate quantiles of filtered detection image.
6. Keep bottom q fraction of filtered image as background mask.
7. Re-estimate background statistics.
8. Repeat steps 2-7, up to K times.

We have found that for a single chemical detector, 7-10 iterations were needed for good results, while when using a library of 12 chemicals, 15-25 iterations were needed. Since, we knew that the plume was at most approximately 40% of the scene, q was set to 0.6. The resulting background masks are shown in Fig. 8

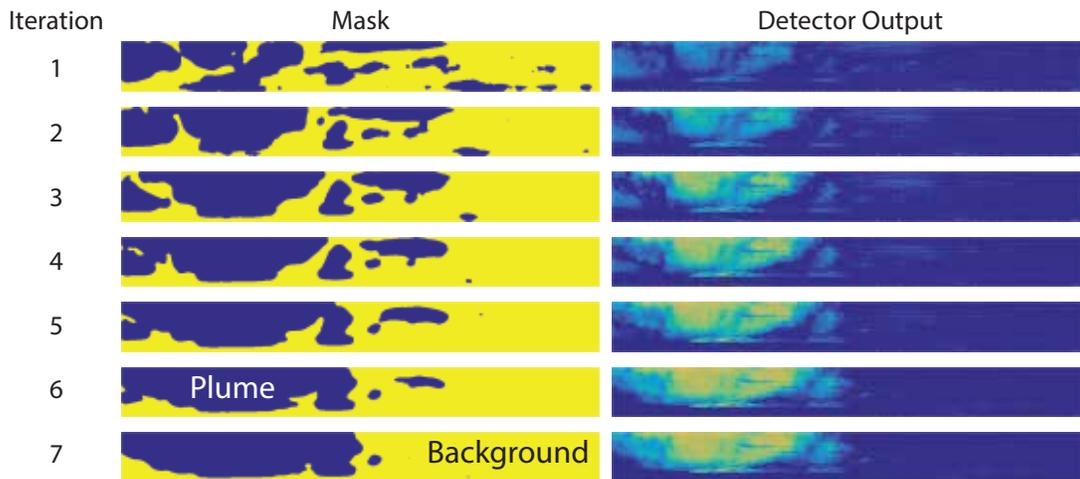


Figure 8: (left) Plume free background estimates at first 7 iterations. (right) Robust detector output after each iteration.

Fig. 9a shows the resulting ROC curves at each iteration of the procedure, for a total of 7 iterations, with the final two ROC curves being nearly identical. The first iteration is a slight improvement on the robust detector described in previous sections, but the later iterations are a vast improvement. The probability of detection at zero false alarm rate is substantially better than all previous examples, and each iteration appears to improve performance. Fig. 9b shows the

fraction of the plume pixels included in the background estimate as a function of iteration. Initially, a relatively large portion of the plume is included in the estimates, leading to substantial performance degradation. However, in each iteration of the procedure, more of the plume is excluded from the background statistics, and performance improves substantially.

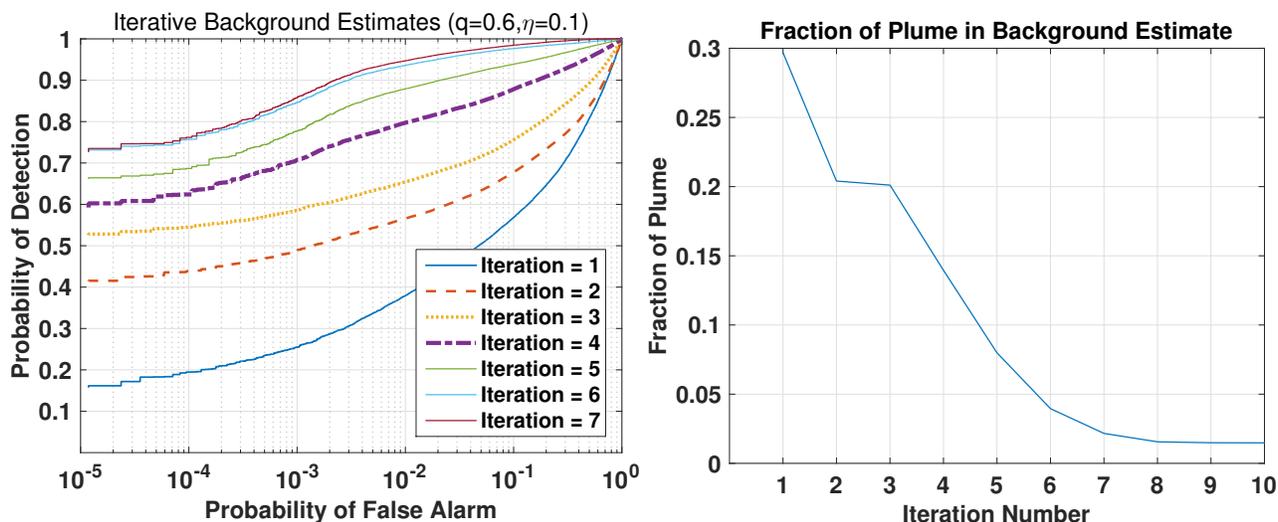


Figure 9: (a) Iteratively taking the bottom q fraction of intensity scores as background, while thresholding the ACE scores at 0.1. (b) Fraction of plume in background estimates, as a function of iteration.

When there is no plume in the scene, removing a portion of the cube as background has the effect of increasing the overall detection scores. This effect can increase the false alarm rate substantially without proper thresholding. Fig. 10 shows the probability of exceedance of the ACE filter bank at different stages of the process. The initial scores when no loading factor or PFBE are used, is shown in solid blue. After the iterative PFBE is run, the slightly heavier-tailed purple curve results. The highest detection score moves from about 0.5 to about 0.6, which must be accounted for when trying to set a CFAR threshold.

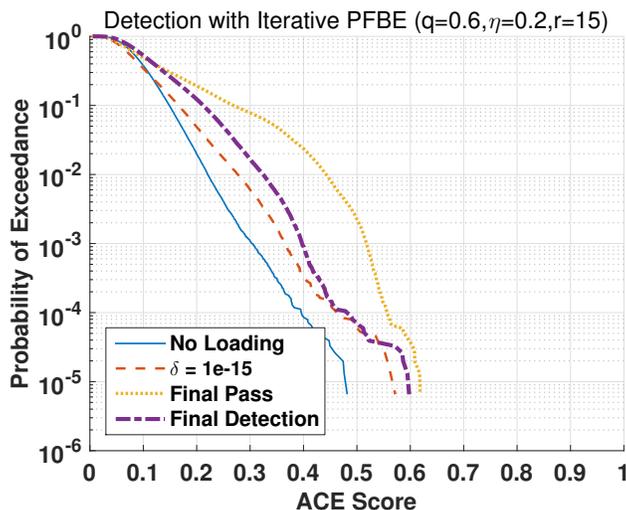


Figure 10: Chemical A, Run 1, Cube # 1. Probability of exceedance for a plume-free cube and ACE detector bank with: (blue) no loading factor and no background estimation, and (orange) a robust covariance inverse. (yellow) Robust detection after PFBE. (purple) Detection without loading factor after PFBE. A disk filter of radius 15 was used and the top 60% of the filtered hits used for iterative background estimation.

When using the full library of chemical signatures, a bank of ACE detectors was used and the maximum detection

score for each pixel was kept. These maximum scores were then thresholded to get a detection map. Some of the signatures in the library produce higher scores for background false alarms than others, resulting in larger background false alarm areas than in the single-detector case. Using the modified procedure on the same cube as in Fig. 8 resulted in the same type of performance as when using a single-gas detector. However, the number of iterations required to achieve the same performance increased by about 2-fold. Fig. ?? shows ROC curves at several different iterations through the process. At termination, the ROC curve is worse than the one for a single chemical. This is primarily due to other library chemicals (aside from the correct one) causing additional false alarms. As shown in Fig. 11, approximately 15 iterations were required to remove most of the plume from the background statistics in this case.

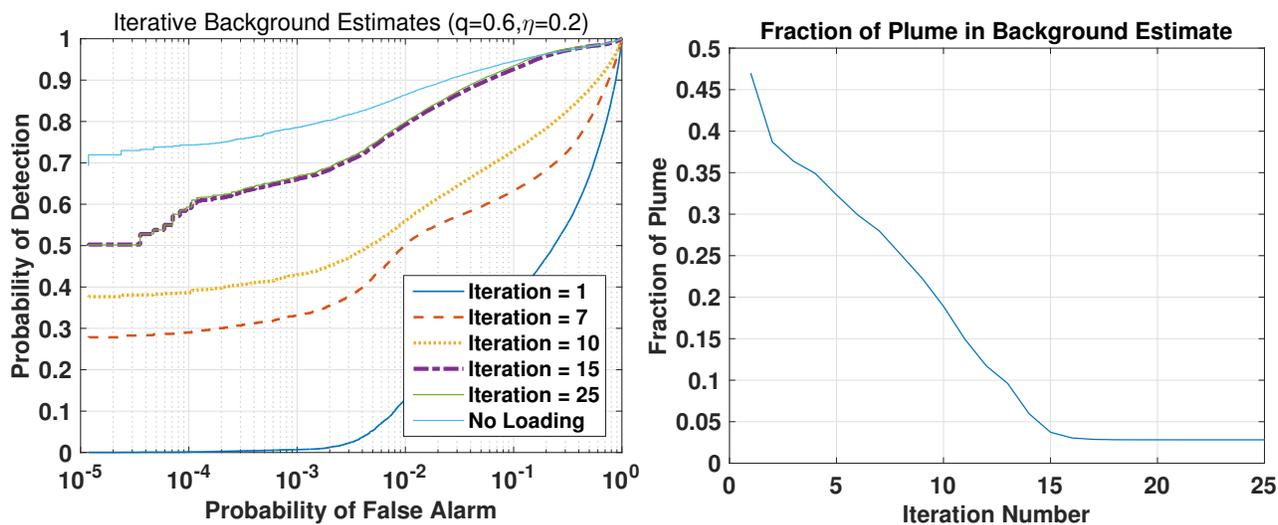


Figure 11: (a) Chemical A, Run 1, Cube # 16 with a library of 12 signatures. ROC curves using iterative background estimation, taking the bottom q fraction of intensity scores as background, while thresholding the ACE scores at 0.2 and applying a disk filter. (light blue) Final detection performance after iteratively improving background estimate. (b) Fraction of the plume included in background estimates as a function of iteration. A library of 12 gases was used along with an ACE detector bank.

6. SUMMARY AND CONCLUSIONS

Plume contamination of the background parameter estimates remains a difficult problem to address for plume detection systems, especially when the background is relatively rapidly changing. This type of situation can occur when the sensor is in motion, or if the weather or atmospheric conditions are changing quickly. The sensor in this paper was stationary, and the background relatively similar for each cube examined. This allowed us to develop a pseudo ground-truth for ROC curve construction, and to quantitatively examine detector performance on a difficult dataset. Without this ground-truth, regions of the image would have to hand-labeled as containing plume, which is impractical for a large number of cubes and pixels. Essentially, we compare a detector with imperfect knowledge of the background to one with perfect knowledge of the background.

The use of spatial point patterns for describing detection images is useful in assessing whether or not there is structure in the detection results. When false alarms are CSR and the plume exhibits spatial clustering, a spatial filter can be applied to separate the high density plume regions from the low-density false alarm regions. We showed that iteratively applying this technique could be used to remove plume contamination before a final pass of the detector.

We showed that using a relatively large loading factor lead to detections being more non-CSR than using no loading factor. When the plume was present, the regions that exhibited clustering were more likely to contain the plume, leading to better PFBE. However, when the cube did not contain plume, the loading factor increased the scores of undesirable background false alarms.

ACKNOWLEDGMENTS

This work was sponsored by the Defense Threat Reduction Agency under Air Force contract FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the author and not necessarily endorsed by the United States Government.

REFERENCES

1. Manolakis, D., Truslow, E., Pieper, M., Cooley, T., and Brueggeman, M., "Detection algorithms in hyperspectral imaging systems: An overview of practical algorithms," *IEEE Signal Processing Magazine* **31**(1), 24–33 (2014).
2. Niu, S., Golowich, S. E., Ingle, V. K., and Manolakis, D. G., "Implications of model mismatch and covariance contamination on chemical detection algorithms," *Proceedings of SPIE* **8048**, 80481E–80481E–12 (May 2011).
3. Manolakis, D., Pieper, M., Truslow, E., Cooley, T., Brueggeman, M., and Lipson, S., "The remarkable success of adaptive cosine estimator in hyperspectral target detection," *Proc. SPIE* **8743**, 874302–874302–13 (2013).
4. Manolakis, D., Lockwood, R., Cooley, T., and Jacobson, J., "Hyperspectral detection algorithms: use covariances or subspaces?," *Proc. SPIE* **7457**, 74570Q–74570Q–8 (2009).
5. Diggle, P., [*Statistical Analysis of Spatial and Spatio-Temporal Point Patterns, Third Edition*], Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Taylor & Francis (2013).
6. Papoulis, A. and Pillai, S., [*Probability, Random Variables, and Stochastic Processes*], McGraw-Hill series in electrical engineering: Communications and signal processing, Tata McGraw-Hill (2002).
7. Diggle, P. J., Moraga, P., Rowlingson, B., and Taylor, B. M., "Spatial and Spatio-Temporal Log-Gaussian Cox Processes: Extending the Geostatistical Paradigm," *Statistical Science* **28**, 542–563 (Nov. 2013).
8. Diggle, P., [*Statistical Analysis of Spatial Point Patterns*], Mathematics in biology, Arnold (2003).
9. Baddeley, A. J. and Turner, R., [*Spatstat: An R Package for Analyzing Spatial Point Patterns*], University of Western Australia. Department of Mathematics and Statistics (2004).
10. Niu, S., Golowich, S. E., Ingle, V. K., and Manolakis, D. G., "Design and evaluation of robust matched filters for chemical agent detection," *Proc. SPIE* **8186**, 81860R–81860R–10 (2011).