

Onboard Cubesat Data Processing for Hyperspectral Detection of Chemical Plumes

James Theiler, Bernard R. Foy, Claira Safi, and Steven P. Love

Intelligence and Space Research Division
Los Alamos National Laboratory, Los Alamos, NM 87545

ABSTRACT

We describe the development and implementation of plume detection algorithms under severe bandwidth and processing constraints imposed by a CubeSat architecture. In particular, two ideas will be presented: one employs onboard processing to reduce the data that is downlinked, and one employs the Sparse Matrix Transform (SMT) to speed up the onboard computation of an approximate Mahalanobis distance.

Keywords: cubesat, satellite, hyperspectral imagery, algorithm, plume detection, sparse matrix transform, Mahalanobis distance, matched filter, adaptive coherence estimator

1. INTRODUCTION

The Targeted Atmospheric Chemistry Observations from Space (TACOS) project aims to launch lightweight CubeSat hardware into orbit, carrying an ultra-compact UV-Visible hyperspectral imager whose purpose is to detect and characterize anthropogenic and volcanic gas-phase emissions. Specific targets include NO₂, SO₂, ozone, and formaldehyde, with sufficient spectral resolution to confidently separate the trace gas signatures from the atmosphere. The TACOS instrument is similar in purpose and design to NASA's OMI (Ozone Monitoring Instrument)¹ but with a much smaller pixel size – 0.4km instead of 12km – suitable for characterizing individual power plants or detecting the low-level emissions of recently-awakened volcanoes.^{2,3} Building on a legacy of compact hyperspectral sensors,⁴ the TACOS imager has custom-designed optics that focus and spectrally disperse a slit of light onto a 320×320 patch of the CCD focal plane. The 320 spectral channels span a 300–500nm wavelength range, with a resolution of 0.6nm.

Routine downlinking of raw hyperspectral data cubes is always a challenge, but the extremely limited bandwidth of CubeSat telemetry makes this particularly so. Our CubeSat bus will have an approximately 16.4MB single-pass downlink capability. For a nominal hyperspectral data cube comprising 2000 (along-track) × 320 (across-track) pixels, each with 320 spectral channels, a full downlink of 16-bit data would require 25 passes to transmit.⁵ We will therefore include on-board processing to produce chemical retrieval images on orbit; then downlink just a few (up to ten, or so) of these processed 2D grayscale images, along with a small sample (up to a thousand or so) of unprocessed full-spectrum pixels, all of which can be fit into a single-pass downlink.

A further challenge is the limited computational power that is available onboard. Although optimized implementations for gas detection algorithms have been developed for sophisticated CPU and GPU platforms,^{6,7} we will be dealing with substantially more limited processing power (see Table 1). Further, with the proposed low-earth orbit altitude of approximately 550km, we will only have about 90 minutes of processing time per collect.

Two of the most computationally expensive steps in our processing chain are the computation of the $d \times d$ covariance matrix from the N pixels (where nominally, $d = 320$ and $N = 2000 \times 320 = 6.4 \times 10^5$) and the computation of the Mahalanobis distance for each of the N pixels. Both of these computations scale as $O(Nd^2)$, and would nominally require over five hours of processing to complete. The covariance matrix computation can be straightforwardly expedited by making the estimate with a subsample of $N_* \ll N$ pixels (called “covariance sparse sampling,” according to Buckland *et al.*⁸). Speeding up the Mahalanobis distance computation is more challenging because it has to be done for *every* pixel. We will describe how we employ the the sparse matrix transform (SMT) of Cao and Bouman⁹ to obtain an accurate approximation with substantially less computational effort.

Microcontroller	ARM Cortex-M7
FPGA	A3P100L
Bus Width	32 bits
Clock Speed	216 MHz
Flash Program Memory	2 MB
Instruction/Data Cache	16 kB
Internal RAM	512 kB
External SRAM	8 MB
External SDRAM	512 MB
External Storage (Flash)	2-4 GB

Table 1. TACOS on-board processor includes a Field Programmable Gate Array (FPGA) and associated microcontroller. The 512MB SDRAM, which is accessed through the FPGA, is used to store the hyperspectral image cube (up to 2500 320×320 CCD frames, corresponding to a single hyperspectral image with 2500 × 320 pixels and 320 spectral channels). The 8MB SRAM is accessed directly by the microcontroller and is used for storing the covariance matrix and other intermediate computations (that require more space than can be fully allocated within the faster 512kB internal RAM).

2. DATA PROCESSING

This discussion begins with a review of various gas-phase plume detection algorithms, along with some of their computational requirements. First among these is the classic adaptive matched filter (AMF).^{10–12} The AMF provides a grayscale image, each pixel of which characterizes a measure of detectability of the plume at that pixel. A set of AMF images will be computed onboard, each image corresponding to a specific chemical species of interest.

For each set of matched filter images, we will also downlink a Mahalanobis distance (also called “global RX”) image.^{13,14} In addition to providing anomaly detection diagnostics, the RX image can be combined with the matched filter images to produce adaptive coherence estimator (ACE) images,^{15,16} as well as more specialized output images that are better adapted to the background statistics. These include the mixture-tuned matched filter,¹⁷ a false-alarm mitigating matched filter,¹⁸ a family of detectors built on an elliptically contoured multivariate-t model of background variability,¹⁹ and a machine learning approach based on matched pairs of real and simulated data.^{20–23}

2.1. Hyperspectral detection of chemical plumes

A satellite looks down on the ground, and does not directly observe a plume, but instead observes the effect of the plume on the spectral image of the ground. In the thermal infrared, there are both emissive and absorptive features, depending on the temperature of the plume and of the ground that is observed through the plume. But for the visible wavelengths employed on the TACOS mission, the effect is primarily absorptive. Particularly for a weak plume, the effect of the plume on the observed spectrum is subtle. But because there are so many distinct spectral channels, this subtle signal can nonetheless still be teased out from the cluttered background.

Although radiative transmission and absorption are fundamentally nonlinear physical processes, the role of the chemical plume in all of this, because we assume the plume is weak, can be treated as a linear additive signal-in-noise problem.^{8,24–26}

To correspond to the d spectral channels in the hyperspectral image, each pixel is associated with a vector $\mathbf{x} \in \mathbb{R}^d$ of radiance values (or reflectance values if the appropriate conversion has been made). Write $\mathbf{s} \in \mathbb{R}^d$ as the target vector; this is the spectral signature of the chemical we are interested in. The hypothesis test we make for each pixel is

$$H_0 : \mathbf{x} = \mathbf{z} \tag{1}$$

$$H_1 : \mathbf{x} = \mathbf{z} + \epsilon \mathbf{s} \tag{2}$$

where \mathbf{z} is the (non-target) background, which we take to be distributed as a Gaussian with mean $\boldsymbol{\mu}$ and covariance R , and ϵ is a measure of the strength of the plume. For this additive model, a likelihood ratio test

leads to the AMF detector¹⁰⁻¹²:

$$\text{AMF: } \mathcal{D}(\mathbf{x}) = \mathbf{s}^\top R^{-1}(\mathbf{x} - \boldsymbol{\mu}) / \sqrt{\mathbf{s}^\top R^{-1} \mathbf{s}} \quad (3)$$

where the denominator is a scalar constant that ensures that the AMF does not depend on the magnitude (only on the direction) of the target vector \mathbf{s} ; it also ensures that $\mathcal{D}(\mathbf{x})$ is dimensionless.

The mean $\boldsymbol{\mu}$ and covariance R are estimated from the observed data cube. Since the observed pixels include both target and background, there is a contamination effect,²⁷⁻³⁰ but for small and/or weak plumes, the effect is small, and in the onboard analysis, we do not attempt to account for it.

From the N pixels in the hyperspectral image cube, we compute mean $\boldsymbol{\mu} = \langle \mathbf{x} \rangle = (1/N) \sum_{n=1}^N \mathbf{x}_n$ with $O(Nd)$ effort (and most of those operations are additions; the number of multiplications is only $O(d)$). On the other hand, the covariance matrix computation $R = \langle (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^\top \rangle = (1/N) \sum_{n=1}^N (\mathbf{x}_n - \boldsymbol{\mu})(\mathbf{x}_n - \boldsymbol{\mu})^\top$ nominally requires $O(Nd^2)$ multiplications, which is a considerably greater computational expense. By taking a small subsample of $N_* \ll N$ pixels, however, this expense is reduced to $O(N_*d^2)$. For matched filter estimates, as long as $N_* \gg d$, this leads to an accurate approximation.¹⁰

Once the covariance matrix is computed, it can be inverted with $O(d^3)$ effort. We currently do this in two steps: first a Cholesky decomposition, followed by inversion of the Cholesky component. That is: $R = LL^\top$ where L is a lower triangular matrix; and then from L^{-1} we can obtain $R^{-1} = (L^{-1})^\top L^{-1}$. Because this is only done once per data cube, this is not a computationally expensive step, but it is a very numerically demanding step. Preliminary work was performed with double precision floating point, but more recent work with 32-bit floating point has identified numerical issues that may require some kind of automated pre-regularization.³¹

From the inverse covariance, the vector $\mathbf{q} = R^{-1}\mathbf{s}$ and scalar quantities $\mathbf{q}^\top \boldsymbol{\mu}$ and $\sqrt{\mathbf{s}^\top R^{-1} \mathbf{s}}$ can be readily computed. The computation of $\mathcal{D}(\mathbf{x})$ then amounts to a dot product ($\mathbf{q}^\top \mathbf{x}$) for each pixel: a total effort of $O(Nd)$ multiplications. Although this computation is not a bottleneck, it could in principle be reduced even further by using a sparse approximation to the matched filter.^{32,33}

An alternative, and actually somewhat more popular (and by most accounts, better³⁴) detector is given by the Adaptive Coherence (sometimes ‘‘Cosine’’) Estimator,^{15,16} and is related to the angle between \mathbf{s} and \mathbf{x} . Specifically:

$$\text{ACE: } \mathcal{D}(\mathbf{x}) = \frac{[\mathbf{s}^\top R^{-1}(\mathbf{x} - \boldsymbol{\mu})]^2}{[\mathbf{s}^\top R^{-1} \mathbf{s}] [(\mathbf{x} - \boldsymbol{\mu})^\top R^{-1}(\mathbf{x} - \boldsymbol{\mu})]} \quad (4)$$

The ‘‘one-sided’’ (aka ‘‘single cone’’) variant of ACE is given by the signed square root:

$$\text{ACE: } \mathcal{D}(\mathbf{x}) = \frac{\mathbf{s}^\top R^{-1}(\mathbf{x} - \boldsymbol{\mu})}{\sqrt{[\mathbf{s}^\top R^{-1} \mathbf{s}] [(\mathbf{x} - \boldsymbol{\mu})^\top R^{-1}(\mathbf{x} - \boldsymbol{\mu})]}} \quad (5)$$

This one-sided variant is sometimes called the normalized matched filter (NMF), but is usually just called ‘‘ACE’’.

A detector that is in some sense ‘‘in between’’ AMF and ACE is based on a generalized likelihood ratio test (GLRT) of an additive target against an elliptically-contoured background;^{35,36} this is the EC-GLRT detector:¹⁹

$$\text{EC-GLRT: } \mathcal{D}(\mathbf{x}) = \sqrt{\frac{(\nu - 1)}{(\nu - 2) + (\mathbf{x} - \boldsymbol{\mu})^\top R^{-1}(\mathbf{x} - \boldsymbol{\mu})}} \frac{\mathbf{s}^\top R^{-1}(\mathbf{x} - \boldsymbol{\mu})}{\sqrt{\mathbf{s}^\top R^{-1} \mathbf{s}}}. \quad (6)$$

Here, $\nu \rightarrow 2$ leads to ACE and $\nu \rightarrow \infty$ leads to AMF. For Gaussian data, $(\mathbf{x} - \boldsymbol{\mu})^\top R^{-1}(\mathbf{x} - \boldsymbol{\mu})$ has mean given by the dimension d of the data. Values of $\nu \approx d$ provide behavior that is ‘‘between’’ that of ACE and AMF. Further, since the equation is derived in the context of a multivariate- t distribution with ν degrees of freedom, one can estimate ν directly from the data; in particular, ν can be estimated from the distribution of Mahalanobis distances that are available in the RX image (*e.g.*, see the Appendix in Ref. [37]).

It would not be practical to compute all of these detectors – AMF, ACE, and EC-GLRT – directly onboard, but we enable the computation of ACE and EC-GLRT on the ground by downlinking one further image, the quadratic RX term, $(\mathbf{x} - \boldsymbol{\mu})^\top R^{-1}(\mathbf{x} - \boldsymbol{\mu})$, that appears in the expressions for ACE and EC-GLRT. And it bears

remarking that the RX term does *not* depend on the target spectrum \mathbf{s} . Thus, only one RX image needs to be computed and downlinked for each scene, no matter how many AMF images (one for each chemical species of interest) are computed and downlinked.

Unfortunately, the computation of the RX image is considerably heavier than the computation of the AMF images. For instance, using a Cholesky decomposition, and writing $R^{-1} = (L^{-1})^T L^{-1}$, we can express the Mahalanobis distance $(\mathbf{x} - \boldsymbol{\mu})^T R^{-1} (\mathbf{x} - \boldsymbol{\mu}) = (\mathbf{x} - \boldsymbol{\mu})^T (L^{-1})^T L^{-1} (\mathbf{x} - \boldsymbol{\mu}) = \|L^{-1}(\mathbf{x} - \boldsymbol{\mu})\|^2$. Since L is lower triangular, the matrix-vector multiplication takes only $d^2/2$ multiplications. But this is still an $O(d^2)$ calculation, and since it must be done for every pixel, the total computation is $O(Nd^2)$. In Section 3, we discuss ways to approximate the Mahalanobis distance at far less cost.

2.2. Matched-Filter-Residual

Foy *et al.*²⁰ argue for a general decomposition of the spectral output into two components, a (signed) matched-filter output component and a (non-negative) residual component.

$$\text{MF}(\mathbf{x}) = \mathbf{s}^T R^{-1} (\mathbf{x} - \boldsymbol{\mu}) / \sqrt{\mathbf{s}^T R^{-1} \mathbf{s}} \quad (7)$$

$$\text{R}(\mathbf{x}) = \sqrt{(\mathbf{x} - \boldsymbol{\mu})^T R^{-1} (\mathbf{x} - \boldsymbol{\mu}) - \text{MF}^2} \quad (8)$$

This two-dimensional space is called matched-filter-residual space, and in this space one can construct many of the classic detection algorithms, including AMF, ACE, and EC-GLRT, in addition to others. In particular, AMF is given by $\text{MF}(\mathbf{x})$, and ACE is monotonically equivalent to $\text{MF}(\mathbf{x})/\text{R}(\mathbf{x})$. By downlinking the matched filter and the Mahalanobis distance images, one can construct the matched-filter-residual decomposition in which those algorithms operate.

2.3. Solid target detection (an aside)

For solid targets, the additive model expressed in Eq. (2) is not representative of the underlying physics; the replacement model $\mathbf{x} = (1 - \alpha)\mathbf{z} + \alpha\mathbf{s}$ provides an alternative in which the target with signature \mathbf{s} occludes a fraction α of a pixel. Some detectors take the replacement model literally, and derive GLRT detectors for specific background distributions.^{38,39} Less formal approaches have a matched-filter “flavor” but exclude pixels that are far from the matched-filter signature even though they have a large projection in that direction.^{17,18} An adaptive residual approach recognizes the $(1 - \alpha)\mathbf{z}$ term as an indication that only a fraction $(1 - \alpha)$ of the mean background should be subtracted, with the value of α being adaptively estimated.⁴⁰ Other models are possible, in which the effect of a target can be quite arbitrary as long as it can be modeled; for instance, one can model exponential Beer’s law behavior in a plume without having to make the linear low-concentration approximation. This leads to a situation in which data is created in matched pairs,²¹⁻²³ and machine learning can be used to address the problem in terms of binary (target/no-target) classification.

Although the TACOS project mission does not include solid target detection, that expansion might be accommodated by including target signatures \mathbf{s} for the solids of interest. By also downlinking $\boldsymbol{\mu}^T \mathbf{x}$, this (in combination with the matched filter image) would enable the computation of terms like $(\mathbf{t} - \boldsymbol{\mu})^T (\mathbf{x} - \boldsymbol{\mu})$ that appear in some of the solid target algorithms. We remark that (like the Mahalanobis distance image), this mean dot product image $\boldsymbol{\mu}^T \mathbf{x}$ does not depend on target signature, so there would only be one such image per hyperspectral data cube, even for multiple target signatures.

3. FAST COMPUTATION OF RX

The computation of RX, or Mahalanobis distance, is nominally $O(Nd^2)$ and (unlike the case for the covariance matrix), that factor of N is inevitable, since we need RX for *every* pixel. The goal then is to reduce the $O(d^2)$ effort required for each pixel.

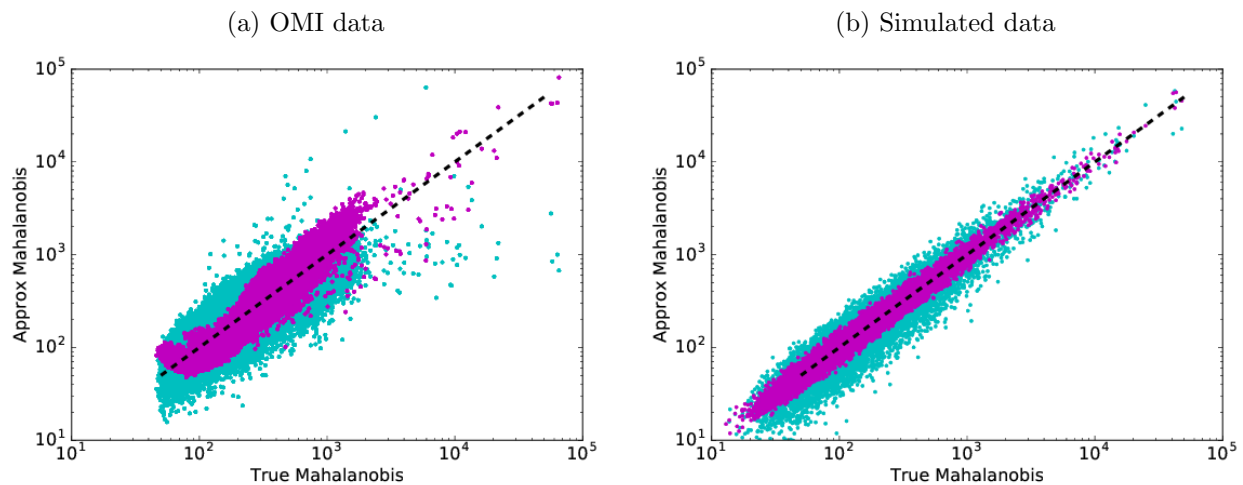


Figure 1. Scatter plots of true vs approximate Mahalanobis distances, with each point corresponding to a pixel. In these plots the lighter (cyan) points correspond to the subspace approximator (with $d_* = 15$), and the darker (magenta) points are for SMT (with $K = 2000$). See the caption of Table 2 for more details. (a) Pixels from an OMI hyperspectral image, spectrally resampled to match the TACOS instrument. (b) Simulated data sample drawn from a multivariate-t distribution, having $\nu = 3$ (very non-Gaussian) and a covariance that matches that of OMI data.

Approximator	OMI	Simulated
Diagonal	1.537	1.480
Subspace ($d_* = 15$)	0.319	0.291
SMT ($K = 2000$)	0.236	0.121
Shrinkage ($\epsilon = 0.00006$)	0.234	0.123

Table 2. Average absolute deviation of the log ratio of approximate to true Mahalanobis distances. For instance, the SMT approximates the Mahalanobis distance on OMI data with an error factor of $\exp(0.236) \approx 1.266$ which corresponds to a 27% error. By contrast, the subspace error factor is about 38% on the same data. Note that the parameters for subspace and SMT were chosen so both correspond to roughly the same computational effort, about 4700 multiplications per pixel (compared to $d^2/2$ for the standard computation, which is over 50,000 multiplications per pixel). Two approximators in this table that are not shown Fig. 1 are a simple diagonal approximator (only $2d = 640$ multiplications) and a shrinkage regularized estimator (same cost as standard computation), which is included to illustrate that the magnitude of error in the SMT approximation corresponds to what would be obtained if the covariance were regularized with a factor of $\epsilon = 0.00006$. The “true” Mahalanobis distances is computed with $\epsilon = 0$.

3.1. Subspace approximation

Consider a low-rank approximation to the covariance matrix given by $\hat{R} = UDU^T$, where U is an orthogonal matrix of size $d \times d_*$ for $d_* \ll d$, and D is a diagonal matrix of (smaller) size $d_* \times d_*$. By construction, \hat{R} is not invertible, but if we employ the pseudoinverse, and write $\hat{R}^{-1} = UD^{-1}U^T$, then we can use $\mathbf{x}^T \hat{R}^{-1} \mathbf{x} = \|D^{-1/2}U^T \mathbf{x}\|^2$ to approximate the Mahalanobis distance. On average, we expect the estimated distance to be a factor d_*/d of (*i.e.*, smaller than) the actual distance. Thus, we take our approximation to be

$$\frac{d}{d_*} \|D^{-1/2}U^T \mathbf{x}\|^2 \quad (9)$$

The matrix $D^{-1/2}U^T$ is of size $d_* \times d$ and can be computed just once, and then applied to all of the pixels, at a cost of d_*d per pixel. Taking the norm of the vector $D^{-1/2}U^T \mathbf{x}$ is a further cost of d_* per pixel, for a total cost of $Nd_*(d+1)$ multiplications. This is much cheaper than the nominal $O(Nd^2)$, but it is an approximation, and one that (even after the d/d_* correction) may have systematic biases.

For pixels that are anomalous because they are of unusually large magnitude (and have large components in the direction of the largest eigenvectors), they will have their anomalousness suppressed by the approximation in Eq. (9). By the same token, pixels that are anomalous because they have unusual direction (and have large components in the direction of the smallest eigenvectors) will have their anomalousness exaggerated.

Note that even though $D^{-1/2}U$ is a $d_* \times d$ matrix, the associated approximation to R^{-1} given by $U^T D^{-1} U$ is still a dense matrix. In the next section, we consider an approach that explicitly creates a sparse approximation to R^{-1} .

3.2. Graphical lasso

A natural idea for speeding up the computation of the RX term is to replace the matrix R^{-1} with an approximation that includes a lot of zero-valued elements. A scheme for doing just that was proposed by Friedman *et al.*,⁴¹ and a nice software package for python was identified (called `skggm`⁴²) for trying this idea out on our data. The approach is an L1-penalized log likelihood:

$$\hat{R}^{-1} = \operatorname{argmax}_{\Psi} \log \det(\Psi) - \operatorname{trace}(R\Psi) - \rho \|\Psi\|_1 \quad (10)$$

where the term $\|\Psi\|_1$ is the sum of the absolute values of the non-diagonal elements of Ψ . As $\rho \rightarrow 0$, the solution just becomes R^{-1} , an accurate but dense inverse. As ρ increases, the approximation becomes poorer but sparser. To be useful, there needs to be a range of values of ρ for which the approximation is good enough to give accurate estimates of RX, but at the same time sparse enough that there is an appreciable gain in computational efficiency in computing RX.

Preliminary studies indicate that there is no such range. In order to get a reasonably sparse solution, a large penalty term is required, and that distorts the accuracy of the approximation, as illustrated in Fig. 2.

We remark that even if we were able to obtain approximations to R^{-1} that were both sparse and accurate, we would also have to incorporate the fairly complicated glasso algorithm into the onboard software. That might be possible, but we found a much better solution with the sparse matrix transform (SMT) described below.

3.3. Sparse matrix transform (SMT)

We were led to the SMT⁹ because it has been employed for a variety of hyperspectral data analyses^{43–47}; in most of those studies, SMT is employed for its favorable regularizing properties, but SMT can also be used to speedup computation.^{48, 49}

Since the covariance matrix R is symmetric and non-negative definite, we can write it using a singular value decomposition: $R = EDE^T$, where E is orthogonal, and D is diagonal, with non-negative elements on the diagonal. We will approximate the covariance, using an approximation for E based on a product of Givens rotations.

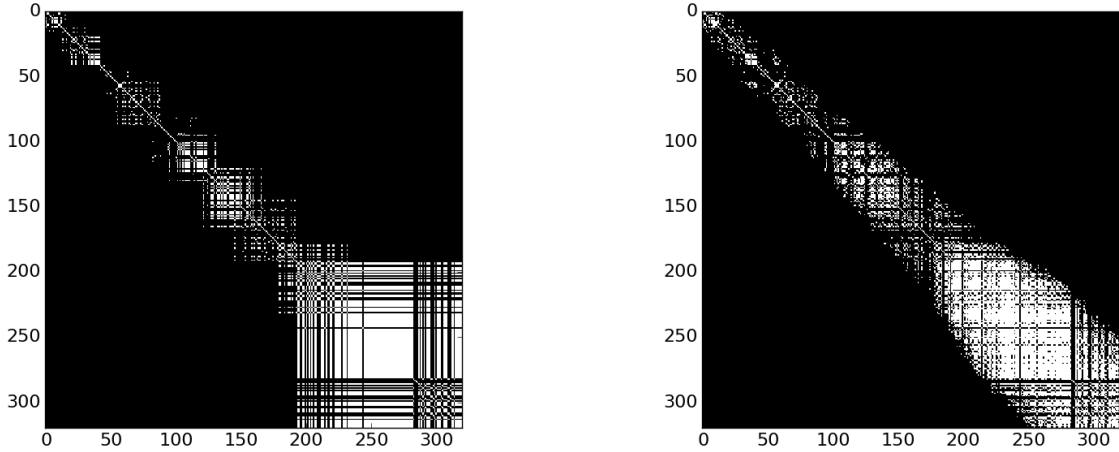


Figure 2. SMT (left) vs Glasso (right) – shown are the inverse covariance matrices with nonzero components in white, and zero-valued components in black. The SMT value for $K = 310$ has been adjusted so that both exhibit approximately the same level of “sparsity” (approximately 3×10^5 nonzero elements, or about 30% of the total). But the SMT is a better approximation, and is computed with many fewer computations ($\sim 10^3$ for SMT vs $\sim 10^5$ for Glasso).

A Givens rotation is a matrix of the form

$$G = \Theta(i, j, \theta) = \begin{bmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \cos \theta & \cdots & \sin \theta & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & -\sin \theta & \cdots & \cos \theta & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix}, \quad (11)$$

which is the identity matrix except for four elements, which define a rotation by an angle θ over two channels, i and j . The Givens rotation is an orthogonal matrix, and it can be shown that *any* orthogonal matrix can be written as a product of $d(d+1)/2$ Givens rotations. Our approach, however, is to use a much smaller product of Givens rotations to *approximate* the orthogonal matrix E .

In particular, write $E \approx \tilde{E} = G_1 G_2 \cdots G_K$ where K is the number of rotations in the approximation. If we define $D^* = \tilde{E}^T R \tilde{E}$, then we have $R = \tilde{E} D^* \tilde{E}^T$. If we choose our Givens rotations so that D^* is nearly diagonal (as shown in Algorithm 1), then $\tilde{D} = \text{diag}(D^*)$ will be a good approximation. Specifically, we write

$$\tilde{D} = \text{diag}(G_K^T \cdots G_2^T G_1^T R G_1 G_2 \cdots G_K). \quad (12)$$

from which

$$R \approx \tilde{R} = \tilde{E} \tilde{D} \tilde{E}^T. \quad (13)$$

Now, if we write $Q = \tilde{D}^{-1/2} G_K^T \cdots G_2^T G_1^T$, then it follows that $Q^T Q = \tilde{R}^{-1} \approx R^{-1}$. Hence, the RX expression is given by

$$\mathbf{x}^T R^{-1} \mathbf{x} \approx \mathbf{x}^T Q^T Q \mathbf{x} = \|\mathbf{Qx}\|^2. \quad (14)$$

And the important point here is that $\mathbf{Qx} = D^{-1/2} G_K^T \cdots G_2^T G_1^T \mathbf{x}$ is an operation that requires $O(K+d)$ operations. Computing the squared norm is another d operations, so the full RX calculation is $O(K+d)$ operations per pixel, for a total of $O(N(K+d))$ for the full image cube. We find from experience (*e.g.*, in Ref. [43]) that taking K to be factor of a few times d is often adequate for obtaining a good approximation to R . That leads to computation of Mahalanobis distance over the full image in $O(Nd)$ time.

Algorithm 1 Find Givens rotations that make up the Sparse Matrix Transform (SMT)

Require: Sample covariance matrix R (will be overwritten)

Require: Number of rotations K

Initialize F so that $F_{ij} = R_{ij}^2 / (R_{ii}R_{jj})$

▷ Magnitude of off-diagonal elements

for $k = 1, \dots, K$ **do**

 Let $(i, j) = \operatorname{argmax}_{ij} F_{ij}$

 ▷ Nominally takes $O(d^2)$, but can be made $O(d)$

 Let $\theta = \frac{1}{2} \operatorname{atan}(-2R_{ij}, R_{ii} - R_{jj})$

 ▷ Angle of rotation

 Let $G_k = \Theta(i, j, \theta)$

 ▷ Satisfies $G_k = \operatorname{argmin}_G \| \operatorname{diag}(G^T R G) \|$

 Update $R \leftarrow G_k^T R G_k$

 ▷ R is becoming progressively more diagonal

 Update F so that $F_{ij} = R_{ij}^2 / (R_{ii}R_{jj})$

 ▷ Note only two rows and two columns need updating

end for

Return Givens matrices G_1, \dots, G_K

Return Diagonal matrix $D = \operatorname{diag}(R)$

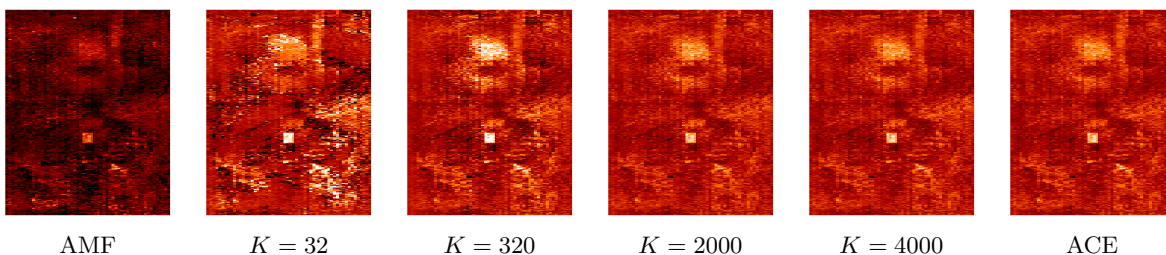


Figure 3. Detection maps for AMF (far left panel) and for ACE based on true Mahalanobis distance (far right panel), and for ACE based on approximate Mahalanobis distance using SMT for various values of K . This is a 320×400 tile from a larger image, obtained from the OMI data archive. The square just below the center of the image is a 20×20 implanted NO_2 plume.

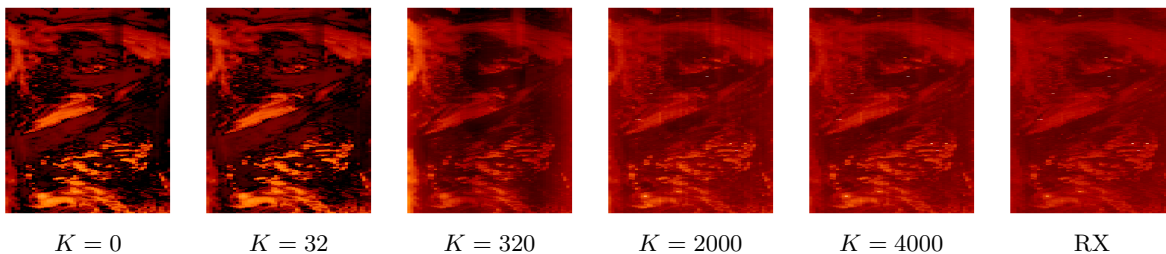


Figure 4. Mahalanobis distance images for the same tiles seen in Fig. 3. Far left panel is RX estimated only using the diagonal elements of the covariance matrix. Far right panel is “true” RX.

Detector	Sigmas	Detector	Sigmas	Detector	Sigmas
AMF	2.963	ACE/SMT $K = 16$	2.783	ACE/SMT $K = 640$	2.723
ACE/SMT $K = 0$	2.045	ACE/SMT $K = 32$	2.934	ACE/SMT $K = 2000$	3.192
ACE/SMT $K = 2$	2.284	ACE/SMT $K = 64$	3.226	ACE/SMT $K = 4000$	3.314
ACE/SMT $K = 4$	2.368	ACE/SMT $K = 160$	3.659	ACE/SMT $K = 16000$	3.343
ACE/SMT $K = 8$	2.566	ACE/SMT $K = 320$	3.679	ACE	3.336

Table 3. Detectability of implanted target, in “sigmas,” defined as the average of the pixels in the implanted target (20×20 pixels) minus the average of the pixels over the whole scene (320×1444 pixels), divided by the standard deviation of the pixels over the whole scene. As K increases, the detectability approaches that of the full ACE detector.

3.3.1. Adequacy of approximation

We performed some experiments using publicly available data obtained from NASA’s OMI mission.¹ Two hyperspectral images were identified, one with an NO_2 plume,⁵⁰ and one with an SO_2 plume.⁵¹ The data were spectrally resampled to match the wavelengths in the TACOS sensor.

In Fig. 1, we compare various Mahalanobis distance estimators, and observe that the SMT estimator provides the best approximation for the least computational effort. Fig. 3 and Fig. 4 provide illustrative images of detection maps, from a data cube that contains an implanted plume.

Table 3 shows that the performance of ACE/SMT (the ACE detector using an SMT-approximated RX as the denominator instead of the true RX) approaches that of ACE as K increases. This is not a comprehensive evaluation of performance, and details will vary depending on where in the scene we implant the target. As it turns out, for this example, the “optimal” SMT detector has $K = 320$, but that value is idiosyncratic. It is not even clear that ACE is the optimal detector (there is a whole family of detectors, the EC-GLRT family, which provides a continuum along which AMF and ACE are the two endpoints).

3.3.2. Computational efficiency

Table 4 describes the computational expense, directly in seconds of wall clock time on our onboard processor, for two approaches to creating the RX and AMF images for a single gas species. The standard/nominal approach requires over five hours to obtain images of size $N = 2000 \times 320$ pixels from data with $d = 320$ spectral channels. The largest bottlenecks are the covariance matrix computation and the Mahalanobis distance computation. In the optimized/approximate approach, we substantially reduce both of the bottleneck times, and obtain a pair of images in less than half an hour. This is for a single chemical species; since we only need to compute the Mahalanobis once, we can add multiple species at a cost of six and a half minutes each.

4. CONCLUSION

In adapting gas-phase chemical detection algorithms to a Cubesat platform, two innovations are suggested. The first is to limit the bandwidth requirements by downlinking only the matched filter images for each gas of interest, and to augment that with a single Mahalanobis distance map, thereby enabling a large suite of possible detection algorithms to be employed after the fact. The second is to make the Mahalanobis distance computation more efficient by employing an approximation based on the sparse matrix transform.

By also downlinking a sample of pixels, some of which are chosen from points with high matched filter values (or high ACE values), and some of which are chosen more randomly from the background, we may be able to perform more extensive processing, to for instance achieve quantification as well as detection, as is currently done with instruments such as OMI,^{52, 53} which have the twin luxuries of downlinking the entire hyperspectral image cube, and performing all the processing on ground-based computers.

5. ACKNOWLEDGMENTS

The research described in this paper was supported by the U.S. Department of Energy (DOE) National Nuclear Security Administration (NNSA) through the Los Alamos Laboratory Directed Research and Development Directed Research (LDRD/DR) program. We are grateful to our colleagues on the TACOS team for their helpful suggestions and insightful questions. We also acknowledge Kevin Mertes for useful discussions.

(a) Standard

Operation	Execution time H:MM:SS		
	$N = 320 \times 2000$	$N = 320 \times 320$	Big-O
Mean vector (μ) calculation	0:01:26	0:00:14	$O(Nd)$
Covariance matrix (R) calculation	3:27:20	0:33:10	$O(Nd^2)$
Cholesky decomposition of R	0:00:07	0:00:07	$O(d^3)$
Covariance matrix inversion	0:00:04	0:00:04	$O(d^3)$
Mahalanobis distance (No SMT)	1:54:40	0:18:21	$O(Nd^2)$
AMF calculation (per chemical)	0:06:30	0:01:02	$O(Nd)$
Total Time (one chemical)	5:30:07	0:52:59	

(b) Optimized

Operation	$K = 2000, N_* = N/100$		
	$N = 320 \times 2000$	$N = 320 \times 320$	Big-O
Mean vector (μ) calculation	0:01:26	0:00:14	$O(Nd)$
Covariance matrix (R) calculation	0:02:04	0:00:20	$O(N_*d^2)$
SMT Decomposition [†] of R	0:00:41	0:00:41	$O(Kd^2)$
Mahalanobis distance (SMT)	0:13:57	0:02:14	$O(NK) + O(Nd)$
Cholesky decomposition of R	0:00:07	0:00:07	$O(d^3)$
Covariance matrix inversion	0:00:04	0:00:04	$O(d^3)$
AMF calculation (per chemical)	0:06:30	0:01:02	$O(Nd)$
Total Time (one chemical)	0:24:50	0:04:42	

Table 4. Timing studies for the onboard processor, using (a) the standard algorithm, and (b) with the proposed speed-ups for computation of the covariance matrix and the Mahalanobis distance image. [†]We remark that the SMT decomposition can be computed with a straightforward algorithm⁹ that requires $O(Kd^2)$ effort; more sophisticated approaches⁴⁸ can reduce that to $O(Kd) + O(d^2)$, but with more bookkeeping. Since the SMT decomposition algorithm is not a bottleneck operation, we have implemented the straightforward algorithm.

REFERENCES

1. “NASA-OMI.” [Online] https://www.nasa.gov/mission_pages/aura/spacecraft/omi.html.
2. Love, S. P., Goff, F., Counce, D., Siebe, C., and Delgado, H., “Passive infrared spectroscopy of the eruption plume at Popocatepetl volcano, Mexico,” *Nature* **396**, 563–567 (1998).
3. Love, S. P., Goff, F., Schmidt, S. C., Counce, D., Pettit, D., Christenson, B. W., and Siebe, C., “Passive infrared spectroscopic remote sensing of volcanic gases: Ground-based studies at White Island and Ruapehu, New Zealand, and Popocatepetl, Mexico,” in [*Remote Sensing of Active Volcanism*], Mouginis-Mark, P., Crisp, J., and Fink, J., eds., *Geophysical Monograph* **116**, 117–138, American Geophysical Union, Washington, D.C. (2000).
4. Love, S. P., Hale, T. C., Jolin, L. J., Barefield, J., Atkins, W., and Tiee, J., “The ORCAS compact long-wave infrared hyperspectral sensor,” in [*Proc. Military Sensing Symposia (MSS) Specialty Group on Passive Sensors*], (2008).
5. The reported quantities are for uncompressed data. Lossless onboard compression could potentially reduce downlink requirements by a substantial factor.
6. Brett, C. J. C., DiPietro, R. S., Manolakis, D. G., and Ingle, V. K., “Efficient implementations of hyperspectral chemical-detection algorithms,” *Proc. SPIE* **8897**, 88970T (2013).
7. Molero, J. M., Garzón, E. M., García, I., and Plaza, A., “Analysis and optimizations of global and local versions of the RX algorithm for anomaly detection in hyperspectral data,” *IEEE J. Selected Topics in Applied Earth Observations and Remote Sensing* **6**, 801–814 (2013).
8. Buckland, K. N., Young, S. J., Keim, E. R., Johnson, B. R., Johnson, P. D., and Tratt, D. M., “Tracking and quantification of gaseous chemical plumes from anthropogenic emission sources within the Los Angeles basin,” *Remote Sensing of Environment* **201**, 275–296 (2017).

9. Cao, G. and Bouman, C. A., "Covariance estimation for high dimensional data vectors using the sparse matrix transform," in [*Advances in Neural Information Processing Systems*], **21**, 225–232, MIT Press (2009).
10. Reed, I. S., Mallett, J. D., and Brennan, L. E., "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerospace and Electronic Systems* **10**, 853–863 (1974).
11. Kelly, E. J., "Performance of an adaptive detection algorithm: rejection of unwanted signals," *IEEE Trans. Aerospace and Electronic Systems* **25**, 122–133 (1989).
12. Robey, F. C., Fuhrmann, D. R., Kelly, E. J., and Nitzberg, R., "A CFAR adaptive matched filter detector," *IEEE Trans. Aerospace and Electronic Systems* **28**, 208–216 (1992).
13. Mahalanobis, P. C., "On the generalised distance in statistics," *Proc. National Institute of Sciences of India* **2**, 49–55 (1936).
14. Reed, I. S. and Yu, X., "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoustics, Speech, and Signal Processing* **38**, 1760–1770 (1990).
15. Scharf, L. L. and McWhorter, L. T., "Adaptive matched subspace detectors and adaptive coherence estimators," in [*Proc. Asilomar Conference on Signals, Systems, and Computers*], (1996).
16. Kraut, S., Scharf, L. L., and Butler, R. W., "The Adaptive Coherence Estimator: a uniformly most-powerful-invariant adaptive detection statistic," *IEEE Trans. Signal Processing* **53**, 427–438 (2005).
17. Boardman, J. W. and Kruse, F. A., "Analysis of imaging spectrometer data using N -dimensional geometry and a mixture-tuned matched filtering approach," *IEEE Trans. Geoscience and Remote Sensing* **49**, 4138–4152 (2011).
18. DiPietro, R. S., Manolakis, D. G., Lockwood, R. B., Cooley, T., and Jacobson, J., "Hyperspectral matched filter with false-alarm mitigation," *Optical Engineering* **51**, 016202 (2012).
19. Theiler, J. and Foy, B. R., "EC-GLRT: Detecting weak plumes in non-Gaussian hyperspectral clutter using an elliptically-contoured generalized likelihood ratio test," in [*Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*], I:221 (2008).
20. Foy, B. R., Theiler, J., and Fraser, A. M., "Decision boundaries in two dimensions for target detection in hyperspectral imagery," *Optics Express* **17**, 17391–17411 (2009).
21. Theiler, J., "Matched-pair machine learning," *Technometrics* **55**, 536–547 (2013).
22. Theiler, J., "Transductive and matched-pair machine learning for difficult target detection problems," *Proc. SPIE* **9088**, 90880E (2014).
23. Ziemann, A., Kucer, M., and Theiler, J., "A machine learning approach to hyperspectral detection of solid targets," *Proc. SPIE* **10644** (2018). In these Proceedings.
24. Hayden, A., Niple, E., and Boyce, B., "Determination of trace-gas amounts in plumes by the use of orthogonal digital filtering of thermal-emission spectra," *Applied Optics* **35**, 2802–2809 (1996).
25. Theiler, J., Foy, B. R., and Fraser, A. M., "Characterizing non-Gaussian clutter and detecting weak gaseous plumes in hyperspectral imagery," *Proc. SPIE* **5806**, 182–193 (2005).
26. Manolakis, D. G. and D'Amico, F. M., "A taxonomy of algorithms for chemical vapor detection with hyperspectral imaging spectroscopy," *Proc. SPIE* **5795**, 125–133 (2005).
27. Theiler, J. and Foy, B. R., "Effect of signal contamination in matched-filter detection of the signal on a cluttered background," *IEEE Geoscience and Remote Sensing Letters* **3**, 98–102 (2006).
28. Theiler, J., Foy, B. R., and Fraser, A. M., "Nonlinear signal contamination effects for gaseous plume detection in hyperspectral imagery," *Proc. SPIE* **6233**, 62331U (2006).
29. Matteoli, S., Diani, M., and Corsini, G., "Effects of signal contamination in RX detection of local hyperspectral anomalies," in [*Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*], 4845–4848 (2012).
30. Matteoli, S., Diani, M., and Corsini, G., "Impact of signal contamination on the adaptive detection performance of local hyperspectral anomalies," *IEEE Trans. Geoscience and Remote Sensing* **52**, 1948–1968 (2014).
31. Theiler, J., "The incredible shrinking covariance estimator," *Proc. SPIE* **8391**, 83910P (2012).
32. Theiler, J. and Glocer, K., "Sparse linear filters for detection and classification in hyperspectral imagery," *Proc. SPIE* **6233**, 62330H (2006).

33. Minet, J., Taboury, J., Goudail, F., Péalat, M., Roux, N., Lonnoy, J., and Ferrec, Y., “Influence of band selection and target estimation error on the performance of the matched filter in hyperspectral imaging,” *Applied Optics* **50**, 4276–4285 (2011).
34. 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 (2013).
35. Manolakis, D., Marden, D., Kerekes, J., and Shaw, G., “On the statistics of hyperspectral imaging data,” *Proc. SPIE* **4381**, 308–316 (2001).
36. Marden, D. B. and Manolakis, D., “Using elliptically contoured distributions to model hyperspectral imaging data and generate statistically similar synthetic data,” *Proc. SPIE* **5425**, 558–572 (2004).
37. Theiler, J., Scovel, C., Wohlberg, B., and Foy, B. R., “Elliptically-contoured distributions for anomalous change detection in hyperspectral imagery,” *IEEE Geoscience and Remote Sensing Letters* **7**, 271–275 (2010).
38. Schaum, A. and Stocker, A., “Spectrally selective target detection,” in [*Proc. ISSSR (International Symposium on Spectral Sensing Research)*], 23 (1997).
39. Theiler, J., Zimmer, B., and Ziemann, A., “Closed-form detector for solid sub-pixel targets in multivariate t-distributed background clutter,” in [*Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*], (2018). Accepted.
40. Theiler, J. and Ziemann, A., “Local background estimation and the replacement target model,” *Proc. SPIE* **10198**, 101980V (2017).
41. Friedman, J., Hastie, T., and Tibshirani, R., “Sparse inverse covariance estimation with the graphical lasso,” *Biostatistics* **9**, 432–441 (2007).
42. “Gaussian graphical models with skggm.” [Online] <https://skggm.github.io/skggm/tour>.
43. Theiler, J., Cao, G., Bachega, L. R., and Bouman, C. A., “Sparse matrix transform for hyperspectral image processing,” *IEEE J. Selected Topics in Signal Processing* **5**, 424–437 (2011).
44. Bachega, L., Theiler, J., and Bouman, C. A., “Evaluating and improving local hyperspectral anomaly detectors,” in [*Proc. 40th IEEE Applied Imagery and Pattern Recognition (AIPR) Workshop*], (2011).
45. Gorelik, N., Blumberg, D., Rotman, S., and Borghys, D., “Target detection using nonsingular approximations for a singular covariance matrix,” *J. Electrical and Computer Engineering* **2012**, 628479 (2012).
46. Velasco-Forero, S., Chen, M., Goh, A., and Pang, S. K., “Comparative analysis of covariance matrix estimation for anomaly detection in hyperspectral images,” *IEEE J. Selected Topics in Signal Processing* **9**, 1061–1073 (2015).
47. Peng, J. and Luo, T., “Sparse matrix transform-based linear discriminant analysis for hyperspectral image classification,” *Signal, Image and Video Processing* **10**, 761–768 (2016).
48. Bachega, L. R., Cao, G., and Bouman, C. A., “Fast signal analysis and decomposition on graphs using the sparse matrix transform,” in [*Proc. ICASSP*], 5426–5429 (2010).
49. Theiler, J., Cao, G., and Bouman, C. A., “Sparse matrix transform for fast projection to reduced dimension,” in [*Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*], 4362–4365 (2010).
50. Mebust, A. K., Russell, A. R., Hudman, R. C., Valin, L. C., , and Cohen, R. C., “Characterization of wildfire NO_x emissions using MODIS fire radiative power and OMI tropospheric NO₂ columns,” *Atmospheric Chemistry and Physics* **11**, 5839–5851 (2011).
51. Yang, K., Krotkov, N. A., Krueger, A. J., Carn, S. A., Bhartia, P. K., and Levelt, P. F., “Retrieval of large volcanic SO₂ columns from the Aura Ozone Monitoring Instrument: Comparison and limitations,” *J. Geophysical Research* **112**, D24S43 (2007).
52. Liu, X., Bhartia, P., Chance, K., Spurr, R. J. D., and Kurosu, T. P., “Ozone profile retrievals from the Ozone Monitoring Instrument,” *Atmospheric Chemistry and Physics* **10**, 2521–2537 (2010).
53. Jindal, P., Shukla, M. V., Sharma, S. K., and Thapliyal, P. K., “Retrieval of ozone profiles from geostationary infrared sounder observations using principal component analysis,” *Q. J. Royal Meteorological Society* **142**, 3015–3025 (2016).