

# Algorithm development with on-board and ground-based components for hyperspectral gas detection from small satellites

James Theiler and Steven P. Love

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

## ABSTRACT

Los Alamos is currently working toward demonstrating Cubesat-based hyperspectral detection of gas-phase chemical plumes, a goal initially pursued in the internally funded Targeted Atmospheric Chemistry Observations from Space (TACOS) project, and now advancing toward space deployment with the NASA-funded Nanosat Atmospheric Chemistry Hyperspectral Observation System (NACHOS). This paper will present a general overview of these projects. Bandwidth considerations prevent full datacube downloads, and so processing algorithms include an on-board processing component to provide matched-filter and RX images for the gases of interest. The downlinked data will additionally include the full spectrum for a small sample of pixels, and one of the challenges for ground-based analysis will be to incorporate these different but incomplete “views” of the datacube into a more physical interpretation/analysis of the downlinked data.

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

## 1. INTRODUCTION

Hyperspectral remote sensing from space provides one way to monitor what is happening on, and what we collectively are doing to, our venerable planet. NASA’s Ozone Monitoring Instrument (OMI)<sup>1</sup> observes the whole earth in the ultraviolet and visible wavelengths (270–500nm), which permits the detection and quantitative characterization of various gas species in the atmosphere. OMI provides an invaluable service, but it is a large and expensive satellite, and it achieves its global coverage by using fairly coarse (12km) pixels.

In the Targeted Atmospheric Chemistry Observations from Space (TACOS) and Nanosat Atmospheric Chemistry Hyperspectral Observation System (NACHOS) projects, we have similar aims. An important difference is that a much smaller (0.4km) pixel size is used, thereby enabling more sensitive detection of smaller sources, such as individual power plants<sup>2,3</sup> or low-level volcanic emissions.<sup>4,5</sup> The main difference, though, is the size of the satellite; where OMI weighs 65kg for the instrument alone, the TACOS/NACHOS projects build on a heritage of compact hyperspectral sensors,<sup>6</sup> and will weigh about 4kg, including both the instrument and the satellite bus. The spectral range of the TACOS/NACHOS imager (300–500nm) is only slightly narrower than OMI, but the resolution (320 channels, corresponding to 0.6nm per channel) is a little higher. The main driver for this is the need to resolve the narrow lines in the spectra of both NO<sub>2</sub> and SO<sub>2</sub>. The CCD focal plane is 320×320, which means that in pushbroom mode, the hyperspectral image will be 320 pixels wide, and may be up to 2000 pixels in the along-track direction. This works out to about 400 Megabytes per image, which, because our satellite is cheap and small, is beyond the downlink capacity for a single overpass (approximately 14MB).

In a previous report,<sup>7</sup> we described a strategy for limiting the data downlink using onboard computation. Instead of downlinking the full data cube, we will compute matched filter outputs for each of the target gases of interest (in addition to NO<sub>2</sub> and SO<sub>2</sub>, we are also interested in tropospheric ozone, formaldehyde, and volcanic BrO)\*. This computation is performed onboard and the downlinked output images are only one band each,

---

\*Of course we are also very interested in CO<sub>2</sub>, but we use NO<sub>2</sub> as a “proxy” gas, since it has a shorter timescale in the atmosphere (on the order of a day, depending on the season, compared to decades and even centuries for CO<sub>2</sub>), which leads to its more nonuniform dispersal in the atmosphere, and this makes it a better target for a narrow field-of-view sensor. Furthermore, the atmospheric background for NO<sub>2</sub> is much smaller than that for CO<sub>2</sub>, as its spectral features lie

compared to the 320 bands of the full datacube. We also downlink an auxiliary Mahalanobis distance band; this has potential value for anomaly detection but it can also be combined with the matched filter output bands to produce outputs associated with a variety of target detectors (adaptive matched filter,<sup>8–10</sup> adaptive coherence estimator,<sup>11,12</sup> and many others<sup>13–18</sup>) as well as the RX anomaly detector.<sup>19,20</sup> The onboard computation of these quantities was facilitated by two approximations: covariance sparse sampling<sup>21</sup> and sparse matrix transform.<sup>22,23</sup>

In addition to the matched filter and Mahalanobis distance band, we will also downlink a small sample of full-spectrum pixels. Some of these pixels will be chosen at random, as background samples, but we will also downlink the pixels with highest matched filter values.

Finally, we will consider downlinking a few more single-band images: a matched filter output treating the mean spectrum as the target signature provides a value that further extends our options for detection algorithms, particularly for solid targets;<sup>24–26</sup> the first (or first few) principal component bands provides valuable information for reconstructing an approximate full-spectrum hyperspectral data cube.

## 2. PHYSICS

The physics of radiative transmission in the ultraviolet and visible is in some ways simpler than the radiative transmission in the infrared, because the radiation is not directly emitted by the materials in the scene, but reflected directly or indirectly from the sun. The equation

$$L = L_g \exp(-\epsilon \mathbf{b}) + L_s \quad (1)$$

describes the radiance  $L$  observed at the sensor, as a function of the ground radiance  $L_g$ , the scattered radiance  $L_s$ , the plume absorption spectrum  $\mathbf{b}$  and the plume strength  $\epsilon$ . The atmosphere also absorbs some radiation, but  $L_g$  and  $L_s$  incorporate this absorption and correspond to the at-sensor radiances due to ground reflection and scattering. Off plume (with  $\epsilon = 0$ ), the radiance is given by  $L_g + L_s$ .

Although  $L_s$  will vary from scene to scene, depending as it does on aerosol content in the atmosphere, it should be relatively constant over the pixels in a given image. To the extent that it can be estimated, it can be combined with the observed  $L$  to produce a value  $L - L_s$  that varies multiplicatively with plume strength  $\epsilon$  and absorption spectrum  $\mathbf{b}$ .

We have spoken here of radiances, but it is possible to work in reflectance space as well. For instance, in this case  $L_g$  would characterize the reflectance of the ground, diminished by the atmospheric absorption from ground to sensor, the so-called *at-sensor reflectance*. But since the actual sensor measurement is a radiance, and since precise calibration of ground reflectance is not a goal of the project (the ground is the nuisance, the plume is the target), we will work in radiance space.

## 3. ALGORITHMS

### 3.1. Matched filters

A popular algorithm for plume detection is the adaptive matched filter. Given a target signature  $\mathbf{t}$ , and a pixel value (measured radiance)  $\mathbf{x}$  whose off-plume statistics are given by covariance matrix  $R$ , we write  $\mathbf{q} = R^{-1}\mathbf{t}$  as the matched filter vector, and  $\mathbf{q}^T(\mathbf{x} - \boldsymbol{\mu})$  as the matched filter output. Here,  $\boldsymbol{\mu}$  is the mean value of  $\mathbf{x}$  over the scene.

Although the matched filter is a popular, and often surprisingly effective, detector, the assumptions behind the derivation of the matched filter do not really match the physical model we have for an absorptive plume. The matched filter is associated with an additive model, and Beers-Law absorption is fundamentally multiplicative.

For an absorption spectrum  $\mathbf{b}$ , and a measure of plume strength (basically, column density)  $\epsilon$ , the effect of plume on a background spectrum  $\mathbf{z}$  is given by  $\exp(-\epsilon \mathbf{b})\mathbf{z}$ , and for a weak plume this can be approximated with

---

in an otherwise relatively transparent and uncluttered region of the electromagnetic spectrum. We are also interested in characterizing aerosols, and our wavelength range is appropriate for distinguishing dark, absorbing, sooty aerosols from primarily scattering aerosols.

$\mathbf{z} - \epsilon \mathbf{b} \cdot \mathbf{z}$ , where the operator ‘ $\cdot$ ’ corresponds to element-wise multiplication. Neglecting atmospheric scattering, we can express plume detection in a hypothesis testing framework:

$$H_o : \mathbf{x} = \mathbf{z} \quad (2)$$

$$H_1 : \mathbf{x} = \mathbf{z} - \epsilon \mathbf{b} \cdot \mathbf{z}, \quad (3)$$

where  $\mathbf{x}$  is the measured spectral radiance, and  $\mathbf{z}$  is the off-plume (or background) radiance. The “additive” term  $\epsilon \mathbf{b} \cdot \mathbf{z}$  is not a constant additive term, since it depends on  $\mathbf{z}$ , but we can approximate  $\mathbf{z}$  with a constant. Since gas-phase chemical plumes tend to have spectra with sharp and distinctive features whereas the solid-phase background tends to have reflectances that vary smoothly with wavelength, a simple choice is to treat  $\mathbf{z}$  as a constant flat spectrum. This scalar constant can be incorporated into  $\epsilon$  so the additive term becomes  $\epsilon \mathbf{b}$ . Thus, the matched filter uses the gas absorption spectrum  $\mathbf{b}$  as the target signature.

We can do a little better than this, with an approximation that is arguably more natural, and that does not make assertions about the shape of the background spectrum. Here, we approximate  $\mathbf{z}$  with its mean  $\boldsymbol{\mu}$ . In this case, we have a matched filter in which the target spectrum is given by  $\mathbf{t} = \mathbf{b} \cdot \boldsymbol{\mu}$ . A potential disadvantage of this approach is that the target signature depends on the data.

To the extent that the plume truly is purely absorptive, a nominally more accurate approach would be to work with the logarithm of the radiance. In this case the multiplicative absorption becomes additive and the matched filter target is the absorption spectrum. Another theoretical advantage is that the small  $\epsilon$  assumption is not necessary. We have not extensively investigated this approach, but are concerned that the additive effect of scattering will lead to problematic results.

### 3.2. Stripes

One thing we have observed with OMI data is the visual appearance of (often subtle) stripes in the along-track direction. These are presumably due to small differences in gain and offset of the individual pixels on the focal plane.

To first order, we can adjust for these small differences by computing a separate mean  $\boldsymbol{\mu}$  for each stripe. Thus, for a pixel in the  $k$ ’th stripe, we can write  $\mathbf{q}^T(\mathbf{x} - \boldsymbol{\mu}_k)$  where  $\boldsymbol{\mu}_k$  is the mean associated with the  $k$ ’th stripe. This can improve the sensitivity of the matched filter, but it is not always advised. (For example, if there is a very bright object in the scene, a glint perhaps, the stripe it appears in will have an unrealistically large  $\boldsymbol{\mu}$ .)

Fortunately, this stripe correction can be carried out offline, on the ground, after the data has been downlinked. If  $\boldsymbol{\mu}$ , the global mean, is used in the onboard processing, and  $\boldsymbol{\mu}_k$  is the mean associated with the  $k$ ’th stripe, then the corrected matched filter can be written

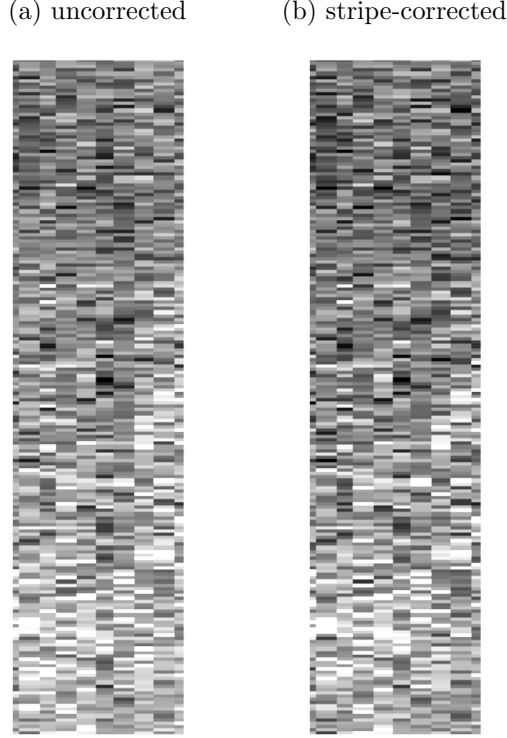
$$\mathbf{q}^T(\mathbf{x} - \boldsymbol{\mu}_k) = \mathbf{q}^T(\mathbf{x} - \boldsymbol{\mu}) - \mathbf{q}^T(\boldsymbol{\mu}_k - \boldsymbol{\mu}). \quad (4)$$

Here the first term is the matched filter output that is downlinked, and the second term is equal to the average of this matched filter value, with the average taken over all the pixels in the  $k$ ’th stripe. In other words, this is a scalar quantity that can be computed from the downlinked data. Further, if there is a bright glint and the user judges that a more robust average be taken by masking out the glint, then this is a decision that can be made on the ground, long after the data acquisition and downlink. Fig. 1 illustrates the effect of stripe correction in a small patch of the image. Although the visual difference is subtle, the results of numerical experiments shown in Table 1 indicate a numerical improvement in detection performance when stripe correction is applied.

### 3.3. Reconstruction

*For what shall it profit a man if  
he shall gain the whole spectrum?*

In this section we consider the question of reconstructing the full spectrum given the limited information that is downlinked for the bulk of the pixels in the hyperspectral image. For the primary objectives of the mission (eg, plume detection), we need to ensure that the specific information we need is downlinked, and our algorithms



**Figure 1.** Stripe correction in Eq. (4) reduces the stripe artifacts in the AMF image.

should make use of that information directly. But for secondary applications that may arise after the acquisition of the data, it may be useful to have a full datacube, even if it is only estimated.

Let  $\mathbf{x} \in \mathbb{R}^d$  be the full  $d$ -component spectrum at a pixel. Let  $\mathbf{t}_1, \dots, \mathbf{t}_k$  be the  $k$  target spectra whose matched filters will be downlinked.<sup>†</sup> We write  $T = [\mathbf{t}_1, \dots, \mathbf{t}_k]$  as the  $d \times k$  matrix of target spectra, so then  $\mathbf{u} = T^T R^{-1}(\mathbf{x} - \boldsymbol{\mu})$  is vector of  $k$  components that is downlinked for each pixel. (Typically,  $k \ll d$ ; otherwise, we'd just downlink the full spectrum.) We will write  $v^2 = (\mathbf{x} - \boldsymbol{\mu})^T R^{-1}(\mathbf{x} - \boldsymbol{\mu})$  as the Mahalanobis magnitude of the pixel  $\mathbf{x}$ . The aim is to reconstruct  $\mathbf{x}$  given only the measurements  $\mathbf{u}$  and  $v^2$ .

We will consider two approaches: a direct deductive approach that explicitly enforces the constraints, and a data-driven approach that uses machine learning to infer  $\mathbf{x}$  as a function of  $\mathbf{u}$  and  $v^2$ .

### 3.3.1. Direct deductive approach

We do not have enough information to fully reconstruct  $\mathbf{x}$  from the measurements  $\mathbf{u}$  and  $v^2$ , so our reconstruction will have two steps. The first step is to identify the space of possible  $\mathbf{x}$  values that are consistent with the measurements, and the second step is to choose a specific  $\mathbf{x}$  from that set.

To begin, we will transform to whitened coordinates

$$\tilde{\mathbf{x}} = R^{-1/2}(\mathbf{x} - \boldsymbol{\mu}) \tag{5}$$

$$\tilde{T} = R^{-1/2}T \tag{6}$$

---

<sup>†</sup>In addition to the specific targets of interest, we might also include in this list  $\mathbf{t} = \boldsymbol{\mu}$  (so as to obtain the matched filter output associated with the mean spectrum  $\boldsymbol{\mu}$ ), and  $\mathbf{t} = \mathbf{s}_k$  for the first few values of  $k$ , where  $\mathbf{s}_k$  is the  $k$ 'th principal component – ie, the  $k$ 'th eigenvector of the covariance matrix  $R$ .

As an ansatz, we will write our reconstructed spectrum  $\widehat{\mathbf{x}} \approx \widetilde{\mathbf{x}}$ :

$$\widehat{\mathbf{x}} = A^\top \mathbf{u} + B^\top \mathbf{z}, \quad (7)$$

where  $A$  and  $B$  are matrices to be determined,  $\mathbf{u}$  is the set of matched filter measurements, and  $\mathbf{z}$  is (for now) wholly unspecified, and represents the degrees of freedom in the solution. We know that  $T^\top R^{-1}(\mathbf{x} - \boldsymbol{\mu}) = \tilde{T}^\top \widetilde{\mathbf{x}} = \mathbf{u}$  for the true spectrum  $\widetilde{\mathbf{x}}$ , so we will impose that on our estimate  $\widehat{\mathbf{x}}$ :

$$\tilde{T}^\top \widehat{\mathbf{x}} = \tilde{T}^\top A^\top \mathbf{u} + \tilde{T}^\top B^\top \mathbf{z} = \mathbf{u}. \quad (8)$$

This leads to the requirements that  $\tilde{T}^\top A^\top = I$  and  $\tilde{T}^\top B^\top = 0$ ; or equivalently:

$$A\tilde{T} = I, \quad (9)$$

$$B\tilde{T} = 0. \quad (10)$$

We can satisfy the first condition by making  $A$  the pseudoinverse<sup>‡</sup> of  $\tilde{T}$ ; that is,

$$A = \tilde{T}^+ = (\tilde{T}^\top \tilde{T})^{-1} \tilde{T}^\top \quad (11)$$

Further, we can take

$$B = I - \tilde{T}\tilde{T}^+ = I - \tilde{T}(\tilde{T}^\top \tilde{T})^{-1} \tilde{T}^\top \quad (12)$$

and note that  $B^\top = B$ .

Thus, Eq. (7), along with Eq. (11) and Eq. (12), specifies a space of solutions  $\widehat{\mathbf{x}}$  that are consistent with the matched filter outputs:

$$\widehat{\mathbf{x}} = A^\top \mathbf{u} + B^\top \mathbf{z} = \tilde{T}(\tilde{T}^\top \tilde{T})^{-1} \mathbf{u} + \left( I - \tilde{T}(\tilde{T}^\top \tilde{T})^{-1} \tilde{T}^\top \right) \mathbf{z} \quad (13)$$

Recall that in addition to the matched filter values, we also downlink the Mahalanobis distance  $v^2 = (\mathbf{x} - \boldsymbol{\mu})^\top R^{-1}(\mathbf{x} - \boldsymbol{\mu}) = \widetilde{\mathbf{x}}^\top \widetilde{\mathbf{x}}$ . Thus, in addition to the constraints in Eq. (8), we want to impose this constraint on  $\widehat{\mathbf{x}}$ . Starting with Eq. (13), we obtain

$$v^2 = (A^\top \mathbf{u} + B^\top \mathbf{z})^\top (A^\top \mathbf{u} + B^\top \mathbf{z}) \quad (14)$$

$$= \mathbf{u}^\top A A^\top \mathbf{u} + \mathbf{z}^\top B A^\top \mathbf{u} + \mathbf{u}^\top A B^\top \mathbf{z} + \mathbf{z}^\top B B^\top \mathbf{z} \quad (15)$$

$$= \mathbf{u}^\top A A^\top \mathbf{u} + \mathbf{z}^\top B B^\top \mathbf{z} = \|A^\top \mathbf{u}\|^2 + \|B^\top \mathbf{z}\|^2 \quad (16)$$

since the cross-terms vanish; eg,

$$\begin{aligned} AB^\top &= (\tilde{T}^\top \tilde{T})^{-1} \tilde{T}^\top \left( I - \tilde{T}(\tilde{T}^\top \tilde{T})^{-1} \tilde{T}^\top \right) \\ &= (\tilde{T}^\top \tilde{T})^{-1} \tilde{T}^\top - (\tilde{T}^\top \tilde{T})^{-1} \tilde{T}^\top \tilde{T}(\tilde{T}^\top \tilde{T})^{-1} \tilde{T}^\top \\ &= (\tilde{T}^\top \tilde{T})^{-1} \tilde{T}^\top - (\tilde{T}^\top \tilde{T})^{-1} \tilde{T}^\top = 0 \end{aligned} \quad (17)$$

If we write  $\mathbf{z} = \zeta \widetilde{\mathbf{z}}$ , where  $\zeta$  is a scalar magnitude, and  $\widetilde{\mathbf{z}}$  is an arbitrary unit-magnitude vector, then Eq. (16) leads to  $\zeta^2 = (v^2 - \|A^\top \mathbf{u}\|^2) / \|B^\top \widetilde{\mathbf{z}}\|^2$ .

This leads to our simplest strategy for reconstructing the full datacube. Let  $\mathbf{z}$  be an arbitrary vector, chosen at random from an isotropic distribution (eg, multivariate Gaussian with covariance equal to the identity matrix); then rescale the magnitude of  $\mathbf{z}$  to be  $\zeta$  and use  $\widehat{\mathbf{x}}$  as given by Eq. (13).

---

<sup>‡</sup>The pseudoinverse of a matrix  $Z$  is written  $Z^+$ . If  $Z$  has  $k$  columns and  $d$  rows, then  $Z^+$  has  $k$  rows and  $d$  columns, and satisfies  $Z^+ Z = I$ , though in general  $ZZ^+ \neq I$ . There may be multiple choices for  $Z^+$  that satisfy this property, but we take the Moore-Penrose pseudoinverse. If  $Z$  is of rank  $k$ , then we can write  $Z^+ = (Z^\top Z)^{-1} Z^\top$ . Although in general  $ZZ^+ \neq I$ , note that  $ZZ^+$  is symmetric:  $(ZZ^+)^\top = ZZ^+$ .

In whitened space, the fraction of “variance” that is specified by the  $A\mathbf{u}$  component is given by  $\|A\mathbf{u}\|^2/v^2$ . This specification fixes the equivalent of  $k$  components out of  $d$ , so one might generally expect this ratio of variances to be on the order of  $k/d$ , which is much less than 1. Further, to the extent that the target directions are uncorrelated with the high-variance directions in the unwhitened data space [as is often the case for gas spectra], the variance explained by  $A\mathbf{u}$  is likely to be a very small fraction of the total variance in the image. This argues for including a few extra target spectra associated with  $\boldsymbol{\mu}$  and/or principal components of the covariance  $R$ .

We can be somewhat less arbitrary in our choice of  $\mathbf{z}$  by adding a further constraint that the choice of  $\mathbf{z}$  minimizes the magnitude of  $\mathbf{x} - \boldsymbol{\mu}$  in the original (unwhitened) data space. This is equivalent to choosing  $\mathbf{z}$  so as to minimize the magnitude of  $\mathbf{z}^\top R \mathbf{z}$  subject to  $\mathbf{z}$  being in the subspace defined by  $B$ . This can be achieved by making  $\mathbf{z}$  in the direction of the smallest (but nonzero) eigenvalue of  $BRB^\top$ . Note that this  $\mathbf{z}$  should be multiplied by either  $+1$  or  $-1$  with equal probability, in order to avoid altering the mean of the data. This is the lowest-magnitude solution that satisfies the criteria for  $\mathbf{u}$  and  $v^2$ , but an undesirable feature is that the covariance matrix will not agree with that of the original data. To achieve that, we go back to our simpler strategy of choosing  $\mathbf{z}$  from an isotropic Gaussian distribution.

### 3.3.2. Machine learning approach

Another idea for a less arbitrary choice of  $\mathbf{z}$  is to exploit the sample of full-spectrum pixels that we *do* have. For each of these full-spectrum pixels  $\mathbf{x}$  we also have the matched filter values  $\mathbf{u}$  and the Mahalanobis measure  $v^2$ . So the naive variant of this idea is to “learn” a function that maps  $\mathbf{u}$  and  $v^2$  into  $\mathbf{x}$ . As we have seen in the previous section, this is potentially a one-to-many map, in that there are many  $\mathbf{x}$  values that are consistent with a given  $\mathbf{u}$  and  $v^2$ . The appeal of machine learning is that we can use data from the image to help decide *which*  $\mathbf{x}$  is most likely to occur for a given  $\mathbf{u}$  and  $v^2$ .

We can, however, leverage the insights gained in the previous section by re-casting the problem in a way that leads to a more physics-informed machine learning. Given a training sample  $\mathbf{x}$  (or, in whitened space,  $\tilde{\mathbf{x}}$ ), we can identify the associated  $\mathbf{z}$  from the definition in Eq. (7):

$$B^\top \mathbf{z} = \tilde{\mathbf{x}} - A^\top \mathbf{u} = \tilde{\mathbf{x}} - \left( (\tilde{T}^\top \tilde{T})^{-1} \tilde{T}^\top \right)^\top (\tilde{T}^\top \tilde{\mathbf{x}}) \quad (18)$$

$$= \tilde{\mathbf{x}} - \tilde{T} (\tilde{T}^\top \tilde{T})^{-1} \tilde{T}^\top \tilde{\mathbf{x}} \quad (19)$$

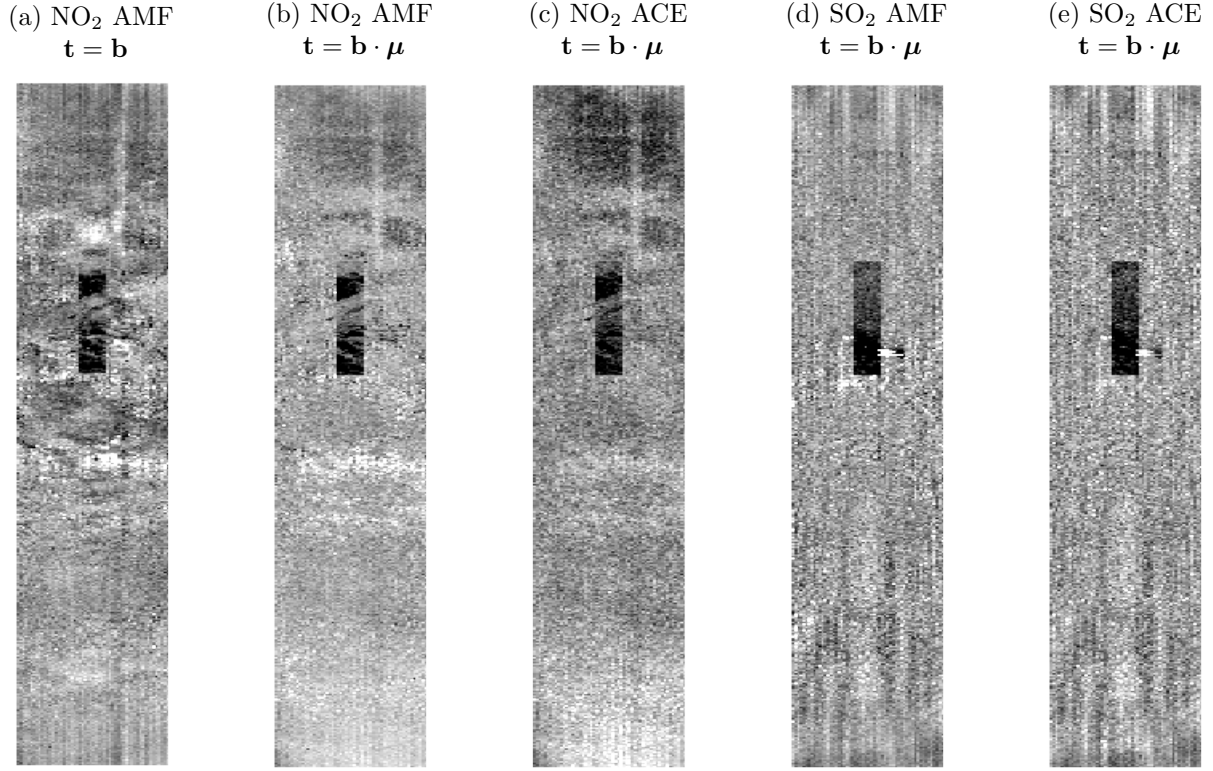
$$= \left( I - \tilde{T} (\tilde{T}^\top \tilde{T})^{-1} \tilde{T}^\top \right) \tilde{\mathbf{x}} \quad (20)$$

$$= B \tilde{\mathbf{x}} \quad (21)$$

Since  $B$  is a projection operator, we can constrain our interest to  $\mathbf{z}$  for which  $\mathbf{z} = B\mathbf{z}$ ; thus,  $\mathbf{z} = B\tilde{\mathbf{x}}$  defines a unique  $\mathbf{z}$ .

This learning-based approach seeks a function  $f$  that provides  $\mathbf{z}$  as a function of  $\mathbf{u}$  and  $v^2$ . Given that function, we can go back and reconstruct  $\hat{\mathbf{x}} = A^\top \mathbf{u} + Bf(\mathbf{u}, v^2)$ . In fact, we can strictly enforce the condition that  $\hat{\mathbf{x}}^\top \hat{\mathbf{x}} = v^2$  by taking  $\hat{\mathbf{x}} = A^\top \mathbf{u} + B\zeta f(\mathbf{u}, v^2)$ , where  $\zeta^2 = (v^2 - \|A^\top \mathbf{u}\|^2) / \|Bf(\mathbf{u}, v^2)\|^2$ .

But a problem with this approach is that, by design,  $\mathbf{z}$  has mean zero, and in this whitened space, the vector  $\mathbf{z}$  is uncorrelated with (though maybe not strictly independent of) the vector  $A\mathbf{u}$ . Thus, attempting to learn the high-dimensional  $\mathbf{z}$  will likely lead to an “average” value that has magnitude near zero, and direction that is noisy (in fact, given that the magnitude is near zero, the direction will be particularly noisy). The magnitude of  $\mathbf{z}$  will be “corrected” when it is multiplied by  $\zeta$ , but since the magnitude is likely to be small to start with, one effect of this multiplication will be to enhance the noise. Another problem is that a deterministic function  $f$  will produce a  $\mathbf{z}$  with nonzero mean, which leads to a reconstructed  $\mathbf{x}$  with a mean unequal to  $\boldsymbol{\mu}$ . Thus, in the end, we rejected this machine learning approach.



**Figure 2.** Plumes are artificially implanted into OMI images, and shown in these stripe-corrected detection images as a dark rectangular patch, indicating that the plumes are absorptive.

**Table 1.** Performance of different detectors. Because the plumes are absorptive, larger negative values are better.

Detector	NO <sub>2</sub> plume				SO <sub>2</sub> plume			
	uncorrected		stripe corrected		uncorrected		stripe corrected	
	$Q_{ave}$	$Q_{med}$	$Q_{ave}$	$Q_{med}$	$Q_{ave}$	$Q_{med}$	$Q_{ave}$	$Q_{med}$
AMF $\mathbf{t} = \mathbf{b}$	-2.117	-1.735	-2.163	-1.776	-2.242	-1.528	-2.285	-1.615
ACE $\mathbf{t} = \mathbf{b}$	-1.606	-1.366	-1.688	-1.398	-2.238	-1.634	-2.352	-1.746
AMF $\mathbf{t} = \mathbf{b} \cdot \boldsymbol{\mu}$	-2.714	-2.044	<b>-2.739</b>	<b>-2.051</b>	-2.718	-2.127	-2.739	-2.097
ACE $\mathbf{t} = \mathbf{b} \cdot \boldsymbol{\mu}$	-2.054	-1.741	-2.145	-1.763	-2.742	-2.223	<b>-2.794</b>	<b>-2.278</b>
log AMF $\mathbf{t} = \mathbf{b}$	-1.565	-1.417	-1.582	-1.433	-1.634	-1.320	-1.649	-1.393
log ACE $\mathbf{t} = \mathbf{b}$	-1.948	-1.567	-1.996	-1.565	-2.184	-1.290	-2.223	-1.324

## 4. NUMERICAL EXPERIMENTS

In this section, we describe a simple experiment in which we implant an absorptive plume into an OMI image and compare the detection performance of different algorithms to find the plume. We use two images, and into one we implant an NO<sub>2</sub> plume and in the other an SO<sub>2</sub> plume. For a given detector  $\mathcal{D}(\mathbf{x})$  we compute two measures of quality:

$$Q_{\text{ave}} = \frac{\text{mean}[\mathcal{D}(\text{on plume})] - \text{mean}[\mathcal{D}(\text{off plume})]}{\text{stdev}[\mathcal{D}(\text{off plume})]} \quad (22)$$

$$Q_{\text{med}} = \frac{\text{median}[\mathcal{D}(\text{on plume})] - \text{median}[\mathcal{D}(\text{off plume})]}{\text{interquartile\_range}[\mathcal{D}(\text{off plume})]} \quad (23)$$

Since the plumes are absorptive, the  $Q$  values are generally negative, with the larger magnitude value corresponding to the better detections. Table 1 lists the performance of the AMF and ACE detectors, based on target signatures  $\mathbf{t} = \mathbf{b}$  and  $\mathbf{t} = \mathbf{b} \cdot \boldsymbol{\mu}$ , where  $\mathbf{b}$  is the absorption spectrum for the indicated chemical, and  $\boldsymbol{\mu}$  is the mean spectrum over the image. Also shown is the performance obtained when the data is logarithmically scaled; in that case only  $\mathbf{t} = \mathbf{b}$  makes sense. These performance values are computed with and without the stripe correction described in Eq. (4).

We observe that stripe correction does consistently improve performance, though not substantially. We also see that  $\mathbf{t} = \mathbf{b} \cdot \boldsymbol{\mu}$  works better than the simpler  $\mathbf{t} = \mathbf{b}$ . We were surprised to observe that the logarithmic scaling did not outperform the linear scaling, given that the target signature is truly additive in the logarithmic space. A tentative explanation is that the logarithm throws off what would otherwise be an approximately Gaussian background distribution, but we have not investigated this explanation any further. In terms of AMF vs ACE, we see that AMF does a little better on the NO<sub>2</sub> plume and ACE is better for the SO<sub>2</sub> plume. This provides further evidence that it is advantageous to downlink both matched filter and Mahalanobis imagery, enabling the AMF vs ACE decision to be made on the ground, after the acquisition.

## 5. ACKNOWLEDGMENTS

This work was supported by the United States Department of Energy (DOE) through the Laboratory Directed Research and Development (LDRD) program at Los Alamos National Laboratory. We are grateful to our colleagues on the TACOS and NACHOS teams for their helpful suggestions and insightful questions. Particular thanks go to Bernard Foy and Clairia Safi, who were instrumental contributors to our original report,<sup>7</sup> and continue to be very actively involved in this work.

## REFERENCES

1. “NASA-OMI.” [Online] [https://www.nasa.gov/mission\\_pages/aura/spacecraft/omi.html](https://www.nasa.gov/mission_pages/aura/spacecraft/omi.html).
2. R. Lindenmaier, M. K. Dubey, B. G. Henderson, Z. T. Butterfield, J. R. Herman, T. Rahn, and S.-H. Lee, “Multiscale observations of CO<sub>2</sub>, <sup>13</sup>CO<sub>2</sub>, and pollutants at Four Corners for emission verification and attribution,” *Proc. National Academy of Sciences* **111**, pp. 8386–8391, 2014.
3. Z. Darynova, A. Maksot, L. Kulmukanova, M. Malekipirbazari, H. Sharifi, M. A. Torkmahalleh, and T. Holloway, “Evaluation of NO<sub>2</sub> column variations over the atmosphere of Kazakhstan using satellite data,” *J. Applied Remote Sensing* **12**, p. 042610, 2018.
4. S. P. Love, F. Goff, D. Counce, C. Siebe, and H. Delgado, “Passive infrared spectroscopy of the eruption plume at Popocatépetl volcano, Mexico,” *Nature* **396**, pp. 563–567, 1998.
5. S. P. Love, F. Goff, S. C. Schmidt, D. Counce, D. Pettit, B. W. Christenson, and C. Siebe, “Passive infrared spectroscopic remote sensing of volcanic gases: Ground-based studies at White Island and Ruapehu, New Zealand, and Popocatépetl, Mexico,” in *Remote Sensing of Active Volcanism*, P. Mougini-Mark, J. Crisp, and J. Fink, eds., *Geophysical Monograph* **116**, pp. 117–138, American Geophysical Union, Washington, D.C., 2000.
6. S. P. Love, T. C. Hale, L. J. Jolin, J. Barefield, W. Atkins, and J. Tiee, “The ORCAS compact long-wave infrared hyperspectral sensor,” in *Proc. Military Sensing Symposia (MSS) Specialty Group on Passive Sensors*, 2008.



7. J. Theiler, B. R. Foy, C. Safi, and S. P. Love, "Onboard cubesat data processing for hyperspectral detection of chemical plumes," *Proc. SPIE* **10644**, p. 1064405, 2018.
8. I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerospace and Electronic Systems* **10**, pp. 853–863, 1974.
9. E. J. Kelly, "Performance of an adaptive detection algorithm: rejection of unwanted signals," *IEEE Trans. Aerospace and Electronic Systems* **25**, pp. 122–133, 1989.
10. F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. Aerospace and Electronic Systems* **28**, pp. 208–216, 1992.
11. L. L. Scharf and L. T. McWhorter, "Adaptive matched subspace detectors and adaptive coherence estimators," in *Proc. Asilomar Conference on Signals, Systems, and Computers*, 1996.
12. S. Kraut, L. L. Scharf, and R. W. Butler, "The Adaptive Coherence Estimator: a uniformly most-powerful-invariant adaptive detection statistic," *IEEE Trans. Signal Processing* **53**, pp. 427–438, 2005.
13. J. Theiler and B. R. Foy, "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)*, p. I:221, 2008.
14. J. W. Boardman and F. A. Kruse, "Analysis of imaging spectrometer data using  $N$ -dimensional geometry and a mixture-tuned matched filtering approach," *IEEE Trans. Geoscience and Remote Sensing* **49**, pp. 4138–4152, 2011.
15. R. S. DiPietro, D. G. Manolakis, R. B. Lockwood, T. Cooley, and J. Jacobson, "Hyperspectral matched filter with false-alarm mitigation," *Optical Engineering* **51**, p. 016202, 2012.
16. B. R. Foy, J. Theiler, and A. M. Fraser, "Decision boundaries in two dimensions for target detection in hyperspectral imagery," *Optics Express* **17**, pp. 17391–17411, 2009.
17. J. Theiler, "Matched-pair machine learning," *Technometrics* **55**, pp. 536–547, 2013.
18. J. Theiler, "Transductive and matched-pair machine learning for difficult target detection problems," *Proc. SPIE* **9088**, p. 90880E, 2014.
19. P. C. Mahalanobis, "On the generalised distance in statistics," *Proc. National Institute of Sciences of India* **2**, pp. 49–55, 1936.
20. I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoustics, Speech, and Signal Processing* **38**, pp. 1760–1770, 1990.
21. K. N. Buckland, S. J. Young, E. R. Keim, B. R. Johnson, P. D. Johnson, and D. M. Tratt, "Tracking and quantification of gaseous chemical plumes from anthropogenic emission sources within the Los Angeles basin," *Remote Sensing of Environment* **201**, pp. 275–296, 2017.
22. G. Cao and C. A. Bouman, "Covariance estimation for high dimensional data vectors using the sparse matrix transform," in *Advances in Neural Information Processing Systems*, **21**, pp. 225–232, MIT Press, 2009.
23. J. Theiler, G. Cao, L. R. Bachega, and C. A. Bouman, "Sparse matrix transform for hyperspectral image processing," *IEEE J. Selected Topics in Signal Processing* **5**, pp. 424–437, 2011.
24. A. Schaum and A. Stocker, "Spectrally selective target detection," in *Proc. ISSSR (International Symposium on Spectral Sensing Research)*, p. 23, 1997.
25. A. Ziemann, M. Kucer, and J. Theiler, "A machine learning approach to hyperspectral detection of solid targets," *Proc. SPIE* **10644**, p. 1064404, 2018.
26. J. Theiler, B. Zimmer, and A. Ziemann, "Closed-form detector for solid sub-pixel targets in multivariate  $t$ -distributed background clutter," in *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 2773–2776, 2018.