

Detection of Spectrally Sparse Anomalies in Hyperspectral Imagery

James Theiler
Space Data Systems Group
Intelligence and Space Research Division
Los Alamos National Laboratory
Los Alamos, NM 87544, USA
Email: jt@lanl.gov

Brendt Wohlberg
Applied Mathematics and Plasma Physics, T-5
Theoretical Division
Los Alamos National Laboratory
Los Alamos, NM 87544, USA
Email: brendt@lanl.gov

Abstract—We present a variant of the classic problem of anomaly detection in hyperspectral imagery. In this variant, the anomalous signatures are assumed to be additive and to exhibit spectra that are sparse – that is, only a few of the many hyperspectral channels are significantly nonzero.

When the background data are Gaussian, and there is no structure in the anomalous signatures, then the optimal detector is given by a Mahalanobis distance and exhibits contours that are ellipsoids. When the desired signature is known, then the solution is given by a matched filter that is specifically optimized for that signature; the contours are parallel planes whose orientation depends on both the desired signature and the covariance of the background. We address an in-between problem, one for which the detailed signature is not known, but a more generic description of the structure is available.

We propose that this solution might have application to the detection of gaseous plumes, when the chemistry of the gas is unknown. Such plumes have approximately additive effect on their backgrounds, and – especially in the thermal infrared “fingerprint region” – tend to have very sparse absorption and emission spectra.

Index Terms—hyperspectral imagery, signal processing, anomaly detection, plume detection, sparse modeling

I. INTRODUCTION

Gaseous plumes, particularly in the infrared, exhibit very distinctive signatures of absorption¹ as a function of wavelength. Hyperspectral imagery enables analysts to exploit this distinctiveness, and to detect specific gaseous chemicals even at low concentrations, using matched filters that are tailored both to the specific structure of the chemical signature and to the statistics of the background clutter [1]–[9].

All of these algorithms, however, assume that the gas signature is known. If the chemical itself is not known, then the usual approach is to attack the data with a large library of gas spectra. The library should not only cover the range of chemicals whose detection might be of interest, it should also include variants of the spectra that are observed in different atmospheric conditions.

An entirely different approach for detecting targets in clutter is to treat the targets as completely unknown, but as unusual

¹Depending on the temperature of the plume relative to its background, the plume may be in absorption or emission; but the spectral shape is the same in both cases.

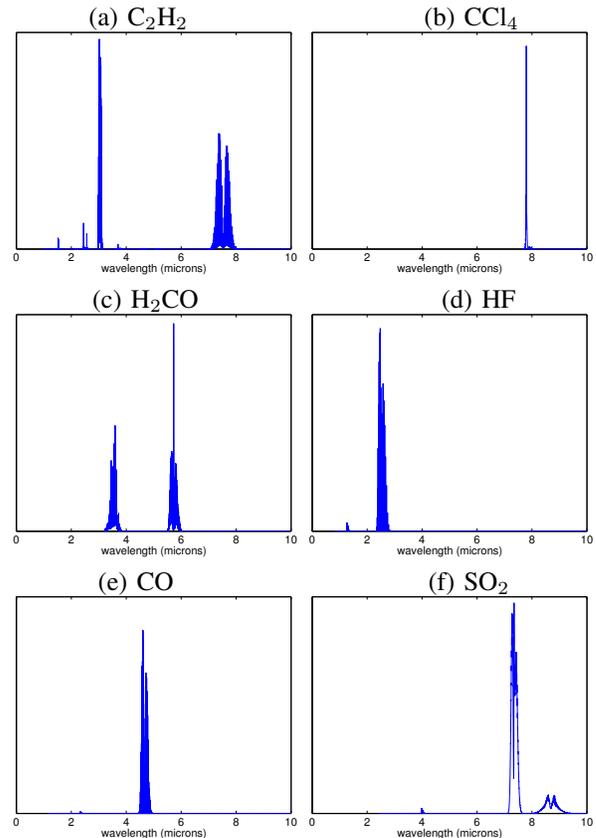


Fig. 1. Infrared absorption signatures for six different gases, chosen more or less at random from the HITRAN high-resolution transmission molecular absorption database [10], illustrate the sparse structure that is typically inherent in gaseous chemical spectra.

compared to the rest of the background. Again, we seek deviations from the background clutter, but in this case we do not know the direction of that deviation. This is the anomaly detection approach, and many algorithms for finding anomalies are variants of the so-called RX algorithm, which is based on Mahalanobis distance [11]–[14].

It is in the nature of anomaly detection that the anomalies are not well-defined, and when they are modeled, then they are generally modeled (with few exceptions [15], [16]) using

a uniform distribution.

One property exhibited by nearly all gaseous chemicals is that the spectrum is composed of a relatively sparse forest of narrow lines; several examples are illustrated in Fig. 1. We propose here to exploit this property and search for anomalies that have this spectrally sparse character.

In a hypothesis testing framework, we write

$$\text{Null } H_0 : \mathbf{x} = \mathbf{z} \quad (1)$$

$$\text{Alternative } H_1 : \mathbf{x} = \mathbf{z} + \mathbf{t} \quad (2)$$

Here, \mathbf{z} is the background, which is distributed with some distribution that is in practice learned from the data, and \mathbf{t} is an additive target. We consider three kinds of targets: In the first case \mathbf{t} is known (up to a scalar multiplier). In the second case, \mathbf{t} is completely unknown. In the third case, which is our main interest for this paper, the only thing known about \mathbf{t} is that it is spectrally sparse; that is, the vector-valued \mathbf{t} has mostly zero-valued elements.

We remark that the equation $\mathbf{x} = \mathbf{z} + \mathbf{t}$ has the flavor of a low-rank plus sparse decomposition [17], [18]. In our work here, the “low-rank” component is treated as Gaussian.² We further remark that the sparseness we seek is in the components of \mathbf{t} , not in the components of \mathbf{x} , which is treated in its full dimensionality. This is in contrast to the problem addressed by Banerjee *et al.* [19], which is to find a sparse set of features to describe \mathbf{x} , and then to use a kernel-based anomaly detector in that lower dimensional space.

II. DERIVATION OF RX FOR ANOMALY DETECTION

One way to derive the RX detector [11] is with the Generalized Likelihood Ratio Test (GLRT), in which the anomalous signal is treated as a nuisance parameter, and then one maximizes the likelihood over all possible values of the nuisance parameters. The likelihood of observing \mathbf{x} when \mathbf{t} is the anomalous target signal is proportional to

$$(2\pi)^{-d/2} |R|^{-1/2} \exp[-(\mathbf{x} - \mathbf{t})^T R^{-1} (\mathbf{x} - \mathbf{t}) / 2] \quad (3)$$

where d is the dimension of the data (number of spectral channels in the hyperspectral image), R is the covariance, and $|R|$ is the determinant of R . When \mathbf{t} is known, the likelihood ratio

$$\mathcal{L}(\mathbf{x}) = \frac{\exp[-(\mathbf{x} - \mathbf{t})^T R^{-1} (\mathbf{x} - \mathbf{t}) / 2]}{\exp[-\mathbf{x}^T R^{-1} \mathbf{x} / 2]} \quad (4)$$

leads to the matched filter

$$\mathcal{M}(\mathbf{x}) = (\mathbf{t}^T R^{-1} \mathbf{t}) / 2 + \log \mathcal{L}(\mathbf{x}) = \mathbf{t}^T R^{-1} \mathbf{x} \quad (5)$$

which is linear in \mathbf{x} .

When \mathbf{t} is not known, we can use the GLRT to write

$$\mathcal{L}(\mathbf{x}) = \frac{\max_{\mathbf{t}} \exp[-(\mathbf{x} - \mathbf{t})^T R^{-1} (\mathbf{x} - \mathbf{t}) / 2]}{\exp[-\mathbf{x}^T R^{-1} \mathbf{x} / 2]} \quad (6)$$

²For hyperspectral data, the Gaussian model typically includes many small eigenvalues, so the low-rank model is not entirely out of place.

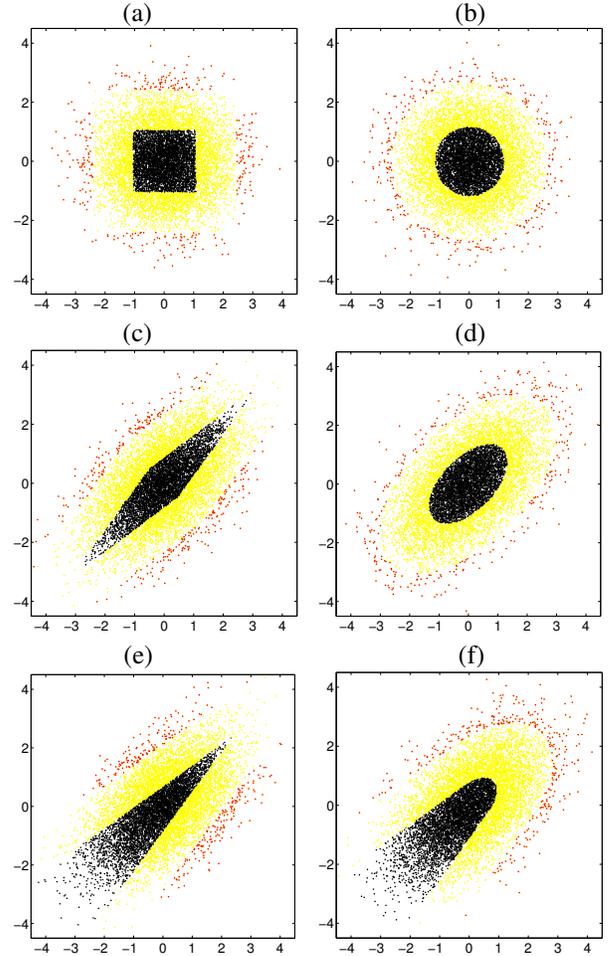


Fig. 2. In this simple $d = 2$ dimensional model, we plot the most anomalous points (top three percent) in red, the least anomalous points (below median) in black, and the rest in yellow. A total of 10^4 points are drawn at random from a Gaussian distribution. In (a,b) the distribution is Gaussian with covariance given by the identity; in (c,d) the Gaussian has a correlation of 0.3 in the off-diagonal component of the covariance matrix. We take $k = 1$ in the leftmost panels (a,c,e), and $k = 2$ in (b,d,e). We note that $k = d$ corresponds to the standard RX algorithm and in (b,d), elliptical contours of anomalousness are seen. For $k = 1$, we see diamond-shaped contours, corresponding to matched filters with respect to the axis directions. In (e,f), we plot the case for anomalies that are restricted to positive values.

The numerator achieves its maximum when $\mathbf{t} = \mathbf{x}$, which leads to

$$\mathcal{L}(\mathbf{x}) = \frac{1}{\exp[-\mathbf{x}^T R^{-1} \mathbf{x} / 2]} \quad (7)$$

and therefore

$$\mathcal{A}(\mathbf{x}) = 2 \log \mathcal{L}(\mathbf{x}) = \mathbf{x}^T R^{-1} \mathbf{x} \quad (8)$$

is the standard RX measure for anomalousness.

III. SPECTRALLY SPARSE ANOMALIES

In the derivation of $\mathcal{A}(\mathbf{x})$ in the previous section, \mathbf{t} was unrestricted, and we found that the likelihood was maximized when $\mathbf{t} = \mathbf{x}$. We suggest two approaches for restricting the target to sparse signatures:

1. Strictly constrain the target to have a fixed number of nonzero elements. Let \mathcal{T}_k correspond to the set of targets \mathbf{t} with k or fewer of the components nonzero; that is, $\mathcal{T}_k = \{\mathbf{t} : \|\mathbf{t}\|_0 \leq k\}$. The likelihood ratio then becomes

$$\mathcal{L}(\mathbf{x}) = \frac{\max_{\mathbf{t} \in \mathcal{T}_k} \exp[-(\mathbf{x} - \mathbf{t})^T R^{-1}(\mathbf{x} - \mathbf{t})/2]}{\exp[-\mathbf{x}^T R^{-1}\mathbf{x}/2]} \quad (9)$$

or equivalently:

$$\mathcal{A}(\mathbf{x}) = \mathbf{x}^T R^{-1}\mathbf{x} - \min_{\mathbf{t} \in \mathcal{T}_k} (\mathbf{x} - \mathbf{t})^T R^{-1}(\mathbf{x} - \mathbf{t}) \quad (10)$$

$$= \max_{\mathbf{t} \in \mathcal{T}_k} [2\mathbf{x}^T R^{-1}\mathbf{t} - \mathbf{t}^T R^{-1}\mathbf{t}] \quad (11)$$

is the measure for anomalousness. The case $k = d$, where d is the number of spectral channels, leads to the standard RX formulation.

The experiments described in Section IV use a matching pursuit [20] algorithm to greedily add components to \mathbf{t} until k components have been added. It is straightforward to modify this algorithm to take advantage of spectra (such as is usually the case with gas spectra) for which the sign of the components of \mathbf{t} are known in advance. Fig. 2 illustrates what the contours of anomalousness look like in the simple case of two-dimensional Gaussian data.

2. Penalize the likelihood function to favor sparse signatures. Here, we can employ an L1 instead of an L0 metric, and consequently achieve a convex optimization. Rather than restrict \mathbf{t} to a fixed number of nonzero elements, we “nudge” it toward sparsity by altering Eq. (6) with a penalty factor:

$$\mathcal{L}(\mathbf{x}) = \frac{\max_{\mathbf{t}} \exp[-(\mathbf{x} - \mathbf{t})^T R^{-1}(\mathbf{x} - \mathbf{t})/2] \exp[-\lambda \|\mathbf{t}\|_1]}{\exp[-\mathbf{x}^T R^{-1}\mathbf{x}/2]} \quad (12)$$

This leads to

$$\mathcal{A}(\mathbf{x}) = \max_{\mathbf{t}} [2\mathbf{x}^T R^{-1}\mathbf{t} - \mathbf{t}^T R^{-1}\mathbf{t} - \lambda \|\mathbf{t}\|_1] \quad (13)$$

as a measure of anomalousness which can be computed as a straightforward quadratic programming problem. The limit $\lambda \rightarrow 0$ leads to the standard RX formulation.

IV. EXPERIMENT

For the experiment shown in Fig. 3, Fig. 4, and Fig. 5, we used the AVIRIS Indian Pines dataset [21], cropped to avoid some of the strong single-pixel anomalies that overpowered the rest of the image. To this background image, along a small 4×21 pixel slice of the image, a gradient of plume was added, weaker on the left and stronger on the right. The detection of this plume shows up in Fig. 5 as a dark horizontal stripe.³ The more sensitive detection observed in Fig. 5(a) is expected since the matched filter in Eq. (5) makes explicit use of the spectrum \mathbf{t} . By contrast, the anomaly-based approaches in the remaining panels do not use explicit knowledge about the simulated gas spectrum shown in Fig. 4. What Fig. 5 further

³We do not know the cause of the light vertical stripe that is observed in Fig. 5(b-h), but speculate that it is a data artifact, caused perhaps by some kind of data interpolation that led to unusually un-anomalous pixel values.

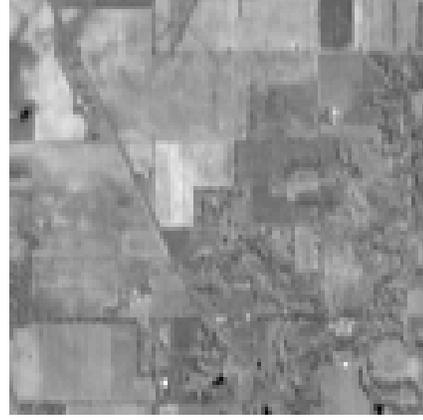


Fig. 3. Cropped Indian Pines dataset. 111×111 pixels, and 200 channels. This broadband image is the sum over all the channels, and although it includes the simulated plume, that plume is too weak to be observed in this projection.

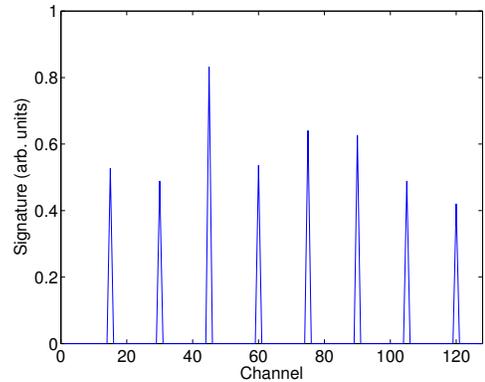


Fig. 4. Sparse spectral structure of simulated gas used in Fig. 5.

illustrates is that the spectrally sparse anomaly detector in Eq. (11), and shown in Fig. 5(c,e,g), provides a more sensitive detection than the traditional RX anomaly detector shown in Fig. 5(b). Thus, we are able to exploit the sparsity of the signal in Fig. 4 without knowing the details of that signature. Also, Fig. 5(d,f,h) exhibits further (albeit moderate) gains in sensitivity by restricting consideration to gas spectra where all the components are the same sign.

V. DISCUSSION

We have described preliminary efforts to extend standard anomaly detection to the case of a sparse additive target, and noted that this may have applications in plume detection.

One limitation of this work is the restriction to Gaussian backgrounds, which are not particularly realistic. Natural extensions are to heavy-tailed backgrounds [22], [23], to more arbitrary global representations [24], to local background estimation [25], or to locally low-rank models that have previously been developed for video data processing [18].

Another direction for future investigation is to use quadratic programming against an L1 penalty instead of matching pursuit to encourage sparseness; this would entail replacing Eq. (11) with Eq. (13) as the measure of anomalousness.

ACKNOWLEDGMENT

This work was supported by the Laboratory Directed Research and Development (LDRD) program at Los Alamos.

REFERENCES

- [1] C. L. Bennett, M. R. Carter, D. J. Fields, and F. D. Lee, "Infrared hyperspectral imaging results from vapor plume experiments," *Proc. SPIE*, vol. 2480, pp. 435–444, 1995.
- [2] A. Hayden, E. Niple, and B. Boyce, "Determination of trace-gas amounts in plumes by the use of orthogonal digital filtering of thermal-emission spectra," *Applied Optics*, vol. 35, pp. 2803–2809, 1996.
- [3] P. V. Villeneuve, H. A. Fry, J. Theiler, B. W. Smith, and A. D. Stocker, "Improved matched-filter detection techniques," *Proc. SPIE*, vol. 3753, pp. 278–285, 1999.
- [4] S. J. Young, "Detection and quantification of gases in industrial-stack plumes using thermal-infrared hyperspectral imaging," Tech. Rep. ATR-2002(8407)-1, The Aerospace Corporation, 2002.
- [5] B. R. Foy, R. R. Petrin, C. R. Quick, T. Shimada, and J. J. Tsee, "Comparisons between hyperspectral passive and multispectral active sensor measurements," *Proc. SPIE*, vol. 4722, pp. 98–109, 2002.
- [6] J. Theiler, B. R. Foy, and A. M. Fraser, "Characterizing non-Gaussian clutter and detecting weak gaseous plumes in hyperspectral imagery," *Proc. SPIE*, vol. 5806, pp. 182–193, 2005.
- [7] D. G. Manolakis and F. M. D'Amico, "A taxonomy of algorithms for chemical vapor detection with hyperspectral imaging spectroscopy," *Proc. SPIE*, vol. 5795, pp. 125–133, 2005.
- [8] C. M. Gittins, "Detection and characterization of chemical vapor fugitive emissions by nonlinear optimal estimation: theory and simulation," *Applied Optics*, vol. 48, pp. 4545–4561, 2009.
- [9] S. Niu, S. E. Golowich, V. K. Ingle, and D. G. Manolakis, "Design and evaluation of robust matched filters for chemical agent detection," *Proc. SPIE*, vol. 8186, pp. 81860R, 2011.
- [10] L. Rothman and S. Tashkun, "HITRAN on the web," Harvard-Smithsonian Center for Astrophysics (CFA), Cambridge, MA, USA; and V.E. Zuev Institute of Atmospheric Optics (IAO), Tomsk, Russia. <http://hitran.iao.ru/>.
- [11] 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*, vol. 38, pp. 1760–1770, 1990.
- [12] D. W. J. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Processing Magazine*, vol. 19, pp. 58–69, Jan 2002.
- [13] A. Schaum, "Hyperspectral anomaly detection: Beyond RX," *Proc. SPIE*, vol. 6565, pp. 656502, 2007.
- [14] S. Matteoli, M. Diani, and G. Corsini, "A tutorial overview of anomaly detection in hyperspectral images," *IEEE A&E Systems Magazine*, vol. 25, pp. 5–27, 2010.
- [15] J. Theiler and D. M. Cai, "Resampling approach for anomaly detection in multispectral images," *Proc. SPIE*, vol. 5093, pp. 230–240, 2003.
- [16] I. Steinwart, D. Hush, and C. Scovel, "A classification framework for anomaly detection," *J. Machine Learning Research*, vol. 6, pp. 211–232, 2005.
- [17] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *J. ACM*, vol. 58, pp. 11:1–11:37, 2011.
- [18] B. Wohlberg, R. Chartrand, and J. Theiler, "Local principal component pursuit for nonlinear datasets," *Proc. ICASSP*, 2012, To appear.
- [19] A. Banerjee, R. Juang, J. Broadwater, and P. Burlina, "Sparse feature extraction for support vector data description applications," *Proc. IGARSS*, pp. 4236–4239, 2010.
- [20] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, pp. 3397–3415, 1993.
- [21] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*, John Wiley & Sons, 2003.
- [22] S. M. Adler-Golden, "Improved hyperspectral anomaly detection in heavy-tailed backgrounds," *Proc. WHISPERS*, vol. 1, 2009.
- [23] J. Theiler and D. Hush, "Statistics for characterizing data on the periphery," *Proc. IGARSS*, pp. 4764–4767, 2010.
- [24] H. Kwon and N. M. Nasrabadi, "Kernel RX: a new nonlinear anomaly detector," *Proc. SPIE*, vol. 5806, pp. 35–46, 2005.
- [25] N. Gorelnik, H. Yehudai, and S. R. Rotman, "Anomaly detection in non-stationary backgrounds," *Proc. WHISPERS*, vol. 2, 2010.

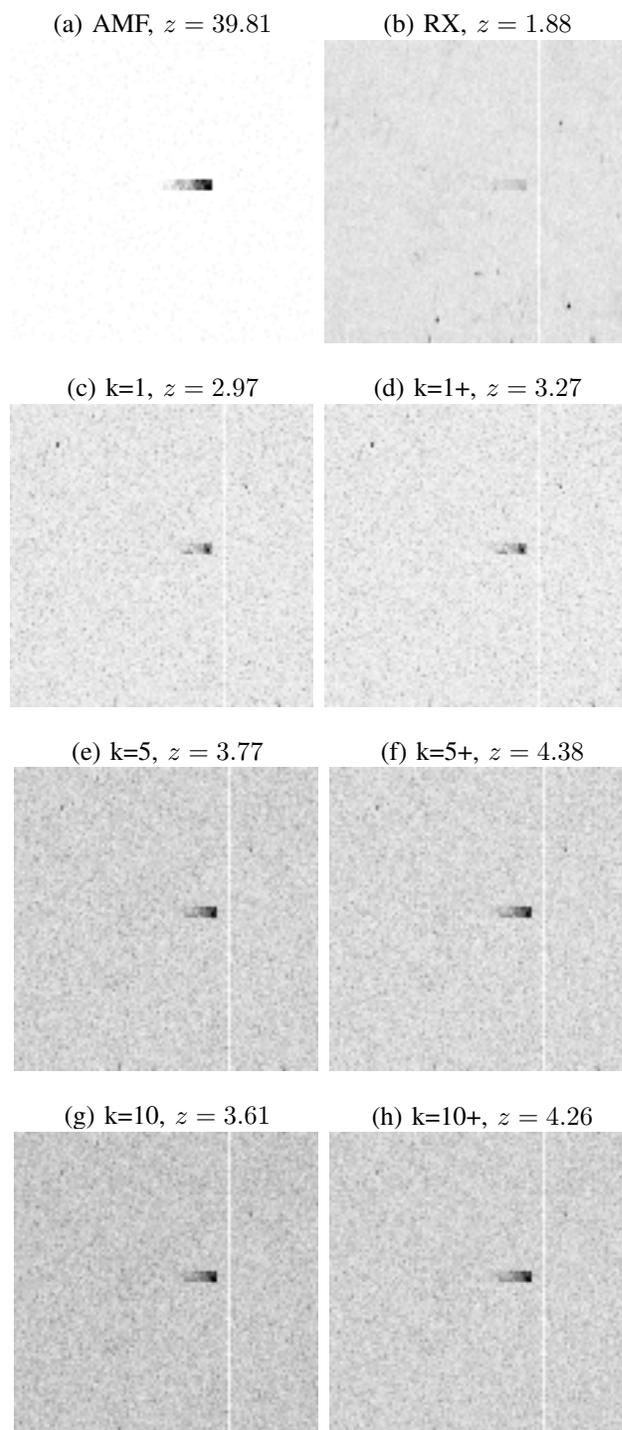


Fig. 5. Detecting a weak plume signature, seen here as a small dark horizontal stripe near the center of the image. (a) The matched-filter detector achieves the strongest detection, but it uses knowledge of the spectral signature. (b) The standard RX anomaly detector uses no knowledge of the spectral signature, and makes no assumptions about it. Its performance is much worse than the matched filter. (c,e,g) The spectrally sparse anomaly detector also does not know the spectral signature, but assumes that it is sparse with k components (shown here with $k = 1, 5, 10$). (d,f,h) The signed spectrally sparse anomaly detector uses the further information that the sign of every component of \mathbf{t} is positive. This is appropriate for emissive plumes (for absorptive plumes, one can assume that every component of \mathbf{t} is negative). For all of the detectors, the reported z -scores are the ratio of mean signal strength to standard deviation of the background.