# Simulation framework for spatio-spectral anomalous change detection

James Theiler, Neal R. Harvey, Reid Porter, and Brendt Wohlberg

Los Alamos National Laboratory, Los Alamos, NM 87545

## ABSTRACT

We describe the development of a simulation framework for anomalous change detection that considers both the spatial and spectral aspects of the imagery. A purely spectral framework has previously been introduced, but the extension to spatio-spectral requires attention to a variety of new issues, and requires more careful modeling of the anomalous changes. Using this extended framework, we evaluate the utility of spatial image processing operators to enhance change detection sensitivity in (simulated) remote sensing imagery.

## 1. INTRODUCTION

One difficulty that arises in the development of all kinds of detection algorithms is the acquisition of adequate quantities of adequately ground-truthed data for training and validation. But this is particularly problematic for anomaly detection and anomalous change detection (ACD), because anomalies are rare and because anomalies defy *a priori* specification. Some datasets have been expressly developed for anomalous change studies[1] and these should be an important part of any research portfolio. But such data is necessarily limited, and if relied upon exclusively, will provide anecdotal and potentially misleading guidance. Even anecdotal guidance is valuable, but we must be careful that in our attempts to optimize generic anomaly detection software, we don't end up instead with the world's best detector of folded green tarps.

Simulation, by contrast, provides not only unlimited quantities of data, but the opportunity to adjust the nature and the variety of that data. This allows the research to more systematically and quantitatively identify which aspects of which algorithms work best (or fail most dramatically) in which situations. The problems with simulation are well known; the main one is we can never be sure that a given simulation is "realistic" enough. There is always a danger that an algorithm that has been optimized to perform well on simulated data will be tripped up on some seemingly insignificant feature of real imagery that was not built into the simulation. Hybrid simulation – based on artificial manipulation of real data – can mitigate some of this problem. For target detection algorithms, for instance, characterizing the background is usually the most challenging part of the task. A simulated dataset that uses real data as a natural background will provide a realistic challenge to the detection algorithm.

Cohen and Rotman[2] used this approach to embed a variety of subpixel "targets" in hyperspectral imagery for the purpose of developing anomaly detection algorithms. A simulation framework for anomalous *changes* was developed[3] in which the anomalies were resampled from the data itself. But this framework treated the pixels in an image as independent samples and made sense only for testing purely spectral change detection algorithms. The most widely used anomalous change detection algorithms[4–7] are purely spectral, so this framework provided a way to compare them to each other in a controlled fashion.[3] There, the aim was to generate corresponding pixels in a pair of images in such a way that the individual pixels were not anomalous, but the relationship between them was anomalous. The aim of the present work is to extend that framework so that we can train and evaluate ACD algorithms that employ spatial processing. We are motivated by the observation that spatial processing can improve detection of spatially extended anomalies; this processing however is inconsistent with the notion, employed in the purely spectral framework, that pixels are independent samples. This is an important component in our efforts to develop algorithms for anomalous change that employ both spatial as well as spectral processing of the image pair. As we will see, however, the extension of this framework to cover spatio-spectral anomalous change detection raises several new issues.

Authors' emails: {jt,harve,rporter,brendt}@lanl.gov

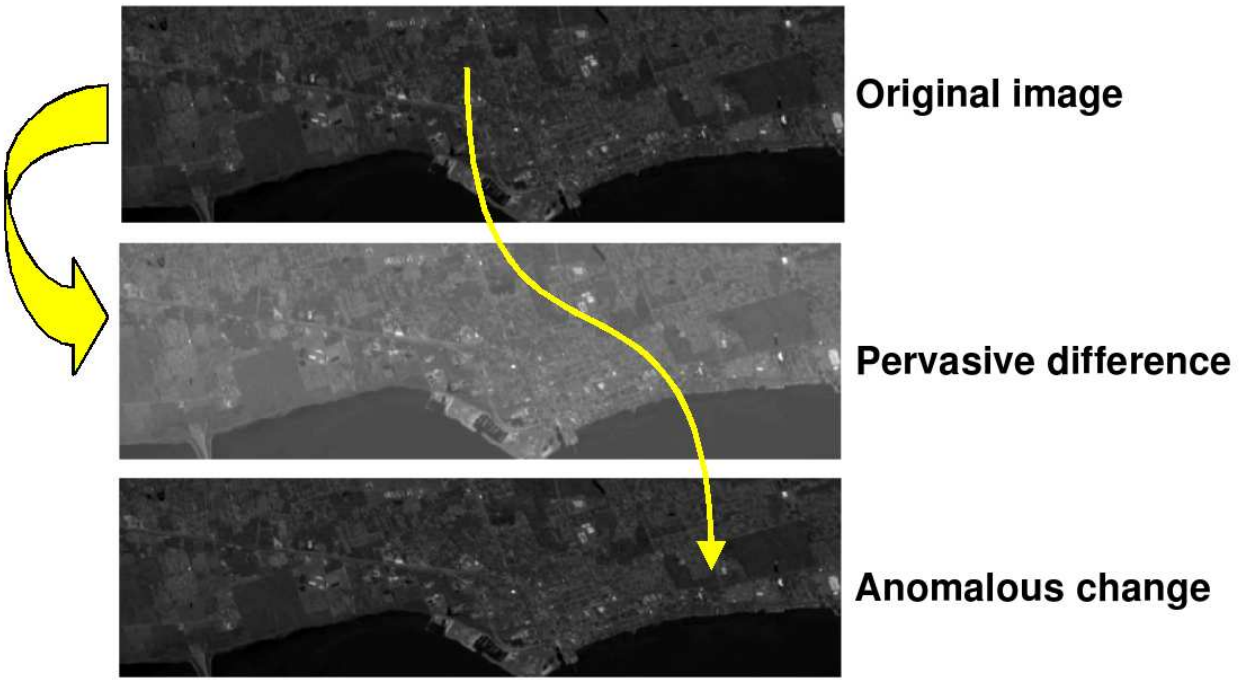


Figure 1. Simulation framework is illustrated on AVIRIS data[9] of the Florida coastline. Starting with a single real image (top panel), pervasive differences are applied to every pixel in the scene (middle panel). These differences might include altering intensity, spatial filtering, deliberate misregistration, *etc.* The anomalous change image (bottom panel) is again produced from the original image, but the change is made at only a single pixel. For the purely spectral simulation framework, the change at that pixel is to replace it with some other pixel randomly chose from somewhere else in the scene. The arrows above illustrate this difference: the fat arrow on the left indicates that all the pixels in the original image are modified to make the pervasive difference image; and the thin arrow in the middle of the image indicates that only a single pixel is replaced in the anomalous change image.

Our paradigm here is that spatio-spectral processing is applied to an image, and that is followed by spectral change detection applied separately to each pixel in the processed image pair. We remark that more integrated approaches are possible, but these approaches will be investigated elsewhere.[8]

In Section 2, the more straightforward *spectral* simulation framework is reviewed, and then in Section 3, the extension to *spatio-spectral* is discussed. In Section 3, we also describe a series of numerical experiments that use the extended framework to compare different spatio-spectral operators. In Section 4 we consider an alternative way to compute performance for spatially extended anomalies, and show how that affects our results. In Section 5, we observe that spatial processing alters the statistical correlations between pixels, which will affect the covariance matrices used in spectral ACD algorithms, and that this changes the the optimal quadratic detector, and Section 6 uses that approach on a dataset with real anomalies. Finally, Section 7 concludes.

## 2. SIMULATION FRAMEWORK FOR SPECTRAL ANOMALOUS CHANGE DETECTION

The aim of anomalous change detection (ACD) is to distinguish actual interesting anomalous changes in the scene from the incidental uninteresting pervasive differences that appear in the images. The aim of ACD *simulation* is to produce both pervasive differences and anomalous changes that are both realistic and adjustable. This is illustrated in Fig. 1 and described in more detail in the subsections below.

### 2.1 Simulating pervasive differences

Two images of the same scene, taken at different times and under different conditions, will inevitably exhibit differences that are *pervasive* over the entire scene. The most realistic simulation of this effect is to actually take

two separate pictures, and in practice this is often quite feasible.

But it is still useful to simulate pervasive differences, both for cases when a second image is not available, and when one wants to control and/or vary properties of those pervasive differences. There is a lot in interest, for instance, in characterizing how robust ACD algorithms are to misregistration.[10–12] The simulation of pervasive differences that was described in Ref. 3 includes both spectral and spatial operators: brightening, smoothing, translating (to simulate the effect of misregistration), and these operators are applied to every pixel in the image.

### 2.2 Simulating anomalous changes

As with pervasive change, there are multiple ways that anomalous changes can be simulated. But, in distinguishing anomalous change from straight anomalies, we have emphasized approaches in which pixels are resampled from the image. The appeal of these resampled pixels is that although they are acting as anomalous *changes*, they are not by themselves anomalous. Indeed, using only spectral detection, they would be virtually undetectable in a single image; it is only in the context of two images that the pixel and its corresponding pixel in the pervasive difference image constitute an anomalous change.

A further appeal of this approach is computational. Fig. 1 illustrates the anomalous pixel as if it were a single pixel. But one can generate a whole image full of anomalous changes by simply scrambling all the pixels in the original image. The pervasive differences are given by pixel pairs generated from the orignal image and the pervasive difference image; and the anomalous changes are given by the pixel pairs generated from the pervasive difference image and the scrambled original image that is now an image of anomalous changes.

Finally, we note that there is a symmetry to this approach. Whether we scramble the pixels in the orignal image and compare that to the unscrambled pixels in the pervasive difference image, or we keep the original image intact and scramble the pixels in the pervasive difference image, we get the same result.

## 3. EXTENDING THE SIMULATION FRAMEWORK TO SPATIO-SPECTRAL ANOMALOUS CHANGE DETECTION

When we considered the purely spectral anomalous change simulation framework, we took the view that anomalous *changes* should be distinguished from straight anomalies. In particular, we chose a framework that generated anomalous changes from components that were themselves unremarkable.

It is acknowledged that interesting changes in real images may not be consistent with this purist point of view. But by making the distinction between anomalies and anomalous changes, we can develop algorithms that target the latter without being distracted by the former. If an analyst decides that the best targets to go after are some combination of anomalies and anomalous changes, then the analyst can decide what that combination is and combine algorithms accordingly.

Where the purely spectral ACD framework involved moving pixels in an image, the spatio-spectral framework will employ "patches" taken from other parts of the image. Analogously with the pixels, these patches are not, by themselves, anomalous, but represent anomalous change. Unfortunately, this analogy breaks down when the radius over which spatial operators act is larger than the anomalous patch – in the context of those operators, the patch is something of an anomaly in its own right. It could be detected even in a single image. In other words, the elegant philosophy used to guide the development of algorithms in the purely spectral case, cannot be applied with the same rigor for spatio-spectral anomalous change detectors. To begin with, we need to consider separately three kinds of anomalies, which are essentially characterized by their size:

1. Small sub-pixel anomalous changes are assumed to be entirely contained within a single pixel.
2. Medium-sized anomalous changes are larger than a pixel, but smaller than the characteristic size of the spatial image processing operators that are considered.
3. Large anomalous changes are larger than the extent of the spatial processing. Being *anomalous*, however, these changes are still much smaller than the size of the image or the image archive over which the search takes place.

### 3.1 Large anomalous changes

Considering the last first, we note that large anomalies are conceptually easier to deal with.

Since the patches are larger than the spatial processing, the effect of spatial processing on the patch is independent of the location of the patch in the image. It follows that the patch can be moved into place *after* the spatial processing. That is, for the purposes of the simulation, one can apply the desired spatial processing to the two images (the original two images, or the original image and its pervasively different variant), and then apply the ordinary pixel shuffling scheme of the purely spectral simulation framework. One consequence of this is that we don't have to specify the size of the anomaly.

In this regime, as Fig. 2 illustrates, we find that smoothing the images invariably improves the ability to identify large anomalies. It is plausible that other spatial processing would give similar results, and our experiment in Fig. 2 roughly bears that out, but it also suggests that other spatial operators are not as effective as simple smoothing.

### 3.2 Medium anomalous changes

The complications that arise when the anomalies are larger than a pixel but smaller than the characteristic radii of the spatial image processing operators are ... messy. They depend on numerous details of the anomaly: size, shape, nature of the boundary between anomaly and background. And the more one has to specify in the "model" of an anomaly, the more one has strayed from the idea that an anomaly is essentially an unspecified entity.

Unfortunately, this is probably the most common type of anomalous change observed in practice. So we cannot simply ignore it because it is ideologically impure. In our simulations, therefore, we have to make some arbitrary and capricious choices.

We have to choose the size of the anomaly. We also have to choose its shape, and we have taken our patches to be square. This is of course an unreasonable assumption (as would be the assumption that it is circular, or octagonal, or...), but since we have to make some kind of choice, we make one that is easy to implement. To get around the assumption that our anomalies are precisely aligned with the grid of pixels, and also to de-emphasize the effect of edges in the anomaly detection problem, we "feather" the edges in the simulation. A square anomaly of size $a$ is centered in a square of $(a+1) \times (a+1)$ pixels; if we keep $a$ to even integer values, then the center of the anomalous square is also the center of a pixel. This is implemented as an anomalous patch of size $a+1$ whose edges are linear combinations of the patch pixel values and the background pixel values. In our simulations, we make a grid of such anomalies, with each anomaly sufficiently well-separated from the others that spatial processing will treat them independently.

In Fig. 3, we show the performance of smoothing on the anomalous change detection; we see that this performance depends on the size of the anomaly – large anomalies are better detected if the image is smoothed with a large radius kernel, but small anomalies are better detected with little or no smoothing.

### 3.3 Small anomalous changes

All of the details that are needed to characterize a medium-sized anomaly can be squeezed down to a single parameter, once the anomaly is small enough to fit inside a single pixel. This parameter is the area fraction of the anomaly compared to the size of the pixel. While this simplifies the task of generating artificial anomalies, it does not avoid the problem that the simulated anomalous *changes* are also anomalies in their own right. As with the medium-sized anomalies, there is an ideological purity that is lost, as well as a symmetry (it matters which image the anomaly is found in); but the plethora of choices that are imposed on the framework in the case of medium-sized anomalies are at least eliminated in this simpler case.

In Fig. 4, we show the effect of three kinds of spatial processing on the performance of various anomalous change detection algorithms. The smoothing is with a square structuring element and is applied to each of the spectral bands in the image. In stacking, the image and the smoothed image are combined to be a multispectral image with twice as many spectral bands, one smoothed and one original. The high-pass filtering is done by subtracting the smoothed image from the original image.

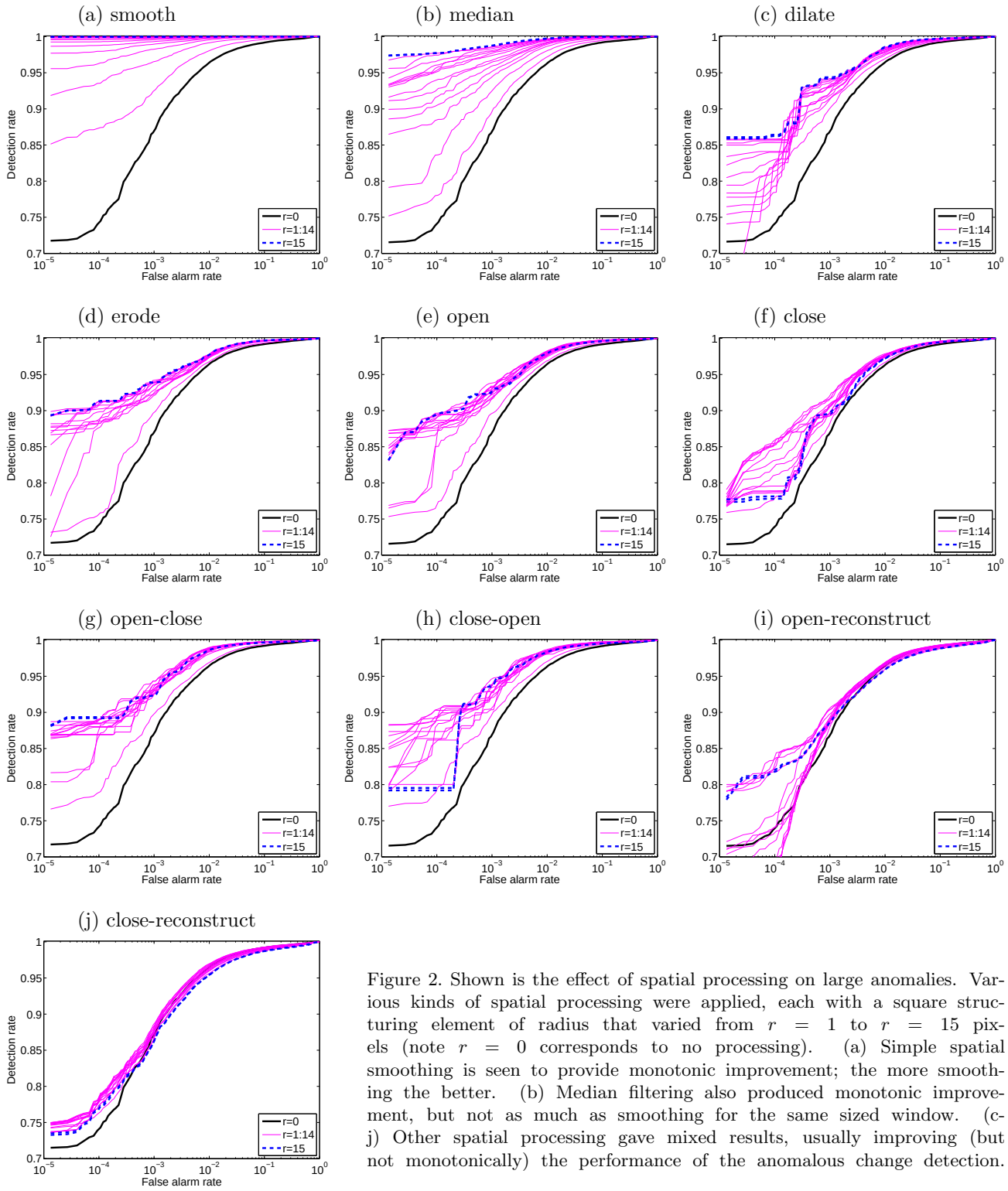


Figure 2. Shown is the effect of spatial processing on large anomalies. Various kinds of spatial processing were applied, each with a square structuring element of radius that varied from $r = 1$ to $r = 15$ pixels (note $r = 0$ corresponds to no processing). (a) Simple spatial smoothing is seen to provide monotonic improvement; the more smoothing the better. (b) Median filtering also produced monotonic improvement, but not as much as smoothing for the same sized window. (c-j) Other spatial processing gave mixed results, usually improving (but not monotonically) the performance of the anomalous change detection.

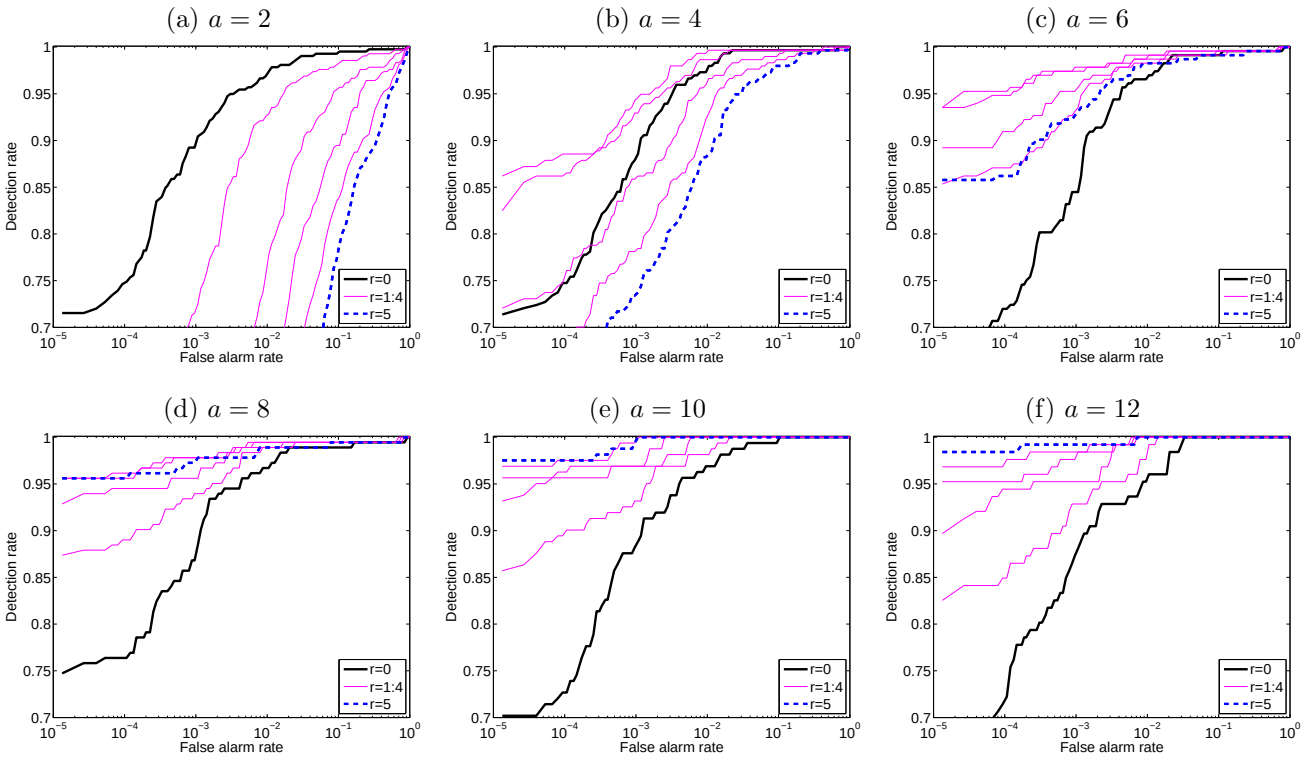


Figure 3. The effect of smoothing on anomalous change detection is shown for a range of anomaly diameters $a$ ranging from 2 to 12 pixels. The smoothing is over a square of radii ranging from $r = 0$ (no smoothing) to $r = 5$ (diameter is 11). The larger the anomaly, the more smoothing helps (this is consistent with the result seen in Fig. 2(a)). In principle, the $r = 0$ curves should be the same for all of these plots; the variation is caused by the randomness in the choice of which points are chosen as anomaly centers.

The broad conclusions are: 1/ straight smoothing always makes things worse (the suppression of the small anomaly far outweighs the suppression of the background variability); 2/ stacking, in which the smoothed data and the original data are treated as if they were different spectral channels, improves performance over the original unprocessed data; 3/ a high-pass filter of the data (which is the difference between the original image and the smoothed image) enhances the subpixel anomalies and improves the ROC curves; in fact, for most algorithms, the high-pass filtered data exhibits better performance than the stacked data; 4/ of the standard algorithms, chronochrome[5] outperformed the "symmetrical" algorithms of covariance equalization,[6] hyper,[7] and subpix-hyper,[13] but chronochrome is asymmetrical, and the "other" chronochrome did not exhibit this performance advantage; 5/ although the subpix detector[13] had the best ROC curve on the original data, its performance for stacked data was remarkably poor; 6/ the "optimal" detector (see Section 5) did indeed exhibit the best performance, and furthermore was able to make better use of the stacked data than any of the other algorithms.

## 4. USE OF DILATION IN EVALUATING AND COMPARING PERFORMANCE

In our simulations so far, we have taken a single pixel at the center of the anomaly as the target, and counted it as a "hit" if that pixel were recognized as an anomalous change. But, because the anomalies are patches, a "hit" anywhere in the patch arguably counts as a detection of an anomalous change. Since the ultimate utility of anomalous change detection is to provide a list of locations to cue an analyst, it is not necessary to hit the precise center of the anomalous patch – it is sufficient to draw the analyst's attention to the presence of an anomalous change. Similarly, it is not necessary to hit every single pixel in the patch. After having been cued, there is little benefit to the analyst in having multiple pixels identified for a single anomalous change.

The idea behind the dilation metric is to take the anomalousness image and dilate it by a fixed radius. In

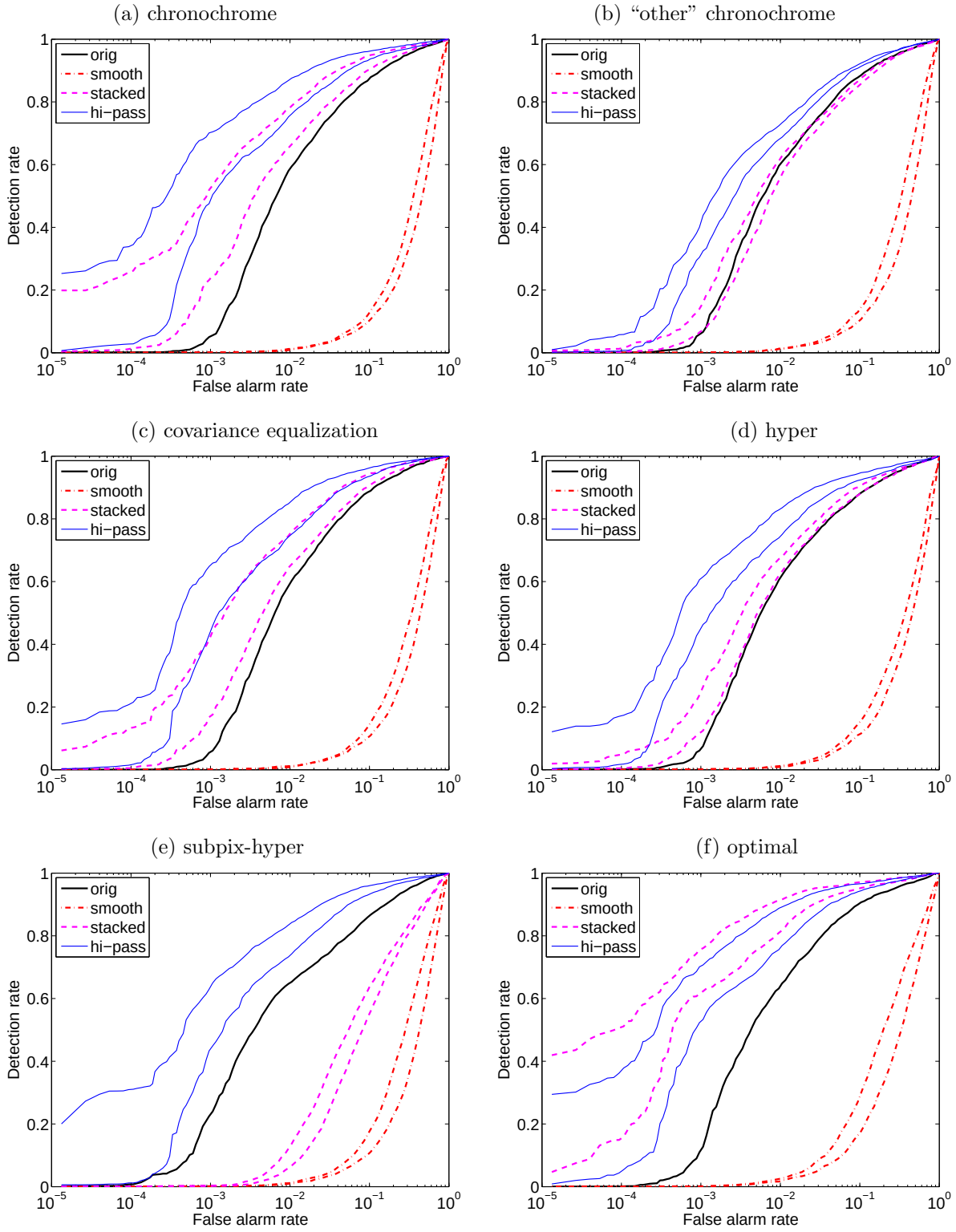


Figure 4. Effect of smoothing, stacking, and high-pass filtering on the ROC curves for simulated subpixel anomalies ($f = 0.25$), using the imagery described in Fig. 1, and shown for six different anomalous change detection algorithms. The smoothing, stacking, and filtering curves are shown for filter radii of $r = 1$ and $r = 2$ and in all cases the $r = 1$ curve is better.

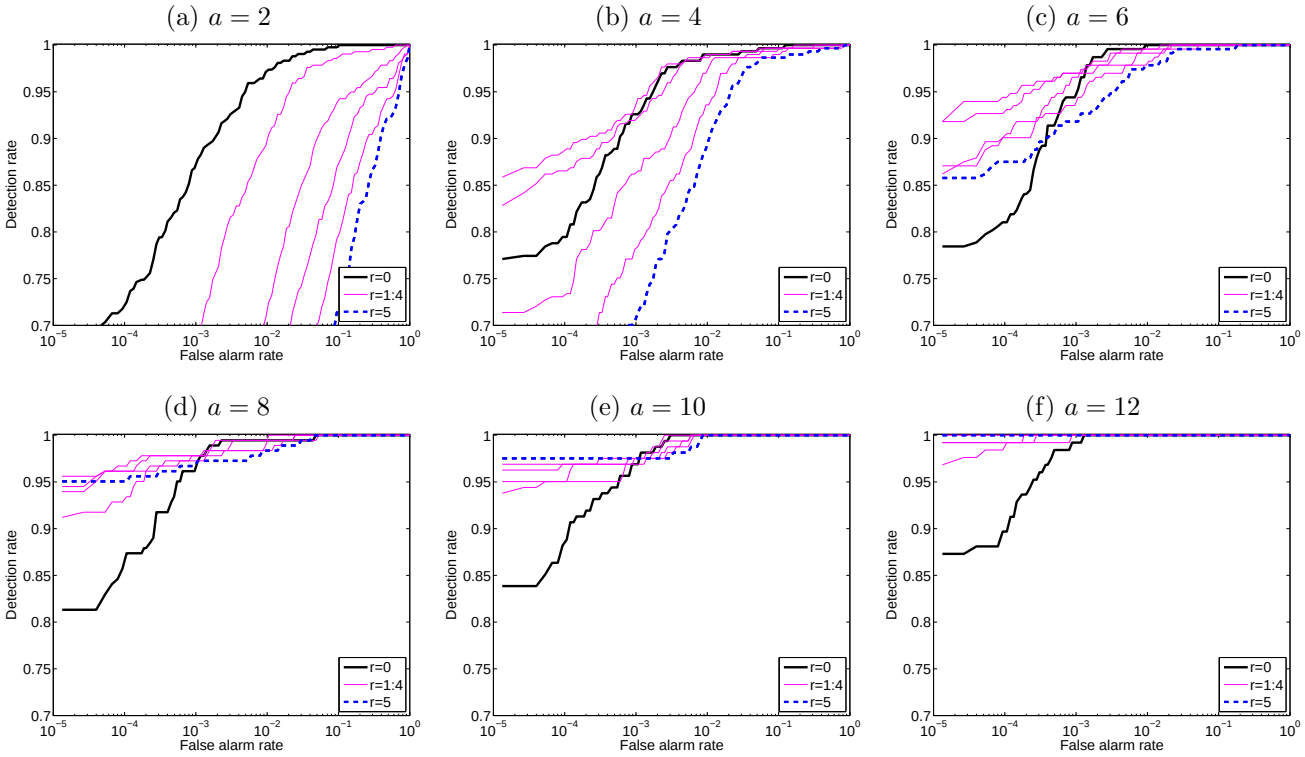


Figure 5. Same as Fig. 3 except that the dilation metric is used to measure performance. With this metric, it is clear that larger anomalies are easier to find, even apart from spatial processing.

this way any "hit" will be expanded by an amount equal to the radius of the dilation. Thus, any hit that is within a radial distance of the center of a target anomalous patch will, after dilation, then overlap the center pixel. Thus, by picking a dilation size appropriate for the size of the anomalies created (or seen) in the data, we can ensure that hits anywhere within the spatial extent of the anomalies will be counted as true detections when calculating the ROC curves.

The effect of including this dilation step is shown in Fig. 5. Although it does not alter our basic conclusions about the relative effects of different spatial processing, it does for instance show that larger anomalies are easier to detect.

## 5. THE PROBLEM WITH HYPER

Let us write $\mathbf{x}$ as a vector-valued pixel in the x-image, and $\mathbf{y}$ as the corresponding pixel in the y-image. It is useful to subtract mean values so that, averaged over the image, $\langle \mathbf{x} \rangle = 0$ and $\langle \mathbf{y} \rangle = 0$. We can describe the pixel pair with a single vector

$$\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}, \tag{1}$$

and in terms of that vector, we can express quadratic anomalous change detectors with an expression of the form[3]

$$\mathcal{A}(\mathbf{z}) = \mathbf{z}^T Q \mathbf{z} \tag{2}$$

where the matrix $Q$ characterizes the detector. When the data can be described by a Gaussian likelihood distribution, then the optimal detector is given by a ratio of likelihoods, and this quadratic detector can be expressed as the difference of two inverse covariance matrices:

$$Q = K_{\text{normal}}^{-1} - K_{\text{anomalous}}^{-1}. \tag{3}$$

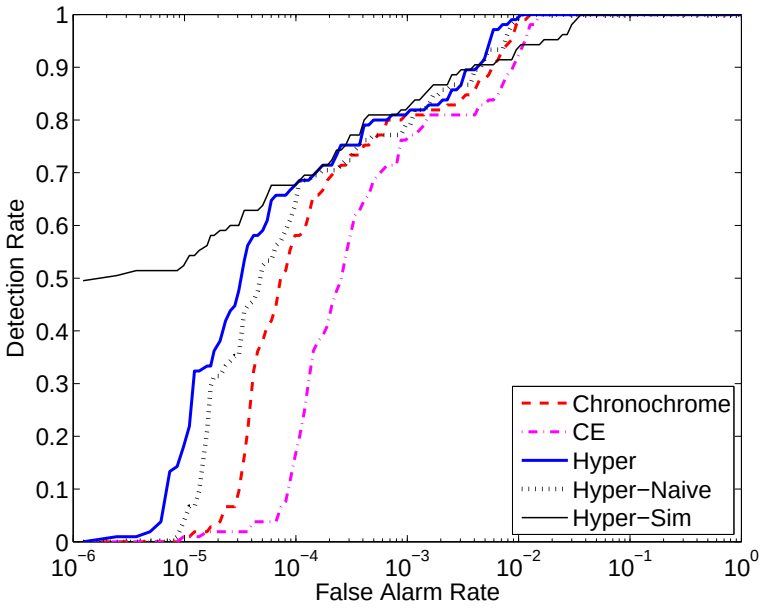


Figure 6. ROC curves illustrate the application of different algorithms to a pair of hyperspectral images which contain "real" anomalies.[1] Chronochrome,[5] covariance equalization (CE),[6] and Hyper[7] are initially applied to the image pair without any spatial processing. After smoothing both images with a $3 \times 3$ kernel, Hyper is naively applied using Eq. (5) for $K_{\text{anomalous}}$. The Hyper-Sim curve is obtained using a $K_{\text{anomalous}}$ that is estimated from simulation.

The covariance matrix associated with the normal data is given by

$$K_{\text{normal}} = \left\langle \mathbf{z}\mathbf{z}^T \right\rangle = \begin{bmatrix} X & C^T \\ C & Y \end{bmatrix}, \tag{4}$$

where $X = \left\langle \mathbf{x}\mathbf{x}^T \right\rangle$, $Y = \left\langle \mathbf{y}\mathbf{y}^T \right\rangle$ and $C = \left\langle \mathbf{y}\mathbf{x}^T \right\rangle$.

When the anomalous changes are produced by the pixel scrambling scheme described in Fig. 1 for the spectral simulation framework, the anomalous y, given by $\hat{\mathbf{y}}$ will have the same covariance as the orignal $\mathbf{y}$ (*i.e.,* $\left\langle \hat{\mathbf{y}}\hat{\mathbf{y}}^T \right\rangle = \left\langle \mathbf{y}\mathbf{y}^T \right\rangle = Y$), but the effect of the scrambling will be to eliminate the cross-covariance; *i.e.,* $\left\langle \hat{\mathbf{y}}\mathbf{x}^T \right\rangle = 0$. Thus, we can simply "write down" $K_{\text{anomalous}}$, given $K_{\text{normal}}$, by setting the cross-covariances to zero:

$$K_{\text{anomalous}} = \begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix}, \tag{5}$$

But the independence of pixels breaks down in the spatio-spectral case (see the example in the Appendix), and the convenience of writing down $K_{\text{anomalous}}$ directly from $K_{\text{normal}}$ is lost. One approach is to go ahead and use Eq. (5) anyway, and this is the standard Hyper algorithm, and its performance is shown in Fig. 4(d). But an alternative is to use simulation to estimate $K_{\text{anomalous}}$, and the performance of this algorithm is seen in Fig. 4(f). This approach, which we call Hyper-Sim, is a more expensive proposition than Hyper, but it can give better performance as seen in Fig. 4(d,f) and in the following section.

## 6. AN APPLICATION WITH REAL ANOMALIES

In Fig. 6, we investigate performance with real anomalous changes using data described in Ref. [1]. As was done there, we projected the hyperspectral data to ten components based on the principal components of a third dataset. Thus, the same projection was applied to both images, and that projection corresponded approximately to principal components of those images. These two images were both taken on October 14, 2005, and in one of the two images a pair of tarps has been laid out in the grass; these are the anomalous changes that we are going to try to detect.

We first applied chronochrome, covariance equalization, and hyper directly to the imagery, without any spatial processing. To generate the additional results we applied spatial smoothing to the images using a flat $3 \times 3$-pixel kernel. The Hyper-Naive result in Fig. 6 is the hyper algorithm applied directly to the smoothed images (as suggested in the large anomalous change case). The Hyper-Sim algorithm is also applied to the smoothed images, but the $K_{\text{anomalous}}$ covariance is different. Specifically, it is estimated from anomalous change images produced within the simulation framework: 1) We sample a regular grid of pixels, where pixels are spaced sufficiently far apart, that the spatial smoothing will not affect adjacent samples. We treat each image independently and randomly swap the pixel values at grid points within each image. This step simulates the spectral component of the anomalous change. 2) The second step is to apply the spatial operator we expect to use, which in our case, is the 3-pixel wide spatial smoothing. 3) The final step is to estimate $K_{\text{anomalous}}$ using pixel values from the grid points. This new covariance leads to the significantly improved result at low false alarm rate, as seen in Fig. 6.

## 7. CONCLUSIONS

This study was motivated, in part, by the observation that spatial processing of imagery can lead to more effective classification of multispectral imagery, and that a systematic scheme for building a spatio-spectral image processing pipeline can be optimized for this task.[14] In asking whether we can extend this idea of an optimized image processing pipeline to the problem of anomalous change detection, we realized that we would first need to find efficient and systematic ways to evaluate any given pipeline. The simulation framework described here is our first cut at providing this systematic evaluation.

While there is a simple, natural, and even symmetrical simulation framework that can be used to evaluate purely spectral ACD algorithms, the extension to spatio-spectral ACD was found to be more complicated. By their very definition, anomalies resist specification. But we found it necessary to specify the characteristic size of the anomalies we wanted to simulate, and we found (not surprisingly) that the performance of spatio-spectral ACD algorithms depended on the size of those anomalies.

For large anomalous changes, the more smoothing the better. Smoothing reduces the variability of the background without washing out the anomalous difference. Meanwhile for medium-sized anomalies, some smoothing helps, but too much washes out the anomalous difference.

It is possible to identify subpixel anomalous changes by purely spectral methods,[13] but our investigation concentrated on enhancing that detection with spatial methods. Although smoothing degraded the detection, effectively "washing out" the anomalies more than reducing background variance, we found that a stack composed of the original image and a smoothed image did enhance the detection. Closer examination revealed that a linear combination of these images resembled a sharpening, or a high-pass filtering, of the image, and using the high-passed images directly also improved performance.

## ACKNOWLEDGMENTS

We are grateful to Joseph Meola and Michael Eismann for generously providing hyperspectral datasets from their change detection experiment. This work was supported by the Laboratory Directed Research and Development (LDRD) program at Los Alamos National Laboratory.

## REFERENCES

[1] Eismann, M. T., Meola, J., and Hardie, R., "Hyperspectral change detection in the presence of diurnal and seasonal variations," *IEEE Trans. Geoscience and Remote Sensing* **46**, 237–249 (2008).

[2] Cohen, Y. and Rotman, S. R., "Spatial-spectral filtering for the detection of point targets in multi- and hyperspectral data," *Proc. SPIE* **5806**, 47–55 (2005).

[3] Theiler, J., "Quantitative comparison of quadratic covariance-based anomalous change detectors," *Applied Optics* **47**, F12–F26 (2008).

[4] Nielsen, A. A., Conradsen, K., and Simpson, J. J., "Multivariate alteration detection (MAD) and MAF post-processing in multispectral bi-temporal image data: new approaches to change detection studies," *Remote Sensing of the Environment* **64**, 1–19 (1998).

[5] Schaum, A. and Stocker, A., "Long-interval chronochrome target detection," in [*Proc. 1997 International Symposium on Spectral Sensing Research*], (1998).

[6] Schaum, A. and Stocker, A., "Hyperspectral change detection and supervised matched filtering based on covariance equalization," *Proc. SPIE* **5425**, 77–90 (2004).

[7] Theiler, J. and Perkins, S., "Proposed framework for anomalous change detection," in [*ICML Workshop on Machine Learning Algorithms for Surveillance and Event Detection*], 7–14 (2006).

[8] Prasad, L. and Theiler, J., "A structural framework for anomalous change detection and characterization," To appear in: *Proc. SPIE* **7341** (2009).

[9] Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), Jet Propulsion Laboratory (JPL), National Aeronautics and Space Administration (NASA) `http://aviris.jpl.nasa.gov/`.

[10] Schaum, A. and Allman, E., "Advanced algorithms for autonomous hyperspectral change detection," in [*33rd Applied Imagery Pattern Recognition (AIPR) Workshop: Emerging technologies and applications for imagery pattern recognition*], 33–38 (2005).

[11] Meola, J. and Eismann, M. T., "Image misregistration effects on hyperspectral change detection," *Proc. SPIE* **6966**, 69660Y (2008).

[12] Theiler, J., "Sensitivity of anomalous change detection to small misregistration errors," *Proc. SPIE* **6966**, 69660X (2008).

[13] Theiler, J., "Subpixel anomalous change detection in remote sensing imagery," in [*IEEE Southwest Symposium on Image Analysis and Interpretation. SSIAI 2008.*], 165–168 (Mar. 2008).

[14] Harvey, N. R., Theiler, J., Brumby, S. P., Perkins, S., Szymanski, J. J., Bloch, J. J., Porter, R. B., Galassi, M., and Young, A. C., "Comparison of GENIE and conventional supervised classifiers for multispectral image feature extraction," *IEEE Trans. Geosci. and Remote Sens.* **40**, 393–404 (2002).

[15] Eismann, M. T., Meola, J., Stocker, A. D., Beaven, S. G., and Schaum, A. P., "Airborne hyperspectral detection of small changes," *Applied Optics* **47**, F27–F45 (2008).

## APPENDIX A. EXAMPLE OF INDUCED CROSS-COVARIANCE BY SPATIAL PROCESSING

In Section 5, we remarked that spatial processing can *induce* correlations that would not be present in the unprocessed data. There is, in particular, an assumption associated with the Hyper algorithm[7] that the cross-covariance of the x-image and y-image is zero for the anomalous pixels. But if the anomaly is smaller than the size of the spatial processing, then the effect of spatial processing is to produce a nonzero cross-covariance, and this leads to the Hyper algorithm being outperformed by other algorithms. To see, with simple but concrete example, how this nonzero cross-covariance arises, consider the following:

Consider a pair of one-dimensional images with $1 \times N$ pixels. Let $x_i$ denote the $i$th pixel in the first image, and $y_i$ be the corresponding pixel in the second image. To keep things simple, let the images be normalized ($\langle x_i^2 \rangle = \langle y_i \rangle^2 = 1$), and without spatial correlation: $\langle x_i x_j \rangle = \langle y_i y_j \rangle = \delta_{ij}$, where $\delta_{ij}$ is the Kronecker delta function: it is one when $i = j$ and zero otherwise. Now, the images *are* correlated with each other (after all, they are two images of the same scene), and in particular, we have $\langle x_i y_i \rangle = c$ for some correlation factor $c < 1$. But the correlation only holds at corresponding pixels: so $\langle x_i y_j \rangle = c\delta_{ij}$.

Let us introduce a single-pixel anomaly into the $y$ image at position $t$; then $\hat{y}_t = y_s$ where $s$ is a random position in the image.

First, before we do any spatial processing, let us look at the covariance matrices associated with this data. Recall that $\mathbf{z} = \begin{bmatrix} x \\ y \end{bmatrix}$, and write

$$K_{\text{normal}} = \left\langle \mathbf{z}_i \mathbf{z}_i^T \right\rangle = \begin{bmatrix} \langle x_i x_i \rangle & \langle x_i y_i \rangle \\ \langle y_i x_i \rangle & \langle y_i y_i \rangle \end{bmatrix} = \begin{bmatrix} 1 & c \\ c & 1 \end{bmatrix} \tag{6}$$

and

$$K_{\text{anomalous}} = \left\langle \hat{\mathbf{z}}_t \hat{\mathbf{z}}_t^T \right\rangle = \begin{bmatrix} \langle \hat{x}_t \hat{x}_t \rangle & \langle \hat{x}_t \hat{y}_t \rangle \\ \langle \hat{y}_t \hat{x}_t \rangle & \langle \hat{y}_t \hat{y}_t \rangle \end{bmatrix} = \begin{bmatrix} \langle x_t x_t \rangle & \langle x_t y_s \rangle \\ \langle y_s x_t \rangle & \langle y_s y_s \rangle \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{7}$$

Notice that the off-diagonal cross-covariance terms of $K_{\text{anomalous}}$ are zero.

Now, consider the case for stacked images, where one layer of the stack is the original image and the second layer is smoothed with a smoothing radius $r = 1$:

$$\mathbf{x}_i = \begin{bmatrix} x_i \\ (x_{i-1} + x_i + x_{i+1})/3 \end{bmatrix}, \tag{8}$$

and similarly for $\mathbf{y}_i$. Then, we can work out that

$$K_{\text{normal}} = \begin{bmatrix} \langle \mathbf{x}_i \mathbf{x}_i^T \rangle & \langle \mathbf{x}_i \mathbf{y}_i^T \rangle \\ \langle \mathbf{y}_i \mathbf{x}_i^T \rangle & \langle \mathbf{y}_i \mathbf{y}_i^T \rangle \end{bmatrix} = \begin{bmatrix} 1 & 1/3 & c & c/3 \\ 1/3 & 1/3 & c/3 & c/3 \\ c & c/3 & 1 & 1/3 \\ c/3 & c/3 & 1/3 & 1/3 \end{bmatrix}. \tag{9}$$

For the anomalous pixels, we have

$$\hat{\mathbf{y}}_t = \begin{bmatrix} \hat{y}_t \\ (y_{t-1} + \hat{y}_t + y_{t+1})/3 \end{bmatrix} = \begin{bmatrix} y_s \\ (y_{t-1} + y_s + y_{t+1})/3 \end{bmatrix}. \tag{10}$$

This gives

$$\langle \hat{\mathbf{y}}_t \hat{\mathbf{y}}_t^T \rangle = \begin{bmatrix} 1 & 1/3 \\ 1/3 & 1/3 \end{bmatrix}, \tag{11}$$

and

$$\langle \hat{\mathbf{y}}_t \hat{\mathbf{x}}_t^T \rangle = \langle \hat{\mathbf{y}}_t \mathbf{x}_t^T \rangle = \begin{bmatrix} 0 & 0 \\ 0 & 2c/9 \end{bmatrix}. \tag{12}$$

Thus we see that the cross-covariance $\langle \hat{\mathbf{y}}_t \hat{\mathbf{x}}_t^T \rangle$ at the anomalous pixels $t$ in the image is now nonzero. This leads to

$$K_{\text{anomalous}} = \begin{bmatrix} 1 & 1/3 & 0 & 0 \\ 1/3 & 1/3 & 0 & 2c/9 \\ 0 & 0 & 1 & 1/3 \\ 0 & 2c/9 & 1/3 & 1/3 \end{bmatrix}. \tag{13}$$

and thus the optimal quadratic coefficient matrix $Q = K_{\text{normal}}^{-1} - K_{\text{anomalous}}^{-1}$ differs from what the Hyper algorithm would prescribe.