Geostatistical and local cluster analysis of high resolution hyperspectral imagery for detection of anomalies

Pierre Goovaerts, Geoffrey M. Jacquez, Andrew Marcus

BioMedware, Inc., 516 North State Street, Ann Arbor, MI 48104, United States
Department of Geography, University of Oregon, United States

Received 13 September 2004; received in revised form 18 December 2004; accepted 27 December 2004

Abstract

This paper describes a new methodology to detect small anomalies in high resolution hyperspectral imagery, which involves successively: (1) a multivariate statistical analysis (principal component analysis, PCA) of all spectral bands; (2) a geostatistical filtering of noise and regional background in the first principal components using factorial kriging; and finally (3) the computation of a local indicator of spatial autocorrelation to detect local clusters of high or low reflectance values and anomalies. The approach is illustrated using 1 m resolution data collected in and near northeastern Yellowstone National Park. Ground validation data for tarps and for disturbed soils on mine tailings demonstrate the ability of the filtering procedure to reduce the proportion of false alarms (i.e., pixels wrongly classified as target), and its robustness under low signal to noise ratios. In almost all scenarios, the proposed approach outperforms traditional anomaly detectors (i.e., RX detector which computes the Mahalanobis distance between the vector of spectral values and the vector of global means), and fewer false alarms are obtained when using a novel statistic $S_2$ (average absolute deviation of $p$-values from 0.5 through all spectral bands) to summarize information across bands. Image degradation through addition of noise or reduction of spectral resolution tends to blur the detection of anomalies, increasing false alarms, in particular for the identification of the least pure pixels. Results from a mine tailings site demonstrate the approach performs reasonably well for highly complex landscape with multiple targets of various sizes and shapes. By leveraging both spectral and spatial information, the technique requires little or no input from the user, and hence can be readily automated.

Keywords: High resolution hyperspectral imagery; Principal component analysis; Factorial kriging

1. Introduction

Spatial data are periodically collected and processed to monitor, analyze, and interpret environmental changes. The recent availability of high spatial resolution hyperspectral (HSRH) imagery offers great potential for enhancing environmental mapping and modelling of spatial systems (Aspinall et al., 2002; Koger et al., 2003; Marcus, 2002; Vaughan et al., 2003). Following Jacquez et al. (2002), HSRH images refer to images with spatial resolution of less than 5 m and include data collected over 64 or more spectral bands. High spatial resolution imagery contains a remarkable quantity of information that could be used to analyze spatial breaks (boundaries), areas of similarity (clusters), and spatial autocorrelation (associations) across the landscape. This paper addresses the specific issue of detecting local anomalies defined as a pixel or small group of pixels that differ in reflectance from surrounding pixels. We focus first on artificial targets with distinct boundaries and dimensions, before applying the technique to the example of disturbed soils. Disturbed soils provide a realistic real world application, because they can indicate a host of disturbance processes ranging from animal burrows to slope erosion to troop movements and land mines (DePersia et al., 1995). A challenge presented by detecting local-scale soil
disturbance is to retain the measurement of fine-scale features (e.g., mineral soil changes, organic content changes, vegetation disturbance related changes, and aspect changes) while still covering large spatial areas. An additional difficulty in remote locations, with military applications, or using historical imagery, is that ground-truth data are often unavailable for the calibration of spectral signatures, and little might be known about the size of the patches to be detected. Regardless of whether it is soil disturbance or some other anomaly, precise and accurate identification typically requires: (1) identification of a potential target of interest, (2) removal of confusion (the environmental setting), and (3) target confirmation. These different steps should be automated as much as possible to allow for the rapid processing of multiple images, while false positives should be reduced to an acceptable level.

Spectral analysis has been the classical approach used in the remote sensing community to identify discrete feature classes, like bare soil (the target or “needle in the haystack”). Spectral analysis approaches range from relatively simple “maximum likelihood classification” techniques found in any introductory remote sensing textbook (e.g., Jensen, 1996) to significantly more complex approaches developed in recent years (Chang, 2003). For example, spectral feature fitting matches image spectra to selected reference spectra from a spectral library (Clark et al., 1990, 1991; Crowley & Clark, 1992; Swayze & Clark, 1995). Spectral unmixing (Boardman, 1989, 1993) determines the relative abundance of materials based on the spectral characteristics of those materials. This approach requires spectral library inputs as well and can be highly accurate, but can fail to work if some spectral end members of the image have not been input as part of the library. Matched filtering (Boardman et al., 1995; Harsanyi & Chang, 1994) performs an unmixing of spectra to estimate the abundance of user-defined endmembers (e.g., bare soil, grass, water, etc.) within each pixel of a scene. This approach has the advantage that it does not require knowledge of all the endmembers within an image scene and can be used to identify single feature types. Mixture tuned matched filtering (Boardman, 1998; Williams & Hunt, 2002) allows the user to map a target object without knowledge of all endmember signatures and reduces the incidence of false positives relative to matched filtering used on its own. In this paper, the proposed classifiers will be compared to anomaly detectors, such as the RX detector or the low-probability detector (LPD), which enable the detection, with no a priori knowledge, of small targets (i.e., with a low probability of occurrence in the image scene) whose signatures are spectrally distinct from their surroundings (Chang & Chang, 2002).

A limitation of all spectral approaches is that they account only for the correlation between spectral bands and neglect the correlation between neighboring pixels (Atkinson, 1999). In particular for detection of local-scale soil disturbances, it is expected that the target pixels exhibit distinct behaviors not only in the spectral space, but also in the physical space where contrasts should be observed with pixels geographically close. A major challenge facing the use of HSRH data is thus the development of new, spatially explicit tools that exploit both the spectral and spatial dimensions of the data. Goovaerts (2002) recently developed a methodology to incorporate both hyperspectral properties and spatial coordinates of pixels in maximum likelihood classification, and demonstrated its benefit in terms of classification accuracy. This approach however relies on the availability of spectral signatures and thus cannot be utilized for the particular application addressed in this paper.

An increase of use of spatial statistics in the analysis of remotely sensed data has occurred in the last decade (Stein et al., 1999). In particular, geostatistics offers a broad range of techniques that allow not only the characterization of multivariate spatial correlation, but also the spatial decomposition or filtering of signal values (Goovaerts, 1997). The approach known as factorial kriging relies on semivariograms to detect multiple scales of spatial variability (i.e., noise and short range or long range variability), followed by the decomposition of spectral values into the corresponding spatial components (Wackernagel, 1998). This technique was first used in geochemical exploration to distinguish large isolated values (pointwise anomalies) from groupwise anomalies that consisted of two or more neighboring values just above the chemical detection limit (Sandjivy, 1984). Ma and Royer (1988) applied the same technique to image restoration, filtering and lineament enhancement, while Wen and Sinding-Larsen (1997) analyzed sonar images. Oliver et al. (2000) used factorial kriging to separate short-range spatial components, which seem to represent patchiness in the ground cover, from long-range components which seem to reflect the coarser pattern in SPOT images imposed by the gross physiography. More recently, Van Meirvenne and Goovaerts (2002) applied factorial kriging to the filtering of multiple SAR images, strengthening relationships with land characteristics, such as topography and land use. None of these studies, however, have addressed the issue of automatic analysis and processing of large series of correlated spectral bands, nor the problem of detecting small anomalous targets in the image scene.

This paper describes a new technique for automatic target detection, which capitalizes on both spatial and spectral bands correlation and does not require any a priori information on the target spectral signature. The technique does not allow discrimination between types of anomalies. This approach combines geostatistical filtering for suppression of image background with local indicators of spatial autocorrelation (LISA), which are used routinely in health sciences for the detection of clusters and outliers in cancer mortality rates (Jacquez & Greiling, 2003). The LISA statistic allows the comparison of an observation (i.e., here a single pixel or small group of pixels) with the surrounding ones, followed by a test procedure to assess whether this difference is significant or not. This approach has been used
recently to detect spatial outliers in soil samples (McGrath & Zhang, 2003), while the LISA has been introduced to quantify the degree of spatial homogeneity in remotely sensed imagery (LeDrew et al., 2004). The novelty of the proposed approach lies in the geostatistical filtering of the image regional background prior to testing the significance of LISA values through randomization, and the development of two new statistics to combine test results across multiple spectral bands.

The approach is illustrated using two case studies: 1) a scene including artificial targets with distinct boundaries and dimensions, and (2) a mine tailings site that has a highly complex landscape with multiple targets of various sizes and shapes. Performance of the method—i.e., probabilities of false alarms versus probabilities of detection—is quantified using ground data and compared to the common RX detection algorithm. Sensitivity analysis is conducted to investigate the impact of spectral resolution, signal to noise ratio (SNR), and kernel detection size on classification accuracy.

2. Methods

Consider the problem of detecting, across an image, single or aggregated pixels that are significantly different from surrounding ones. The information available consists of K variables (i.e., original spectral values or combinations of those) recorded at each of the N nodes of the image, \{zk(u)i, i=1, ..., N; k=1, ..., K\}, where u is the vector of spatial coordinates of the ith pixel. In this section, we describe first a non-spatial anomaly detector, then the geostatistical methodology to account for the spatial pattern of autocorrelation.

2.1. The RX detector

The RX detector developed by Reed and Yu (1990) computes at each pixel \( u \) the Mahalanobis distance between the vector of spectral values at \( u \), \( Z(u) \), and the vector of global means \( \mu \):

\[
\delta_{RXD}(u) = (Z(u) - \mu)^T C^{-1} (Z(u) - \mu) \tag{1}
\]

where \( C \) is the \( K \times K \) variance–covariance matrix between the spectral bands, \( Z(u)=[z_1(u), ..., z_K(u)] \), and \( \mu=[\mu_1, ..., \mu_K] \). Assuming that each variable \( z_k \) has been rescaled to a zero mean and unit variance, the variance–covariance matrix \( C \) in expression (1) is now the correlation matrix \( R \), while \( \mu \) is the null vector. Then, following Chang and Chang (2002), the RXD statistic becomes:

\[
\delta_{RXD}(u) = Z(u)^T R^{-1} Z(u) = \sum_{k=1}^{K} \frac{1}{\lambda_k} y_k^2(u) \tag{2}
\]

where \( \lambda_k \) are the eigenvalues of the correlation matrix \( R \) and \( y_k(u) \) are the principal component (PC) scores at location \( u \). In other words, the detection statistic is a linear combination of PC scores where more weight is given to the last principal components, the ones with the smallest variance or eigenvalue \( \lambda_k \). Indeed, if the image contains few target pixels (i.e., small probability of occurrence), it is likely that these targets will not show up in the major principal components, but rather in the minor components that explain a small proportion of the global variance and are associated with small eigenvalues. This phenomenon was observed and demonstrated in Chang and Heinz (2000). This weighting of the inverse of the PCs is also shared by signal identification methods, which aim to divide the at-sensor radiance received from a pixel into signal and noise or clutter components: orthogonal subspace projection (Harsanyi & Chang, 1994), orthogonal background suppression (Hayden et al., 1996), and matched filters (Funk et al., 2001). The risk, however, is to give too much importance to minor noisy components; hence, in practice, the RXD statistic incorporates only a smaller subset of the first \( t \) components:

\[
\delta_{RXD}(u) = \sum_{k=1}^{t} \frac{1}{\lambda_k} y_k^2(u) \text{ with } t \ll K \tag{3}
\]

The need to determine a priori the intrinsic dimensionality of the data set, hence the \((K-t)\) eigenvalues to be discarded in the analysis (Chang, 2003), can be a weakness of the approach. Another limitation is that the classification of pixel \( u \) as a target or not is made independently of the spectral properties of surrounding pixels.

2.2. Geostatistical methodology

In the RXD approach, principal component analysis is used as an indirect way to remove or attenuate the image background signature in order to facilitate the detection of anomalies. In this paper, we use the pattern of spatial autocorrelation to filter the background signal. Then, at each location across the filtered image, the value of a detection kernel whose size corresponds to the expected size of an anomaly is compared to neighborhood values and flagged as an anomaly if its value is significantly higher or lower than surrounding pixel values.

2.2.1. Geostatistical filtering

The first step involves removing from each image, which can be the original spectral bands or principal component bands, the low-frequency component or regional variability. For the kth image, the low-frequency component, denoted \( m_k \), is estimated at each location \( u \) as a linear combination of the n surrounding pixel values:

\[
m_k(u) = \sum_{i=1}^{n} \hat{\lambda}_{ik} \times z_k(u_i) \text{ with } \sum_{i=1}^{n} \hat{\lambda}_{ik} = 1 \tag{4}
\]

where \( \hat{\lambda}_{ik} \) is the weight assigned to the ith observation in the filtering window of size \( n \). Expression (4) is equivalent to a kernel smoothing. The main feature of this filtering
The technique is that the weights $\lambda_{ik}$ are tailored to the spatial pattern of correlation displayed by each image and quantified using the semivariogram, which is estimated as:

$$\hat{\gamma}_k(h) = \frac{1}{2N(h)} \sum_{x=1}^{N(h)} \left[ z_k(u_x + h) - z_k(u_x) \right]^2$$  \hspace{1cm} (5)

where $N(h)$ is the number of data pairs separated by the vector $h$. The experimental semivariograms are here computed in four different directions (row, column, diagonals) and a model is fitted automatically using weighted least-square regression (Pardo-Iguzquiza, 1999). The semivariogram model is then used to solve the following system of linear equations and compute the weights $\lambda_{ik}$:

$$\sum_{j=1}^{n} \lambda_{jk} \gamma_k(u_i - u_j) + \mu_k(u) = 0 \quad i = 1, \ldots, n$$  \hspace{1cm} (6)

$$\sum_{j=1}^{n} \lambda_{jk} = 1$$

where $\gamma_k(u_i - u_j)$ is the semivariogram of the $k$th image for the separation vector between $u_i$ and $u_j$, and $\mu_k$ is a Lagrange multiplier that results from minimizing the estimation variance subject to the constraint that the estimator is unbiased (i.e., the expected prediction error is zero). System (6) is known as “kriging of the local mean” in the geostatistical literature (Goovaerts, 1997).

### 2.2.2. Detection of anomalies using the local Moran’s I

The second step scans each filtered image, looking for local values that are significantly lower or higher than the surrounding values and thus might indicate an anomaly. This procedure requires the definition of:

1. A detection kernel, whose size corresponds to the expected size of the anomalies,

2. A LISA (Local Indicator of Spatial Autocorrelation) neighborhood, which includes the pixels surrounding the detection kernel, and

3. A target area which is the area to be analyzed.

An example of these three parameters is provided in Fig. 1. The detection of local anomalies is based on local Moran’s I, which is the most commonly used LISA statistic (Anselin, 1995). This statistic is computed for each pixel $u$ and spectral variable $z_k$ as:

$$\text{LISA}_k(u) = \bar{r}_k(u) \left[ \frac{1}{J} \sum_{j=1}^{J} r_k(u_j) \right]$$  \hspace{1cm} (7)

where $\bar{r}_k(u)$ is the average value of the residuals, $r_k(u) = z_k(u) - m_k(u)$, over the detection kernel centered on pixel $u$, and $J$ is the number of pixels in the LISA neighborhood (e.g., $J=12$ for the $2 \times 2$ detection kernel in the example of Fig. 1). Moran’s I can be interpreted as a local and spatially weighted form of Pearson’s correlation coefficient. Since the residuals have zero mean, the LISA statistic takes negative values if the kernel average is much lower or higher than the surrounding values, which indicates negative local autocorrelation and presence of spatial outliers. For example, the LISA value will be negative if the kernel average is below the global zero mean, while the neighborhood average is above the global zero mean, or if the converse occurs. Clusters of low or high values, which correspond to the presence of positive local autocorrelation, will lead to positive values of the LISA statistic (i.e., both kernel and neighborhood averages are jointly above zero or below zero).

In addition to the sign of the LISA statistic, its magnitude informs on the extent to which kernel and neighborhood values differ. To test whether this difference is significant or not, a Monte Carlo simulation is conducted, which consists

---

Fig. 1. Illustration of key parameters used in the geostatistical detection procedure. The LISA (Local Indicator of Spatial Autocorrelation) statistics compare the averaged reflectance within the detection kernel to the averaged reflectance of neighborhood pixels.
of sampling randomly and without replacement from the target area and computing the corresponding simulated neighborhood averages. This operation is repeated many times (e.g., 1000 draws) and these simulated values are multiplied by the detection kernel average $\bar{r}(u)$ to produce a set of 1000 simulated values of the LISA statistic at $u$. This set represents a numerical approximation of the probability distribution of the LISA statistic at $u$, under the assumption of spatial independence. The observed LISA statistic, LISA$_d(u)$, can then be compared to the probability distribution, allowing the computation of the $p$-value, which is the probability that this observed value could be exceeded:

$$p_k(u) = \text{Prob}\{L > \text{LISA}_d(u) | \text{randomization}\}$$

Large $p$-values thus indicate large negative LISA statistics, corresponding to small values surrounded by high values or the reverse (anomalies). Conversely, small $p$-values correspond to large positive LISA statistics, which indicates clusters of high or low values.

The last step is to combine the $K$ $p$-values computed for the set of $K$ images. Two novel statistics were developed to summarize for each node $u$ the information provided by the $K$ bands and to detect target pixels:

1. Average $p$-value over the subset of $K'$ bands that display negative LISA statistics:

$$S_1(u) = \frac{1}{K'} \sum_{k=1}^{K'} i(u; k) p_k(u) \quad \text{and} \quad K' = \sum_{k=1}^{K} i(u; k)$$

with $i(u; k)=1$ if LISA$_d(u)<0$, and zero otherwise. Large $S_1$ values indicate local anomalies (i.e., sample LISA statistic in the left tail of the distribution).

2. Average absolute deviation of $p$-values from 0.5 through the $K$ bands:

$$S_2(u) = \frac{1}{K} \sum_{k=1}^{K} |p_k(u) - 0.5|$$

Large $S_2$ values indicate either clusters or anomalies (i.e., sample LISA in either tails of the distribution).

The different steps of the analysis are fully automated. For example, the entire processing of 25 principal component bands for the first scene displayed in Fig. 2 (131×69 pixels) takes 16.0 s on a Pentium 3.20 GHz.

2.3. Receiver operating characteristics curve

Target detection requires applying a threshold to the maps of statistics $\delta_{RXD}$, $S_1$, or $S_2$ and classifying as anomalies all pixels exceeding this threshold. Instead of selecting a single threshold arbitrarily, it is better to select a series of thresholds and see how the proportion of pixels correctly or incorrectly classified evolves. This information can then be summarized in receiver operating characteristics (ROC) curves that plot the probability of false alarm versus the probability of detection (Swets, 1988). ROC curves will be used to compare the performances of various detection methods under different spectral resolutions, signal to noise ratios, and kernel detection sizes.

3. Field area and data sets

3.1. Field area

All data used in this study were collected in the northern boundary area of Yellowstone National Park, Wyoming and Cooke City, Montana, a small town just northeast of the park. This study focused on two areas: a set of four tarps marking vegetation field sites near a footbridge on Soda Butte Creek, and mine tailings near Cooke City. Probe-1 data collected in the same area were used in several other studies; further descriptions of the field area and procedures are contained in those reports (Goovaerts, 2002; Legleiter et al., 2002; Marcus, 2002; Marcus et al., 2003; Maruca & Jacquez, 2002).

3.2. Data sets

Data were collected on August 2 and 3, 1999 using the Probe-1 sensor, a 128-band hyperspectral system operated by Earth Search Systems. In order to avoid midday cloud buildup, images were acquired at approximately 10:30 a.m. mountain daylight time, 3 h prior to solar noon. The solar azimuth was 68° east of south and the solar altitude was approximately 44.4°. Data were not converted to reflectance values or atmospherically corrected, thus simulating more closely the real time processing demands one might encounter when applying the detection algorithms in a hostile environment where ground calibration data are unavailable (e.g., for detection of land mines).

The Probe-1 sensor is a cross-track scanner with a 60° field of view and average full width at half maximum (FWHM) band widths ranging from 16 nm in the visible to 13 nm in the near infrared to 17 nm in the shortwave infrared spectra (ESSI, 2004). Spectral coverage ranges from 438 to 2507 nm. Data are 11-bit radiometric resolution. This sensor uses four 32-element linear detector arrays (one Si and three InSb). Energy is separated into discrete spectra by 4 dispersive grating spectrometers. The instrument has a signal to noise ratio that exceeds that of any satellite sensor. The Probe-1 is designed to be operated on a stabilized camera mount in a twin-engine aircraft. To obtain 1 m resolution data, the Probe-1 sensor was mounted on an A-Star Aerospatiale helicopter flying approximately 600 m above the ground.
The images were degraded in two ways in order to investigate the robustness of the approach with respect to spectral resolution and signal to noise ratio. The data were first spectrally resampled to 2–3 times lower resolutions, by simply selecting fewer bands in a systematic way (e.g., every other band is selected for reduction of the spectral resolution).
Fig. 4. Maps of the first two principal components for the tarp scene, and the results of the geostatistical filtering of the regional background. Images are derived from the original, unaltered Probe-1 imagery.

Fig. 5. Maps of the three detection statistics computed from 84 principal components before (left column) and after (right column) filtering of the regional background. Images are derived from the original, unaltered Probe-1 imagery.
resolution by two). Noise was added to simulate 50:1 and 100:1 SNRs, according to: $R_{sn}(\lambda) = R_s(\lambda)[1 + \{N(0,1)/SNR(\lambda)\}]$, where $R_{sn}(\lambda)$ is the simulated, noisy spectrum, $R_s(\lambda)$ is the spectrum that has been spectrally resampled, $N(0,1)$ is a Gaussian random number with a zero mean and unit variance, and $SNR(\lambda)$ is the simulated signal-to-noise ratio. PCA was conducted on degraded spectral values and up to the 84 first PCs were used in the subsequent analysis. PCA is a commonly used approach to highlight anomalies as these pixels covary differently than dominant image components (Olsen et al., 1997; Richards, 1994). Analysis of PC bands is also computationally less intensive since the data are condensed into fewer bands. Last, PCs can be used for the computation of the RX detector through expression (3) as well as input to the geostatistical procedure.

We selected a sagebrush vegetation test plot as the initial site for testing the detection algorithms (Fig. 2, top map). Ground cover in the plot consisted of sage brush, senesced grasses, forbs, and soil, as well as 4 blue plastic tarps. The tarps were 2 by 2 m each (i.e., 16 pixels total), mark the corners of the plot, and appear as white pixels in the scene of Fig. 2 (131×69 pixels). The tarps provide a simple target for testing the algorithms because they have reflectances that are markedly dissimilar from that of the surrounding sage and dry grass (Fig. 3). Linear spectral unmixing (Boardman, 1993) was performed on the tarp data and an index of map purity was computed for each of these 16 target pixels to determine the effects of mixed pixels on the detection algorithms.

To confirm the robustness of the methodology for detecting actual disturbed soils, we next analyzed a larger...
(270×145 pixels) and more complicated scene (Fig. 2, bottom picture). This mine tailings site provides a much more realistic setting than the tarp site because of the presence of multiple targets of various sizes and types (e.g., moist soils, bare soils, 4 to 6 in. PCV pipe) and because of the similarity of the target pixel and background material reflectance values (Fig. 3). The total number of disturbed soil target pixels in the mine tailings is 228. Both the sagebrush and tailings sites were flat, so we did not apply slope corrections to adjust for potential variations in reflectance due to topography.

4. Results and discussion

The methodology described in Section 2 was applied to the original and the degraded imagery for both the tarp and mine tailings sites. A sensitivity analysis was

Fig. 7. Receiver operating characteristics (ROC) curves obtained for the three types of detection statistics and four subsets of principal components (first 84 or 25 PCs, 10 and 4 PCs with autocorrelation exceeding 0.25 or 0.5, respectively). The RXD, $S_1$, and $S_2$ statistics are computed from the PC values before and after geostatistical filtering of the regional background. Lower left plots show the purity of pixels according to their order of detection when using all 84 principal components.
performed to investigate the influence of a series of parameters on the detection ability of the technique measured by the ROC curves: number of principal components included in the analysis, size of the detection kernel, signal-to-noise ratio, spectral resolution, and geostatistical filtering of noise.

4.1. The tarp site

The analysis was first performed on the simplest scene with 4 square targets (the tarp) of 4 pixels each. Each image of principal components was decomposed into maps of local means and residuals or filtered values. The filtering was performed using expression (4) and a $5 \times 5$ window centred on the pixel being filtered (i.e., $n=25$). Fig. 4 shows an example for the first 2 principal components. The original PC values are decomposed into the background values $m(u)$ and the residuals or filtered values $r(u)=z(u)−m(u)$. These images illustrate how the removal of regional variability, which might represent different soil or vegetation types, highlights the location of target pixels in the filtered images.

The information provided by either filtered or non-filtered sets of 84 principal components was then summarized using the statistics: $δ_{RXD}$ (the aspatial RXD statistic of Chang & Chang, 2002), $S_1$ (the average p-value over the subset of $K'$ bands that display negative LISA statistics at that node), and $S_2$ (average absolute deviation of p-values for the LISA statistic from 0.5 through the K bands at that node) (Fig. 5). High-valued pixels indicate the presence of local anomalies for $S_1$ and clusters or anomalies for $S_2$. This figure clearly illustrates the benefit of the geostatistical filtering and use of statistic $S_2$, which increases the similarity with the actual image of tarp pixels displayed at the top of Fig. 2. The impact of the filtering is less pronounced for the RXD statistic, although the group of high-valued pixels in the upper left corner is somewhat attenuated.

The histograms displayed in Fig. 6 indicate that the distributions of statistics $S_1$ and $S_2$ are approximately symmetric, while the RXD statistic is characterized by the presence of a few extreme values and a large coefficient of variation. Bottom scatterplots indicate that $S_1$ and $S_2$ are strongly correlated with each other but exhibit little relationship with the RXD statistic. One should thus expect that the two sets of statistics will lead to the identification of different sets of pixels. Differences between the spatial ($S_1$ and $S_2$) and aspatial (RXD) statistics are due to the fact that the RX detector considers each location independently of its neighbors, while the power 2 and division by eigenvalues used in expression (3) makes this statistic very sensitive to extreme values, in particular those found in the noisy last principal components. Accounting for the neighborhood average in the calculation of local Moran’s I, as well as the computation of p-values through randomization leads to a more uniform distribution for statistics $S_1$ and $S_2$.

The final step is to apply a threshold to the maps of statistics $δ_{RXD}$, $S_1$, and $S_2$, and classify as targets all pixels exceeding this threshold. A series of thresholds (probabilities of detection) are defined as $t/T$ with $t=1, \ldots, T$ and $T$ is the total number of target pixels in the scene. For each threshold, the pixels classified as targets are compared to ground data to compute the proportion of misclassified pixels (probability of false alarms). These two sets of probabilities are then plotted to generate the receiver operating characteristics (ROC) curve. Fig. 7 (left top graph) shows an example of ROC curves for detection using each of the three types of statistics computed from filtered or non-filtered images. Lower left graphs show the effects of pixel purity on order of detection using the different statistics. The main conclusions are:

- The filtering and use of statistic $S_2$ allows the detection of all tarp pixels with a probability of false alarms not exceeding 0.20.
- Using the $S_1$ or $S_2$ statistics, detection of 60% of the tarp pixels can be done with a small probability of false alarm (vertical part of the ROC curve). Other pixels are more difficult to detect and generate an increase in the

![Autocorrelation for one pixel lag](image)

![Kernel 1x1 (filtered scene)](image)

Fig. 8. Plot of spatial correlation (lag=1 pixel) versus the order of the principal component. Bottom graph shows, for all PCs, the log ratio of average values of statistics $δ_{RXD}$, $S_1$, and $S_2$ for tarp pixels and background pixels. Note that the numerator and denominator variables (i.e., tarp or background) are always selected such that the ratio exceeds one.
proportion of false alarms, especially if no filtering is performed and only anomalies are searched (i.e., use of statistic $S_1$).

- The highest proportion of false alarms is produced by the RX detector and this rate is not reduced by the filtering procedure, which confirms conclusions drawn from the maps of Fig. 5.

- The order of detection of the 16 target pixels depends on the statistic used. In particular for the filtered scene, the last pixels detected using $S_1$ and $S_2$ are the least pure ones while these pixels are the first ones to be detected using $\delta_{\text{RXD}}$.

Sensitivity analyses were conducted to assess how the methodologies respond to:

1. Principal component rank,
2. The selection of a subset of principal components based on the strength of spatial correlation for the first lag (i.e., correlation between neighboring pixels exceeds a threshold value for all the selected PCs),
3. Choice of detection kernels of various sizes,
4. Signal to noise ratio and spectral resolution.

The effects of principal component rank order on the $\delta_{\text{RXD}}$, $S_1$, and $S_2$ statistics are shown in Fig. 8. As

Fig. 9. Receiver operating characteristics (ROC) curves obtained for three types of detection kernel, two signal to noise (SN) ratios, and three spectral resolutions (WV). The three spectral resolution ROC curves are based on, from left to right: the first 25 PC bands, one half of the bands, and one third of the bands. The RXD, $S_1$, and $S_2$ statistics are computed from the PC values before and after geostatistical filtering of the regional background.
expected, the spatial correlation of the image decreases as the rank of the principal component increases (Fig. 8, top graph). To determine if this affected target detection, the statistics were computed for each principal component separately, then the ratio of each statistic’s average for tarp and background pixels was plotted versus the rank/order of the principal component (Fig. 8, bottom graph). Clearly, differences between tarp and background pixels tend to attenuate as the order of the principal component increases. The effect of PC rank is particularly obvious for the RX detector, which contradicts the common practice of giving more weight to the principal components of high order. The large difference between averaged RXD values for target and background pixels is caused mainly by a few target pixels that have extreme spectral values and are located in the upper tail of the highly positively skewed $\delta_{RXD}$ histogram of Fig. 6 (top graph). Thus, although this difference is much larger than for statistics $S_1$ and $S_2$, the detection of all 16 target pixels will lead to more false alarms for the RXD statistic, as shown in the ROC curves of Fig. 7 (top graph).

Given the low information level of the last PCs and the CPU time (54.5 s on a Pentium 3.20 GHz) of processing all 84 bands, it is worth investigating the performances of the different detection approaches using fewer variables. Subsets of principal components were retained based on a spatial correlation threshold of 0.5 or 0.25 (Fig. 8, top graph). A third subset of the 25 first PCs was also used following Marcus (2002), who found that this number leads to the best classification scores for another HSRH scene in Yellowstone.

The ROC curves for the three subsets of PCs are displayed in the right column of Fig. 7. Using fewer PCs causes more false alarms for the detection of the first pixels; that is the initial part of the ROC curve is more detached from the vertical axis, in particular for the non-filtered scene. Yet, the total proportion of false alarms required for the detection of all 16 pixels can be lower; for example 17.1% versus 20.6% for $S_2$ (filtered scene). The benefit of using fewer PCs is particularly pronounced for the RX detector, which is in agreement with the quick drop in the discriminatory power observed beyond the 7th PC (Fig. 8, bottom graph). In fact for the smaller subset of 4 PCs, $\delta_{RXD}$ and $S_2$ statistics produce comparable proportions of false alarms, although the use of statistic $S_2$ with the filtered scene yields the best results in all situations. All ROC curves computed hereafter will be based on the first 25 PCs, thereby providing a balance between shorter CPU time (16.0 s on a Pentium 3.20 GHz) and slightly more false alarms.

All results presented so far were obtained using a detection kernel of one pixel, without any prior information.

![PC1 (raw values)](image1)
![PC2 (raw values)](image2)
![PC1 (filtered)](image3)
![PC2 (filtered)](image4)
![PC1 (background)](image5)
![PC2 (background)](image6)

**Fig. 10.** Maps of the first two principal components for the mine tailings scene and the results of the geostatistical filtering of the regional background.
regarding the size of the object to be detected. The benefit of tailoring the detection kernel to the size of the object was investigated by performing the classification and computing the ROC curves for three types of kernel: \(1 \times 1\), \(2 \times 1\), and \(2 \times 2\). For the RX detector, expression (3) is applied to principal components values averaged over the kernel. Fig. 9 (top row) shows that the use of kernels \(2 \times 1\) and \(2 \times 2\) improves detection performances of statistics \(d_{RXD}\) and \(S_1\), while more false alarms occur when using statistic \(S_2\). Indeed, statistic \(S_1\) searches for local anomalies of size equal to the kernel, while \(S_2\) detects both clusters and anomalies. The overall best performance of statistic \(d_{RXD}\) for kernel \(2 \times 2\) emphasizes the need to have precise information on target size and shape for this common target detector to outperform the Moran’s I-based statistics.

The impact of the signal-to-noise ratio was investigated by adding a given proportion of noise to reflectance values before performing the principal component analysis. Fig. 9 (middle row) shows the ROC curves obtained for increasing levels of noise (SNR=100:1 to SNR=50:1). As intuitively expected, noisy signals tend to blur the detection of anomalies, causing more false alarms in particular for the detection of the last pixels. This increase in the proportion of false alarms is less pronounced for \(S_1\) and \(S_2\) than \(d_{RXD}\), which reflects a greater robustness of the spatial statistics with respect to the presence of noise in the data.

The last test consisted of investigating how a decrease in spectral resolution would affect the quality of the detection. Fig. 9 (bottom row) shows the ROC curves obtained for the first 25 PCs computed from: (1) the original set of 84 spectral bands, (2) one half of this set (WV2, every other band is retained), and (3) one third of all 84 bands (WV3, one every other 2 bands is retained). As for the signal to noise ratio, ROC curves indicate poorer performances when using the degraded image, in particular in the RX detector. Again the use of statistic \(S_2\) with the filtered scene yields the best results in all situations.

4.2. The mine tailings site

The mine tailings site (Fig. 2) provides a more realistic setting than the tarp site because of the presence of multiple targets of various sizes and types (e.g., moist soils, bare soils, 4 to 6 in. PCV pipe, etc.) and because of the similarity of the target pixel and background material reflectance values (Fig. 3). As with the tarp site, the first 84 principal components were decomposed into maps of regional background and residuals or filtered values. Fig. 10 shows an example for the first 2 principal components. These images, as with the tarp site (Fig. 4), illustrate how the removal of regional variability, which represents different vegetation types and gravels, highlights the location of target pixels of bare soil in patches and along the road.

![Fig. 11. Maps of the three detection statistics computed from the first 25 principal components before (left column) and after (right column) filtering of the regional background. Only the pixels within the tailings site are mapped.](image-url)
The information provided by either filtered or non-filtered sets of 25 principal components was summarized using the statistics $d_{RXD}$, $S_1$, and $S_2$ (Fig. 11). Because only the tailings site was field surveyed for disturbed soils, pixels

Fig. 12. Plot of spatial correlation (lag=1 pixel) versus the order of the principal component (mine tailings site). Middle graph shows, for all the PCs, the log ratio of average values of statistics RXD, $S_1$, and $S_2$ for target pixels and background pixels. Note that the numerator and denominator variables (i.e., target or background) are always selected such that the ratio exceeds one. The location of target pixels is displayed in the bottom graph.

The information provided by either filtered or non-filtered sets of 25 principal components was summarized using the statistics $d_{RXD}$, $S_1$, and $S_2$ (Fig. 11). Because only the tailings site was field surveyed for disturbed soils, pixels

Fig. 13. Receiver operating characteristics (ROC) curves obtained for three types of detection statistics and four subsets of principal components of decreasing size at the mine tailings site. All 84 PC bands were retained in the top graph, while the second graph shows results with the first 25 PC bands. In the lower two graphs, the 19 and 7 PC bands with spatial autocorrelation greater than 0.25 and 0.5, respectively, were retained. The RXD, $S_1$, and $S_2$ statistics are computed from the PC values before and after geostatistical filtering of the regional background.
outside this area were masked out and not considered in the subsequent analysis. High pixel values indicate the presence of local anomalies for $S_1$ and clusters or anomalies for $S_2$. As with the tarp site (Fig. 4), Fig. 11 illustrates for the mine tailings the benefit of the geostatistical filtering and use of the $S_2$ statistic in particular. For all statistics, the filtering removes some large-scale features, such as the areas of high values observed in the upper left and mid-lower right of the non-filtered scene.

Sensitivity analysis indicates that the autocorrelation does not drop below 0.10 until the 30th principal component (Fig 12, top graph). In contrast, the correlation between neighboring pixels in the less complex tarp scene was generally smaller than 0.10 for the 15th and higher PCs (Fig. 8). The higher spatial autocorrelation leads to differences between target and background pixels that are smaller than those observed for the tarp scene, but still tend to decrease as the order of the principal component increases (Fig. 12, middle graph).

ROC curves were computed to determine how the number of principal components affected the outcome with the full set of 84 PCs, the first 25 PCs, and subsets based on a spatial correlation threshold of 0.25 (19 PCs) or 0.5 (7 PCs). Fig. 13 indicates that the benefit of filtering the regional variability increases as fewer principal components are used. For the three largest subsets, the use of statistic $S_2$ with the filtered scene yields the smallest proportion of false alarms. As with the tarp site, the RX detector starts

---

**Fig. 14.** Receiver operating characteristics (ROC) curves obtained for three types of detection kernel, two signal to noise (SN) ratios, and three spectral resolutions (WV). The RXD, $S_1$, and $S_2$ statistics are computed from the PC values before and after geostatistical filtering of the regional background.
performing at a level comparable to the spatial statistics when only the few spatially correlated PCs are used in the analysis. All ROC curves computed hereafter will be based on the first 25 PCs, thereby providing a balance between shorter CPU time (29.2 s on a Pentium 3.20 GHz for the processing of the 13,134 non-masked pixels) and slightly more false alarms.

The benefit of tailoring the detection kernel to the size of the objects was investigated by performing the classification and computing the ROC curves for three types of kernel besides the $1 \times 1$ kernel used earlier: $2 \times 1$, $1 \times 2$, and $2 \times 2$ (Fig. 14, top row). As the size of the kernel increases, the proportion of false alarms decreases for the RX detector, while it increases for the spatial statistics $S_1$ and $S_2$. Thus, the $\delta_{RXD}$ statistic ends up outperforming the $S_1$ statistic, in particular for the non-filtered scene. Even better performances were observed for kernels $3 \times 3$ and $4 \times 4$ (results not shown), which suggests that the higher detection power of RXD statistic is caused by the smoothing of $\delta_{RXD}$ values within the kernel instead of a better match between kernel size and target size. The larger kernel size also masks key autocorrelation patterns, lessening the ability of $S_1$ and $S_2$ statistics to detect local changes in spatial pattern.

The impact of the signal-to-noise ratio and spectral resolution was quantified using a procedure similar to the one applied to the tarp site. Image degradation through the addition of noise or reduction of the number of spectral bands causes an increase in the proportion of false alarms. As with the tarp site, statistics $S_1$ and $S_2$ seem to be more robust with respect to noisy signals. In all situations, the use of statistic $S_2$ with the filtered scene yields the best results (Fig. 14, 2 bottom rows).

5. Conclusions

This paper presented and demonstrated the efficacy of spatially explicit approaches for detecting anomalies and patches on high spatial resolution hyperspectral imagery. The innovative technique uses principal component analysis to reduce dimensionality of the imagery, employs geostatistical filtering to remove regional background and enhance local signal, applies a Local Indicator of Spatial Autocorrelation to identify anomalies, and combines the $p$-values across all spectral bands through two novel statistics. Analyses were conducted using tarps and disturbed soils in mine tailings at two locations in or near Yellowstone National Park. Results from the tarp site evaluated the ability of the method to detect regular patches on a simple landscape. Analysis of the tailings site evaluated detection capability on a complex landscape with multiple targets of various sizes and shapes. Following our results, a Pentium 3.20 GHz would allow the processing of a 1000×1000 scene including 25 bands within 18 min.

Although the proposed approach is more CPU intensive than the common RX detector, it generally leads to fewer false alarms, in particular in the presence of noisy signals. One of the main limitations of the RX detector is the tendency to assign too much weight to the uninformative and noisy PCs of high order. Better results were generally obtained when incorporating fewer PCs in the computation of the RXD statistic, but its implementation in practice suffers from the fact that no ground data will likely be available to assess the appropriate number of PCs to be used. The only situation where statistic $S_2$ did not outperform the alternative approaches is when precise information about the size of the target pixels was used in the definition of the kernel. Still, the benefit of geostatistically filtering the regional background was systematic and helped reduce the proportion of false alarms for both conventional and spatial detection statistics. By leveraging both spectral and spatial information, this novel approach requires little or no input from the user, and hence can be readily automated. This technique could be useful in a large range of applications where field information cannot be readily obtained, such as identifying potential locations of buried landmines or toxic waste, locating disturbed areas in remote settings, or finding targets on historic images for which ground data are not available.

Acknowledgements

Portions of the field work and analysis were supported by grants from the NAVAIR SBIR Phase I N02-172 program and from the U.S. Environmental Protection Agency. Hyperspectral data were provided by W. Andrew Marcus through a grant from NASA EOCAP, Stennis Space Center. Degradation of the imagery to generate lower signal to noise ratios was conducted by Amanda Warner. Kerry Halligan collected and conducted spectral resampling of the ground-based ASD reflectance data shown in Fig. 3. Statistical analysis by Drs. Goovaerts and Jacquez was funded by TerraSeer.

References


