

Detection of the Number of Signal Sources in the Hyperspectral Data

Majid Mohamady Oskouei, Rashed Poormirzaee

Abstract—This study investigates and compare different methods to estimate the number of signal sources in the hyperspectral data. To achieve an accurate map of mineral distributions in the study area by means of the spectral analysis of Hyperion data, the number of endmembers was computed by different methods. This process is also known as determination of virtual dimensionality of the image. Estimation of Virtual Dimensionality of data or in other words, the number of detectable endmembers from data is very important task in hyperspectral imagery. If this estimated number doesn't meet the reality, final estimation of mineral abundances will be erroneous. The results established that principle component analysis underestimates the virtual dimensionality of data. This is reasonably due to lower abundances of some minerals on the earth surface that will be considered as unimportant principle components because of their lower energy fraction in the total radiance measured at sensor. The higher order statistical method on the other hand, showed better performance. This method uses Neyman–Pearson detection theory and its estimation is more realistic.

Keywords—Hyperion, Hyperspectral imagery, Principle component analysis, Signal processing, Virtual dimensionality.

I. INTRODUCTION

Estimation of Virtual Dimensionality (VD) of data or in other words, the number of detectable endmembers from data is very important. In the case of having considerable mistake in determination of VD, final estimation of mineral abundances will be erroneous. In such a condition, determination of minerals in the image is meaningless. Two groups of methods were investigated to perform this task. One based on variance evaluation of channels (PCA and MNF) which have been applied for long time in various researches (e.g. [1]-[5]) and another based on high order statistical method which has found its way in image data analysis recently. The study is performed on the Hyperion scene from Erongo complex, Namibia.

A. Hyperion description

The objective of the Hyperion instrument is to provide high quality calibrated data that can support evaluation of hyperspectral technology for Earth observing missions.

Majid Mohamady Oskouei, Mining Engineering Faculty, Sahand University of Technology, Tabriz, Iran (e-mail: moskouie@yahoo.com). (corresponding author, phone: +98(0)4123459299; fax: +98(0)4123459299; e-mail: moskouie@yahoo.com).

Rashed Poormirzaee, Mining Engineering Faculty, Sahand University of Technology, Tabriz, Iran (e-mail: Rashed.poormirzaee@gmail.com).

Hyperspectral images can be defined as images with a high spectral resolution, typically 100 to 300 different wavelengths [6]. The Hyperion is a pushbroom instrument. Each image frame taken in this "push broom" configuration captures the spectrum of a line 30 m long by 7.5 km wide (perpendicular to the satellite motion). Frames are then combined to form a two dimensional spatial image with a complete spectral signature for each pixel. Hyperion has a single telescope and two spectrometers, one visible/near infrared (VNIR) spectrometer and one short-wave infrared (SWIR) spectrometer. Each pixel views a 30 m x 30 m region of the ground. The swath length of each image depends on the duration of the collect and is commanded by the spacecraft [7].

B. Study Area

The area which is located in northwest of Namibia and includes Erongo Complex with a diameter of approximately 35 km which is one of the largest Cretaceous anorogenic complexes in that country. The centre of the complex is located approximately on 21°40' S and 15°38'E.

It represents the eroded core of a caldera structure with peripheral and central granitic intrusions. Surrounding the outer granitic intrusions of the Erongo Complex is a ring dyke of olivine dolerite, which locally reaches some 200 m in thickness and has a radius of 32 km (Fig.1 and Fig.2). The ring dyke weathers easily and is therefore highly eroded. However, it can be easily identified on aeromagnetic data and satellite images. The central part of the Erongo complex consists of a layered sequence of volcanic rocks, which form prominent cliffs rising several hundred meters above the surrounding basement. The interior of the complex is deeply eroded, giving access to the roots of the structure. The basement rocks consist of mica schists and meta-greywackes of the Kuiseb Formation and various intrusions of granites. In the southeast, the rocks of the Erongo Complex overlie the Triassic Lions Head Formation, which consists of conglomerates, gritstone, arkose with interbedded siltstone and mudstone, and quartz arenite.

The base of the Erongo Complex consists of a series of flat-lying basaltic lava flows and interbedded pyroclastic rocks. These basal volcanics are exposed throughout the entire complex and may originally have had an even wider distribution. With some 300 m thickness, the thickest layers of the basal volcanics are located in the southeastern part of the complex. The rock compositions range from tholeiitic, fine-grained basalt and basaltic andesite to andesite. Most basalts are considerably altered and commonly amygdaloidal, with vesicles filled with calcite and chalcedony. The alteration also

caused local growth of quartz, actinolite, epidote and chalcopyrite. Plagioclase is sericitised and saussuritised, and Clinopyroxene is replaced partially or fully by chlorite and epidote.

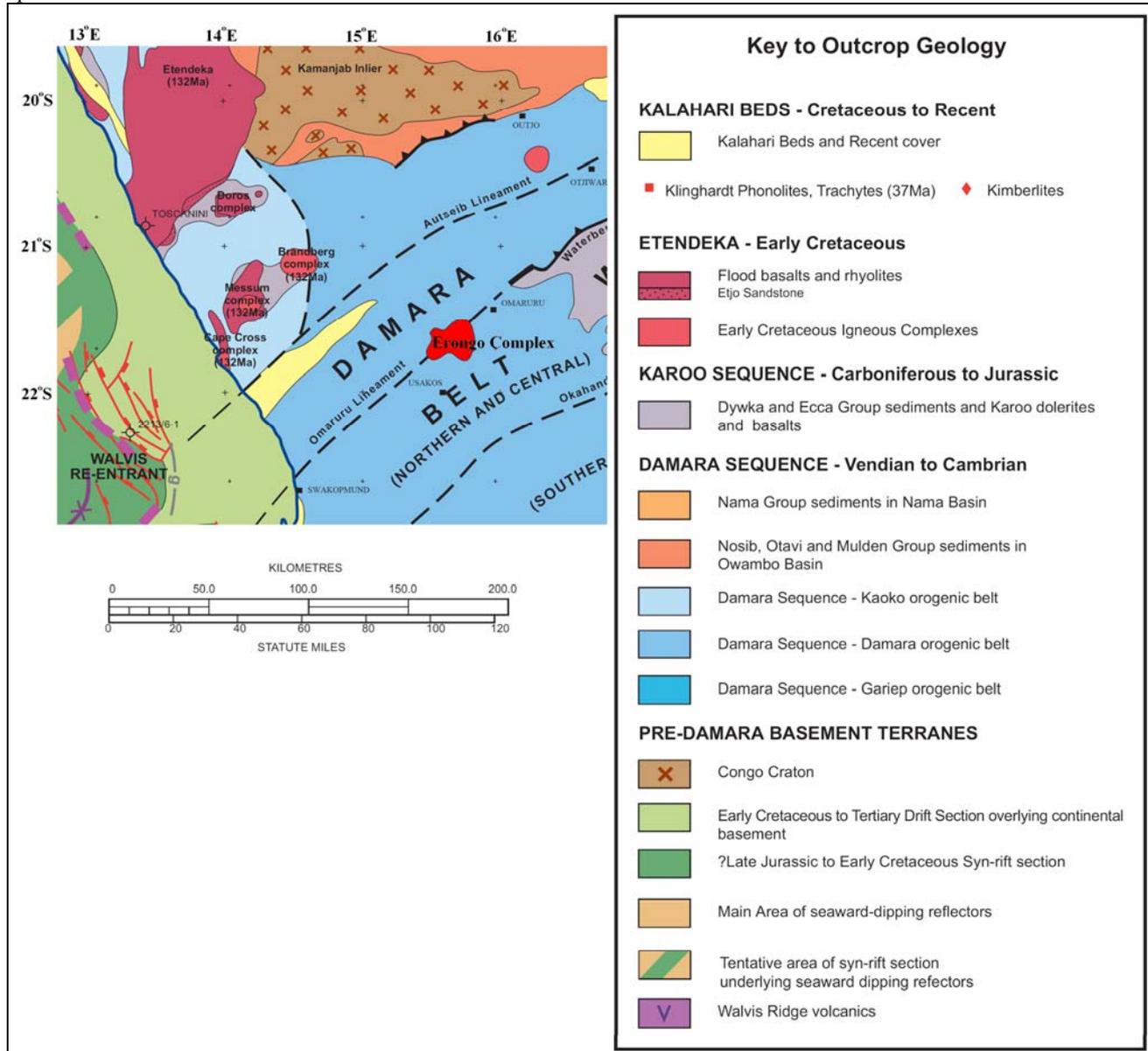


Fig.1: Location of Erongo Complex and surrounding geological formations.

The basal volcanics are followed by a sequence of felsic volcanic units, which have been subdivided by Pirajno into four phases. The first phase is characterized by minor eruptions of mafic-intermediate lavas and major ash-flow tuffs of intermediate to felsic composition. These so-called Erongorus ash-flow tuffs occur mainly in the north, the northwest and the west of the complex, but the units are absent in the east. They are generally altered and characterized by a devitrified groundmass with phenocrysts of quartz and altered K-feldspar.

The Erongorus tuffs are overlain by the Ombu ash-flow tuff sequence, which is volumetrically the main rock type of the complex and forms the most prominent cliffs. The Ombu tuffs are generally more quartz-rich than the Erongorus tuff units, but the most striking difference is that the Ombu tuffs frequently contain abundant sizable (cm- to dm) fragments of basement rocks. Compositionally, the Ombu tuff units are of rhyodacitic and rhyolitic composition [8].

I. DATA CORRECTION

Electromagnetic waves travelling from the energy source to the earth and from the earth to the sensor interact with a variety of materials, including various gases, water vapour, and aerosols. It is therefore necessary to compensate for these undesirable effects, which is known as atmospheric correction. We therefore used ACORN mode 1 for the atmospheric correction of the dataset.

Quality of Hyperion datasets is related to existence of bad pixels and bad bands. For example a dead detector will make a dark sample column respected to it with 0 or filled with constant DN for all pixels. This problem could be corrected by resampling using adjacent pixels. It is noteworthy that during processing, they preferably would be masked but for any visual output the resampling results in better performance.

Other problem about pixels is smile effect. According to Jupp and Datt [9] this is a systematic effect that has been noticed in Hyperion processing and application involves problems in the spectral alignment of the VNIR array. It has been called spectral smile since its consequence is that the central wavelength of a band varies with spatial position across a line in a smoothly curving pattern. However, since the peak of the curve in Hyperion tends to be near the middle of the line it could perhaps be better called (and for more than one reason) spectral frown. In their report, they have collected the findings of different researchers. They showed that among options for its removal only wavelength interpolation, in which the smile is removed from each vertical column by interpolating to a standard set of central wavelengths prior to atmospheric correction, could reduce it and maintain spectral integrity.

Unfortunately, interpolation was unable to remove all of the smile effect just to reduce it. Therefore there was a great deal of discussion about how to live with this smile effect and most people have worked without a specific correction in place. The good points for the situation are that:

- It is only significant in the VNIR. The SWIR has little to no "smile" effect; and

- It only affects indices, ratios or derivatives near to sharp atmospheric features or the residuals from such features after atmospheric correction.

Whereas we used SWIR channels in most of feature detection processes, therefore fortunately this problem will not be a crucial in our work.

Determination of bad bands in our work could be the most important task of quality assessment of the dataset. A short definition for bad bands refers to some bands in the dataset which have insignificant and inconsiderable information or no information to be extracted. The dataset produced by USGS so only 198 bands from 242 (VNIR: 8-57, SWIR: 77-224) available channels have already been selected by supplier according to different calibration parameters.

Some bands having large amount of noise degrades the interpretability of data and should then categorized as bad bands. This noise might be produced by numerous factors including thermal effects, sensor saturation, quantization errors and transmission errors.

The quality of digital remote sensing data is directly related to the level of signal to system noise ratio (SNR). This is a dimensionless number that describes overall system radiometric performance. System noise is tied to sensor design and takes into account factors such as detector performance/sensitivity, spatial/spectral resolution, and noise characteristics of the system electronics. Theoretically, SNR ratio for Hyperion is 190 to 40 as the wavelength increases [10]. Although the noise levels for a given sensor are generally fixed, for remote sensing data acquisition, the signal portion of the SNR is affected by other external factors such as solar zenith angle, atmospheric attenuation and scattering, and surface reflectance, which modify the signal available to the sensor.

One common means for determining an approximate SNR for remote sensing data is to use a mean/standard deviation method. This approach requires definition of a spectrally homogeneous area, calculation of the average spectrum for that area, and determination of the spectrally distributed standard deviation for the average spectrum. SNR calculated using this method are representative of those that can be extracted directly from the data, however, SNR for bright targets may be underestimated because of homogeneity issues at higher SNR (increasing SNR may result in breakdown of apparently homogeneous areas into multiple materials, and new homogeneous areas must be selected). Slightly higher SNR values could probably be obtained through direct analysis of the data dark current signal, an "Instrument SNR", however, this is not always possible. SNR calculated using the mean/standard deviation method, an "Environmental SNR," is sensitive to acquisition conditions as mentioned above and, thus, should be considered a lower limit on performance.

Analysis of approximately 14 Hyperion scenes from around the world using the mean/standard deviation SNR method shows that there is a strong relationship between the acquisition time of year and the SNR of the Hyperion data. Calculated SNR for Hyperion SWIR data are higher in the summer and lowest in the winter. This has a direct effect on spectral mineral mapping, with lower SWIR SNR resulting in extraction of less detail [11]. As a result of this process, uncalibrated channels (channels 1-7 and 225-242) in addition to those affected strongly by water vapour absorption (940nm, 1400nm, 1900nm) are listed in bad bands list of the data for next processing tasks. Besides that, the bands which possess considerable less SNR ratio comparing to vicinity channels are also considered as bad bands.

After assessing of quality of data bands (signal to noise ratio), we used up only good bands (166 of total 242 bands) to avoid any possible mistakes during preprocessing and subsequent steps. As a final pre-processing task, the data were polished using a geostatistical algorithm presented by the author [12]. This algorithm reduces the total error in terms of

discrimination of the endmembers.

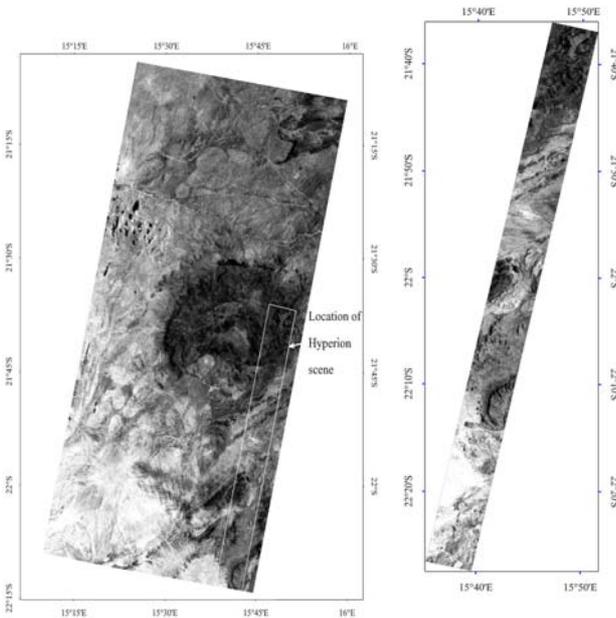


Fig.2: Erongo complex can be easily identified in the middle of this ASTER scene (left), and the image in the right hand belongs to the Hyperion image of Erongo.

II. METHODOLOGY

A. Principle component analysis

To estimate the VD of the Hyperion data, PCA transform available on ENVI platform was used by rotating the majority of information content (variance) into fewer bands. The specific method of PCA is completed by rotating the original data into a set of axes that maximizes the variability in the first few axes ([13];[14]). This is done by finding a new coordinate system in hyper-dimensional vector space where the data exhibits no correlation. Basically, the covariance matrix in the new coordinate system is diagonal or uncorrelated. A linear transformation, U , that transforms the original hyperspectral data, X , into the new coordinate system Y must then be calculated.

Use either SI (MKS) or CGS as primary units. (SI units are strongly encouraged.) English units may be used as secondary units (in parentheses). This applies to papers in data storage.

The original covariance matrix, Σ_x , becomes the diagonalized covariance matrix, Σ_y . The x solution to this problem becomes a generalized y eigenvalue problem of the form:

$$\Sigma_x U = U \Sigma_y \quad (1)$$

where the eigenvalues are the diagonal elements of Σ_y , and the eigenvectors form the columns of U [15]. The original data X is then multiplied by the eigenvectors of the original data covariance Σ_x . The PCA transformation is then computed by:

$$Y = UTX \quad (2)$$

Often times PCA is used to either reduce the dimensionality of the data processed in a target detection algorithm, or to reduce the noise in the hyperspectral imagery. Typically to reduce the dimensions of a data set a selection of a certain number of PCA bands is made so that certain percentage of the total variance in the data is preserved. The results which introduce principle components of the data are listed in table 1 with respect to eigenvalues amount. Fig.3 demonstrates the result in diagram in which the eigenvalues portions for each eigenvector (0 to 100%) are plotted. First 13 eigenvectors includes approximately 99% of sum of all the eigenvalues. Therefore a preliminary judgment about VD or the number of detectable endmembers by this method is 13.

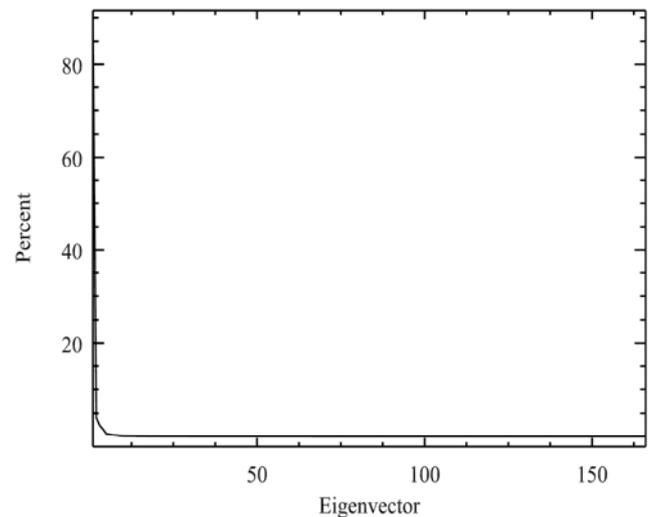


Fig.3: The eigenvalues proportions for each eigenvector calculated by PCA

A. Minimum noise fraction

The MNF transformation is a linear transformation related to principle components. It determines the inherent dimensionality of the data, segregates noise in the data, and reduces the computational requirements for subsequent processing[16]. Noise is undesired information that contaminates an image. Noise appears in images from variety sources[17].

Table 1: Eigenvalues calculated for principle components of data by PCA ENVI.

PCA1	2.62E+08	89.29766
PCA2	11869887	4.051156
PCA3	6949261	2.371762
PCA4	4499289	1.535594
PCA5	1510253	0.515445
PCA6	1193729	0.407416
PCA7	711518.3	0.242839
PCA8	629891.3	0.21498
PCA9	366060.5	0.124935
PCA10	273020.7	0.093181

It is defined as a two-step cascaded PCA. The first step, based on an estimated noise covariance matrix, is to decorrelate and rescale the data noise, where the noise has unit variance and no band-to-band correlations. The next step is a standard PCA of the noise-whitened data. The inherent dimensionality of the data is also determined by examination of the final eigenvalues and the associated images by using MNF transformation from ENVI. The data space can be divided into two parts: one part associated with large eigenvalues and coherent eigen images, and a complementary part with near-unity eigenvalues and noise-dominated images. The results of the MNF transform are presented in table 2 and related plot (Fig.4). In contrary to PCA, none of MNF eigenvectors poses such a high eigenvalue. For that eigenvalue number 1 is 13.68% of sum of the all eigenvalues, however, even 40th eigenvector has a considerable value. This outcome could be as a result of the application of algorithm presented in the last chapter, because the noise in the dataset was whitened in all bands. But the corollary that could be realized from this outcome is important:

Probably some of other eigenvectors beside those 13 introduced by PCA have also some extractable information. As a matter of fact this is a reality when the goal of study is to survey minerals anomalies, as some low grade minerals on the surface of study field are not able to be monitored in high value eigenvectors in principle component analysis. Therefore in the case of mineral study by hyperspectral data, using other

complementary and non-linear methods which are not based on only variance role is necessary.

Num	Eigenvalue	Eigenvalues percentage
MNF1	75.49992	13.68934
MNF2	54.43878	9.870621
MNF3	28.981	5.254717
MNF4	27.61613	5.007246
MNF5	23.13752	4.195202
MNF6	13.02569	2.361766
MNF7	10.07678	1.827081
MNF8	8.866292	1.6076
MNF9	7.55313	1.369503
MNF10	7.153453	1.297035

Table 2: Eigenvalues calculated for principle components of data by MNF ENVI

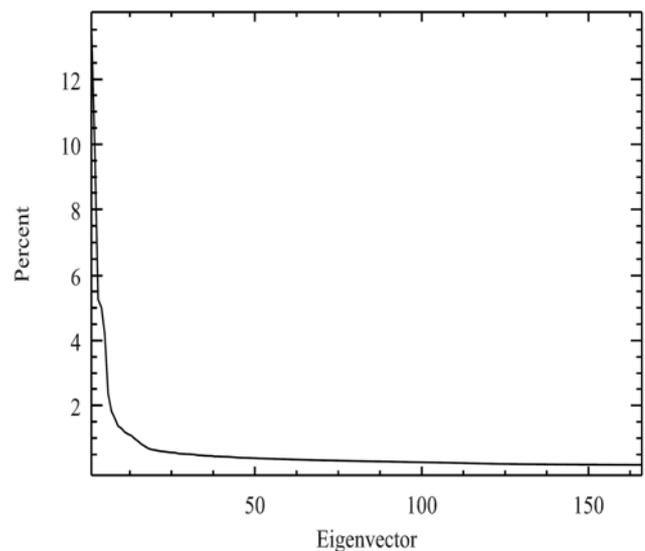


Fig.4: The eigenvalues proportions for each eigenvector calculated by MNF

B HFC method

This is a method presented by Harsanyi, Farrand, and Chang (HFC) [18] and it uses Neyman–Pearson detection theory for estimation the number of endmembers. This idea is described in [18] and [19], but we will have a brief description here as it necessary to evaluate the reliability of the method.

Table 3: VDs calculated by HFC method for different false alarm probabilities (Pfa).

P _{fa}	10 ⁻¹	10 ⁻²	10 ⁻³	10 ⁻⁴	10 ⁻⁵
VD	43	38	37	37	37

eigenvalues generated by the sample correlation matrix and the

Let the sample covariance matrix be denoted by correlation eigenvalues and covariance eigenvalues, respectively. Since the component dimensionality is equal to the total number of eigenvalues, each eigenvalue specifies a component dimension and provides an indication of the significance of that particular component in terms of energy or variance. If there is no signal source contained in a particular component, the corresponding correlation eigenvalue and covariance eigenvalue in this component should reflect only the noise energy, in which case, correlation eigenvalue and covariance eigenvalue are equal. This fact provides us with a base from which we can formulate the difference between the correlation eigenvalue and its corresponding covariance eigenvalue as a binary composite hypothesis testing problem. The null hypothesis represents the case of the zero difference, while the alternative hypothesis indicates the case that the difference is greater than zero. When the Neyman–Pearson test is applied to each pair of correlation eigenvalue and its corresponding covariance eigenvalue, the number of times the test fails indicates how many signal sources are present in the image. In other words, a failure of the Neyman–Pearson test in a component indicates a truth of the alternative hypothesis, which implies that there is a signal source in this particular component.

Using this approach, we can estimate the virtual dimensionality with the receiver operating characteristic analysis for evaluating the effectiveness of the decision. It first calculates the sample correlation matrix $R_{L \times L}$ and sample covariance matrix $K_{L \times L}$ and then finds the difference between their corresponding eigenvalues. Let $\{\lambda'_1 \geq \lambda'_2 \geq \dots \geq \lambda'_L\}$ and $\{\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L\}$ be two sets of eigenvalues generated by $R_{L \times L}$ and $K_{L \times L}$, called correlation eigenvalues and covariance eigenvalues, respectively. By assuming that signal sources are non-random unknown positive constants and noise is white with zero mean, we can expect that

$$\lambda'_1 > \lambda_l \text{ for } l=1,2,\dots,VD \quad (3)$$

$$\text{and } \lambda'_1 = \lambda_l \text{ for } l=VD+1,\dots,L \quad (4)$$

More specifically, the eigenvalues in the l^{th} spectral channel can be related by

$$\lambda'_1 > \lambda_l > \sigma_{nl}^2 \text{ for } l=1,2,\dots,VD \quad (5)$$

$$\text{and } \lambda'_1 = \lambda_l = \sigma_{nl}^2 \text{ for } l=VD+1,\dots,L \quad (6)$$

Where σ_{nl}^2 is the noise variance in the l^{th} spectral channel.

In order to determine the VD, Harsanyi *et al* formulated the problem of determination of VD as a binary hypothesis problem as follows:

$$H_0: z_l = \lambda'_1 - \lambda_l = 0 \text{ versus } H_1: z_l = \lambda'_1 - \lambda_l > 0$$

for $l=VD+1,\dots,L$ (7)

Where the null hypothesis H_0 and the alternative hypothesis H_1 represent the case that the correlation eigenvalue is equal to its corresponding covariance eigenvalue and the case that the correlation eigenvalue is greater than its corresponding covariance eigenvalue, respectively. In other words, when H_1 is true (i.e., H_0 fails), it implies that there is an endmember contributing to the correlation eigenvalue in addition to noise, since the noise energy represented by the eigenvalue of $R_{L \times L}$ in that particular component is the same as the one represented by the eigenvalue of $K_{L \times L}$ in its corresponding component. Despite the fact that the λ_l and λ'_1 are unknown constants, we can model each pair of eigenvalues λ'_1 and λ_l under hypotheses H_0 and H_1 as random variables by the asymptotic conditional probability densities given by

$$p_0(z_l) = p(z_l|H_0) \approx N(0, \sigma_{z_l}^2),$$

for $l=1,2,\dots,L$ (8)

and

$$p_1(z_l) = p(z_l|H_1) \approx N(\mu_l, \sigma_{z_l}^2),$$

for $l=1,2,\dots,L$ (9)

respectively, where μ_l is an unknown constant and the variance $\sigma_{z_l}^2$ is given by

$$\sigma_{z_l}^2 = \text{Var}[\lambda'_1 - \lambda_l] = \text{Var}[\lambda'_1] + \text{Var}[\lambda_l] - 2\text{Cov}(\lambda'_1 - \lambda_l) \quad (10)$$

Eventually, they defined the false-alarm probability and detection power (i.e., detection probability) by using above mentioned equations and some approximations as follow:

$$P_F = \int_{\tau_1}^{\infty} P_0(z) dz \quad (11)$$

$$P_D = \int_{\tau_1}^{\infty} p_1(z) dz \quad (12)$$

A Neyman–Pearson detector for $\lambda'_1 - \lambda_l$, denoted by $\delta_{NP}(\lambda'_1 - \lambda_l)$, in the binary composite hypothesis testing problem can be obtained by maximizing the detection power P_D in (12), while the false-alarm probability in (11) is fixed at a specific given value, α , which determines the threshold value in (11) and (12). So, a case of $\lambda'_1 - \lambda_l > T_1$ indicating that $\delta_{NP}(\lambda'_1 - \lambda_l)$ fails the test, in which case there is signal energy assumed to contribute to the eigenvalue λ'_1 in the l^{th} data dimension. It should be

noted that the test for hypothesis must be performed for each of L data dimensions. Therefore, for each pair of $\lambda'_1 - \lambda_j$, the threshold T is different and should be L -dependent, i.e., T_L .

Virtual dimensionalities of the image were calculated using various amounts of false alarm probabilities are listed in table 3.

C Estimation of the number of endmembers by PPI and N-D visualizer

Decreasing in P_{fa} will increase the threshold amount which is used in comparison of λ'_1 and λ_j , but actually after a certain magnitude it doesn't have significant effect on threshold. Therefore as it mentioned by the innovators of the method (Harsanyi et al. [18]; Chang and Du [19]) 10^{-4} seems to be a reasonable amount for that. 37 detectible endmembers for the Hyperion data is final result of this section. As discussed before pixel purity index (PPI) is a process to determine most pure pixels in the scene. All other pixels' signal amounts could be written as a linear combination of these pure pixels. In the case of mineral exploration due to small abundances of minerals, however, finding pure pixels is somewhat impossible. Therefore the purest pixels themselves are consisted of some minerals. This fact will conduct us to two points of view:

We intend to determine the minerals so it should be defined how many minerals are comprised in each purest pixel. The result will guide us to estimate the number of detectible minerals or the number of endmembers in the image data which is the goal of this stage of research.

We intend to determine the groups of minerals like alteration mineral groups. This will create new opportunities and new challenges and will be discussed next.

In this research we applied ENVI's well known procedure to determine purest pixels. This procedure uses MNF to dimensionality reduction after which the pixel purity index is computed by repeatedly projecting n -dimensional scatter plots onto a random unit vector. The extreme pixels in each projection (those pixels that fall onto the ends of the unit vector) are recorded and the total number of times each pixel is marked as extreme is noted. A pixel purity image is created in which the DN of each pixel corresponds to the number of times that pixel was recorded as extreme.

To be sure about not to miss any extractable information in data we used 30 first bands of MNF transform to produce pixel purity index image. The PPI is calculated by repeatedly collapsing n -dimensional scatter plots (30 dimensions for endmember selection) of MNF data onto a random 2-dimensional unit vector. In each projection, the pixels in the scatter plot that fall at or near the ends of the vector (highest degree of variance) are identified as extreme pixels. A threshold value of 2 to 3 times the variance in the data set is suggested as a starting point for the selection of extreme pixels. As the PPI threshold increases, the number of selected

pixels increases, and the probability that mixed pixels will be selected increases.

For each iteration ENVI records which pixels are marked as extrema. After multiple iterations, the pixels identified as extreme the highest number of times are considered pure. It is important to note that if a target does not exist in the image as 100% cover, the PPI could mark a mixed pixel as being pure. If this is the case, the identified mixed pixels are still considered to represent endmembers, because they most closely represent the spectral reflectance of pure target reflectance.

Interactively adjusting the number of PPI iterations and the PPI threshold will produce different results in the PPI calculation. Fig.5 illustrates a plot of the number of pixels selected versus the number of iterations with the use of 2.5 as PPI threshold by iterating 100,000 times. Typically setting a high PPI threshold results the selection of more pixels including impure pixels as extremes. Alternatively, setting a very low PPI threshold (e.g. a threshold of one) will not select many mixed pixels, but will also not select all purest pixels. For this study, 100,000 iterations at a PPI threshold of 2.5 produce approximately 27,000 pixels marked as extreme in at least one projection.

The PPI algorithm was able to distinguish purest pixels almost uniformly distributed all over the scene. It is noteworthy that some pixels along first and last columns and also on a few other columns which were selected as pure pixels were ignored before subsequently processes.

The purest pixels indexed by PPI were to be clustered by N-Dimensional visualizer method in ENVI. This algorithm in conjunction with the MNF transform and PPI results can be used to locate, identify, and cluster the purest pixels and most extreme spectral responses in a data set. The N-Dimensional visualizer is an interactive tool to use for selecting the endmembers in n -space. Spectra can be thought of as points in an n -dimensional scatter plot, where n is the number of bands. The coordinates of the points in n -space consist of " n " values that are simply the spectral radiance or reflectance values in each band for a given pixel. The distribution of these points in n -space can be used to estimate the number of spectral endmembers and their pure spectral signatures. When using the N-Dimensional Visualizer, you can interactively rotate data in n -D space, select groups of pixels into classes, and collapse classes to make additional class selections easier. The selected classes can be exported to Regions of Interest (ROIs).

At first automatic clustering by ENVI N-D Visualizer module was performed to obtain an overall understanding about purest pixels signatures. By this method the number of classes is equal to the number of MNF image bands (30 in this case). More investigation and comparison of mean spectrum of classes conducted us to merge some of classes produced by N-D Visualizer; however, a few new classes were distinguished by rotating and projecting pixels in 3,4,5,6 dimensional spaces. Finally 11 independent classes were identified. As formerly discussed these purest pixel are always mixed of a few minerals themselves, as a matter of fact if we want to

make a decision about the number of detectable minerals by this method, it could be sophisticated. But the most important conclusion of applying this algorithm is that the number of endmembers could be more than 13 which were resulted by PCA method.

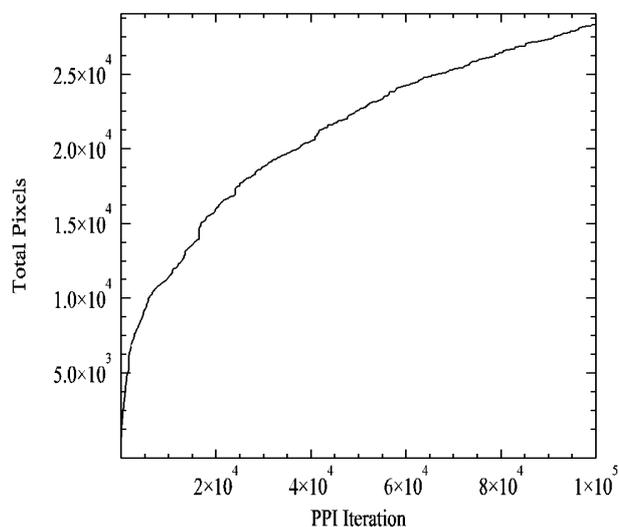


Fig.5: The performance of PPI and selected pixels as purest pixels versus iteration

IV Conclusion

The huge volumes and rates of data generated by hyperspectral sensors demand expensive processors with very high performance and memory capacities. Dimensionality reduction is, therefore, a relevant first step in the hyperspectral data processing chain. With very high spectral resolution hyperspectral sensors which can extract many unknown material substances which cannot be provided by a prior knowledge, estimation of virtual dimensionality of hyperspectral data is more problematic than that of multispectral data.

According to the results of two methods applied for estimation of virtual dimensionality on a set of Hyperion data, more and reliable VD can be computed by high-order statistic based methods such as HFC in comparison to PCA.

REFERENCES

- [1] Yetkin E, Toprak V, Suezzen ML, Alteration Mapping By Remote Sensing: Application To Hasandağ – Melendiz Volcanic Complex. *ISPRS*, 12-13 July, Istanbul, Turkey, 2004.
- [2] Miao X, Gong P, Swope S, Pu R, Carruthers R, Anderson GL, Heaton JS, Tracy CR, Estimation of yellow starthistle abundance through CASI-2 hyperspectral imagery using linear spectral mixture models, *Remote Sensing of Environment* 101, 2006, p. 329–341.
- [3] Goovaerts P, Geoffrey MJ, Marcus A, Geostatistical and local cluster analysis of high resolution hyperspectral imagery for detection of anomalies, *Remote Sensing of Environment* 95, 2005, pp. 351–367.
- [4] Hubbard BE, Crowley JK, Mineral mapping on the Chilean–Bolivian Altiplano using co-orbital ALI, ASTER and Hyperion imagery: Data dimensionality issues and solutions, *Remote Sensing of Environment* 99, 2005, pp. 173 – 186.
- [5] Xiang Zhu, Yingyi Chen, Daoliang Li, Soil Environmental Quality Assessment in Sustainable Rehabilitation of Mine Waste Area: Establishing an Integrated Indicator-based System, *WSEAS Transactions On Information Science And Applications*, Issue 3, Volume 6, 2009.
- [6] Mielikainen, Pekka Toivanen, Francisco Urendes Porcel, Reducing the Compression Time of the Spectral DPCM for Lossless Compression of 3D Hyperspectral Sounding Data, 6th WSEAS Int. Conf. on AUTOMATION & INFORMATION (ICAI'05), Buenos Aires, Argentina, 2005, 503-235.
- [7] Barry P, *EO-1 Hyperion Science Data User guide*, TRW Space, defense and Information System, 2001.
- [8] Schneider, Gondwanaland Geopark, *The Gondwanaland Geopark Project*, financial support from Unesco (Windhoek Cluster Office), 2004, pp. 31-33.
- [9] Jupp DLB, Datt B, 2004, "Evaluation of the EO-1 Hyperion Hyperspectral Instrument and its applications at Australian validation sites 2001-2003", CSIRO Earth Observation Centre Report 2004, pp. 18-20.
- [10] Pearlman J, Segal C, Liao L, Carman S, Folkman M, Browne B, Ong L, Ungar S, 2000, "Development and operations of the eo-1 hyperion imaging spectrometer", In: proceedings of SPIE, Earth observation system 4135, pp. 243-254.
- [11] Kruse FA, Boardman JW, Huntington JF, 2003, "Comparison of Airborne Hyperspectral Data and EO-1 Hyperion for Mineral Mapping" *IEEE Transactions on Geoscience and Remote Sensing* 41, pp. 1388-1400.
- [12] Oskouei MM, Busch W, A geostatistically based preprocessing algorithm for hyperspectral data analysis, *GI Science & Remote Sensing* 45 (3), 2008, 356–368.
- [13] Ientilucci E, Comparison and Usage of Principal Components Analysis (PCA) and Noise Adjusted Principal Component (NAPC) or Maximum Noise Fraction (MNF), *Centre for Imaging Science - Rochester Institute of Technology*, 2003.
- [14] Discriminative Feature Extraction based on PCA Gaussian Mixture Models, Hossein Marvi, Proceedings of the 5th WSEAS Int. Conf. on DATA NETWORKS, COMMUNICATIONS & COMPUTERS, Bucharest, Romania, October 16-17, 2006.
- [15] Staab B, Investigation of noise and dimensionality reduction transforms on hyperspectral data as applied to target detection, Chester F, Carlson Centre for Imaging Science, *Rochester Institute of Technology*, 2005.
- [16] Kruse FA, Richardson LL, and Ambrosia VG, 1997, "Techniques Developed for Geologic Analysis of Hyperspectral Data Applied to Near-Shore Hyperspectral Ocean Data", In: Fourth International Conference on Remote Sensing for Marine and Coastal Environments, Orlando, Florida, USA.
- [17] Examining Filtration Performance on Remotely Sensing Satellite Images, Yas Abbas Alsultanny, Nidal Shilbayeh, Proceedings of the WSEAS International Conference on Speech, Signal and Image Processing, Malta, September 1-6, 2001.
- [18] Harsanyi JC, Farrand W, Chang CI, Determining the number and identity of spectral endmembers: An integrated approach using Neyman–Pearson eigenthresholding and iterative constrained RMS error minimization, *Proceeding of 9th Thematic Conference of Geologic Remote Sensing*, 1993.
- [19] Chang CI, Du Q, Estimation of number of spectrally distinct signal sources in hyperspectral imagery, *IEEE Transaction on Geoscience and Remote sensing* 42(3), 2004, pp. 608-619.