INT. J. REMOTE SENSING, 1985, VOL. 6, NO. 6, 883–896

Standardized principal components

ASHBINDU SINGH†

Department of Geography, University of Reading, Reading RG6 2AU, England

ANDREW HARRISON‡

NERC Thematic Information Services, Holbrook House, Station Road, Swindon SN1 1DE, England

(*Received 12 May 1984; in final form 2 July 1984*)

Abstract. In remote sensing, principal components analysis is usually performed using unstandardized variables. However, the use of standardized variables yields significantly different results. In this paper principal components of two LANDSAT MSS subscenes were separately calculated using both methods. The result indicate substantial improvement in signal-to-noise ratio and image enhancement by using standardized variables in the principal components analysis.

1. Introduction

The principal components transformation, also referred to as the eigenvector transformation, the Hotelling transformation and the Karhunen–Loève (K–L) transformation in the remote-sensing and pattern-recognition literature, is a multivariate statistical technique, which is often used for determining the underlying statistical dimensionality of the image data set (Ready and Wintz 1973), for image enhancement (Gillespie 1980), for digital change detection (Byrne *et al.* 1980) and for characterizing seasonal changes in cover types (Townshend *et al.* 1984). The technique essentially consists of choosing uncorrelated linear combinations of the variables in such a way that each successively extracted linear combination, called a principal component, has a smaller variance. If the variables have significant linear intercorrelations, the first few components will account for a large part of the total variance.

Mathematically, if $\mathbf{X}^{\mathrm{T}}=[X_1,\ldots,X_N]$ is an N-dimensional random variable with mean vector $\mathbf{M}$ and covariance matrix $\mathbf{C}$, then a new set of variables, say, $Y_1, Y_2, \ldots, Y_N$, known as principal components, can be expressed by:

$$\begin{aligned} Yj &= a_{1j}X_1 + a_{2j}X_2 + \cdots + a_{Nj}X_N \\ &= a_j^{\mathrm{T}}X \end{aligned}$$

where T denotes transpose of a matrix and $a_j^{\mathrm{T}}=[a_{1j},\ldots,a_{Nj}]$ are the normalized eigenvectors [i.e., $a_j^{\mathrm{T}}a_j=1$] of the variance–covariance matrix. By denoting the (N

† Member of the Indian Forest Service, Forest Department, Government of Manipur, Imphal, India.

‡ Now with the Remote Sensing Unit, Department of Geography, University of Bristol, University Road, Bristol BS8 1SS.

$\times N$) matrix of eigenvectors by $\mathbf{A}$ and the $[N \times 1]$ vector of principal components by $\mathbf{Y}$, we can write,

$$\mathbf{Y} = \mathbf{A}^{\mathrm{T}} \mathbf{X}$$

The $(N \times N)$ co-variance matrix of $\mathbf{Y}$, $\mathbf{C}$, is given by:

$$\mathbf{C} = \begin{bmatrix} \lambda_1 & \cdots & & 0 \\ \vdots & \lambda_2 & & \vdots \\ & & \lambda_3 & \\ 0 & \cdots & & \lambda_N \end{bmatrix}$$

where λ_i are eigenvalues of matrix $\mathbf{C}$. The matrix is diagonal as the components have been chosen to be uncorrelated and $\lambda_1 > \lambda_2 > \cdots > \lambda_N$.

The principal components transformation has several characteristics, some of which are of special interest in remote sensing:

(1) The total variance is preserved in the transformation i.e.

$$\sum_{i=1}^{N} \sigma_i^2 = \sum_{i=1}^{N} \lambda_i$$

where σ_i^2 are variances of the original variables.

(2) It minimizes the mean square approximation errors.

(3) This is the only transformation that generates uncorrelated coefficients (Moik 1980). In a geometrical sense, it rotates the highly correlated features in N-dimensions to a more favourable orientation in the feature space, orthogonal to each other, such that the maximum amount of variance is accounted for in decreasing magnitude along the ordered components. The process has been viewed as an information compression into a smaller number of components from the large number of features by discarding redundant information into higher-order components.

The technique was first developed by Hotelling (1933) for his work in educational psychology. The K–L transformation was introduced to pattern recognition by Watanabe (1965) and since then research efforts have developed in two parallel directions. In the statistical literature interest has been in the area of sampling theory and inference procedures (Kendall *et al.* 1983), whereas in pattern recognition the main concern has been with feature extraction methods such as information compressibility, and the relationship with the Fourier transform (Devijver and Kittler 1982).

Computationally, no fast algorithm exists for calculating the transform and this is a serious drawback because of the amount of computation involved (Gonzalez and Wintz 1977).

2. Covariance or correlation matrix?

In remote-sensing applications principal components are usually calculated from a variance–covariance matrix (Donker and Mulder 1976, Byrne *et al.* 1980, Townshend 1984). It is common practice in psychology and ecology to standardize the covariance matrix by dividing by the appropriate standard deviations and hence to reduce it to the correlation matrix. In effect, the procedure reduces all the variables to equal importance as measured by scale. Clearly, principal components are not invariant under the linear transformation including separate scaling of the original variables. Thus, the principal components of the covariance matrix are not the same as those of the correlation matrix. However, the principal components of the correlation matrix are invariant under separate scaling of the original variables. Whether standardization is desirable is, in the ultimate analysis, to be decided on non-statistical grounds. From the statistical view point it is a nuisance, especially in sampling investigations, because it complicates the distributional theory (Kendall *et al.* 1983). The circumstances governing the choice of matrix are reviewed by Morrison (1967) and Davis (1973).

In the context of remote sensing a case can be made for using a standardized or non-standardized matrix. A non-standardized form may be justified on the basis of possible differences in radiometric resolution between the spectral bands of a sensor. This may occur as the maximum number of levels into which a signal may be quantized depends on the signal-to-noise ratio (SNR) and the confidence level that can be assigned when discriminating between levels (Slater 1980). Where this is true, the differences in variances between spectral bands should arguably be retained since they indicate the relative amount of useful signal. Another possibility not considered further in this paper, would be to convert the digital counts to radiance values using appropriate calibration curves and then to use the variance–covariance matrix derived from the radiances. Since each band would have the same physical units ($W M^{-2} sr^{-1}$), the derivation of the components from a non-standardized matrix can be justified. On the other hand one may not wish to make the assumption that the relative importance of the underlying components should be controlled by the relative magnitude of the SNR in each band or even assume that the information contained within each band is necessarily a function of the precision with which the spectral data are sensed. For example, sensors with narrow bands can usually record data with less precision than sensors with broad spectral bands, but the former may be more important because of their ability to detect specific absorption features.

In remote sensing using LANDSAT MSS and other scanner data, the loadings of the eigenvectors have been used in the past to describe the relative contributions of each original band to the transformed components. As far as LANDSAT MSS data are concerned, using a variance–covariance matrix one finds that the first principal component (PC1) usually has all the four bands positively weighted on the basis of the size of their standard deviations, whereas principal component two (PC2) is the difference between the infrared bands and the visible bands (Donker and Mulder 1976, Lodwick 1979). The last two components contain very little variance and are dominated by noise (Santisteban and Munoz 1978). When the variance–covariance matrix is used, if the variances across all the spectral bands are uniform then the acquisition of one component reduces the uncertainty about the images by a factor of $1/N$ where N is the number of bands used, otherwise the components are weighted by the variances of the bands.

Obviously, there is no *a priori* reason why each band should not contribute equal weight to the analysis. In fact, this has important implications for the methods of change detection using multitemporal principal components analysis. Essentially, there are two factors which contribute to the variance between multitemporal images. The substantial sources of variation between the images are due to external conditions, such as atmospheric transmission, angle of Sun incidence and differences between detector calibration procedures of sensors. Changes in land cover introduce variances which are confined to smaller portions of the scene and these variances are orthogonal to that of the former (Byrne *et al.* 1980). One way to minimize the external variances is to standardize all the data from different bands to a standard deviation of one. Then each band would contribute equal variance. There is no reason why proportionate changes in one band should not be considered as important as proportionate changes in the other bands. Because of the low response of reflected light in absolute terms from one band as compared with the others, it does not imply that similar relative changes are less important (Lodwick 1979). In order to investigate the differences resulting from the use of the covariance or correlation matrix in the principal components transformation an experiment was conducted with LANDSAT MSS data.

3. Principal component analysis

Two 256 × 256 pixel LANDSAT MSS subscenes were chosen for the analysis. Subscene 1 was extracted from a LANDSAT-1 CCT and subscene 2 from a LANDSAT-2 CCT, both relating to the north-eastern part of India (path 145, row 042). The digital-image processing was done on a microcomputer-based image processing system housed in the Department of Geography, University of Reading, with indigenously developed software. A description of the hardware configuration and software system can be found in Harrison and Singh (1984). Both the subscenes were pre-processed to reduce six line banding, data dropouts and bit slips and the grey values were converted to 8 bit integers for all spectral bands. The principal components analysis was performed using both the variance–covariance matrix and the correlation matrix.

4. Results and discussion

Tables 1 and 2 give the variance–covariance matrices and the correlation matrices for the two MSS subscenes. The magnitudes of eigenvalues and eigenvectors are depicted in figures 1–3. A comparison of the eigenvectors in figures 2 and 3 show that a significant difference results from standardizing the variables to unit variance and it seems reasonable to regard these standardized variables as a better basis for carrying

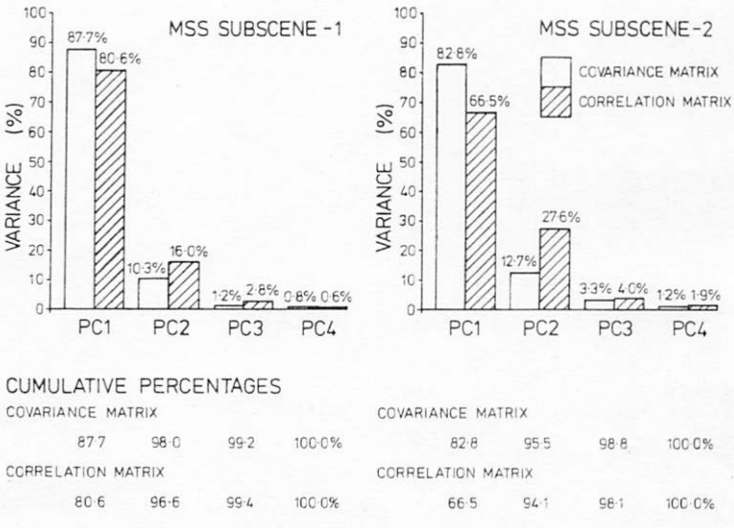

Figure 1. Magnitude of eigenvalues for MSS subscenes.

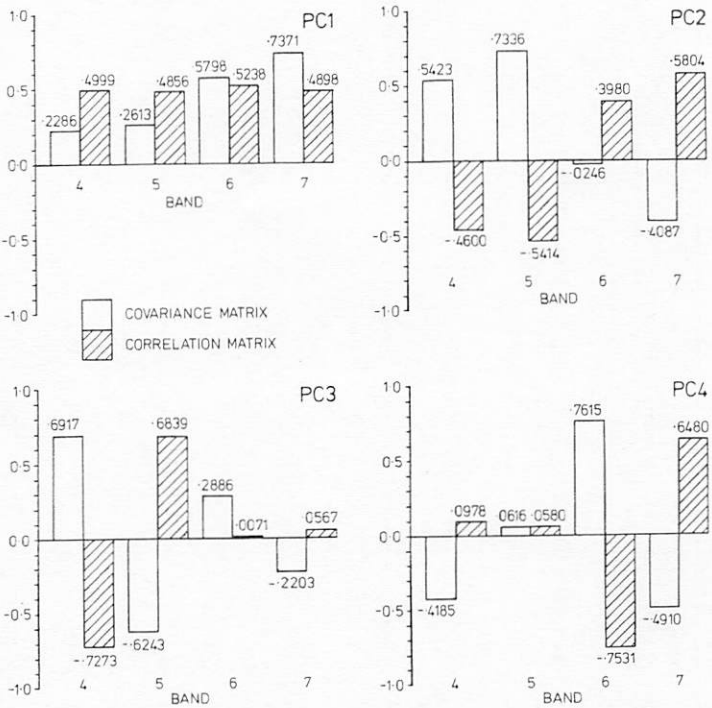

Figure 2. Magnitude of eigenvectors for MSS subscene 1.

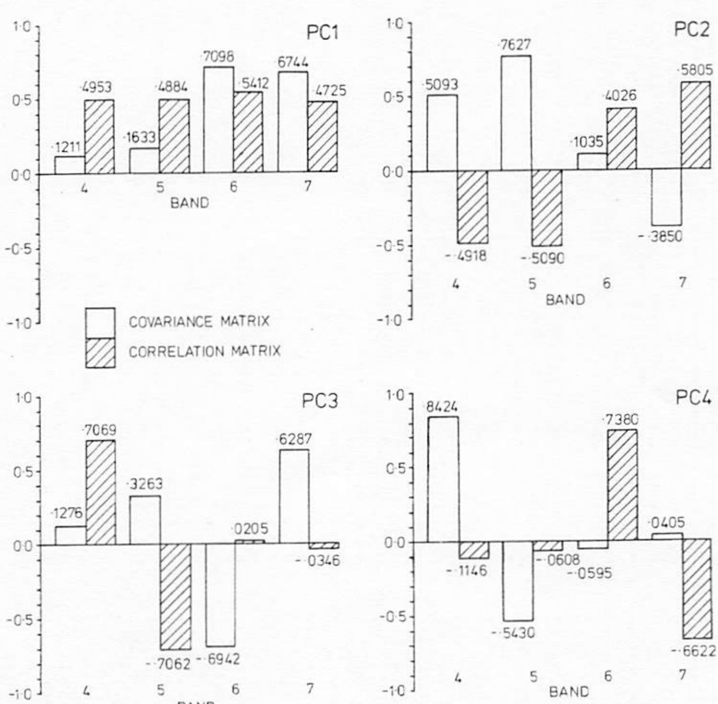

Figure 3. Magnitude of eigenvectors for MSS subscene 2.

Table 1. Covariance matrix, correlation matrix and eigenvalues for MSS subscene 1.

MSS band	4	5	6	7
4	70·03	74·62	96·48	105·28
5		101·10	109·04	117·33
6			253·44	313·67
7				418·47

Eigenvalues

1	2	3	4
739·42	87·21	9·42	6·99

Correlation matrix

MSS Band	4	5	6	7
4	1·00	0·89	0·72	0·62
5		1·00	0·68	0·57
6			1·00	0·96
7				1·00

Eigenvalues

1	2	3	4
3·22	0·64	0·11	0·03

out principal components analysis. The major eigenvector (PC1) is now more uniformly weighted across all spectral bands rather than being heavily dominated by the infrared band. The second eigenvector (PC2), while still describing the difference between the red and infrared bands, is now inverted. Figures 4–7 show the principal component images of subscene 1 using the covariance matrix and figures 8–11 show the principal component images, for the same subscene, resulting from use of the correlation matrix. Visual inspection of the first principal component images (figures 4 and 8) reveals that the image obtained from the correlation matrix has better contrast. The significant improvement in image enhancement, however, occurs in the second principal component image using the correlation matrix (figure 9) which, in comparison with the corresponding image computed from the covariance matrix (figure 5), appears to be more useful for visual interpretation purposes. While there are visual differences between the third and fourth principal component images using both methods they contain very little information. On the basis of these observations, therefore, it is worthwhile to compare the relative performance of using either the covariance or correlation matrix in principal components analysis.

Ready and Wintz (1973) have demonstrated that significant improvements in SNR can be obtained from noisy multispectral data using principal components. The eigenvectors of the spectral covariance matrix were shown to be insensitive to additive noise, while the eigenvalues increased by an amount corresponding to the additive noise variance. They concluded that since λ_N is approximately zero for noiseless correlated data, the value of λ_N computed from noisy data provides an

Table 2. Covariance matrix, correlation matrix and eigenvalues for MSS subscene 2.

Covariance matrix

MSS Band	4	5	6	7
4	23·67	27·40	32·75	20·04
5		45·33	43·54	26·40
6			190·89	165·77
7				179·56

Eigenvalues

1	2	3	4
364·01	56·20	14·28	4·95

Correlation matrix

MSS band	4	5	6	7
4	1·00	0·84	0·49	0·31
5		1·00	0·47	0·29
6			1·00	0·90
7				1·00

Eigenvalues

1	2	3	4
2·65	1·11	0·76	0·08

approximation to the noise variance. This remains equally true for eigenvalues computed using the correlation matrix since standardizing the variables has no effect on the variance of uncorrelated randomly distributed noise. Reference to figure 1 shows that noise variance estimates using λ_N for both MSS subscenes are similar within 1 per cent variance whichever method of transformation to principal components is used.

In practice, random noise is not uniformly distributed throughout multispectral data as the dominance of atmospheric noise and optical/detector noise varies according to spectral bandwidth; for instance, MSS 7 has a significantly higher SNR than the other MSS spectral bands (Norwood 1974) largely due to its comparative immunity from atmospheric scattering. The eigenvectors computed using standardized variables are equally sensitive to all spectral bands irrespective of the distribution of variance or SNR in the original data and provide a set of unbiased eigenvalues. The benefits of this can be demonstrated in terms of the relative SNR improvement over the maximum SNR in the original spectral bands achieved by principal components analysis using the two methods.

The SNR improvement (ΔSNR) achieved by principal components analysis can be calculated from the eigenvalues of the original spectral band variances (Ready and Wintz 1973)

$$\Delta \mathrm{SNR} = \lambda_1 / \sigma_x^2$$

Figure 4. First principal component image of MSS subscene 1 using covariance matrix.

Figure 5. Second principal component image of MSS subscene 1 using covariance matrix.

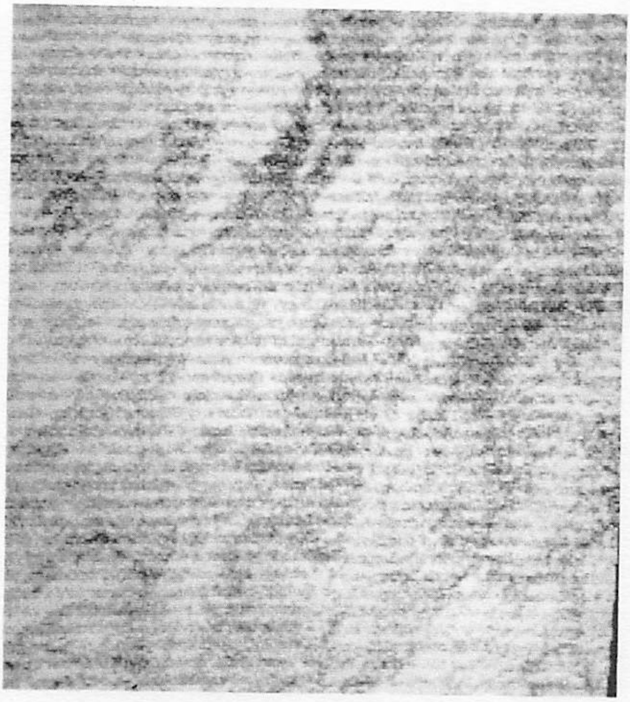

Figure 6. Third principal component image of MSS subscene 1 using covariance matrix.

Figure 7. Fourth principal component image of MSS subscene 1 using covariance matrix.

Figure 8. First principal component image of MSS subscene 1 using correlation matrix.

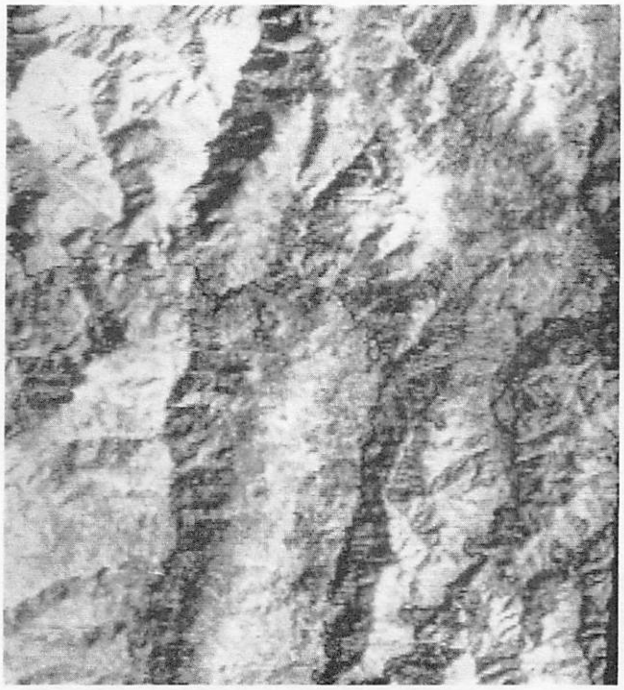

Figure 9. Second principal component image of MSS subscene 1 using correlation matrix.

Figure 10. Third principal component image of MSS subscene 1 using correlation matrix.

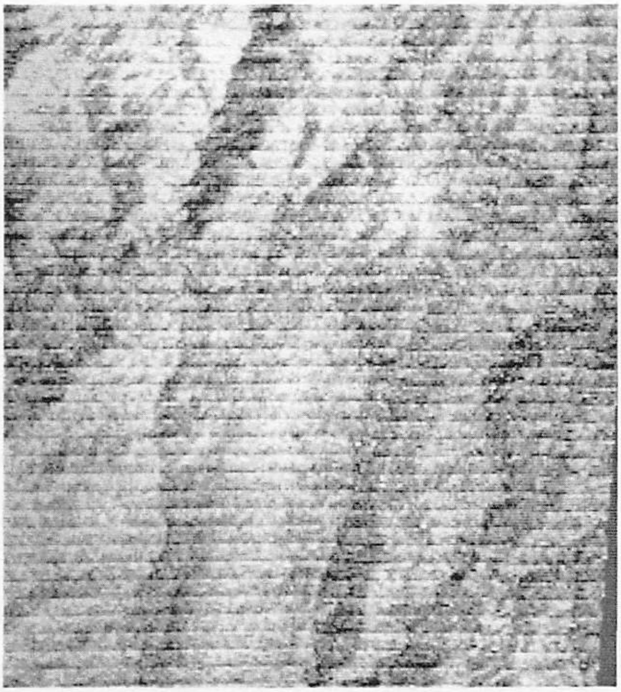

Figure 11. Fourth principal component image of MSS subscene 1 using correlation matrix.

where

$$\sigma_x^2 = \max\,[\sigma_{x_1}^2, \sigma_{x_2}^2, \ldots, \sigma_{x_N}^2]$$

In the first MSS subscene using the eigenvalues derived from the covariance matrix

$$\Delta\mathrm{SNR} = 739{\cdot}42/418{\cdot}47 = 2{\cdot}47\,\mathrm{dB}$$

and using the eigenvalues derived from the correlation matrix

$$\Delta\mathrm{SNR} = 3{\cdot}22/1{\cdot}0 = 5{\cdot}08\,\mathrm{dB}$$

A 2·61 dB difference in SNR improvement over maximum original SNR is therefore realized with eigenvectors computed from the correlation matrix. This result is repeated using the results of principal components analysis from the second MSS subscene. Thus, using eigenvalues derived from the covariance matrix

$$\Delta\mathrm{SNR} = 364{\cdot}01/190{\cdot}89 = 2{\cdot}80\,\mathrm{dB}$$

and using eigenvalues derived from the correlation matrix

$$\Delta\mathrm{SNR} = 2{\cdot}65/1{\cdot}0 = 4{\cdot}23\,\mathrm{dB}$$

resulting in a 1·43 dB difference in SNR improvement between the two approaches in favour of the correlation matrix method. These results verify the improved visual discrimination noted in PC1 and confirm the enhancement of SNR improvement through the use of standardized variables.

The improved visual appearance of PC2 would be expected from an examination of the eigenvalues in figure 1, where the variance of PC2 as a percentage of the total variance in the data is greater where PC2 is derived from the correlation matrix. It is interesting to note that the second eigenvalue (PC2) from the correlation matrix of the second MSS subscene is greater than the maximum original spectral band variance

$$\Delta\mathrm{SNR} = 1{\cdot}10/1{\cdot}0 = 0{\cdot}41\,\mathrm{dB}$$

This increased SNR improvement of the second eigenvector (PC2) relates to the magnitude and orientation of the major eigenvector (PC1) to which the former is orthogonal and the latter optimal for all spectral dimensions when computed from standardized variables via the correlation matrix. The magnitude and orientation of the second eigenvector is not, therefore, biased by the major eigenvector and able to explain an increased percentage of the total image variance. The eigenvector loadings for PC2 computed from the correlation matrix, shown in figures 2 and 3, testify to this.

It has often been argued that because the information content of digital images is invariant under a unitary transform and the variance of a variable is a measure of its information content the compression strategy is to discard variables with low variances (Wintz 1972). However, in some studies using LANDSAT Thematic Mapper bands it has been found that the higher-order components contained valuable information for land-cover discrimination in comparison with some of the lower-order components, which had no obvious information content (Townshend 1984). In this connection it must be pointed out that principal components analysis is an exploratory technique of constructing new artificial variables which do not necessarily have any physical meaning or significance. They are simply linear combinations of variables that can be measured, but they themselves cannot, in general, be observed directly (Kshirsagar 1972). They provide a different perspective for viewing the original data within the feature space and are optimal in the sense of minimizing mean square error approximation. Naturally, the greater the number of principal components chosen, the better will be the performance of the new variables in explaining the internal relationship of the original variables. In any case, images contain a significant amount of structure that cannot be accounted for by their means, covariances and first-order probability density function (Wintz 1972).

In the situation where it is necessary to replace p images by $k<p$ images without loss of information the efficiency of any choice of k linear functions will depend on the extent to which these functions will permit reconstruction of the N original images. A method of reconstructing the original images is to determine the best linear predictor where the efficiency of prediction of each k linear function may be measured by the residual variances σ_i^2 and where $\sum \sigma_i^2$ is an overall measure of predictive capability. The best choice of linear functions which minimizes $\sum \sigma_i^2$ are the first k principal components (Rao 1973).

5. Conclusion

It has been found that a significant improvement in SNR and image enhancement is realized by employing the correlation rather than the variance–covariance matrix in the principal components analysis. Hence, it may well be desirable to construct co-ordinate axes which decorrelate variates of the feature vector on the basis of the correlation with other variables. However, it must be cautioned that if the off-diagonal elements of the correlation matrix are approximately equal then the interpretation value of a principal components analyses of this matrix would be dubious (Seal 1964).

Acknowledgments

The authors are grateful to Dr. J. R. G. Townshend, Department of Geography, University of Reading, and Dr. D. C. Mason, NERC Thematic Information Services, for reading the initial draft of this paper and making very valuable suggestions. Andrew Harrison acknowledges the permission of the Director of NERC Scientific Services to publish this paper and Ashbindu Singh is thankful to the Commonwealth Scholarship Commission in the U.K. and the Forest Department, Government of Manipur, India, respectively, for their financial support and sponsorship.

References

Byrne, G. F., Crapper, P. F., and Mayo, K. K., 1980, Monitoring land cover change by principal component analysis of multitemporal Landsat data. *Remote sensing Environ.*, **16**, 3.

Davis, J. C., 1973, *Statistics and Data Analysis in Geology* (New York: Wiley).

Devijver, P. A., and Kittler, J., 1982, *Pattern Recognition: A Statistical Approach* (Englewood Cliffs, New Jersey: Prentice Hall).

Donker, N. H. W., and Mulder, N. J., 1976, Analysis of MSS digital imagery with the aid of principal components transform. *ITC J.*, **3**, 434.

Gillispie, A. R., 1980, Digital techniques of image enhancement. In *Remote Sensing in Geology*, edited by B. S. Siegal and A. R. Gillispie (New York: Wiley), pp. 139–226.

Gonzalez, R. C., and Wintz, P., 1977, *Digital Image Processing* (Reading, Massachusetts: Addison-Wesley).

Harrison, A., and Singh, A., 1984, Microcomputer based digital image processing systems in remote sensing. *Photogramm. Engng remote Sensing* (submitted).

Hotelling, H., 1933, Analysis of a complex of statistical variables into principal components. *J. educ. Psychol.*, **24,** 417, 498.

Kendall, M. A., Stuart, A., and Ord, J. K., 1983, *The Advanced Theory of Statistics*, Vol. 3, 4th edition (London: Charles Griffen).

Kshirsager, A. M., 1972, *Multivariate Analysis* (New York: Marcel Dekker).

Lodwick, G. D., 1979, Measuring ecological changes in multitemporal Landsat data using principal components. *Proceedings of the 13th International Symposium on Remote Sensing of the Environment*, Ann Arbor, Michigan, pp. 1131–1141.

Moik, J. G., 1980, *Digital Processing of Remotely Sensed Images*. NASA SP-431, Washington DC.

Morrison, D. F., 1967, *Multivariate Statistical Methods* (New York: McGraw-Hill).

Norwood, V. T., 1974, Balance between resolution and signal to noise ratio in scanner design for Earth resources system. *SPIE Proc. Scanners Images Earth Obs.*, **51,** 37.

Rao, C. R., 1973, *Linear Statistical Inference and Its Applications*, 2nd edition (New York: Wiley).

Ready, P. J., and Wintz, P. A., 1973, Information extraction, SNR improvement and data compression in multi-spectral imagery. *I.E.E.E. Trans. Commun.*, **21,** 1123.

Santisteban, A., and Munoz, L., 1978, Principal components of a multispectral image. Application to a geological problem. *IBM Jl Res. Dev.*, **22,** 444.

Seal, H. L., 1964, *Multivariate Statistical Analysis for Biologists* (London: Methuen).

Slater, P. N., 1980, *Remote Sensing: Optics and Optical Systems* (Reading, Massachusetts: Adison-Wesley).

Townshend, J. R. G., 1984, Agricultural land cover discrimination using Thematic Mapper spectral bands. *Int. J. remote Sensing*, **5,** 681.

Townshend, J. R. G., Goff, T. E., and Tucker, C. J., 1984, Multitemporal dimensionality of images of the normalized difference vegetation index at continental scales. *I.E.E.E. Trans Geosci. remote Sensing* (submitted).

Watanabe, S. 1965, Karhunen–Loève expansion and factor analysis, theoretical remarks and applications. *Transactions of the Fourth Prague Conference on Information Theory*, Prague Czechoslovakia.

Wintz, P. A., 1972, Transform picture coding. *Proc. Inst. elect. electron. Engrs*, **60,** 809.