B. C. Forster
School of Surveying
Centre for Remote Sensing
University of New South Wales
Sydney, Australia

Principle and Rotated Component Analysis of Urban Surface Reflectances

Principal component analysis allows a data set of measurements which are correlated with each other to be transformed into one where the variables are mutually uncorrelated.

Landsat MSS response data consists of measurements which are correlated with each other. Principal component analysis allows this data set to be transformed into one where the variables are mutually uncorrelated. In addition, the method allows the determination of the number of linearly independent sources of variation within the Landsat data set that can effectively summarize the data without a significant loss of information. Because of correlation, this will be generally less than the number of variables. A good description of the principal component method can be found in Mather (1976).

A number of researchers have used principal component analysis for the analysis of rural areas, including Donker and Mulder (1976), Austin and Mayo (1978), and Lodwick (1979); for land-use change detection, for example, Bryne and Crapper (1979) and Bryne *et al.* (1980); and for mapping fire burns and vegetation regeneration (Richards and Milne, 1984). Generally, it has been found that the first two principal components of a Landsat data set contribute on the order of 98 percent of the total variance. The first principal component is seen as representing a general brightness component, and the second a general greeness component or, as Donker and Mulder (1976) suggest, "the designation color is a better one."

This paper examines the results of the application of principal component analysis to urban residential data by comparing the form to that determined in rural studies, by explaining the nature of the components, and by suggesting new explanatory, rotated orthogonal axes. The Landsat data analyzed was derived from 2800 pixels from 70 sampling areas over residential areas of the city of Sydney, Australia. These areas comprised essentially single family detached housing with housing density ranging from 5 to 20 houses per hectare. For specific details of the sampling procedure, see Forster (1983). Landsat digital count values were converted to radiance using known sensor characteristics (G.S.F.C., 1982). Radiance values were then converted to reflectance using an atmospheric correction procedure (Forster, 1981) so that the results were temporally independent. The procedure was essentially derived from a model proposed by Turner and Spencer (1972), where the attenuating effect of aerosols is estimated from visibility measures taken at the time of the satellite overpass. For each pixel the cover proportions of grass, trees, and various urban materials were also determined. On average, each pixel contained 36 percent grass surface, 23 percent roof material (primarily red terracotta tiles), 16 percent evergreen tree cover, 14 percent road material (bitumen), 5 percent concrete material, and 6 percent other roof material, water, and bare soil.

The data were analyzed using an SPSS (Nie *et al.*, 1975) factor analysis program, without rotation, which is equivalent to a principal component analysis when the communalities are set to one, i.e., all variables are retained in the analysis. The coefficient score matrix was determined, with each component normalized to one. These coefficients represent the weights to be given to each variable so that composite scales can be built that represent the theoretical dimensions associated with the respective principal components. The coefficient matrix, with the respective percentage of variance contributed by each component, is shown in Table 1.

The results of Table 1 clearly indicate the basic band sum, visible/infrared difference common to principal component analysis. Austin and Mayo (1978) indicated that their first principal component represented approximately 90 percent of the variance, and a similar proportion was determined by

Photogrammetric Engineering and Remote Sensing,
Vol. 51, No. 4, April 1985, pp. 475-477.

0099-1112/85/5104-0475$02.25/0

TABLE 1. UNROTATED NORMALIZED COEFFICIENT SCORE MATRIX FOR REFLECTANCE DATA

	PC1	PC2	PC3	PC4
R4	0.565	−0.397	0.723	−0.018
R5	0.548	−0.430	−0.655	0.295
R6	0.529	0.473	−0.176	−0.683
R7	0.318	0.659	−0.135	−0.668
% Variance	50.7	43.3	3.8	2.2
Cumulative %	50.7	94.0	97.8	100.0

Ri = Reflectance, Band i; PCi = Principal Component, i.

Donker and Mulder (1976). This represents a major difference from the results of this study where, from Table 1, the first component (PC1) contributes to only 50.7 percent of the variance while the second component (PC2) contributes 43.3 percent. In addition, the final two components (PC3 and PC4) contribute a larger percentage of the variance, a total of 6 percent, compared to less than 1 percent for both of the cited results.

While vegetative cover is the primary cover in rural areas, this is not the case in urban residential areas, where a large proportion of the cover is non-vegetative material. Urban residential areas have a different visible/infrared signature from vegetation, so that, while brightness will still be a major contributor to the variance, the visible/infrared difference will be nearly as important. Because of the great diversity found in an urban scene when compared to a rural scene, the third and fourth components may also be marginally significant contributors to the variance.

The correlations between the four derived components and the various cover variables were calculated in order to determine whether the components represented some underlying structure of the cover data. The correlation matrix is given in Table 2. Surfaces represented are house (predominantly red/brown tiled roofs), road (bitumen surfaces), concrete, tree (predominantly eucalypts), and grass (healthy mown grass). Immediately apparent is that non-vegetative surfaces primarily correlate with the second component, while the vegetative surfaces, particularly tree cover, are relatively equally correlated with the first two components. This suggested that the sum of the vegetative surfaces would be positively correlated with PC2, and a correlation of 0.76 was calculated. Further, the negative and positive correlation of tree and grass cover, respectively, with PC1 and their positive correlation with PC2 indicated that these two variables were essentially orthogonal. This was confirmed by a calculated low correlation of −0.13 between these two cover variables. It followed, therefore, that rotated components could represent a tree and a grass dimension.

TABLE 2. CORRELATION MATRIX OF UNROTATED PRINCIPAL COMPONENTS AND COVER VARIABLES

	PC1	PC2	PC3	PC4
House	+0.20	−0.46	−0.22	−0.08
Road	+0.01	−0.62	0.00	−0.04
Concrete	+0.27	−0.24	−0.15	−0.05
Tree	−0.47	+0.49	+0.01	+0.02
Grass	+0.19	+0.51	+0.18	+0.10

TABLE 3. VARIMAX ROTATED NORMALIZED COEFFICIENT SCORE MATRIX FOR REFLECTANCE DATA

	PR1	PR2	PR3	PR4
R4	0.727	−0.036	−0.698	−0.080
R5	0.663	−0.002	0.714	−0.173
R6	−0.151	0.638	−0.017	0.707
R7	0.100	0.769	0.046	−0.682
% Variance	47.7	45.9	3.7	2.7
Cumulative %	47.7	93.6	97.3	100.0

An orthogonal varimax rotation was applied to the four-band Landsat reflectance data. This method of rotation is widely used (Nie *et al.*, 1975) and defines a simple component as one with only ones and zeros in the columns. The rotated matrix of the normalized scores is given in Table 3. It should be noted that the rotated PC1 will no longer contribute to the maximum variance. Rotated components are given the symbol PRi, i = 1 to 4.

A quite different data structure has resulted, with the first component being strongly dependent on the visible reflectance or visible brightness, the second on the infrared brightness, while the third and fourth rotated components have a similar structure to the unrotated components (see Table 1).

The correlation matrix of rotated components with all cover variables is given in Table 4. An additional cover variable is included—Green—which represents the surface percentage covered by vegetative material (i.e., grass plus tree percentage cover).

Tree cover percentage was calculated to have a correlation of −0.66 with PR1, and grass cover percentage was calculated to have a correlation of 0.52 with PR2. The first rotated component was, therefore, considered to represent a tree cover dimension and the second rotated component a grass cover dimension. The sum of the vegetative surfaces was found to be equally correlated with PR1 and PR2, with correlations of −0.58 and 0.50, respectively. The total vegetative cover could, therefore, be rep-

TABLE 4. CORRELATION MATRIX OF ROTATED COMPONENTS WITH COVER VARIABLES

	PR1	PR2	PR3	PR4
House	0.42	−0.25	0.21	0.15
Road	0.37	−0.50	0.01	0.06
Concrete	0.34	−0.03	0.15	0.11
Tree	−0.66	0.12	−0.03	−0.08
Grass	−0.13	0.52	−0.16	−0.14
GREEN	−0.58	0.50	−0.15	−0.16

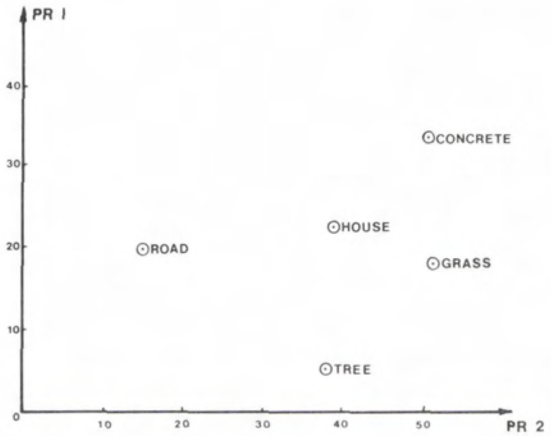

FIG. 1. Location of various residential cover surfaces in the rotated PR1 and PR2 system.

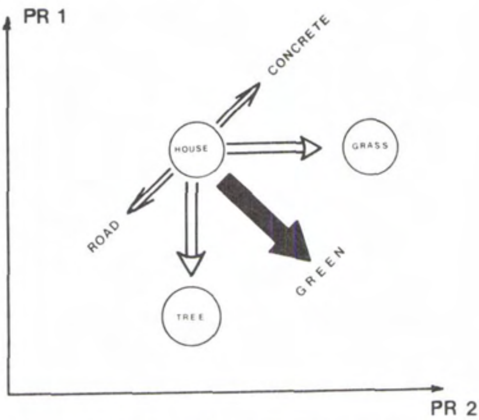

FIG. 2. Schematic relationship between various urban residential cover types with respect to rotated components PR1 and PR2.

resented by an axis at approximately 45° to the orthogonal negative direction of PR1 and the positive direction of PR2. That is, ((PR2) + (−PR1)) should be highly correlated with total vegetative cover, showing that this direction parallels the original PC2. A correlation coefficient of 0.76 was determined, which is the same as previously calculated for PC2 and vegetative cover percentage.

Using the scores of Table 3 as weighing coefficients and substituting a calculated 100 percent reflectance value for each cover (Forster, 1981), the coordinates of each cover in the rotated system were determined. These values are plotted in Figure 1. It can be seen that house percentage cover represents the reflectance of an average non-vegetative surface and that the 100 percent non-vegetative surface reflectances define an axis approximately at right angles to the green dimension, previously mentioned, and approximately parallel to a line defining the 100 percent vegetative surfaces. Concrete cover lies at one end of this axis and road cover at the opposite end. All of these relationships are shown schematically in Figure 2.

Research into the application of principal component analysis and rotated components to urban studies is being continued, and further extended to multi-temporal scene analysis of land-use change, and to Landsat Thematic Mapper data from urban areas (Forster, 1984).

References

Austin, M. P., and K. K. Mayo, 1978. Some results from principal component analysis on Landsat imagery. *Proceedings, U.S.-Australia Workshop on Image Processing Techniques for Remote Sensing*, Canberra.

Bryne, G. R., and P. F. Crapper, 1979. An example of the detection of changes between successive Landsat images by numerical methods in an urban area. *Proceedings, 1st Australasian Conference on Remote Sensing*, Sydney.

Bryne, G. R., P. F. Crapper, and K. K. Mayo, 1980. Monitoring land cover change by principal component analysis of multi-temporal Landsat data. *Remote Sensing of Environment*, 10, 175-184.

Donker, N. H. W., and N. J. Mulder, 1976. Analysis of M.S.S. digital imagery with the aid of principal component transformation. *Proceedings, Comm. VII, 13th Congress I.S.P.*, Helsinki.

Forster, B. C., 1981. Some measures of urban residential quality from Landsat multispectral data. *Unisurv S18*, University of N.S.W., Sydney.

———, 1983. Some urban measurements from Landsat data. *Photogrammetric Engineering and Remote Sensing*, Vol. 49, No. 12, 1693-1707.

———, 1984. Urban feature extraction from Thematic Mapper Data. Presented paper, Commission VII, 15th International Congress of Photogrammetry and Remote Sensing, Rio de Janeiro.

Goddard Space Flight Center, 1982. *Landsat 4 to ground station interface description*. Document No. 435-D-400, Supplementary Information, G.S.F.C., August, 1982.

Lodwick, G. D., 1979. Measuring ecological changes in multitemporal Landsat data using principal components. *Proceedings, 13th Inter. Symp. on Rem. Sens. of Environment*, Ann Arbor.

Mather, P. M. 1976. *Computational Methods of Multivariate Analysis in Physical Geography*. John Wiley and Sons, London.

Nie, N. H., G. Hadlai Hull, J. G. Jenkins, K. Steinbrenner, and D. H. Bent, 1975. *Statistical package for the Social Sciences* (Second Edition). McGraw Hill, New York.

Richards, J. A., and A. K. Milne, 1984. Mapping fire burns and vegetation regeneration by classification of multi-temporal Landsat MSS image data. *Proceedings, 3rd Australasian Remote Sensing Conference*, Queensland.

Turner, R. E., and M. M. Spencer, 1972. Atmospheric model for correction of spacecraft data. *Proceedings of the 8th International Symposium on Remote Sensing of Environment*, Vol. II, Ann Arbor, 895-934.

(Received 13 June 1983; revised and accepted 30 October 1984)