Application of Principal Components Analysis to Change Detection

Tung Fung and *Ellsworth LeDrew*
Department of Geography, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

ABSTRACT: Principal components analysis (PCA) has been applied for land-cover change detection with multitemporal Landsat Multispectral Scanner (MSS) data. Previous work found that the higher order principal components were able to account for land-cover changes. In the computation of principal components, the eigenvectors used for transformation can be derived from a covariance matrix (with non-standardized data) or a correlation matrix (with standardized data); and from the total area or a subset area of specific land-cover types. In this paper, we examine the effect of using different types of matrices for principal components transformation with special emphasis on its application in land-cover change detection in the Kitchener-Waterloo-Guelph area, Ontario, Canada. It is found that standardized principal components are more accurate than the non-standardized components because of their better alignment along land-cover changes in the multiemporal data structure. Statistics extracted from the total study area are also better and more reliable than those extracted from the subset area. It is concluded that principal components analysis is scene dependent, and the use of this technique requires a careful appraisal of the eigenstructures and images of the principal components.

INTRODUCTION

PRINCIPAL COMPONENTS ANALYSIS (PCA), is a commonly used technique for remote sensing image analysis. It has been used for determining the underlying dimensions of remotely sensed data (Ready and Wintz, 1973; Anuta *et al.*, 1984), data enhancement for geological applications (Gillespie, 1980; Santisteban and Munoz, 1978), and land-cover change detection (Lodwick, 1979; Byrne et al., 1980; Richards, 1984; Singh, 1986).

Multispectral data acquired from most remote sensors exhibit high interband correlations. Taking Landsat Multispectral Scanner (MSS) imagery as an example, band 1 (green) and band 2 (red) are highly correlated because of the relatively low reflectance of vegetation. Bands 3 and 4, the two infrared bands, are highly correlated because of the high reflectance of vegetation. Similarly, for the Thematic Mapper (TM) data, the first three visible bands (TM1, TM2, and TM3) are also highly correlated (Townshend, 1984). Data processing with all of the spectral bands therefore involves a certain degree of redundancy. This increases the cost of data processing, especially when addressing change detection issues which involve images of more than one date.

To alleviate this, PCA has been used as a data compression technique (Ingebritsen and Lyon, 1985). Uncorrelated linearly transformed components are derived from the original data such that the first principal component accounts for the maximum possible proportion of the variance of the original data set, and subsequent components account for the maximum proportion of the unexplained residual variance, and so forth. PCA has the characteristics of preserving the total variance in the transformation and minimizing the mean square approximate errors. For these reasons, it is an attractive data reducing technique.

In this paper, we examine the effect of using principal components transformation in land-cover change detection. The principal components are based on the eigenvectors of the covariance or correlation matrices. These matrices may be extracted from a subset area or the total study area of a Landsat MSS image. An experiment is undertaken to examine image enhancement for change detection based on standardized versus non-standardized, and total versus specific land cover principal components. Eigenstructures and each individual image of the principal components so derived are compared to evaluate their information content for land-cover change detection.

THE COMPUTATIONAL PROCEDURE

The entire principal components transformation can be divided into three steps (Richards, 1986):

- derivation of the variance-covariance matrix,
- computation of the eigenvectors, and
- linear transformation of the data set.

Mathematically, given an N-dimensional variable $\mathbf{X}$ with mean vector $\mathbf{M}$ and variance-covariance matrix $\mathbf{C}_x$, $\mathbf{C}_x$ can be estimated as

$$\mathbf{C}_x = \frac{1}{K-1} \sum_{i=1}^{K} (\mathbf{X}_i - \mathbf{M})(\mathbf{X}_i - \mathbf{M})^{\mathrm{T}}$$

where K is the number of pixels. Each principal component $\mathbf{Y}_j$ is expressed as

$$\begin{aligned} \mathbf{Y}_j &= a_{1j}X_1 + a_{2j}X_2 + \ldots + a_{Nj}\ X_N \\ &= \mathbf{a}_j^{\mathrm{T}}\mathbf{X} \end{aligned}$$

where $\mathbf{a}_j^{\mathrm{T}}$ is the transpose of the normalized eigenvectors of the variance-covariance matrix $\mathbf{C}_x$ of $\mathbf{X}$. Thus, the entire transformation can be written as

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

where $\mathbf{A}$ is the matrix of eigenvectors which diagonalizes the covariance matrix $\mathbf{C}_x$ of $\mathbf{X}$ such that the covariance matrix $\mathbf{C}_y$ of $\mathbf{Y}$ is

$$\mathbf{C}_y = \mathbf{A}\mathbf{C}_x\mathbf{A}^{\mathrm{T}}$$

whose diagonal elements are the eigenvalues $\boldsymbol{\lambda}_N$ of $\mathbf{C}_x$ i.e.,

$$\mathbf{C}_y = \begin{bmatrix} \lambda_1 & & & & 0 \\ & \lambda_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ 0 & & & & \lambda_N \end{bmatrix}$$

where $\lambda_1 > \lambda_2 > \ldots > \lambda_N$.

MATRIX OF WHAT?

In the computational procedure, the covariance matrix $\mathbf{C}_x$ of the original data set $\mathbf{X}$ can be the covariance matrix of the total number of pixels or an estimate of the matrix from a sample of pixels. For example, statistics can be extracted from a decimated image for a sample of pixels of every m pixels and m lines (Anuta *et al.*, 1984), or through the use of training sets. Training sets of a combination of different land-cover units are selected so that the statistics extracted can be used to provide an estimate of the actual total variance. It may also be the covariance matrix of a specific land-cover unit so that the rotated components can highlight the predetermined features of interest (Duggin *et al.*, 1986).

PHOTOGRAMMETRIC ENGINEERING AND REMOTE SENSING,
Vol. 53, No. 12, December 1987, pp. 1649–1658.

0099-1112/87/5312–1649$02.25/0

In addition, the principal components can also be standardized to zero value mean and unit variance with the use of the correlation instead of the covariance matrix (Singh and Harrison, 1985). This is especially useful in multitemporal analysis because standardization can minimize the differences due to atmospheric conditions or Sun angles. Different eigenvectors derived from the two matrices certainly produce different principal components.

THE STUDY AREA

The study area is composed of the cities of Kitchener, Waterloo, Cambridge, and Guelph and the surrounding rural landscape (approximately 80° 10′ W to 80° 40′ W longitude and 43° 20′ N to 43° 40′ N latitude) (Figure 1). The total area covered is about 790 sq km, which is equivalent to 388 by 600 pixels at 80-m spatial resolution. The twin cities of Kitchener-Waterloo have experienced a rapid urban development, especially along the western fringe of the cities. The rural land is mainly used for agricultural production. Forest, wetland, and gravel pits are other land-cover types found in the rural area.

DATA DESCRIPTION AND DATA PREPROCESSING

Three Landsat MSS images dated 19 August 1981, 10 April 1984, and 24 August 1984 of Path 19 and Row 30 were used for the analysis. The August images were used for change detection purposes. The April and August, 1984 images were used for demonstration of the variability of the principal components. Aerial photographs taken during April and May 1980 and 1985 were also available for reference. They were useful for the identification of rural to urban land conversion, but the early spring condition of the photos was of little help for crop identification. Field work was carried out during August, 1986 to check actual land-cover changes. The field data collected were compared with aerial photos and the images. Nine major land-cover changes were identified:

- from cropland to construction sites,
- from cropland to residential area,
- from construction sites to residential area,
- from corn to pasture,
- from corn to bare soil (harvested grain fields),
- from pasture to corn;
- from pasture to bare soil;
- from bare soil to corn, and
- from bare soil to pasture.

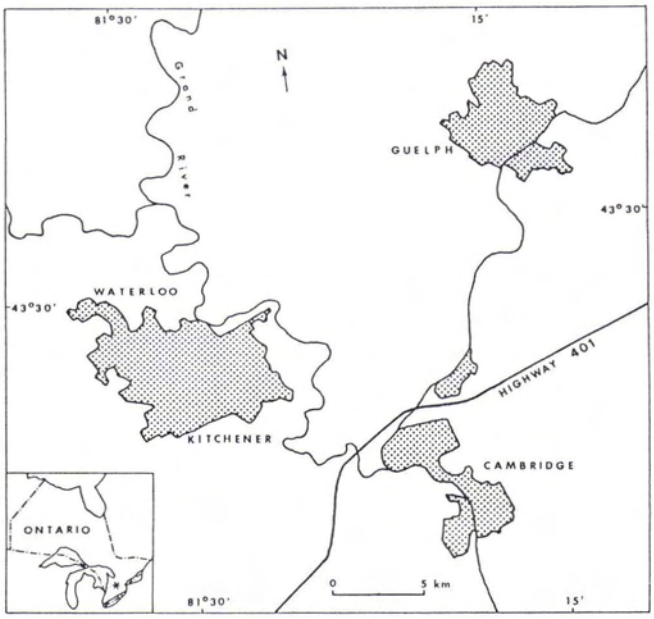

FIG. 1. The study area.

The first three types were related to rural to urban land conversion. The rest of the changes were associated with crop type changes.

Data preprocessing of the three image data sets included radiometric correction and geometric registration. Radiometric correction was used to compensate for the differences in calibration of the sensors of different satellites as well as the differences in solar elevation. All of the data were converted to reflectance values (Robinove, 1982; Markham and Barker, 1987). Assuming that atmospheric differences were substantial sources of variance and they should therefore be evident as orthogonal principal components to the variances related to land-cover changes (Byrne *et al.*, 1980), atmospheric correction was not performed.

The 1984 August image was selected as the master reference image for geometric registration. Twenty-three and 24 control points were used for registrating the 1981 August and 1984 April images, respectively. Applying the nearest neighbor resampling routine, the standard errors were 0.25 pixel (X-direction) and 0.26 pixel (Y-direction) for the 1981 August image, and 0.44 pixel (X-direction) and 0.29 pixel (Y-direction) for the 1984 April image.

STANDARDIZATION VERSUS NON-STANDARDIZATION OF THE MATRICES

SINGLE DATE DATA

Using a single date Landsat MSS image for illustration, Singh and Harrison (1985) compared the use of non-standardized (the use of covariance matrix) versus standardized (the use of correlation matrix) data for rotation. Non-standardized principal components were justified because of the possible differences in radiometric resolution between the spectral bands. They found that the non-standardized PC1 (first principal component) usually had positive loadings of eigenvectors on all of the spectral bands of Landsat MSS. The non-standardized PC2 was the difference between the visible and infrared bands of MSS. These two components were able to account for over 95 percent of the total variance of a single date MSS data. The third and fourth components usually contained noisy elements with low eigenvalues.

Standardization meant that each band would have equal variance. Singh and Harrison (1985) found that the standardized PC1 was similar to the non-standardized one in terms of the eigenvector loadings and signs. The standardized PC2 was also a contrast between the visible and infrared bands, but the signs of the loadings were the reverse of that of the non-standardized data. The authors found that both the standardized PC1 and PC2 were visually more enhanced as compared with the non-standardized ones. This was verified by comparing the values of signal-to-noise ratio (SNR) improvement (Ready and Wintz, 1973) as expressed by

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

$$\text{where } \sigma_x^2 = \max(\sigma_{x_1}^2, \sigma_{x_2}^2, \ldots, \sigma_{x_N}^2)$$

$$\sigma_{x_i}^2 = \text{variance of original band } X_i$$

The ΔSNR of standardized PCs were higher than the non-standardized ones. Their examination of eigenvalues also showed that the standardized PC2 had a higher proportion of total variance than the non-standardized PC2.

In this paper, the eigenstructures of the 1981 and 1984 August Landsat MSS data were extracted separately for both the standardized and non-standardized data based on all pixels of the study area (Tables 1 and 2). It can be shown that statistics from both dates are very similar. From Table 1 we see that the non-standardized PC1s are heavily loaded on the infrared bands. This is attributed to the predominately agricultural landscape of the study area with a great variety of vegetation types, ranging

from high infrared reflecting pasture to low reflecting vegetated wetland. The PC2s, in contrast, are loaded mainly on the visible bands. These two components account for about 98 percent of the total variance in both images. The PC3s and PC4s are differences between the individual visible or infrared bands.

In contrast to the findings of Singh and Harrison (1985), the standardized PCs are clearly distinct from the non-standardized ones. The standardized PC1s are evenly loaded on all of the spectral bands and represent a contrast between the visible and infrared bands. The negative loadings on the infrared bands suggest that the PC1s are negative greenness measures in which high biomass objects are dark in tone instead of bright in tone as is the case in most vegetation indices. The PC2s are positively loaded on all the bands which is interepreted as a brightness measure. A comparison between the components shows that the non-standardized ones are in essence a summary of either the visible or the infrared bands whereas the standardized ones are better in describing the underlying data dimension of MSS data of this particular study area.

Multidate Data for Change Detection: Separate Rotation

For land-cover change detection, principal components analysis has been applied in two ways. Lodwick (1979) and Singh (1986) performed two separate rotations for two single-date images and then compared the derived PCs with other digital change detection techniques such as image differencing or regression analysis. This kind of application assumes that the PCs so derived represent similar spectral information. This is possible only if the areas of change occupy a small proportion of the total area.

Table 1. Eigenstructure of 19 August 1981 Data

Non-standardized PCs				
	component			
band	1	2	3	4
1	−0.12	0.57	−0.05	0.81
2	−0.20	0.74	−0.31	−0.56
3	0.44	0.37	0.81	−0.14
4	0.86	0.07	−0.50	0.05
eigenvalues	46.99	11.58	0.57	0.30
% variance	79.06	19.48	0.95	0.51
Standardized PCs				
	component			
band	1	2	3	4
1	0.50	0.50	−0.69	0.14
2	0.53	0.43	0.72	0.10
3	−0.41	0.63	−0.02	−0.66
4	−0.55	0.41	0.05	0.73
eigenvalues	2.43	1.48	0.05	0.03
% variance	60.87	37.03	1.27	0.83

Table 2. Eigenstructure of 24 August 1984 Data

Non-standardized PCs				
	component			
band	1	2	3	4
1	−0.12	0.60	−0.21	0.76
2	−0.18	0.72	−0.17	−0.65
3	0.58	0.34	0.74	0.02
4	0.78	0.00	−0.62	−0.05
eigenvalues	32.93	8.39	0.56	0.37
% variance	77.95	19.86	1.32	0.87
Standardized PCs				
	component			
band	1	2	3	4
1	0.48	0.52	−0.70	0.04
2	0.52	0.46	0.70	0.16
3	−0.44	0.59	0.10	−0.67
4	−0.55	0.41	−0.03	0.73
eigenvalues	2.44	1.44	0.08	0.04
% variance	61.11	35.90	2.06	0.92

Table 3. Eigenstructure of 10 April 1984 Data

Non-standardized PCs				
	component			
band	1	2	3	4
1	0.27	0.58	0.19	0.74
2	0.33	0.64	−0.50	−0.49
3	0.59	−0.03	0.72	−0.37
4	0.69	−0.50	−0.45	0.26
eigenvalues	31.11	4.36	0.35	0.26
% variance	86.23	12.07	0.98	0.73
Standardized PCs				
	component			
band	1	2	3	4
1	0.49	−0.52	−0.67	−0.19
2	0.50	−0.45	0.73	−0.03
3	0.52	0.36	−0.11	0.77
4	0.48	0.63	0.03	−0.61
eigenvalues	3.29	0.61	0.06	0.03
% variance	82.36	15.37	1.58	0.69

This technique is illustrated by comparing the PCs derived from two images of different seasons in this study. The 10 April 1984 MSS image is compared with the 24 August 1984 image. Differences between seasons can be expresssed by the lack of field crops on ground, appearance of unmelted snow, predominance of wet bare soil in the agricultural land, and leafless deciduous woodland on the April image. In contrast, the August image represents predominantly vegetation subjects.

A comparison between the two eigenstructures (Tables 2 and 3) illustrates that both the non-standardized and standardized PCs of the different dates contain quite different information. Especially for the standardized PCs, the April PC1 is a brightness measure while the August PC1 is a greenness measure. In contrast, the standardized April PC2 is a greenness measure but the August PC2 is a brightness measure. In other words, the information content of PCs derived for the two dates reverses in order due to the large difference in land cover between the two dates, i.e., the difference in scene content. The non-standardized PCs are less extreme in contrast but the loadings and signs indicate that the April PCs are not merely a summary of either the visible or infrared bands. For instance, the April non-standardized PC2 is equally loaded on bands 1, 2 (positive), and 4 (negative).

Considerable error would certainly be involved if other change detection techniques such as image differencing are applied using PCs of different information content. Even if images of the same season are used, provided that there is an area of significant change occupying a large proportion of the study area, the PCs obtained will also be different in information content. This renders the separate rotation questionable. A careful comparison between the individual PC images derived and their eigenstructures is essential before any other techniques can be applied.

Multidate Data for Change Detection: Merged Rotation

Another approach for applying PCA for change detection is using a single rotation of the merged multidate data. Both Byrne *et al.* (1980) and Richards (1984) demonstrated that higher order PCs (e.g., PC3 and PC4) were able to account for land-cover changes. Both studies used covariance matrices (i.e., non-standardized data) in deriving eigenvectors for rotation. Richards (1984) also noted that this approach should be applied to a study

TABLE 4. EIGENSTRUCTURE OF MULTITEMPORAL NON-STANDARDIZED PCS

	component							
band	1	2	3	4	5	6	7	8
81 1	−0.10	0.18	0.42	−0.31	−0.04	−0.05	0.33	0.76
2	−0.16	0.30	0.52	−0.45	−0.23	−0.25	−0.19	−0.51
3	0.37	−0.18	0.33	−0.29	0.54	0.56	−0.13	−0.09
4	0.71	−0.46	0.17	−0.03	−0.34	−0.37	0.09	0.02
84 1	−0.07	−0.01	0.39	0.46	−0.10	0.25	0.67	−0.33
2	−0.09	−0.07	0.48	0.56	−0.14	0.07	−0.60	0.23
3	0.33	0.51	0.08	0.30	0.55	−0.47	0.04	−0.02
4	0.46	0.61	−0.16	0.02	−0.45	0.43	−0.07	0.05
eigenvalues	60.96	19.19	16.28	3.47	0.58	0.52	0.37	0.26
% variance	59.98	18.88	16.02	3.41	0.57	0.51	0.37	0.26

TABLE 5. EIGENSTRUCTURE OF MULTITEMPORAL STANDARDIZED PCS

	component							
band	1	2	3	4	5	6	7	8
81 1	0.38	0.34	−0.26	−0.39	−0.14	−0.68	0.19	0.07
2	0.39	0.29	−0.36	−0.35	0.16	0.67	−0.08	0.18
3	−0.26	0.46	0.36	−0.42	−0.04	0.05	−0.28	−0.58
4	−0.37	0.32	0.42	−0.19	0.01	0.04	0.27	0.69
84 1	0.38	0.33	0.27	0.43	−0.68	0.17	0.00	0.01
2	0.38	0.29	0.35	0.36	0.69	−0.09	0.16	−0.11
3	−0.28	0.44	−0.37	0.38	0.11	−0.18	−0.59	0.22
4	−0.37	0.33	−0.41	0.23	−0.03	0.14	0.66	−0.30
eigenvalues	3.69	2.37	1.22	0.52	0.08	0.05	0.04	0.03
% variance	46.09	29.68	15.24	6.53	1.03	0.58	0.46	0.38

area where a large proportion of the area belonged to areas of no change, i.e., areas of high correlation. Thus, these areas of no change, or common variance of the two dates, could be explained by the first two PCs as they sought to account for the maximum possible variance of the multidate data. Areas of land-cover changes should occupy only a minor proportion of the entire scene so that the minor components (PC3 and PC4) would pick up their variance.

Ingebritsen and Lyon (1985) applied similar techniques with standardized PCs. They noted that the common variances of the two different dates could be expressed in two PCs and named them brightness and greenness accordingly. Land-cover changes were detected in minor PCs which were changes in brightness and changes in greenness.

THE EXPERIMENT

Two principal component transformations, one based on the covariance matrix and the other based on the correlation matrix, were performed for the merged 1981 and 1984 August data set. Tables 4 and 5 illustrate the eigenstructures of these two rotated components. Figures 2 and 3 show the first four PCs of both transformations.

For the non-standardized components, PC1 is positively loaded on the infrared bands and negatively loaded on the visible bands. But it is more heavily loaded on the infrared bands, which suggests that it is mainly a summary of the multidate infrared reflectance. Comparing the PC1 (Figure 2a) with the original single date band 4 images illustrates that any high infrared reflecting objects in either date are also bright in the PC1. PC2 (Figure 2b) is loaded positively on the 1981 red band and the 1984 infrared band but is loaded negatively on the 1981 band 4. It is thus a contrast between a negative vegetation index of 1981 and the 1984 infrared bands. It enhances changes among vegetation objects and can be named as an increase in greenness measure. PC3 (Figure 2c) is mainly loaded on the visible bands and cannot account for any land-cover changes. PC4 (Figure 2d) is negatively loaded on the 1981 bands but positively loaded on the 1984 bands. It measures changes in brightness. Any areas with an increase in brightness (such as changes from cropland to bare soil or construction sites) give a bright tone in this component and vice versa. The rest of the higher order components are mainly contrast between either bands 1 and 2 or bands 3 and 4 of the images and contain little useful information for land-cover changes.

The standardized components are found to have a better alignment along the object of interest. PC1 is a negative measure of greenness which is evenly loaded on each band with positive loadings for visible bands and negative loadings for infrared bands (Table 5 and Figure 3a). Thus, high biomass objects are dark and low biomass objects are bright in this component. PC2 (Figure 3b) is a measure of brightness as the loadings are all positively and evenly distributed. PC3 (Figure 3c) is a decrease in greenness measure. The 1981 visible bands are negatively loaded and 1981 infrared bands are positively loaded. The 1984 loadings show reversed signs from the 1981 counterparts. Changes among the crop types are highlighted in this component. It appears to be a negative of the non-standardized PC2 but they contain similar information. However, the loadings of this standardized PC3 indicate that it has a better alignment to the land-cover changes. PC4 is a measure of change of brightness (Figure 3d). Objects with an increase in brightness are bright in this image while those with a decrease in brightness, such as changes from construction sites to residential areas, are dark in this image. Visual interpretation indicates that the standardized PC4 and non-standardized PC4 (Figures 3d and 2d) appear to be identical to each other.

TOTAL VERSUS SUBSET AREA

In this section of the paper we examine the effect of using statistics extracted from the total study area versus those ex-

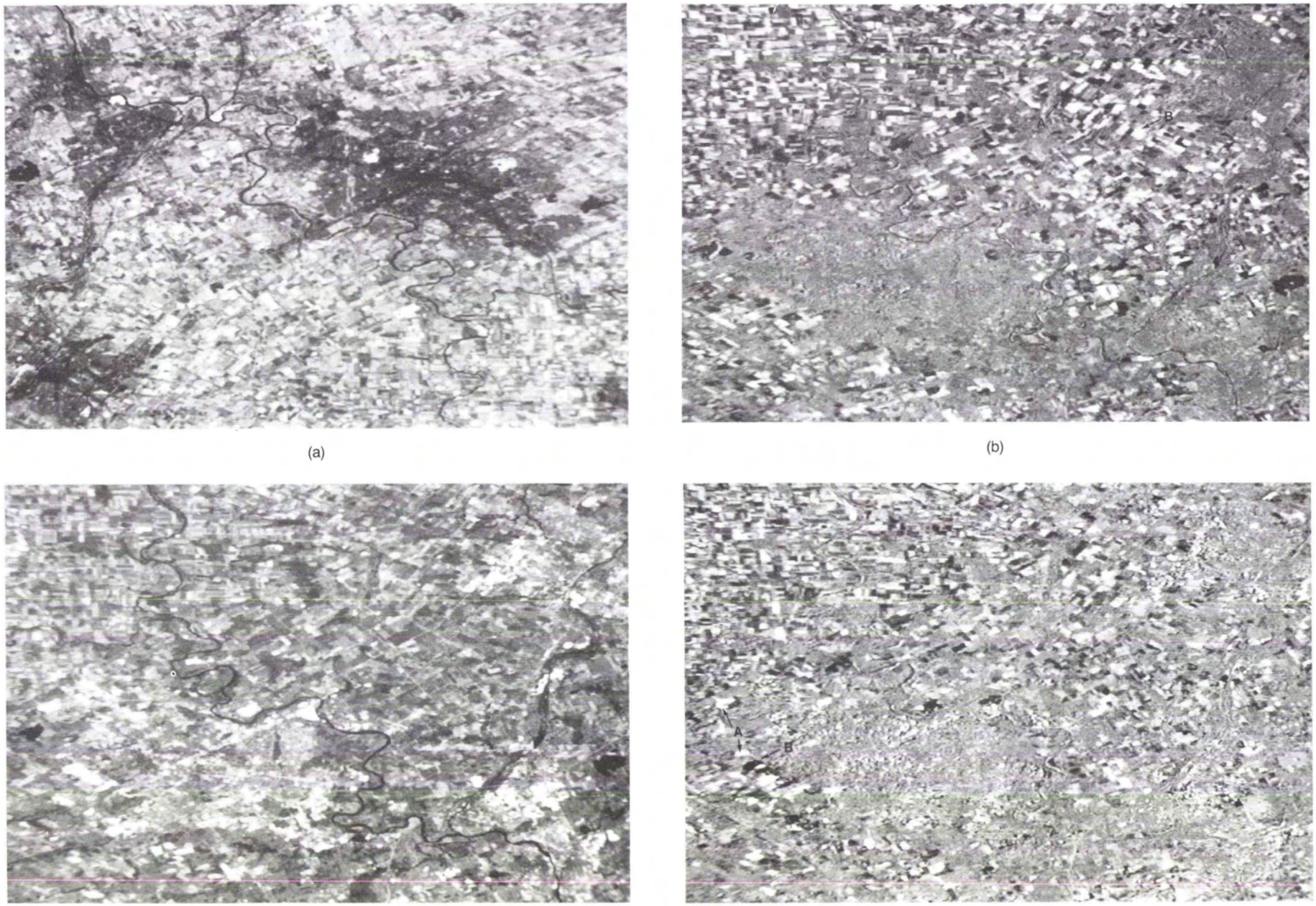

FIG. 2. Multitemporal non-standardized principal components. (a) PC1. A summary of the multidate infrared bands. (b) PC2. A change in greenness measure. A = changes from pasture to grain. B = changes from corn to pasture. (c) PC3. A summary of the multidate visible bands. (d) PC4. A change in brightness measure. A = changes from cropland to construction sites. B = changes from construction sites to residential land.

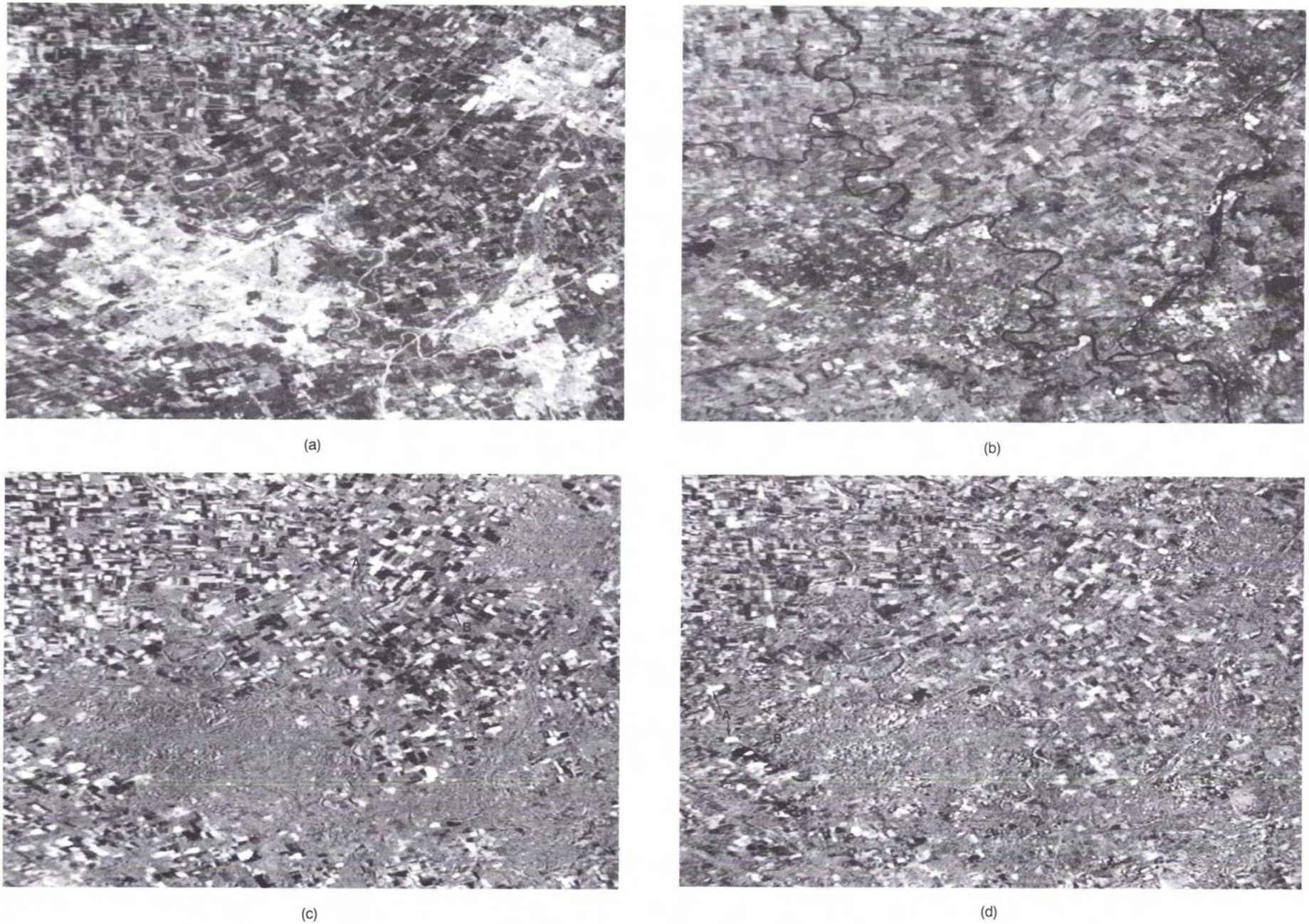

FIG. 3. Multitemporal standardized principal components. (a) PC1. A negative measure of greenness. (b) PC2. A brightness measure. (c) PC3. A change in greenness measure. A = changes from pasture to grain. B = changes from corn to pasture. (d) PC4. A change in brightness measure. = changes from cropland to construction sites. B = changes from construction sites to residential land.

TABLE 6. EIGENSTRUCTURE OF VEGETATION TO VEGETATION CHANGES (V-V) PCS

Non-standardized PCs								
	component							
band	1	2	3	4	5	6	7	8
81 1	0.00	−0.05	0.52	−0.04	−0.17	0.15	−0.03	0.82
2	−0.05	−0.09	0.68	−0.07	−0.33	0.32	0.01	−0.57
3	0.48	0.24	0.44	−0.17	0.58	−0.38	0.00	−0.08
4	0.68	0.50	−0.19	0.12	−0.39	0.30	0.01	0.02
84 1	0.01	−0.03	0.12	0.57	−0.14	−0.33	0.73	0.00
2	0.04	−0.07	0.12	0.59	−0.18	−0.37	−0.68	−0.04
3	−0.38	0.52	0.10	0.43	0.44	0.44	−0.04	0.01
4	−0.40	0.64	0.06	−0.31	−0.36	−0.45	−0.00	−0.01
eigenvalues	80.78	9.21	2.20	1.23	0.61	0.58	0.26	0.21
% variance	84.96	9.68	2.32	1.29	0.64	0.61	0.27	0.22
Standardized PCs								
	component							
band	1	2	3	4	5	6	7	8
81 1	−0.00	0.61	−0.40	0.20	0.05	−0.65	−0.01	0.06
2	−0.19	0.59	−0.31	−0.09	−0.03	0.72	−0.00	0.05
3	0.47	−0.02	−0.20	0.46	0.06	0.17	−0.01	−0.71
4	0.46	−0.12	−0.12	0.48	0.08	0.18	−0.04	0.70
84 1	0.13	0.39	0.65	0.24	−0.59	0.00	0.05	−0.00
2	0.29	0.33	0.47	−0.19	0.74	0.02	0.04	−0.01
3	−0.46	−0.01	0.19	0.44	0.22	0.03	−0.71	−0.04
4	−0.47	−0.06	0.10	0.48	0.23	0.04	0.70	−0.00
eigenvalues	3.89	1.89	1.19	0.43	0.32	0.19	0.04	0.03
% variance	48.67	23.58	14.93	5.43	4.05	2.44	0.55	0.35

TABLE 7. EIGENSTRUCTURE OF VEGETATION TO NON-VEGETATION CHANGES (V-N) PCS

Non-standardized PCs								
	component							
band	1	2	3	4	5	6	7	8
81 1	−0.03	0.07	0.02	0.55	−0.23	0.06	0.01	0.80
2	−0.09	0.10	0.02	0.67	−0.40	0.09	−0.07	−0.60
3	0.52	−0.04	0.05	0.44	0.71	−0.13	−0.04	−0.07
4	0.84	−0.09	−0.00	−0.17	−0.49	0.10	0.04	0.01
84 1	0.04	0.44	−0.36	0.03	0.01	−0.23	0.79	−0.04
2	0.09	0.53	−0.57	−0.07	−0.00	−0.14	−0.60	0.04
3	0.05	0.58	0.34	−0.09	0.16	0.72	0.04	−0.00
4	0.04	0.40	0.65	−0.07	−0.13	−0.61	−0.10	0.01
eigenvalues	24.18	9.31	7.12	2.59	0.59	0.55	0.43	0.21
% variance	53.77	20.69	15.83	5.76	1.32	1.22	0.95	0.46
Standardized PCs								
	component							
band	1	2	3	4	5	6	7	8
81 1	0.28	0.38	0.15	0.52	−0.68	0.05	0.08	0.10
2	0.26	0.49	0.08	0.40	0.72	−0.01	−0.02	0.10
3	0.15	−0.49	0.21	0.47	0.08	−0.02	−0.09	−0.67
4	0.10	−0.60	0.14	0.31	0.08	0.01	0.05	0.72
84 1	0.51	−0.07	−0.43	−0.08	−0.08	−0.51	−0.52	0.05
2	0.47	−0.12	−0.50	−0.06	0.05	0.22	0.68	−0.10
3	0.50	−0.03	0.33	−0.35	−0.01	0.63	−0.34	0.02
4	0.31	0.01	0.60	−0.34	0.01	−0.54	0.37	−0.03
eigenvalues	2.37	2.22	1.57	1.38	0.17	0.13	0.11	0.07
% variance	29.57	27.69	19.63	17.29	2.06	1.61	1.32	0.82

tracted from a subset of the area. Horler and Ahern (1986) used two principal components transformations in their study of forestry with TM data. The first principal component rotation was based on a sample of general land-cover types and the second one was derived from softwood statistics. They noted that the general PC1 was a measure of brightness which was a weighted

TABLE 8. EIGENSTRUCTURE OF NON-VEGETATION TO VEGETATION CHANGES (N-V) PCS

Non-standardized PCs								
				component				
band	1	2	3	4	5	6	7	8
81 1	0.10	0.29	−0.43	0.01	0.02	0.15	0.82	−0.13
2	0.14	0.38	−0.70	−0.07	−0.01	−0.38	−0.43	0.09
3	0.25	0.50	0.10	−0.03	0.05	0.76	−0.30	0.00
4	0.32	0.57	0.55	−0.01	−0.03	−0.49	0.14	0.00
84 1	−0.02	0.05	−0.02	0.62	−0.27	0.06	0.08	0.73
2	−0.07	0.08	−0.03	0.66	−0.31	−0.02	−0.13	−0.66
3	0.59	−0.28	−0.05	0.37	0.66	−0.05	−0.02	−0.03
4	0.67	−0.33	−0.06	−0.21	−0.62	0.04	0.01	−0.02
eigenvalues	22.11	15.55	6.35	1.59	0.59	0.45	0.31	0.27
% variance	46.84	32.94	13.45	3.36	1.24	0.94	0.65	0.57
Standardized PCs								
				component				
band	1	2	3	4	5	6	7	8
81 1	0.50	0.11	−0.18	−0.41	−0.03	−0.70	−0.23	0.04
2	0.47	0.10	−0.20	−0.48	0.00	0.70	−0.05	−0.09
3	0.53	−0.02	−0.08	0.37	−0.01	−0.07	0.75	0.10
4	0.43	−0.07	−0.00	0.65	0.01	0.12	−0.61	−0.08
84 1	0.13	0.36	0.65	−0.04	−0.66	0.03	−0.01	0.05
2	0.10	0.50	0.45	−0.01	0.74	0.00	0.00	0.07
3	0.15	−0.52	0.44	−0.15	0.13	−0.08	0.08	−0.68
4	0.13	−0.57	0.33	−0.14	0.10	0.07	−0.08	0.71
eigenvalues	2.90	2.34	1.26	1.02	0.28	0.08	0.06	0.06
% variance	36.22	29.20	15.69	12.75	3.53	1.05	0.80	0.77

sum of all bands. The general PC2 was a greenness measure as it represented the contrast between the visible (TM1, TM2, and TM3) and near-infrared (TM4) bands. The general PC3 was a contrast between the middle-infrared (TM5 and TM7) bands and the first four shorter wavelength bands.

In their softwood transformation, however, the PC2 was essentially a 'blueness' component as it was heavily weighted on TM1 (0.94) which indicated that this PC was important for softwood identification. The softwood PC3 was also different from that of the general transformation because it was mainly a contrast between the near-infrared (TM4) and the middle-infrared bands (TM5 and TM7).

Duggin *et al.* (1986) also used principal components analysis to study the urban features in Washington D.C. with multidate TM data. They defined forest, dense urban, downtown, and airport as the training areas for four transformations. Their intention in using training areas was to 'rotate the measurement axis . . . to make the selected land-use type more separate from "everything else"' (p.98). They also found that 'for each land-use type typified by a training area, there is a substantial difference between the eigenvector for a given principal component evaluated over the training area and the eigenvectors for the other training areas, and that for the whole image' (p.104).

To examine the effect of using total versus subset statistics for PC transformation, an unsupervised classification map of 32 spectral classes was used to derive the statistics. Spectral classes were used instead of training sites because of the small field size in the image. Also, the use of training sites might not produce a sufficient number of pixels to estimate the class variance. Four categories of land cover changes were identified. They are

- vegetation to vegetation changes (V-V),
- vegetation to non-vegetation changes (V-N),
- non-vegetation to vegetation changes (N-V), and
- non-vegetation to non-vegetation changes (N-N).

In Tables 6, 7, 8, and 9, the eigenstructures of the non-standardized and standardized data are presented. Only the first four PCs are discussed because the fifth and other higher order PCs have low eigenvalues and account for small percentages of the total variance.

For the V-V PCs (Table 6), both the standardized and non-standardized PC1s can reveal changes among vegetation objects (e.g., changes among corn, pasture, and grain). The standardized PC3 is also a measure of changes in brightness. Other PCs in both transformations are not related to land-cover changes. The standardized PCs are again better than the non-standardized PCs for the provision of more change information.

In the V-N transformations, no change can be detected from any one of the non-standardized PCs (Table 7). The loadings show that each PC is heavily weighted on the visible and infrared bands of one date or the other. A similar situation is found in the standardized PCs with the exception of PC4 which shows the changes of vegetation to non-vegetation objects, the objects of interest in this rotation.

In the N-V transformations (Table 8), only the non-standardized PC2 is able to highlight changes from vegetation to no-vegetation objects. The rest of the PCs are heavily loaded on one date or the other but do not explain the differences between the two dates.

In the N-N transformations (Table 9), both the standardized and non-standardized PC1s are the contrast between the 1981 visible and 1984 infrared bands, which render both images difficult to interpret. Only the standardized PC4 can account for the changes among vegetation objects.

To compare the results of all those PCs highlighting land-cover changes, thresholding was applied to produce binary images with areas of no change assigned to a 0.0 value and areas of change to a 1.0 value. The threholded images are checked against the air photos and field data collected. Table 10 summarizes the accuracy of the thresholded PCs. PCs accounting for greenness changes are more accurate than those accounting for brightness changes. This is because crops type changes (changes in greenness) occupy a larger areal proportion of the study area than changes due to rural to urban land conversion and changes

TABLE 9. EIGENSTRUCTURE OF NON-VEGETATION TO NON-VEGETATION CHANGES (N-N) PCS

Non-standardized PCs								
				component				
band	1	2	3	4	5	6	7	8
81 1	0.48	0.03	−0.01	0.34	0.52	0.14	−0.06	0.61
2	0.61	−0.08	−0.13	0.45	−0.45	−0.17	−0.26	−0.33
3	0.30	0.48	−0.10	0.01	0.10	0.15	0.74	−0.30
4	0.05	0.81	−0.12	−0.28	−0.13	−0.10	−0.46	0.15
84 1	−0.09	0.18	0.52	0.25	0.51	−0.37	−0.18	−0.43
2	−0.11	0.18	0.65	0.28	−0.49	−0.03	0.24	0.38
3	−0.37	0.18	−0.11	0.51	0.00	0.69	−0.21	−0.19
4	−0.40	0.12	−0.50	0.46	0.00	−0.55	0.20	0.19
eigenvalues	10.78	7.16	5.79	4.80	0.71	0.53	0.44	0.40
% variance	35.24	23.39	18.91	15.69	2.31	1.72	1.43	1.32
Standardized PCs								
				component				
band	1	2	3	4	5	6	7	8
81 1	0.50	0.14	−0.02	0.43	−0.29	−0.28	−0.39	−0.49
2	0.49	0.02	−0.03	0.48	0.30	0.33	−0.03	0.58
3	0.37	0.32	0.47	−0.10	0.02	−0.06	0.71	−0.17
4	0.09	0.34	0.57	−0.42	0.04	0.08	−0.58	0.19
84 1	−0.17	0.61	−0.26	0.08	−0.64	0.21	0.09	0.27
2	−0.17	0.59	−0.30	0.05	0.63	0.13	−0.06	−0.34
3	−0.42	0.17	0.31	0.46	0.12	−0.63	0.02	0.28
4	−0.36	−0.12	0.46	0.43	−0.09	0.59	−0.02	−0.31
eigenvalues	2.74	1.92	1.60	1.16	0.21	0.14	0.12	0.11
% variance	34.19	23.99	19.96	14.52	2.68	1.78	1.50	1.38

between bare soil and crops (changes in brightness). For the changes in greenness PCs, the standardized PC3 extracted from the statistics of the total study area has the highest accuracy of 88.43 percent. Both the non-standardized PC2 and standardized PC3 from the statistics for the study area are more accurate than the other PCs computed from the subset statistics. The three changes in brightness PCs are similar with an accuracy of from 65 to 67 percent.

To summarize, the use of a subset defined in terms of specific land-cover types does not always provide unequivocal change information. This is because the eigenvectors for rotation are only derived from the specifically selected subset of the entire data. These eigenvectors may not always yield greater separability between the specific land cover and other land-cover types. How the other land-cover units are involved in the rotation is unknown. This renders the use of subset statistics less certain in providing useful information than the use of the statistics extracted from the entire data set for this region.

CONCLUSION

In this paper, we have reviewed the concepts in principal components transformation and evaluated this application to land-cover changes in the Kitchener-Waterloo-Guelph area. It is found that the eigenvectors used for rotation can vary significantly depending upon whether standardized or non-standardized data, and the whole data set or a subset of it, are used. Statistics extracted from a data subset are not recommended for land-cover change detection because of the great variability and uncertainty of the unextracted part of the data. Principal components extracted refer only to the subset data and cannot be claimed to be the 'principal components' of the entire study area. The eigenstructure derived from the entire data set is found to be more valid to use for land-cover change detection. This study confirms earlier work (Byrne *et al.*, 1980; Richards, 1984) which suggested that minor components can detect land-cover changes. It is also found that standardized principal components computed from the eigenvectors of the correlation matrix provide more accurate information for change detection than do non-standardized principle components derived from the covariance matrix.

Because Richards (1984) has noted that areas of change should occupy only a minor proportion of the entire study area, it is interesting to know how this proportion would affect the results of change detection. How large should this proportion of areas of change be in order to yield minor PCs which can detect changes? If this proportion is large, should we increase the study area at the expense of increasing data processing time and cost? Or could the first or second PCs pick up the variance of the changes?

Clearly, principal components analysis is a scene dependent technique. We do not know the exact nature of the principal components derived without an examination of the eigenstructure and visual inspection of the image. Thus, even though it is a powerful data reducing technique, it should be used only with a thorough understanding of the characteristics of the study area to avoid drawing any faulty conclusions.

ACKNOWLEDGMENTS

This research has been supported by NSERC operating and infrastructure grants to E. LeDrew, a WATDEC grant of computing equipment to E. LeDrew, and a BILD grant (Board of Industrial Leadership and Development) to R. Newkirk and E. LeDrew. The authors are also grateful to the Canadian Commonwealth Scholarship and Fellowship Committee for the financial support to T. Fung. This is paper 87-01 of the Earth-Observations Laboratory of the Institute for Space and Terrestrial Sciences.

REFERENCES

Anuta, P.E., L.A. Bartolucci, M.E. Dean, D.F. Lozano, E. Malaret, C.D. McGillem, J.A. Valdes, and C.R. Valenzuela, 1984. Landsat-4 MSS and Thematic Mapper Data Quality and Information Content

TABLE 10. ACCURACY OF THE PRINCIPAL COMPONENTS

Component	Accuracy (%) area of no change	area of change	scene	(95% confidence range)*
Changes in Greeness PCs				
Total- PC2 (N)	84.40	75.88	80.59	(76.93-83.79)
Total - PC3 (S)	96.45	78.51	88.43	(85.36-90.02)
VV - PC1 (N)	83.33	69.74	77.25	(73.42-80.68)
VV - PC1 (S)	80.14	76.75	78.63	(74.86-81.96)
VN - PC4 (S)	65.60	36.40	52.55	(48.21-56.85)
NV - PC2 (N)	78.01	66.23	72.75	(68.72-76.43)
NN - PC4 (S)	70.92	75.88	73.14	(69.13-76.80)
Changes in Brightness PCs				
Total- PC4 (N)	84.04	45.61	66.86	(62.66-70.81)
Total - PC4 (S)	84.40	46.49	67.45	(632.7-71.37)
VV - PC3 (S)	78.01	50.00	65.49	(61.26-69.49)

Note:
(N) = Non-standardized PCs.
(S) = Standardized PCs.
Total no. of sample = 510.
No. of correct sample for area of no change = 282
No. of correct sample for area of change = 228
* Computed according to Hord and Brooner (1976)

Analysis, *I.E.E.E. Transactions on Geoscience and Remote Sensing*, GE-22 (3), 222–235.

Byrne, G.F., P.F. Crapper, and K.K. Mayo, 1980. Monitoring Land-Cover Change by Principal Component Analysis of Multitemporal Landsat Data, *Remote Sensing of Environment*, 10, 175–184.

Duggin, M.J., R. Rowntree, M. Emmons, N. Hubbard, A.W. Odell, H. Sakhavat, and J. Lindsay, 1986. The Use of Multidate Multichannel Radiance Data in Urban Feature Analysis, *Remote Sensing of Environment*, 20 95–105.

Gillispie, A.R., 1980. Digital Techniques of Image Enhancement, *Remote Sensing in Geology*, B.S. Siegal and A.R. Gillispie, (eds), John Wiley & Son, New York, pp. 139–226.

Hord, R.M., and W. Brooner, 1976. Land-Use Map Accuracy Criteria, *Phtogrammetric Engineering and Remote Sensing*, 42 (5), 671–677.

Horler, D.N.H., and F.J. Ahjern, 1986. Forestry Information Content of Thematic Mapper Data, *International Journal of Remote Sensing*, 7 (3), 405–428.

Ingebritsen, S.E, and R.J.P. Lyon, 1985. Principal Component Analysis of Multitemporal Image Pairs, *International Journal of Remote Sensing*, 6 (5), 687–696.

Lodwick, G.D., 1979. Measuring Ecological Changes in Multitemporal Landsat Data Using Principal Components, *Proceedings of the 13th International Symposium on Remote Sensing of Environment*, Ann Arbor, Michigan, pp. 1131–1141.

Markham, B.L., and J.L. Barker, 1987. Radiometric Properties of U.S. Processed Landsat MSS Data, *Remote Sensing of Environment*, 22, 39–71.

Ready, P.J., and P.A. Wintz, 1973. Information Extraction, SNR Improvement, and Data Compression in Multispectral Imagery, *I.E.E.E. Transactions on Communications*, COM-21 (10), 1123–1130.

Richards, J.A., 1984. Thematic Mapping from Multitemporal Image Data Using the Principal Components Transformation, *remote Sensing of Environment*, 16, 35–46.

———, 1986. *Remote Sensing Digital Image Analysis*, Springer-Verlag, Berlin.

Robinove, C.J., 1982. Computation of Physical Values for Landsat Digital Data, *Photogrammetric Engineering and Remote Sensing*, 48 (5), 781–784.

Santisteban, A., and L. Munoz, 1978. Principal Components of a Multisepectral Image: Application to a Geological Problem, *IBM Journal of Research and Development*, 11 (5) 444–454.

Singh, A., 1986. Change Detection in the Tropical Forest Environment of Northeastern India Using Landsat, *Remote Sensing and Tropical Land Management*, (M.J. Eden and J.T. Parry eds), John Wiley & Son, Chichester, pp. 237–254.

Singh, A., and A. Harrison, 1985. Standardized Principal Components, *International Journal of Remote Sensing*, 6 (6), 883–896.

Townshend, J.R., 1984. Agricultural Land-Cover Discrimination using Thematic Mapper Spectral Bands, *International Journal of Remote Sensing*, 5 (4), 681–698.

(Received 17 April 1987; revised and accepted 13 August 1987)

Forthcoming Articles

M. X. Borengasser, E. F. Kleiner, P. Vreeland, F. F. Peterson, H. Klieforth, and *J. V. Taranik*, Geological and Vegetational Applications of Shuttle Imaging Radar-B, Mineral County, Nevada.

Albert K. Chong and *S. A. Veress*, Cost Estimates of Photogrammetric Related Services Using Electronic Spreadsheets.

Paul F. Hopkins, Ann L. Maclean, and *Thomas M. Lillesand*, Assessment of Thematic Mapper Imagery for Forestry Applications under Lake States Conditions.

Kurt Kubik and *Dean Merchant*, Photogrammetric Work without Blunders.

Fang Lei and *H. J. Tiziani*, A Comparison of Methods to Measure the Modulation Transfer Function of Aerial Survey Lens Systems from the Image Structures.

Warren L. Philipson and *William L. Teng*, Operational Interpretation of AVHRR Vegetation Indices for World Crop Information.