Ordination of multispectral imagery for multitemporal change analysis using Principal Components Analysis

Joseph M. Piwowar, University of Regina
Andrew A. Millward, University of Waterloo

Abstract: Early change analysis studies established the fundamental basis for applying the Principal Components Analysis (PCA) transformation to remote sensing images acquired on two dates. There are an increasing number of studies, however, which extend this basis to longer image time series with little concern for its appropriateness. In particular, when multispectral and multitemporal data are used in the same analysis, the components may be difficult to interpret since they would contain not only temporal variation, but spectral changes as well. In this paper we sought to establish an appropriate ordination technique to condense the multispectral information from each date prior to multitemporal PCA. Multispectral PCA and Normalized Difference Vegetation Index (NDVI) ordination approaches were applied to a series of four Landsat and SPOT multispectral images spanning a twelve year period. We found that the NDVI technique provides superior results because it produces annual composites with a strong physical basis.

Introduction

Remote sensing has a key role to play in environmental monitoring because it is the only source of data from which we can view the entire planet and monitor changes in the nature of the surface of the Earth through time in a consistent, integrated, synoptic and numerical manner (LeDrew, 1992). As our concern for changes to the Earth's environment heightens we must begin to look for new analysis tools to help us identify where and when these changes are occurring. One such technique that has been used is Principal Components Analysis (PCA). Although the application of

PCA in change detection studies - analyses between *two* image dates - has been thoroughly examined (e.g., Fung and LeDrew, 1987), its use in multitemporal analyses - investigations between *many* image dates - has developed *ad hoc* without much attention paid to its appropriateness. In particular, when multispectral and multitemporal data are used in the same analysis, the components may be difficult to interpret since they would contain not only temporal variation, but spectral changes as well (Eastman and Fulk, 1993). The objective of this paper is to examine several options for addressing this potential for multispectral - multitemporal confusion.

Principal Components Analysis

Many remotely sensed images have significant inter-band correlation that, if not accounted for, can interfere with accurate and timely information extraction. For example, spectral response from a feature that is measured at green wavelengths is typically highly correlated to that feature's response in the red spectral region. Similarly, since many Earth features do not move much there is significant spectral correlation between images acquired days, months, or even years apart. Principal Components Analysis is a mathematical transformation that can remove much of this redundancy (Jensen, 1996). Given a multi-band (multispectral or multitemporal) data set, a PCA will create a new image with fewer, uncorrelated bands, called *components*.

Although PCA will generate the same number of components as there are input variables, a key characteristic of the method is the concentration of the original data's variance into the first components. Thus, there should be a point at which it can be determined that most of the original scene variance has been accounted for, leaving only noise in the remaining components, which can subsequently be discarded. This "cut-off" point can be quite subjective and a variety of evaluation techniques have been devised to have it quantitatively determined (McGarigal *et al.*, 2000). In practical applications with remote sensing imagery, however, the statistical contributions from very small (relative to the entire remote sensing scene), but important, change regions do not typically pass most significance thresholds. We suggest that for multitemporal image analysis the utility of a component should be based more on the analyst's ability to ascribe meaning to the observed spatial and temporal patterns than on blanket statistical tests. This is the approach followed below.

Traditionally, PCA has been applied for image enhancement and to remove inter-channel redundancy (Singh and Harrison, 1985; Tangestani and Moore, 2001), however, it has also been effectively used in two-date change detection studies (e.g., Franklin *et al.*, 2000; Li and Yeh, 2002). PCA can also be a powerful technique for information extraction across many dates (Eastman and Fulk, 1993; Piwowar and LeDrew, 1996; Young and Anyamba, 1999). When applied to multidate imagery, the PCA transformation should isolate the highest differences in image brightness in the first components and more statistically smaller changes in the lower components (Rundquist and Di, 1989).

Most sensor data shows wide differences in their dynamic range, even between simultaneously acquired spectral bands or between different dates. When used in a PCA, bands with higher data ranges tend to dominate the results. This is known as *non-standardized* PCA and can be useful for some applications where a particular band emphasis is desired. Alternatively, the original data can be normalized prior to PCA, thereby giving equal weighting to each input band. Such *standardized* PCAs have been shown to be effective at isolating variations in multitemporal analyses, so they were used throughout our analyses (Eastman and Fulk, 1993; Singh and Harrison, 1995; Fung and LeDrew, 1987; Young and Anyamba, 1999).

Imagery

We examined multispectral satellite imagery acquired of a rapidly urbanizing coastal city in Hainan Province, Southern China. In total, four images were obtained from three different satellite sensors to form a twelve-year chronology (Table 1). All of the imagery was acquired during the months of November and December to minimize the possibility of identifying changes that could be due to differences in the phenological stages of vegetation. To facilitate inter-annual comparisons, the images were co-registered to the 1991 SPOT image with a mean RMS error of

Table 1: *Optical satellite imagery analyzed.*

Year	Sensor	Bands Used
1987	Landsat 5 TM	2, 3, 4
1991	SPOT HRV	1, 2, 3
1997	SPOT HRV	1, 2, 3
1999	Landsat 7 ETM+	2, 3, 4

less than one pixel. Only the bands from similar spectral regions (i.e., green, red, and near-IR) were used to avoid biasing the PCA results from one year with spectral information that was not available from the other years' data.

Ordination Techniques

In order for any multitemporal image analysis to be effective: (a) precise co-registration of each image must be guaranteed; (b) the data must be univariate at each temporal instance; and (c) there must be some normalization of the data values between time slices (Piwowar and LeDrew, 1995). This paper is couched in an examination of the last two criteria. Specifically, given four multispectral images acquired over a twelve year period, we evaluated three *ordination* approaches - methods of reducing the spectral dimensionality of each image to render them univariate and normalized at each temporal instance.

We approached the ordination issue in three ways. For the first two methods, we reasoned that since PCA is a prime ordination technique in itself, it could be applied to the imagery for each year to condense the multispectral information independently (McGarigal *et al.*, 2000). That is, we determined if PCA could be used to address the multispectral - multitemporal confusion concerns by analyzing each year independently (to condense the multispectral information) and then input the resulting components into a second principal components analysis to highlight the multitemporal characteristics. For the third ordination technique, the normalized difference vegetation index (NDVI) was computed from the visible red and near-IR bands from each date before multitemporal PCA processing.

Technique 1: Split Annual PCA Ordination:

When applying PCA to multispectral optical imagery of a typical vegetated terrestrial landscape the first two components will highlight the differences between the visible and near-IR spectral regions (Byrne *et al.*, 1980; Rundquist and Di, 1989). The analysis of our imagery followed this pattern: the first component was consistently highly correlated with the visible bands and the second PC loaded heavily on the near-IR channel. Thus we were able to use PC 1 as an ordination for the first two bands and PC 2 as a representative of the third band. We then used the four PC 1 images (one from each year) in a second principal components analysis to highlight the multitemporal characteristics evident in the visible spectral

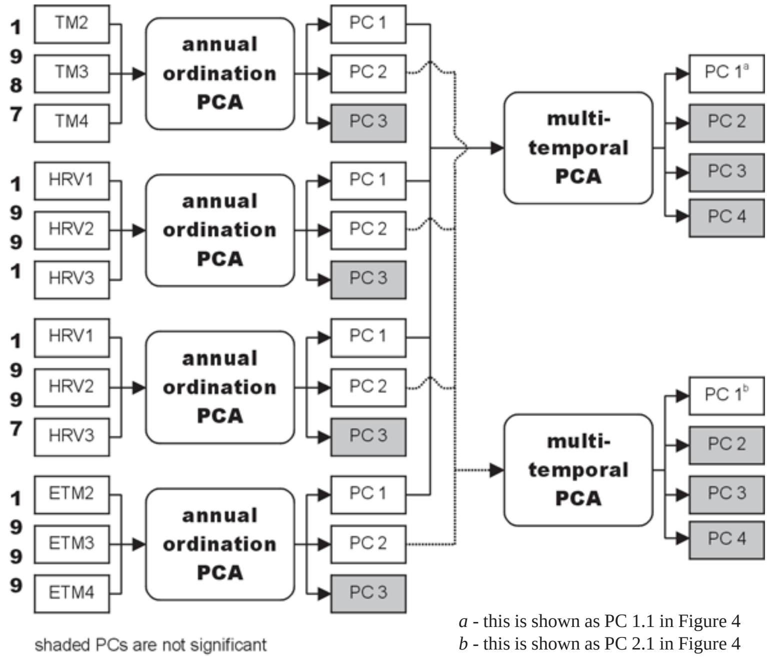

Figure 1: *Multitemporal PCA from Split Annual Ordination PCA Flowchart.*

data (Figure 1). Similarly, the temporal changes in the near-IR imagery were isolated with a multitemporal PCA of the four annual PC 2 images.

Technique 2: Joint Annual PCA Ordination:

Instead of examining PC 1 (representing primarily the visible spectral region) and PC 2 (representing the near-IR) separately, for Technique 2 we analyzed the first and second components from each year together (8 input bands in total) in a multitemporal PCA (Figure 2).

Technique 3: NDVI Ordination:

The spectral reflectance from typical terrestrial landscapes is dominated by the characteristics of the vegetated land cover. The normalized difference vegetation index (NDVI) is a mathematical combination of the visible red and near-IR bands that has been found to be a sensitive indicator of the presence and condition of green vegetation (Townshend, 1994) and is potentially an effective ordination technique. For multitemporal analysis, the NDVI also has the advantage of helping

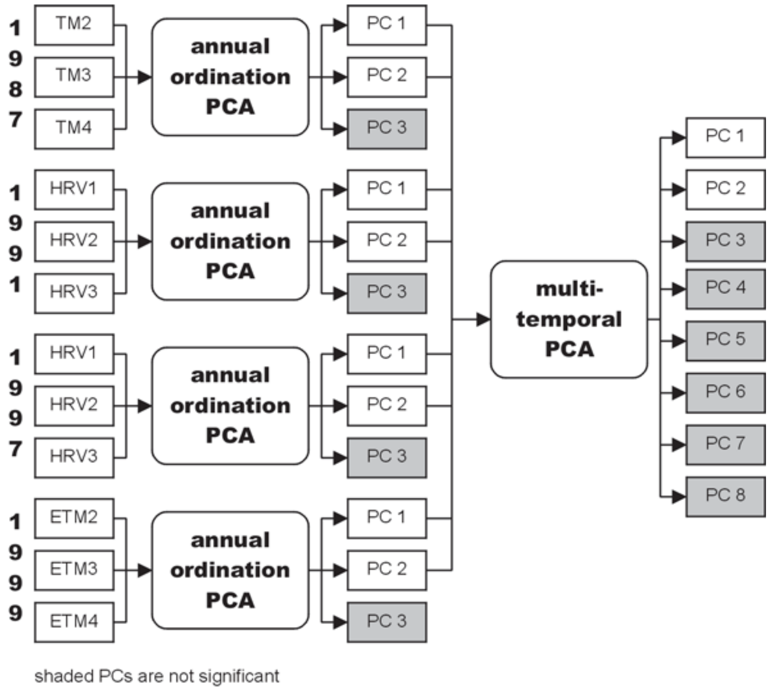

Figure 2: *Multitemporal PCA from Joint Annual Ordination PCA Fowchart.*

to compensate for extraneous factors such as differences in scene illumination and atmospheric conditions. In Technique 3, we used the four NDVI images (one computed from each year) in a second principal components analysis to highlight multitemporal changes (Figure 3).

Results

The results presented below are interpreted through an examination of the principal component images and their associated loadings plots. Whereas the component images identify the *spatial* arrangement of the change patterns, the loadings plots indicate their *temporal* domains. That is, the images show us *where* strong spatial patterns were occurring while the plots report *when* these patterns were strongest.

Multitemporal PCA from Split Annual Ordination:

Standardized principal components were calculated from the individual years' PC 1 data.[1] The first two components created from the

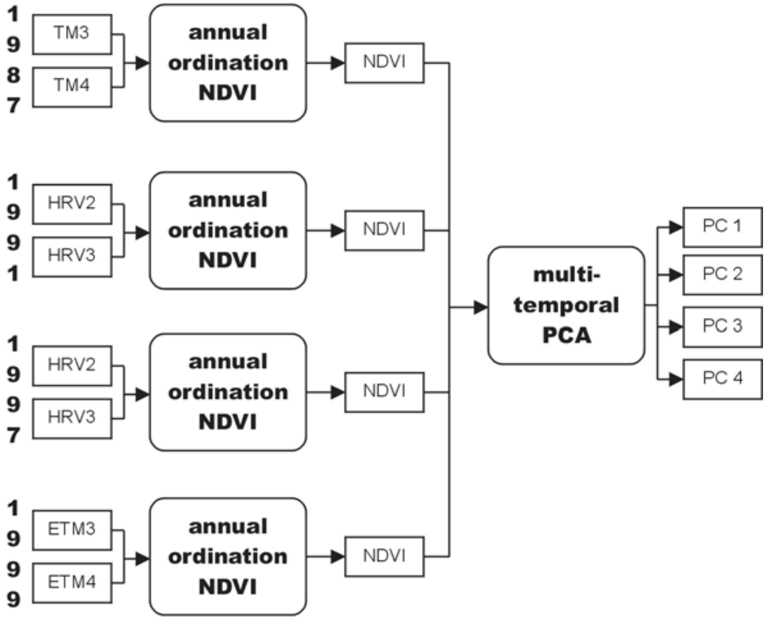

Figure 3: *Multitemporal PCA from NDVI Ordination Flowchart.*

PCA of the multitemporal analysis are shown in Figure 4 as PC 1.1 and PC 1.2. The loadings plot for PC 1.1 shows strong positive correlation between this component and the first components from all years except 1991. Thus, this component can be considered as the integrated average of the source data and any real change information is lost in the transformation. Since these data are drawn primarily from the visible bands of the original imagery, the tonal ranges in the PC 1.1 image follow the typical patterns evident in visible imagery: urbanized areas have strong reflectance; vegetated regions exhibit moderate reflectance; and generally low reflectance from water bodies. There is little detail in any of these areas, however.

From the loadings for PC 1.2 we see that this component is strongly correlated only with 1991. However, the correlation is *negative*, suggesting that this component shows typical image detail that was *not* evident in 1991. This is borne out in a comparison of the PC 1.2 and 1991 images (not included here) that show the extensive road network developed in the latter half of the 1990s and clearly shown in the PC 1.2 image was absent in 1991.

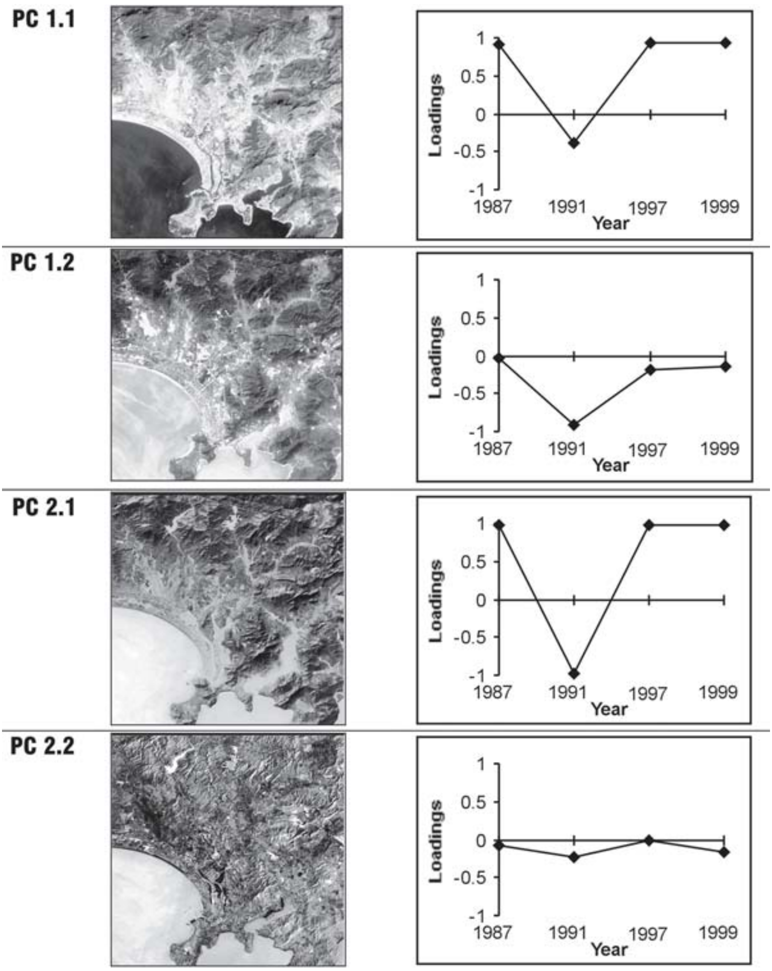

Figure 4: *Multitemporal Principal Components from Split Annual Ordination PCA.*

Figure 4 also shows the first two components calculated from the individual years' PC 2 data, labelled as PC 2.1 and PC 2.2. The associated loadings plots show that PC 2.1 is strongly positively correlated with 1987, 1997 and 1999, and strongly negatively correlated with 1991. The PC 2.1 image highlights the land-water dichotomy evident in near-IR imagery. PC 2.2 is not significantly correlated at any year. The PC 2.2 image highlights localized differences in land development between the first two and last two dates.

The strength of the apparent anomaly in 1991 remains a mystery and we are continuing our attempts to arrive at a suitable explanation.

Multitemporal PCA from Joint Annual Ordination PCA:

Instead of splitting the individual years' PC 1 and PC 2 results into separate multitemporal analyses (as was done in the previous section) we repeated the multitemporal PCA including both the individual years' PC 1 and PC 2 data jointly. The first four components thus derived are shown in Figure 5. The first component loads strongly positive with the near-IR data (PC 2s) from the individual years, with 1997 as the lone anomaly, and moderately negative with the visible imagery (PC 1s). The PC 1 image is thus an integrated average of the differences between the visible and near-IR data and consequently shows poor contrast and detail. The second component loads strongly with the visible band (PC 1) from 1997 and produces a striking contrast between vegetated uplands and non-vegetated valleys and coastal areas. The third and fourth components are not strongly correlated with any year and progressively show more localized changes. For example, PC 3 appears to highlight major land development changes occurring between 1991 and 1997, while variations in scene illumination dominate PC 4.

Multitemporal PCA from NDVI Ordination:

Recall that the first component derived during PCA is typically an integrated average of the input data. Although this trait has been shown during some of the analyses above, the results have not been consistent or conclusive. The first component derived from a multitemporal PCA of the NDVI calculated for each of the four data years, however, has strong loadings across all years (Figure 6). This is the hallmark of an integrated average and is exemplified in the PC 1 image. This scene shows, in detail, the average vegetation characteristics in the temporal sequence. Little detail is allocated to areas of consistently low NDVI for each year.

The loadings from the second component clarify the interpretation of an otherwise obscure pattern in the PC 2 image. The plotted loadings identify this component as the declining trend in NDVI across the four dates. The brighter areas in the PC 2 image are associated with higher loadings, hence more vegetative greenness in 1987 than in 1999. Conversely, the darker regions are areas where there has been a vegetative increase between these years.

The loadings for PC 3 show a slight increase from 1997 to 1999. This is reflected in the PC 3 image with brighter tones in areas of higher NDVI in 1999 and darker tones where there was more green vegetation in 1997.

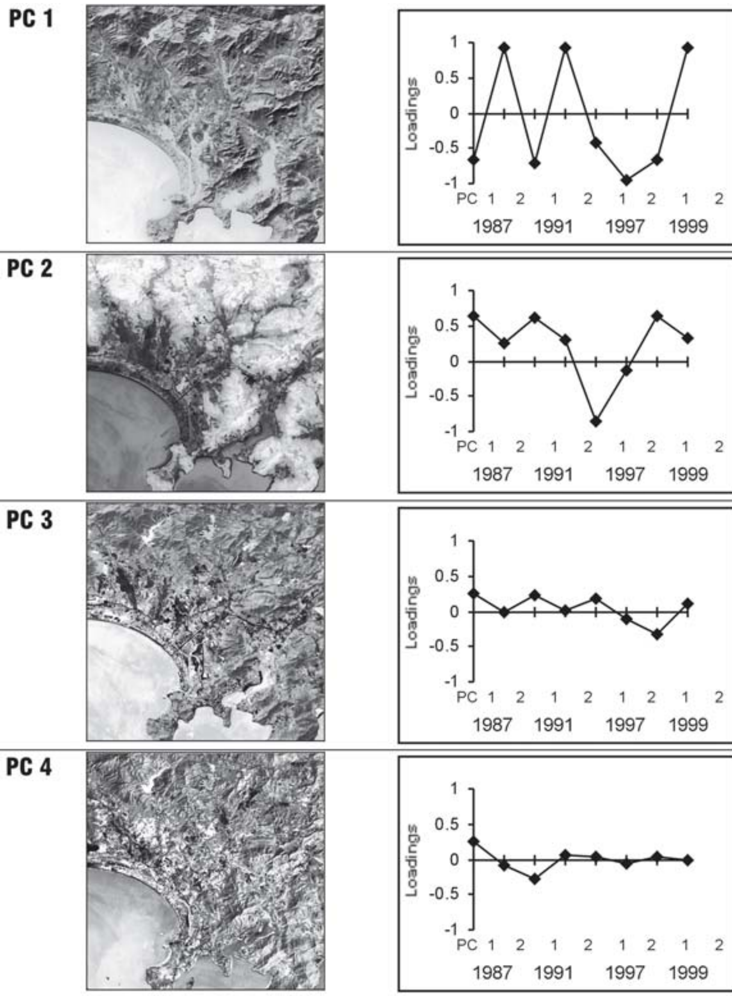

Figure 5: *Multitemporal Principal Components from Joint Annual Ordination PCA.*

Discussion

Principal components do not tell us about the possible mechanisms that are creating the observed patterns; they simply describe the major spatial relationships in the data. Thus, they do not tell us what the relationships mean or what is causing them. To take the analysis to the

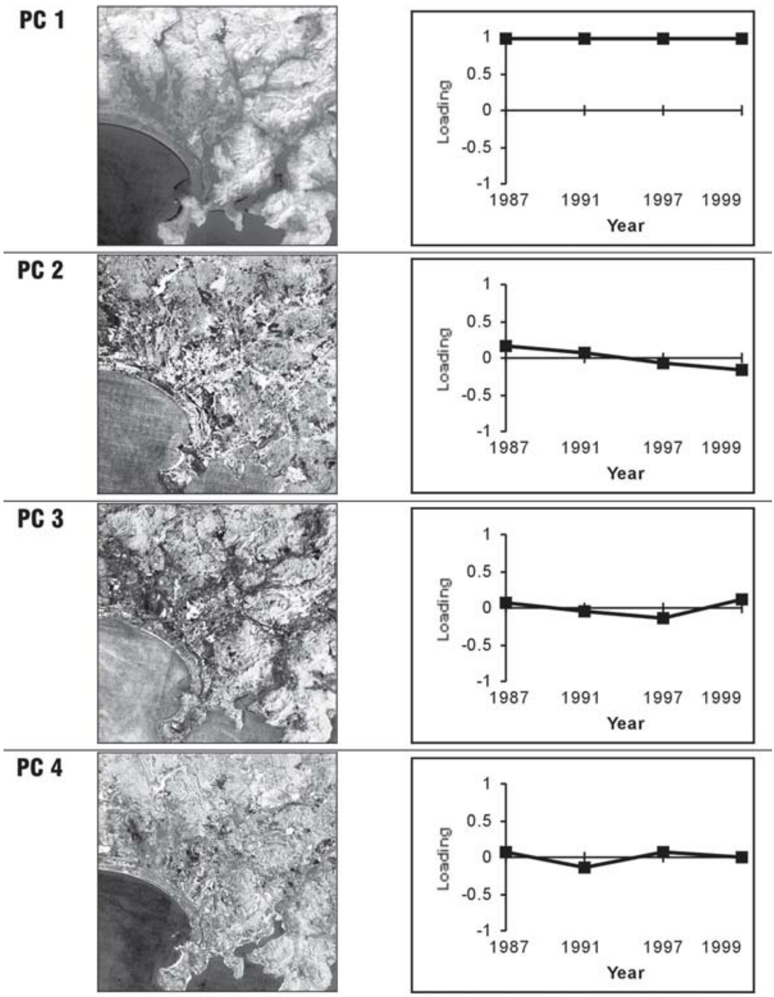

Figure 6: *Multitemporal Principal Components from Annual NDVI Ordination.*

next stage we must try to associate the observed patterns with other factors to interpret their meaning.

In the first technique discussed above, multitemporal components were derived separately from the first and second components of annual ordination PCAs. In both cases (shown in Figure 4), the first multitemporal component was identified as an integrated average while PC 2 isolated changes between 1987-1991 and 1997-1999. The third and fourth components (not shown in this report) isolated more localized changes.

Although this method was effective, interpretation of the components was hampered by the need to examine two different result sets and by apparently conflicting component loadings and images (e.g., PC 1.2).

The first component derived from the multitemporal PCA of joint annual ordination PCA (PC 1 in Figure 5) is a clear case of the potential problems arising from mixed spectral-temporal analysis that Eastman and Fulk (1993) warned about. This component loads highly positive with most of the annual PC 2s and moderately negative with all of the annual PC 1s. The accompanying component image, however, is representative of neither extreme: it is an average of the two, conveying little useful information. Similar difficulties arose when attempting to interpret the later components.

The first two techniques are plagued with a lack of a strong physical basis: we cannot say what the first component from the annual ordination actually represents. In a typical change analysis exercise we want to be able to relate our analysis findings with ground-based observations. For example, we know that the first component is supposed to be an integrated average, but what does an integrated average look like if you were standing on the ground? The multitemporal PCA from NDVI does have a strong physical basis, on the other hand. We know, for example, that higher NDVI values in the imagery are directly related to increased green vegetation vigour on the ground. In the present study, we saw an overall decrease in NDVI through the twelve year study period, presumably as a result of increased urbanization in the region. Thus, the interpretation of the multitemporal components of annual NDVI images was clear and concise.

Conclusions

In this paper we have examined three ordination techniques for the reduction of the dimensionality of multispectral remote sensing imagery prior to their inclusion in multitemporal principal components analysis. In the first two approaches, PCA was applied to the spectral bands from each date individually. This was based on the principle that, by definition, PCA has the potential to be an excellent ordination technique. We found that the subsequent components created from the multitemporal PCA were difficult to interpret, however, because the annual ordination components did not have a strong physical basis. Using the NDVI transformation on the individual images, on the other hand, produced consistent and easily interpreted results because the NDVI is not an abstract value. Further, the NDVI accounts for interscene illumination and atmospheric differences that are frequent obstacles in inter-scene comparisons. Thus we find that performing a PCA with the NDVI temporal bands is simpler to use and produces more robust results than the annual ordination PCA.

Acknowledgements

Thank-you to Phil Howarth for his support and encouragement during the developmental stages of this project. Financial support for the satellite imagery was generously provided from a CIDA Canada-China Higher Education grant awarded to the University of Waterloo.

References

BYRNE, G.F., P.F. CRAPPER, and K.K. MAYO 1980 'Monitoring land-cover change by principal component analysis of multitemporal landsat imagery' *Remote Sensing of Environment* 10, 175-184

EASTMAN, J.R. and M. FULK 1993 'Long squence time series evaluation using standardized principal components' *Photogrammetric Engineering & Remote Sensing* 59, 991-996

FRANKLIN, S.E., L.M. MOSKAL, M.B. LAVIGNE and K. PUGH 2000 'Interpretation and Classification of Partially Harvested Forest Stands in the Fundy Model Forest Using Multitemporal Landsat TM Digital Data' *Can. J. Remote Sensing* 26, 318-333

FUNG, T. and E. LEDREW 1987 'Application of principal components analysis to change detection' *Photogrammetric Engineering & Remote Sensing* 53, 1649-1658

JENSEN, J.R. 1996 *Introductory Digital Image Processing: A Remote Sensing Perspective* 2nd ed. (Upper Saddle River, NJ: Prentice-Hall)

LEDREW, E. 1992 'The Role of Remote Sensing in the Study of Atmosphere-Cryosphere Interactions in the Polar Basin' *Can. Geographer* 36, 336-350.

LI, X. and A.G. YEH 2002 'Urban Simulation Using Principal Components Analysis and Cellular Automata for Land-Use Planning' *Photogrammetric Engineering & Remote Sensing* 68, 341-351

MCGARIGAL, K., S. CUSHMAN and S. STAFFORD 2000 *Multivariate Statistics for Wildlife and Ecology Research* (New York: Springer-Verlag)

PIWOWAR, J. M. and E. F. LEDREW 1995 'Climate Change and Arctic Sea Ice: Some Observations from Hypertemporal Image Analysis' *Proceedings, Third Thematic Conference on Remote Sensing for Marine and Coastal Environments*, Seattle, 18-20 September 1995, 625-636

PIWOWAR, J. M. and E. F. LEDREW 1996 'Principal components analysis of arctic ice conditions between 1978 and 1987 as observed from the SMMR data record' *Canadian J. Remote Sensing* 22(4), 390-403

RUNDQUIST, D.C. and L. DI 1989 'Band-moment analysis of imaging-spectrometer data' *Photogrammetric Engineering & Remote Sensing* 55(2), 203-208

SINGH, A. and A. HARRISON 1985 'Standardized principal components' *Int. J. Remote Sensing* 6(6), 883-896

TANGESTANI, M.H. and F. MOORE 2001 'Comparison of Three Principal Component Analysis Techniques to Porphyry Copper Alteration Mapping: A Case Study, Meiduk Area, Kerman, Iran' *Can. J. Remote Sensing* 27(2), 176-182

TOWNSHEND, J.R.G. 1994 'Global data sets for land applications from the AVHRR: an introduction' *Int. J. Remote Sensing* 15, 3319-3332

YOUNG, S.S. and A. ANYAMBA 1999 'Comparison of NOAA/NASA PAL and NOAA GVI Data for Vegetation Change Studies over China' *Photogrammetric Engineering & Remote Sensing* 65(6) 679-688

[1] All image processing was completed using ENVI 3.4 software.