Exploring ordination as a method for normalizing disparate datasets: implications for digital change detection

Andrew A. Millward* and Joseph M. Piwowar
Waterloo Laboratory of Earth Observations
Department of Geography, University of Waterloo
Waterloo, Ontario N2L 3G1
*Corresponding author: email - aamillwa@fes.uwaterloo.ca

Abstract

Researchers interested in digital change detection are well aware that the present plethora of earth observation satellites brings with it a mixed blessing. The evolution of satellite technology has evolved rapidly with regard to changes in the spatial and spectral resolution of acquired data. This has revealed numerous research possibilities, with associated unique challenges, that were previously impossible. Potential implications of this are immediately apparent when research questions require an effective transition from data to knowledge. Specifically, this research employs principal components analysis (PCA), a multivariate ordination technique, to perform multi-date change analysis. The application of PCA offers a method for standardization of disparate image types, a necessary step in any digital change-detection procedure.

Change detection is now a common procedure in studies employing remotely sensed data to study differences between two dates. However, interest is gaining in its potential use as an analytical tool for identifying and monitoring environmental change in disparate data sets, acquired at multiple time intervals. The reasons for this are twofold: scientists and concerned members of the public are increasingly interested in understanding anthropogenic influences on the environment; and the archive of satellite data continues to grow permitting longer and more sophisticated analyses of environmental change.

This research uses satellite imagery acquired of a rapidly urbanizing coastal city in Southern China. Four images obtained from three different satellite sensors (Landsat 5 TM, SPOT HVR, Landsat 7 EMS+) form a twelve-year chronology. PCA is used to help elucidate coastal urbanization by offering one method of normalizing differences in the spatial and spectral resolution of these data. This paper contributes to the development of methods for normalizing disparate datasets, an essential requirement for conducting meaningful investigations of environmental change.

Keywords: change detection; principal components analysis; urbanization

Introduction

When employing remotely sensed imagery, analysis of change is concerned with two basic types of data: qualitative and quantitative. Qualitative data represent differences in kind (types of land cover) while quantitative data represent differences to the spatial configuration of change (location and magnitude)(Eastman and McKendry, 1991). It is perhaps valid to state that investment on the part of researchers in digital change detection techniques has not been adequate, given the potential importance of these methods to the study of environmental change. The goal of this research is to determine whether PCA is a useful and robust technique for combining data of disparate spectral and temporal resolutions for the purpose of detecting coastal land-use and land-cover change.

Our ability to map, monitor, and infer land-cover and environmental change, until recently, has been hampered by a lack of synoptic historical data. Remotely sensed imagery has been used to perform change detection and time-series analysis of land cover features in regions experiencing rapid anthropogenic change (Howarth and Boasson, 1983; Jensen, 1996; Ridd and Liu, 1998; Mas, 1999). In most cases, these data represent the longest chronology of quantitative data available to analyze change. Understanding the nature of qualitative change to land features, as well as the quantitative extent and pattern of change, is essential to the investigation of the consequences of anthropogenic alternations to the environment.

In many circumstances, the identification of true change from natural variation is a difficult task (Fung and LeDrew 1988; Eastman and McKendry, 1991; Ridd and Liu, 1998). This research introduces yet another challenge to standard digital change detection procedures, the element of integrating data of disparate spectral and spatial resolution. Very strong spatial registration is needed between, or among, images in order for corresponding pixels to match up in the change analysis procedure (Eastman and McKendry, 1991; Jensen, 1996). In the situation described herein, data from three unique satellite platforms are considered (Table 1). These data differ in their spatial and spectral resolution; however, the spatial resolutions are similar enough to consider each data type to be of medium resolution. To normalize data that are characterized by variance in spatial resolution, researchers (Jensen, 1996; Lillesand and Kiefer, 2000) have employed techniques that are collectively referred to as geometric correction. Thus, while it is clear that all change detection techniques require highly accurate geometric registration, a systematic technique for normalizing spectral and temporal resolution remains outstanding.

One significant problem when comparing multitemporal data is that chronological time does not necessarily match development time for the phenomenon being measured (Batista *et al*., 1997; Lobo *et al*., 1997; Mas, 1999; Brown *et al*., 2000). If for example, the rainy season comes a month late one year, a whole sequence of months will then be out of synchronization with respect the growing cycle of a normal year. This may then appear in the land-cover difference measurement as a substantial change. However, this is not a change *per se*, but a mismatch in the growth cycle between the two years. Similarly, sensors that have marked differences in spectral resolutions (i.e., SPOT and Landsat) represent challenges to traditional change detection procedures. This is of particular concern when employing digital change detection procedures such as image differencing, change vector analysis, and band-three-overlay. See Jensen (1996) for a review of these standard change detection techniques.

There is evidence in the literature however, that PCA may be able to identify seasonal and phenological variations in land-cover features (Eastman and McKendry, 1991; Piwowar and LeDrew, 1996; Anyamba, 2000). Further, it is well documented that PCA is effective at removing redundancy in spectral data, which is reported to be common in multispectral data sets (Jensen, 1996; Lillisand and Kiefer, 2000).

Principal Components Analysis (PCA)

Ordination techniques have as their main goals: 1) to reduce the number of variables within a database and 2) to attempt to elucidate the structure in the relationship among variables (i.e., to classify/describe them more easily). Thus, ordination can be thought of as a data reduction or structure detection method. Essentially with PCA we are attempting to express many factors (channels of data) by fewer factors in an effort to increase interpretability. Extracting PCs is essentially a variance maximizing rotation of the original variable space.

Principal Components Analysis (PCA) is an ordination technique that applies a linear transformation to the original data. It has been shown to be of notable value in the analysis of remotely sensed imagery (Jensen, 1996; Lillesand and Kiefer, 2000). Traditionally its application has been applied for image enhancement and channel reduction (e.g., Taylor, 1974; Singh and Harrison, 1985; Rundquist and Di, 1989), however, it has also been effectively used in terrestrial change detection studies (e.g., Lodwick, 1979; Byrne *et al.*, 1980; Eastman and Fulk, 1993; Piwowar and LeDrew, 1996).

PCA is a procedure for transforming a set of correlated variables into a set of uncorrelated ones. The transformation is a simple rotation of the axis in the original data to new ones that are orthogonal to each other (Figure 1). These components are determined by collapsing the observed variances for n-spectral bands into principal components (PCs) so as maximize the variance in smaller (lower-order components). The PCs that evolve from a transformation are uncorrelated. The net effect of employing this procedure is to remove redundancy that is often present within spectral bands of digital imagery.

In multitemporal studies, the technique offers a useful exploratory tool to analyze the interrelationships between regularly sampled multitemporal remote sensing data sets (Townshend *et al.*, 1985). For example, when images of different dates are compared with a view to monitoring change, there will be a high correlation between them: those parts of the scene which show an absence of correlation are of interest as they represent areas for the investigation of change (Byrne and Crapper, 1980).

The first principal component (PC1) accounts for the most variance in the original data; each successive component accounts for a smaller amount of the variance than its predecessor (Figure 1). Higher components (i.e., those beyond the first one or two) thus define more localized anomalies and are representative of less "information" from the source data and

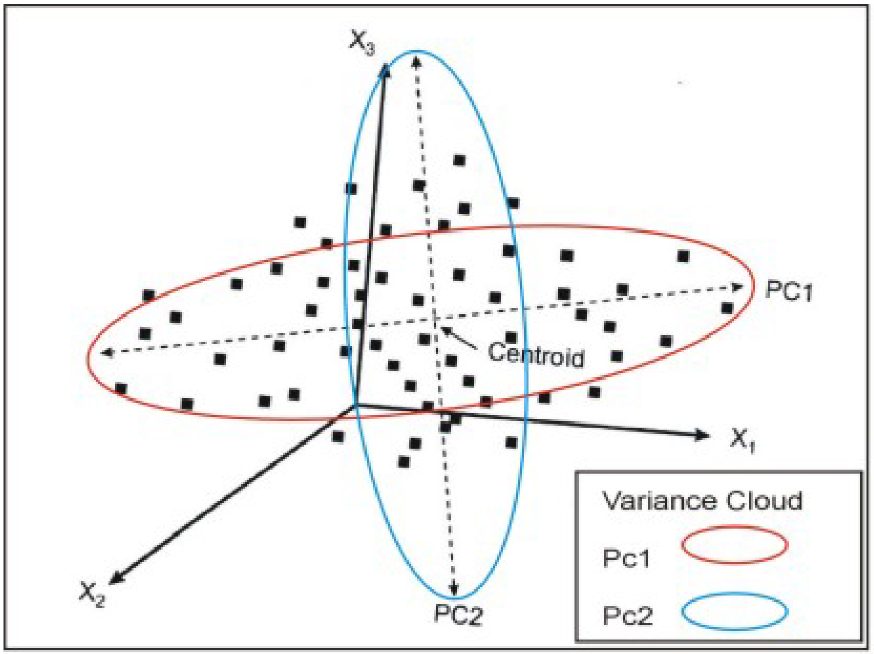

Figure 1: Simplified Example of Centroid Transformation

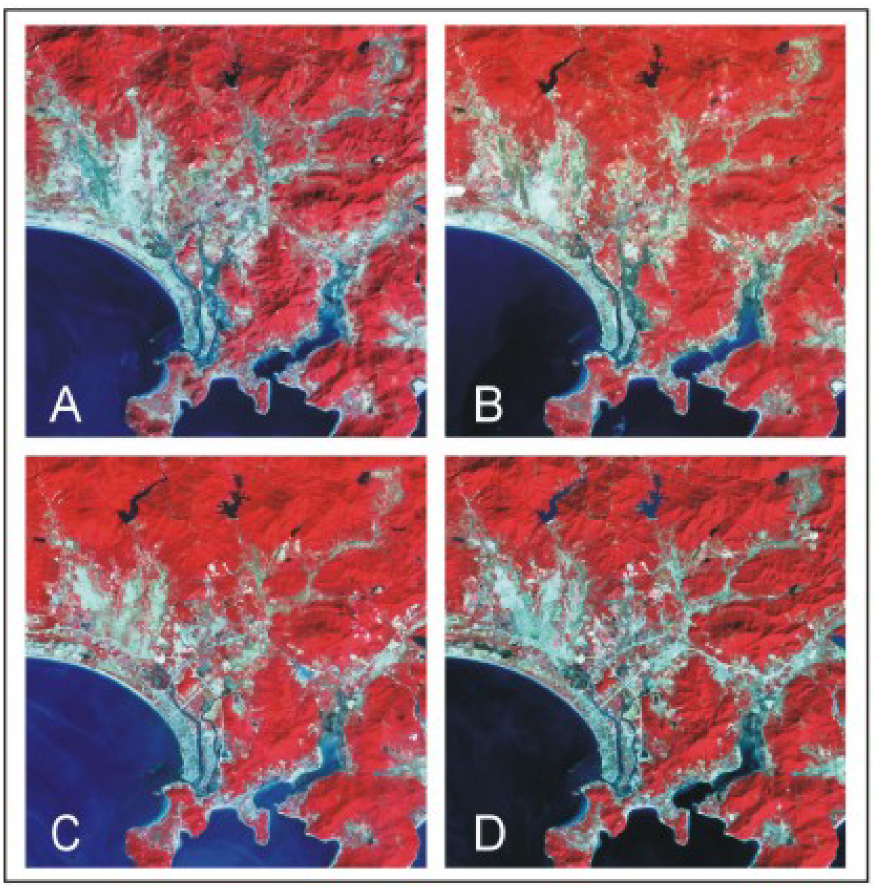

Figure 2: False Colour Images: a) 1987 b) 1991 c) 1997 d) 1999

more "noise". The decision regarding the number of components to include in subsequent analyses can be quite subjective, but is generally judged by performance rather than theoretical considerations (Davis, 1973).

The relative weightings of the original bands in each component produced are called *loadings*. The component loadings indicate the correlation between each component and the original channels on a scale of -1 to +1. A high positive loading means the original data are similar to the component, a negative loading indicates an inverse spatial pattern between the original data and the component. A loading near zero means there is little similarity. When plotted against time, loadings can be interpreted as an indicator of the temporal persistence of that given spatial mode. Extremes in the component loadings (both positive and negative) can be used to identify the temporal intervals when the spatial pattern (or it's inverse) characterized by a given component is present in the time series data.

Principal components are influenced by the variances found in the original data, with those variables having a higher variance imposing a greater influence on the generated components. Thus, for typical optical sensor data, since the near IR bands tend to exhibit greater variances than the visible channels, the first components will represent more of the near IR data and with lesser contributions from the visible bands. This would be the result if the components were generated from the variance-covariance matrix; this is known as the *unstandardized* approach. It is possible, however, to have the contributions from each of the input bands weighted evenly, regardless of their relative variance differences. In this case, known as the *standardized* approach, the components are created from the correlation matrix. There has been some discussion (e.g., Singh and Harrison, 1985; Fung and LeDrew, 1987) whether the standardized or unstandardized components should be used. Most authors agree that standardized components (where data from each input channel is weighted equally) produce the most accurate results (Eastman and Fulk, 1993). However, for multitemporal studies, there is an argument made (Roberts et al., 1994) for the use of unstandardized PCA, since this technique may more accurately reveal changes to vegetation response across a time series of data.

A PCA's success, when applied to RS data, should be critiqued based upon its ability to explain the multivariate phenomena under scrutiny. The caution in employing standardized PCA in this circumstance is that, the output PCs can become disconnected from real-world reflectance changes (e.g., between time period vegetation response), since each input band received an equal weighting.

Data

An historical analysis of environmental change requires the synthesis of disparate sources of data at varying spatial and temporal scales. Due to its unbiased nature, and repeat coverage, remotely sensed data represents an excellent means to quantify land cover and ecological change. To date, four images of Southern China have been collected, and are described in Table 1 and illustrated in Figure 2. To assign ground coordinates (i.e., UTM grid projection) to the satellite data, 1976 Russian military topographic maps were acquired for the study region at a spatial scale compatible with the remotely sensed imagery.

Table 1: Optical Satellite Imagery Analyzed

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

Methods

Initial data preparation centered on the selection of one image that would be registered to the Russian topographic maps. The 1991 SPOT image was selected for several reasons that included: the oldest image with the highest spatial resolution, and of the two SPOT images this one had a sensor look-angle that most closely approximated nadir. This image, referred to hereafter as the *base image*, was geometrically corrected using a second order polynomial shift, and bilinear spectral resampling. See Jensen (1996) for a review of geometric correction procedures.

Image to image registration was then performed to match the 1997 SPOT image to the geometrically corrected base image. Each of the Landsat images obtained for 1987 and 1999 were then tied to the base image individually using similar geometric correction procedures, with the additional step of resampling the 30-metre spatial resolution acquired by the Landsat sensor to a 20-metre resolution. It should be noted that this procedure did not increase the spatial accuracy of the Landsat data, only change its pixel dimensions. This step was considered necessary as it standardized the spatial resolution of the pixels prior to conducting a PCA.

In our research, we used PCA on each image (Table 1) independently to reduce spectral redundancy. All PCs were saved to 32-bit image channels so as to preserve numeric output in real format. This process revealed three PCs for each of the four temporally distinct images. When the factor loadings for each image were analyzed, they revealed a trend toward separation of near infrared (NIR) variance, which loaded highly in all PC1s (Figure 3). Similarly, the majority of variance contained in the visible bands loaded heavily in all PC2s (Figure 3). All PC1s were coded with their associated image year and saved to a single file. A similar file containing PC2s was also generated.

Results

To reveal change across our four time periods (1987, 1991, 1997, 1999) we employed the visual change detection 'band overlay' procedure with our two new images composed of all the PC1s and all the PC2s respectively (Figure 4). Since the image containing PC1s was dominated by variance values originating from the NIR portion of the spectrum, we made the assumption that changes revealed between components from change analysis would be due to variations in vegetation presence and vigor. Conversely, the image containing PC2s was dominated by variance originating in the visible portion the electromagnetic spectrum, revealed changes as a result of coastal development including roads and buildings. Major changes observed across the twelve-year chronology are summarized in Table 2.

Table 2: Major changes to the southern coast of Hainan

Time Period	Observed Change
1987 to 1991	• The dam to the upper right of the image is constructed and flooded, while the dam present in the centre is filled with greater amounts of water exemplified by the red area present to its north (NIR image) • A small piece of coastline within the city appears to have become vegetated during this period as is evidenced by a bright blue patch to the right of the urban core (NIR image) • A high blue response in and around the urban core indicates the presence of cleared land for construction sites (VISIBLE image) • A major highway is completed in the center-north portion of the urban core. It is evident as a solid bright blue line running north-south (VISIBLE image)
1991 to 1997	• Vegetation appears to have degraded appreciably in and around the urban core evidenced by the strong red shades found in the NIR dominated image • Locations around the river network appear to have had an increase in vegetative cover, a likely response to draining wetlands and shallow ponds in this area (NIR image) • Construction of new roads to the periphery of the urban core is clearly evident by the bright blue lines (VISIBLE image) • Land clearance and construction continue to dominate change in the urban core (VISIBLE image)
1997 to 1999	• A small regeneration in vegetation is evident in the urban core as the red dominated city (1987-1997) is finally beginning to show a prominence of blue, suggesting the presence of newly planted trees and the maturation of established vegetation (NIR image) • Vegetation clearance appears to have occurred along a portion of the coastline to the northwest of the city core as well as in a patch along one of the river branches which is known to be newly created saltation ponds (NIR image) • The majority of construction appears to have concluded in the urban core, but the creation of a new highway linking the east of the image to the urban core is clearly visible. Additionally, as small pocket of development appears to be emerging to the east of the urban core (VISIBLE image)

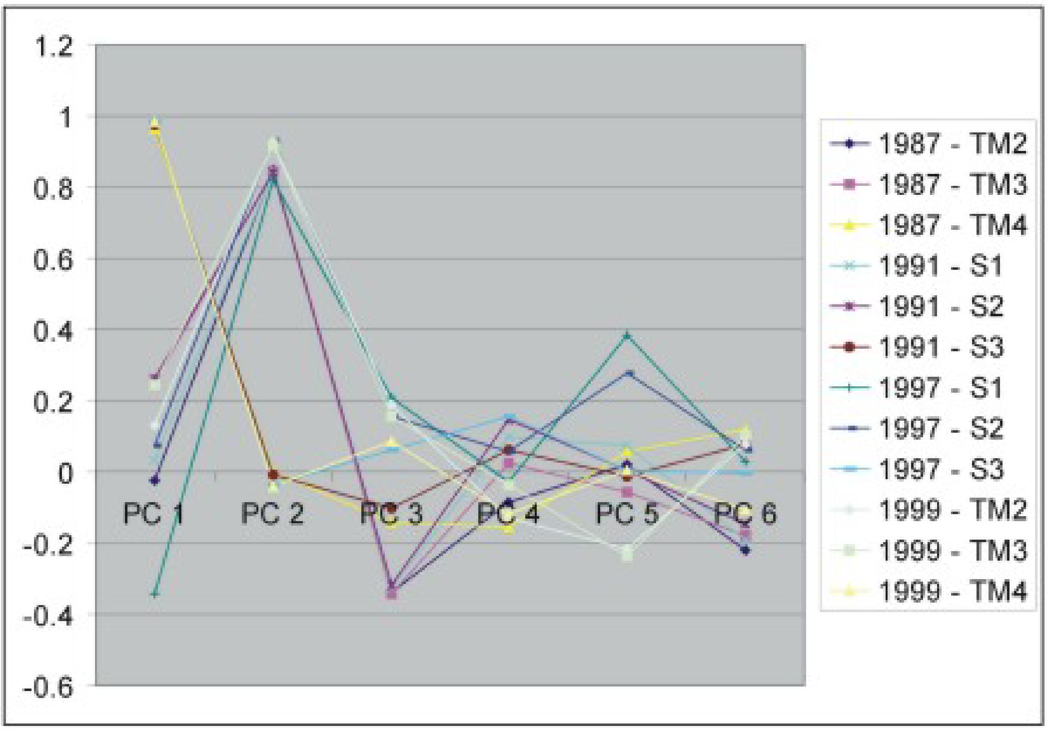

Figure 3: Multitemporal Principal Component Factor Loadings

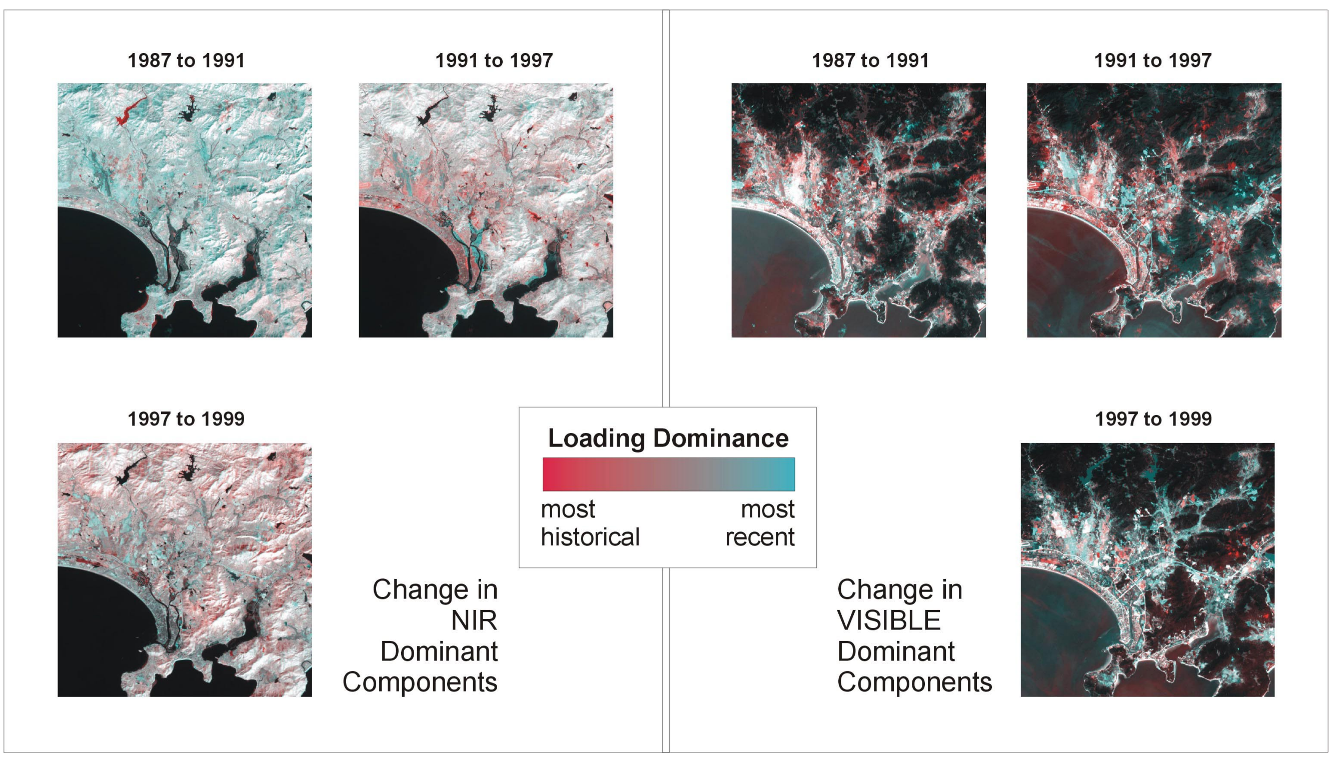

Figure 4: Principal Component Change Images

Discussion

Environmental disturbance and ultimate change is an important process, often influenced heavily by a landscape's existing spatial heterogeneity. Differential exposure to disturbance, coupled with historical influences, contributes to the spatial mosaic of observable land features (Turner, 1989). Estimating the cumulative impacts of disturbances is critical to the conservation of sensitive habitats and to environmental quality. An essential feature of historical investigations of environmental change is a time sequence of measurements, or data containing meaningful information that can be extracted to reveal pattern and process (Swetnam *et al*., 1999) . However, much of our historical digital data was obtained and different spatial, spectral and temporal resolutions than data being collected at present. Obviously then, methods are required to normalize disparate data such as those acquired from SPOT and Landsat satellites.

It is evident that PCA has the potential to be a powerful advocate for the monitoring of change by transforming spectrally and spatially dissimilar imagery into a form that can be compared in a meaningful and efficient fashion. Although in its preliminary stages, this research lays the foundation for planners and cartographers alike, to enhance their ability to produce geographic products that more accurately capture environmental change.

Conclusions

This research has revealed that PCA can make valuable contribution to normalizing disparate data sets so that more meaningful comparisons of historical and contemporary data can be made. This represents a needed theoretical contribution not only to remote sensing and data integration,
but more broadly, it opens the door to a more procedural approach to producing urgently needed maps of environmental change.

Principal component analysis (PCA) has had a long history of use in remote sensing. However, what constitutes important information in the original data, let alone the evolved components, is dependent upon the researcher, and upon the ultimate object of the analysis. In this research we were interested in elucidating and time stamping changes to the coastal environment in the form of urbanization. We feel that the methods employed herein were successful at highlighting significant changes in this capacity. Nevertheless, this approach to change detection is meant to be one tool that could be used in conjunction with others such as a temporal change analysis of the normalized difference in vegetation index (NDVI).

Acknowledgements

We would like to thank Dr. Philip Howarth of the University of Waterloo for his support and encouragement during the developmental stages of this project. Financial support need to purchase the data used in this study was generously provided as part of a Canadian International Development Agency (CIDA) Canada-China Higher Education grant awarded to the University of Waterloo.

References

Anyamba A (2000) Time series change analysis using standardized principal components analysis. University of South Carolina, Department of Geography, Remote Sensing Laboratory, Columbia. pp 13

Batista GT, Shimabukuro YE and Lawrence WT (1997) The long-term monitoring of vegetation cover in the Amazonian region of northern Brazil using NOAA-AVHRR data. *International Journal of Remote Sensing* 18(15): 3195-3210

Brown DG, Duh J-D and Drzyzga SA (2000) Estimating error in an analysis of forest fragmentation change using North American Landscape Characterization (NALC) data. *Remote Sensing of Environment* 71(1): 106-117

Byrne GF and Crapper PF (1980) Land cover change detection by principal components analysis of multitemporal MSS data: the presence of clouds. *Proceedings, 14th International Symposium on Remote Sensing of Environment*, San Jose Costa Rica, 1175-1382

Byrne GF, Crapper PF and Mayo KK (1980) Monitoring land-cover change by principal component analysis of multitemporal Landsat imagery. *Remote Sensing of Environment* 10: 175-184

Davis JC (1973) *Statistics and Data Analysis in Geology*. New York, John Wiley & Sons

Eastman JR and Fulk M (1993) Long sequence time series evaluation using standardized principal components. *Photogrammetric Engineering and Remote Sensing* 59(8): 1307-1312

Eastman JR and McKendry JE (1991) *Change and Time Series Analysis*, Volume 1 of Explorations in Geographic Information Systems Technology, Geneva Switzerland: United Nations Institute for Training and Research, 85 p

Fung T and LeDrew E (1987) Application of principal components analysis to change detection. *Photogrammetric Engineering and Remote Sensing* 53(12): 1649-1658

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

Howarth PJ and Boasson E (1983) Landsat digital enhancements for change detection in urban environments. *Remote Sensing of Environment* 13: 149-160

Lillesand TM and Kiefer RW (2000) *Remote Sensing and Image Interpretation* (4th ed.)., New York, John Wiley & Sons

Lobo A, Ibanez Marti JJ and Gimenez-Cassina, C (1997): Regional scale hierarchical classification of temporal series of AVHRR vegetation index. *International Journal of Remote Sensing* 18(5): 3167-3193

Lodwick GD (1979) Measuring Ecological Changes in Multitemporal Landsat Data Using Principal Components, *Proceedings, 13th International Symposium on Remote Sensing of Environment*, Ann Arbor MI, pp. 1131-1141

Mas JF (1999) Monitoring land-cover changes: a comparison of change detection techniques. *International Journal of Remote Sensing* 20(1): 139-152

Piwowar JM and LeDrew, EF (1996): Principal components analysis of arctic ice conditions between 1978 and 1987 as observed from the SMMR data record. *Canadian Journal of Remote Sensing* 22(4): 390-403

Ridd MK and Liu J (1998) A comparison of four algorithms for change detection in an urban environment - a remote sensing perspective. *Remote Sensing of Environment* 63(2): 95-100

Roberts M, C Wells and TW Doyle (1994) Component analysis for interpretation of time series NDVI Imagery. *American Society for Photogrammetry and Remote Sensing and the American Congress on Surveying and Mapping Annual Convention*, Reno Nevada, pp. 538-550

Rundquist DC and Di L (1989) Band-Moment Analysis of Imaging-Spectrometer Data. *Photogrammetric Engineering & Remote Sensing* 55(2): 203-208

Singh A and Harrison A (1985) Standardized principal components. *International Journal of Remote Sensing* 6(6): 883-896

Swetnam TW, Allen CD and Betancourt JL (1999) Applied historical ecology: using the past to manage the future. *Ecological Applications* 9(4): 1189-1206

Taylor MM (1974) Principal components colour display of ERTS imagery. *Proceedings of the 2nd Canadian Symposium on Remote Sensing*, Guelph Ontario, pp. 295-313

Townshend JRG, Goff TE and Tucker CJ (1985) Multitemporal dimensionality of Images of normalized difference vegetation index at continental scales. *IEEE Transactions on Geoscience and Remote Sensing* GE-23(6): 888-895

Turner GT (1989) Landscape ecology: the effect of pattern on process. *Annual Review of Ecology and Systematics* 20(2): 171-197