Long Sequence Time Series Evaluation Using Standardized Principal Components

Abstract

The potential of using Standardized Principal Components for the analysis of long time series of spatial environmental data is assessed using a series of 36 monthly AVHRR-derived NDVI images for Africa for the years 1986-88 as an illustration. The first component is found to represent the characteristic NDVI regardless of the season. The second, third, and fourth components relate to seasonal changes in NDVI. The fifth and sixth components uncover a sensor-related drift in the NDVI values due to successively later equatorial crossings of the NOAA-9 satellite. The seventh and eighth components illustrate NDVI anomalies related to significant El Niño/ Southern Oscillation (ENSO) events, primarily in southern Africa. The technique is shown to be a comprehensive indicator of change events in time series data that is sensitive to periodic and aperiodic events alike.

Introduction

Principal Components Analysis (PCA) using Unstandardized Components has long been used in remote sensing as a data compression tool. The first two components of Standardized PCA have also been used for land-cover classification, but with mixed results (Tucker *et al.*, 1985; Townshend *et al.*, 1987). However, a recent study has suggested the potential of using PCA and Standardized Components as a tool for the analysis of change in spatial time series data (Eastman, 1992a). In that study, a PCA procedure capable of analyzing up to 12 bands was used to study artificial data sets and monthly AVHRR-derived Normalized Difference Vegetation Index[1] (NDVI) imagery for Africa over the course of a year and for the same month over four years. The results showed that the technique is able to identify both cyclic seasonal elements of change and isolated change events. In this study, a modification of the PCA procedure allowing a large number of bands to be analyzed is used to study the potential of Standardized Principal Components for the analysis of long time series data sets.

Principal Components Analysis undertakes a linear transformation of a set of image bands to create a new band set with images that are uncorrelated and are ordered in terms of the amount of variance explained in the original data (Johnston, 1980, pp. 127-158; Mather, 1987, pp. 206-218). Most commonly, the technique has been used in remote sensing as a procedure for data compression by discarding minor components with little explanatory value. With Unstandardized PCA, the transformation coefficients are developed by computing the principal eigenvectors of the variance/covariance matrix. In this way, bands with higher variability contribute more to the development of the new component images. With Standardized PCA (Singh and Harrison, 1985) the eigenvectors are computed from the correlation matrix. The effect is to force each band to have equal weight in the derivation of the new component images and is identical to converting all image values to standard scores (by subtracting the mean and dividing by the standard deviation) and computing the Unstandardized Principal Components of the results.

Eastman (1992a) has shown that, when the image data set consists of a single variable time series of environmental data, the first standardized component indicates the characteristic value of that variable while the second and all remaining standardized components represent change elements of successively decreasing magnitude. In addition, both Eastman (1992a) and Fung and LeDrew (1987) indicate that Standardized PCA appears to be more effective than Unstandardized PCA in the analysis of change in multi-temporal image data sets.

Long Sequence Time Series PCA

In this study, the PCA procedure of the IDRISI software system (Eastman, 1992b) was modified to allow the computation of up to 62 components in order to examine the utility of the approach for the investigation of long time series data. To facilitate data entry, image file names are entered by means of a standard IDRISI time series file—a simple ASCII file of image names that is used in a variety of time series procedures such as image display sequencing and time profiles. Output consists of the component images and a set of data tables. Full tables of the variance/covariance and correlation matrices, eigenvalues and eigenvectors, and component loadings are provided for the first 12 components. In addition, a data file consisting of the eigenvalues and the component loadings for all components is created in a format suitable for input into a spreadsheet.

To test the technique, a 36-month sequence of AVHRR-derived NDVI data was analyzed for the continent of Africa. The data were extracted from the NGDC Monthly Generalized Global Vegetation Index data set within the NOAA-EPA Global Ecosystems Database (NOAA-EPA, 1992). The data set used consisted of 10-minute resolution raster data sets of NDVI for the months of January 1986 through December 1988, scaled by NOAA to an 8-bit integer range. An ocean mask image from the U.S. Navy Fleet Numerical Oceanographic Center Global Elevation data set, also within the NOAA-EPA Global Ecosystems Database (NOAA-EPA, 1992), was used to mask water areas. Assigning the value zero to all water areas assures that these regions are forced to show up in the first

[1]The NDVI index is derived by dividing the difference between the infrared and red images by the sum of the infrared and red images, i.e., NDVI = (IR − R)/(IR + R).

Photogrammetric Engineering & Remote Sensing, Vol. 59, No. 6, June 1993, pp. 991–996.

0099-1112/93/5906–991$03.00/0

J. Ronald Eastman
Michele Fulk
The Clark Labs for Cartographic Technology and Geographic Analysis, Clark University, Worcester, MA 01610.

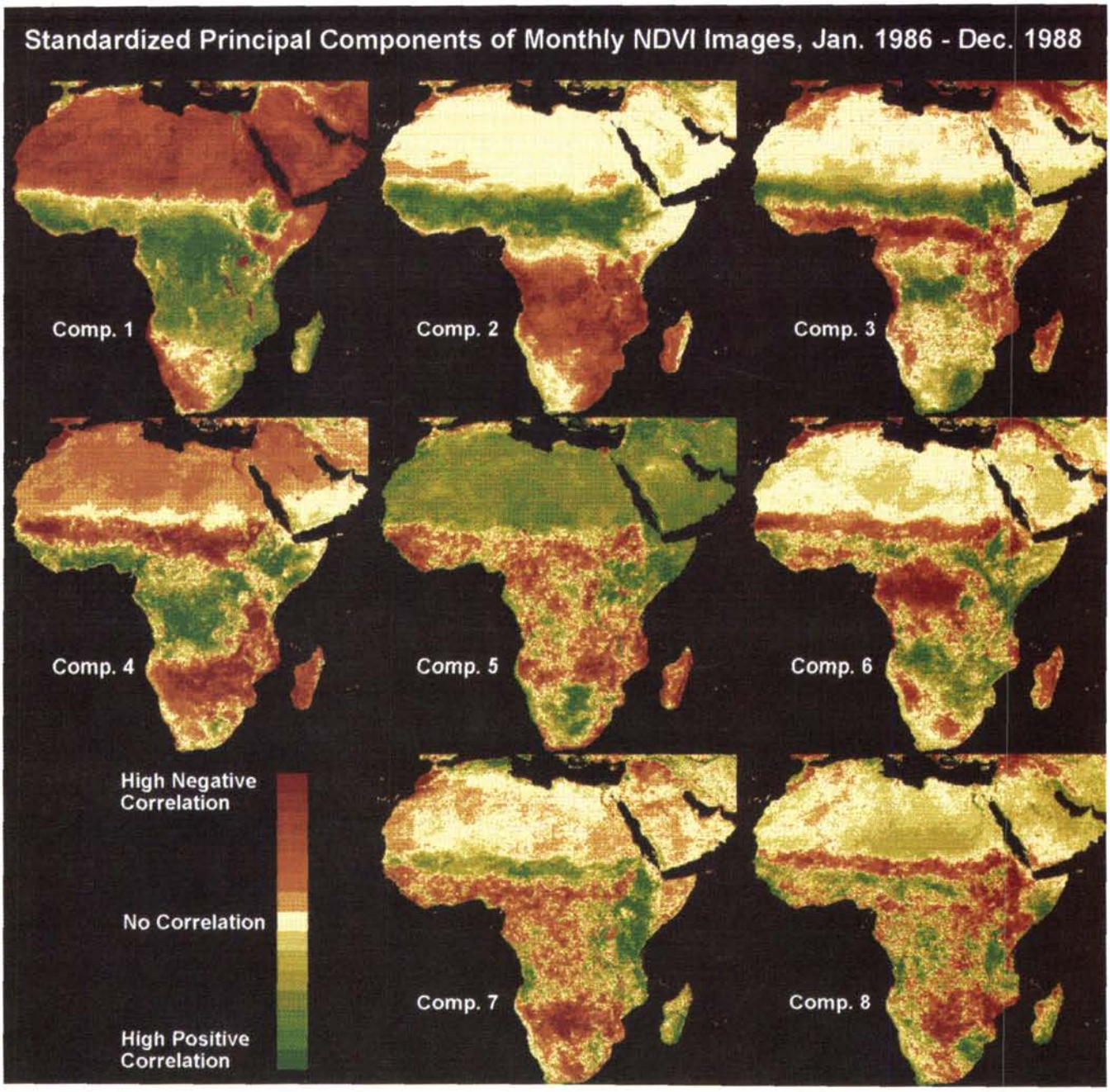

Plate 1. Standardized Principal Component Images 1 through 8 derived from monthly NDVI data, January 1986 to December 1988.

component, thus removing them from all change components.

The 36-Month Africa Time Series Experiment

The interpretation of the results of using standardized PCA on 36 monthly NDVI images for Africa is based on the examination of the component images (Plate 1) and the graphs of the component loadings (Figure 1). The loading charts illustrate the correlation between each of the 36 monthly images and the component being diagrammed. For example, if a month shows a strong positive correlation with a specific component, it indicates that that month contains a latent (i.e., to some extent hidden or unapparent) spatial pattern that has strong similarity to the one depicted in the component image. Similarly, a strong negative correlation indicates that the monthly image has a latent pattern that is the inverse of that shown (i.e., with positive and negative anomalies reversed). To enhance the visualization of the

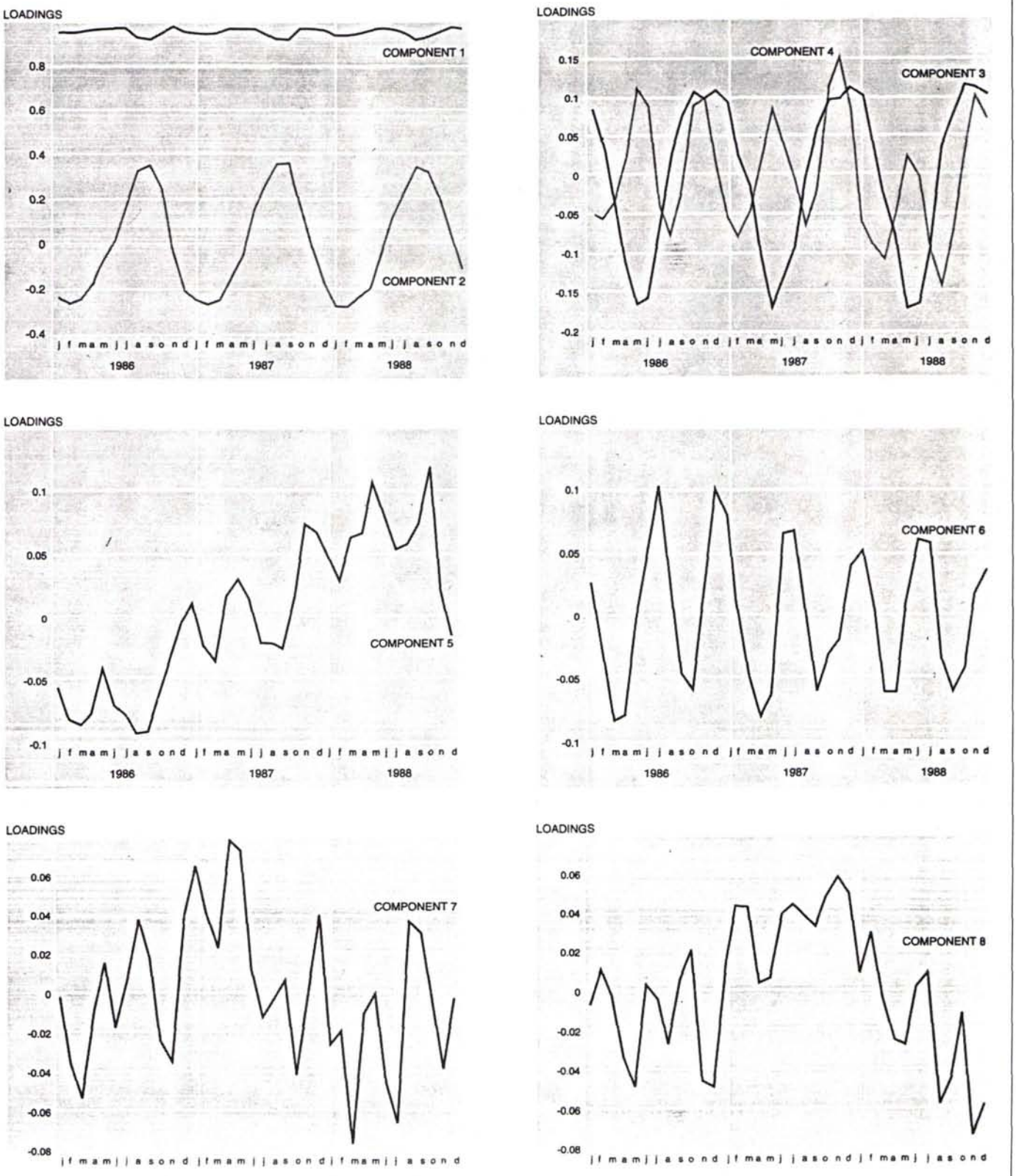

Figure 1. Loadings (*Y*-axis) of the original monthly images (*X*-axis) on the first eight standardized principal components. Loadings can equally be thought of as the correlation between the original images and the derived components.

components in Plate 1, a linear contrast stretch was applied to the output images of the PCA software program. In the case of the first component in Plate 1, the minimum and maximum values were used as the bounds of the stretch. For all other components, however, stretches were forced to be symmetrical about the value 0 (the *no change* value) with a mild saturation (1 percent maximum) in the tails.[2] This no change position has a yellow color while positive and negative anomalies show up in increasing levels of green or red/brown respectively. In all cases, water areas were masked to have a black color.

The first chart in Figure 1 illustrates the loadings from the first two Standardized Principal Components. Component 1 (Plate 1 and Figure 1) clearly represents the characteristic NDVI integrated over all seasons—the loadings are high and very consistent over the entire period. In effect, this component indicates that the major element of variability in NDVI is that which occurs spatially.

Component 2 illustrates the first change component — the most prevalent element of variability in NDVI that is uncorrelated with the characteristic pattern in Component 1. As can be seen by the loadings in Figure 1, this component shows an annual cycle indicating that this second major element of variability in NDVI in Africa is that caused by the winter/summer dichotomy. Summer months correlate positively with the component while winter months correlate negatively.[3] The negative correlation thus indicates that the winter months tend to have an inverse pattern to the summer pattern shown.

The second chart in Figure 1 illustrates the loadings from the third and fourth components. Component 3, like Component 2, also shows an annual cycle, but this time it indicates areas that undergo strong changes in the late spring and autumn periods. In Plate 1, Component 3, quite prominent positive and negative anomalies are seen in the Guinea-Congolic/Zambesian and Guinea-Congolic/Sudanic vegetative transition zones (White, 1983). These areas are south and north of the central equatorial forest region, respectively. These areas experience maximum vegetation peaks in the late fall and late spring, respectively, as a result of the coincidence of the sun's latitudinal position and the rainy season. In addition, the image shows a quite intriguing positive anomaly across the Sahel and into the Ethiopian highlands. The pattern in the Ethiopian highlands is easily explained—the area normally experiences wet and dry periods at these times. However, the interpretation of the pattern in the Sahel is not so obvious. The late autumn (the time at which the positive anomaly is strongest) is normally a time of significant decrease in biomass, yet the image for Component 3 indicates that the vegetation index is abnormally high for this time of year. This is also the time of the Harmattan winds, a prevailing wind pattern out of the Sahara that leads to significant amounts of dust in the Sahel. Although further data would be needed to confirm this, the evidence in the data suggests that atmospheric dust is leading to a greater NDVI measurement than anticipated. The greater attenuation of the shorter red wavelengths compared to the infrared wavelengths would naturally lead to a larger NDVI measurement than expected. Figure 2 illustrates a time series profile over the 36-month sequence of NDVI measurements for a representative area in this region. As can be seen, the expected sinusoidal shape of the annual NDVI pattern is broken by a distinct shoulder in the periods from November until March, the period in which the Harmattan winds are normally experienced. The suggestion is that this shoulder results from an abnormally high apparent NDVI as a result of atmospheric dust. In addition, the pronounced drop in NDVI just before the late summer green-up period may relate to the cleansing of atmospheric dust by early rains. These rains would remove dust from the air, but the vegetation would not yet have had time to respond. Thus, we see a drop to the minimum NDVI level, followed by an increase as the green-up period proceeds.

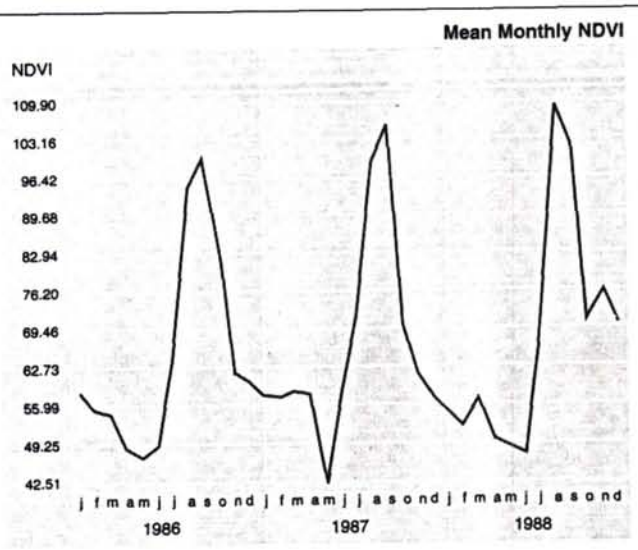

Figure 2. Time series profile of monthly NDVI for a representative site in the Sahel affected by the positive anomaly in Component 3, January 1986 to December 1988.

As can be seen in Figure 1, Component 4 shows a pronounced and consistent semi-annual sequence. Examination of the component image (Plate 1) clearly shows that this illustrates the regions subject to a double precipitation maximum due to the double crossing of the Inter-Tropical Convergence Zone (compare, for example, to Figure 18 in WMO (1984) that maps these areas).

Components 5 and 6 are particularly interesting. The graph of loadings for Component 5 (Figure 1) shows a progressive trend over the three-year period. The image of Component 5 indicates a high positive anomaly in desert areas (note the Sahara) and the lakes of East Africa. Clearly this is illogical because it suggests that desert and water areas are increasing in NDVI over time. A time series profile for representative locations within these regions (Figure 3), however, confirms this trend in the data. As indicated in Tateishi and Kajiwara (1992), the NOAA-9 satellite experienced a progressive delay in the time of equatorial crossing (from 14:20 in December 1984 to 16:10 in November 1988), leading to successively shallower solar angles and, hence, longer atmospheric paths. Tateishi and Kajiwara (1992) state that the effect of this is to diminish, through scattering, the shorter red wavelength reflectances more than the longer infrared wavelength reflectances. As a result, the NDVI would ap-

[2]A 1 percent saturation forces the highest 1 percent of data cells to take on the brightest color and the lowest 1 percent to become the darkest color, regardless of actual value. Because the stretch was forced to be symmetrical about 0 (no change), the stretch was undertaken using end points such that a maximum of 1 percent would be saturated in any tail.

[3]Because Africa spans the equator, seasonal terms (winter, summer, etc.) for the northern hemisphere have been arbitrarily adopted throughout this article.

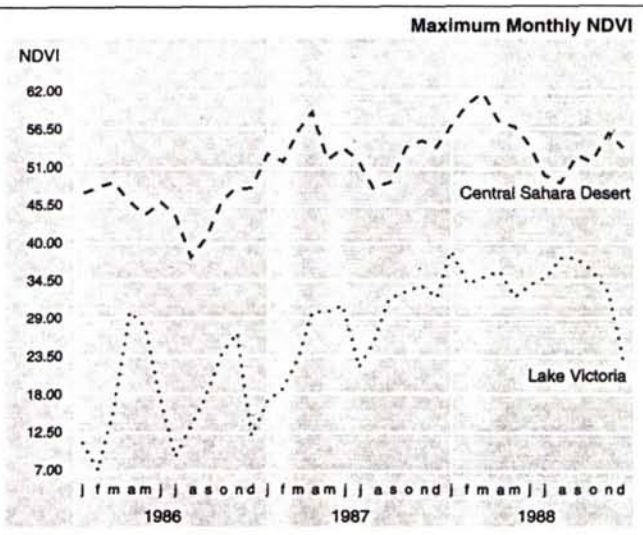

Figure 3. Time series profiles of monthly NDVI for sites in the central Sahara Desert and Lake Victoria, January 1986 to December 1988.

pear to increase in these areas. Component 5 has thus detected a known and significant false trend in the NDVI data.[4]

In addition to the systematic trend in Component 5, the loadings also show a somewhat irregular semi-annual cycle such as that found in Component 4. Examination of the time profiles in Figure 3, however, clearly illustrate that the loading pattern is a composite of the cycles associated with an annual cycle in the Sahara and the semi-annual cycle associated with the precipitation maximum of the lakes of equatorial Africa. Component 6 would also appear to relate to this sensor effect. As indicated by Tateishi (personal communication, 1992) this progressive drift in the NDVI output has the effect of decreasing the range of NDVI in forested regions. This is illustrated by the decreasing amplitude of the cyclic pattern of loadings on Component 6. The semi-annual cycle is typical of that associated with the double precipitation maximum of equatorial forest areas (such as we see in Component 4), and the amplitude is clearly decreasing. Similarly, the component image (Plate 1) shows a strong negative anomaly in heavily forested regions of the continent.

The analysis thus shows that, after the general geographical and major seasonal effects, the next major agent of variability in the NDVI data over the three-year period was system-related variability in the output of the AVHRR-derived index itself. It should be noted that, once a change component has been created, the effects of that component are then held constant as subsequent components are calculated (Johnston, 1980, pp. 139-141). To the extent that Components 5 and 6 are able to describe these sensor system effects, later components would thus be free of their influence.

Components 7 and 8 (Plate 1 and Figure 1) are also of considerable interest. Both would appear to relate to El-Niño/Southern Oscillation (ENSO) events spanning the 1986-88 period. Late 1986 and 1987 experienced a pronounced ENSO warm phase, followed in 1988 by a sharp transition to an ENSO cold phase (WMO, 1989). ENSO warm phases are typically associated with drought in southern Africa and enhanced precipitation in equatorial East Africa (WMO, 1984; 1987; 1989). The southern Africa ENSO pattern is clearly visible in the images of Components 7 and 8 in Plate 1. Similarly, the loading charts for these components show positive correlations with 1987 and negative correlations with 1986 and 1988. It would appear that the two components show the spatial progression of the drought. For example, in Component 7, with a peak correlation in early 1987, the area most strongly affected in southern Africa is Botswana. However, in Component 8, with a peak correlation in late 1987, the area most strongly affected has moved towards the east coast. As noted in WMO (1989), the pattern for equatorial east Africa did not show a typical ENSO pattern during this particular ENSO episode. We see this in the more typical moist pattern for east Africa in Component 7, giving way to a drier pattern in Component 8.

The procedure developed for this analysis is capable of producing as many component images as there are original bands in the data. However, examination of Components 9 through 12 did not show any significant regional effects. Rather, more localized changes appear to be brought out. As a result, the analysis was stopped after the eighth component. It is worth noting, however, that no clear guidelines exist for when to stop an analysis. In PCA, the strength of a component (as reflected both in the eigenvalue and in the range of loadings on that component) will be determined by both the magnitude of the variability it explains and the area over which that variability occurs. Thus, change elements in time series analysis will be area weighted. Small magnitude changes may come out in early components if they affect large areas. Conversely, large effects may come out in later components if they occupy only a small portion of the area analyzed. Thus, it would not be unreasonable to examine minor components (such as those beyond component 8 in this analysis) if very localized effects were of interest.

Conclusions

It would appear from the above illustration that the ability of Standardized Principal Components to uncover significant change events over long time series is very strong. The ENSO-related precipitation patterns represent the most significant anomalies to take place over Africa during the 1986-88 period. Not surprisingly, these were lower in magnitude than the effects attributable to seasonal changes. It is also interesting that the procedure picked up significant anomalies in the output of the sensor system itself, and that these were also seen to be of greater magnitude than the ENSO events. However, as noted above, components are effectively area weighted. The sensor drift effect is in fact quite small but affects broad areas of the image (particularly the Sahara). The ENSO effects occupy substantially less area but have a dramatic effect on natural ecosystems.

Standardized Principal Components Analysis would thus appear to be a remarkably comprehensive tool for the analysis of anomalies and trends in long time series data. It is clearly very effective in isolating periodic seasonal effects. However, it is equally effective in isolating trends in value and variability (as with the sensor drift problem) and isolated anomalous events. Given our general lack of techniques for the abstraction of significant change events in long time series image data, the technique should prove to be a major tool in the areas of remote sensing and GIS.

[4]It is interesting to note the drop at the end of the sequence back to approximately the zero correlation level. The majority of the November 1988 data and all of the December 1988 data were actually derived from NOAA-11, not NOAA-9.

References

Eastman, J. Ronald, 1992a. Time Series Map Analysis Using Standardized Principal Components. *ASPRS/ACSM/RT 92 Technical Papers, Volume 1: Global Change and Education*, 3-8 August, Washington, D.C., pp. 195–204.

———, 1992b. IDRISI: A Grid-Based Geographic Analysis System. Version 4.0. Clark University, Worcester, Massachusetts.

Fung, T., and LeDrew, E. 1987. Application of Principal Components Analysis to Change Detection. *Photogrammetric Engineering & Remote Sensing*, Vol.53, No. 12, pp. 1649–1658.

Johnston, R.J., 1980. *Multivariate Statistical Analysis in Geography.* Longman, New York.

Mather, P.M., 1987. *Computer Processing of Remotely-Sensed Images.* John Wiley and sons, New York.

NOAA-EPA, 1992. *Global Ecosystems Database Project CD-ROM.* National Oceanic and Atmospheric Administration, National Geophysical Data Center, Boulder, Colorado.

Singh, A., and A. Harrison, 1985. Standardized Principal Components. *International Journal of Remote Sensing*, Vol. 6, No. 6, pp. 883–896.

Tateishi, R., and K. Kajiwara, 1992. *Global Land Cover Monitoring by AVHRR NDVI Data. Earth Environment*, Vol. 7, pp. 4–14.

Townshend, J., C. Justice, and B. Kalb, 1987. Characterization and Classification of South American Land Cover Types Using Satellite Data. *International Journal of Remote Sensing*, Vol. 8, No. 8, pp. 1189–1207.

Tucker, C., J. Townshend, and T. Goff, 1985. African Land-Cover Classification Using Satellite Data. *Science*, Vol. 227, No. 4685, pp. 369–375.

White, F., 1983. *The Vegetation of Africa: A Descriptive Memoir to Accompany the UNESCO/AETFAT/UNSO Vegetation Map of Africa.* UNESCO, Paris.

WMO, 1984. *The Global Climate System: A Critical Review of the Climate System During 1982-1984.* World Meteorological Organization, Geneva.

———, 1987. *The Global Climate System: Autumn 1984 - Spring 1986.* World Meteorological Organization, Geneva.

———, 1989. *The Global Climate System: Climate System Monitoring, June 1986-November 1988.* World Meteorological Organization, Geneva.

(Received 26 November 1992; accepted 21 January 1993)

J. Ronald Eastman
J. Ronald Eastman is Director of the Clark Labs for Cartographic Technology and Geographic Analysis and Associate Professor of Geography in the Graduate School of Geography at Clark University. He is also the author of the IDRISI Geographic Information System which is distributed on a nonprofit basis by the Clark Labs. His recent research interests include change and time series analysis with remotely sensed images, error and uncertainty, and decision support tools for GIS.

Michele Fulk
Michele A. Fulk is Public Outreach Representative and Research Associate at The IDRISI Project, Clark University Graduate School of Geography, where she has worked since January of 1991. She is also pursuing a master's degree in International Development at Clark University. Her research interests include change and time series analysis and applications of GIS and image processing to problems of development and resource management.

LIST OF "LOST" CERTIFIED PHOTOGRAMMETRISTS

We no longer have valid addresses for the following Certified Photogrammetrists. If you know the whereabouts of any of the persons on this list, please contact ASPRS headquarters so we can update their records and keep them informed of all the changes in the Certification Program. Thank you.

Jack R. Anthony
Dewayne Blackburn
Gerard Borsje
Albert Brown
Eugene Caudell
Robert Denny
Leo Ferran
Robert Fuoco
Franek Gajdeczka
George Glaser
William Grehn, Jr.
Louis T. Harrod
F.A. Hildebrand, Jr.
James Hogan
William Janssen
Lawrence Johnson
Spero Kapelas
Andre J. Langevin
Harry J. Miller
Marinus Moojen
Gene A. Pearl
Sherman Rosen
Lane Schultz
Keith Syrett
William Thomasset
Conrad Toledo
Robert Tracy
Lawrence Watson
Tad Wojenka