On the Use of the Principal Component Analysis (PCA) for Evaluating Vegetation Anomalies from LANDSAT-TM NDVI Temporal Series in the Basilicata Region (Italy)

Antonio Lanorte, Teresa Manzi, Gabriele Nolè, and Rosa Lasaponara(✉)

Institute of Metodologies for Enviromental Analysis, National Research Council of Italy, Tito Scalo (PZ), Italy
rosa.lasaponara@imaa.cnr.it

Abstract. In this paper, we present and discuss the investigations we conducted in the context of the MITRA project focused on the use of low cost technologies (data and software) for pre-operational monitoring of land degradation in the Basilicata Region. The characterization of land surface conditions and land surface variations can be efficiently approached by using satellite remotely sensed data mainly because they provide a wide spatial coverage and internal consistency of data sets. In particular, Normalized Difference Vegetation Index (NDVI) is regarded as a reliable indicator for land cover conditions and variations and over the years it has been widely used for vegetation monitoring. For the aim of our project, in order to detect and map vegetation anomalies ongoing in study test areas (selected in the Basilicata Region) we used the Principal Component Analysis applied to Landsat Thematic Mapper (TM) time series spanning a period of 25 years (1985-2011).

Keywords: Satellite based analysis · Land degradation · Principal component analysis · GIS · Spatial variation · Vegetation index · Basilicata

1 Introduction

In relation to the land cover changes that can occur over large areas, the use of remote sensing data is an essential tool for change monitoring and mapping [3], [13].

Furthermore, remote sensing has been recognized worldwide as an effective, accurate and economical method to monitor changes in land cover from global down to a local scale [9], [14].

Many studies use remote sensing to monitor and evaluate the impact of urban growth on agricultural land [10], [15] as well as, to a lesser extent, the phenomena of agricultural neglect [1].

The Basilicata Region is characterized by a significant hydrogeological risk level and ongoing land degradation processes which have been particularly evident in the 30 last years. The so-called "erosion" processes affect large areas and are frequently strongly related to land degradation processes.

This paper is focused on the results we obtained from investigation conducted using the Principal Component Analysis applied to Landsat Thematic Mapper (TM)

O. Gervasi et al. (Eds.): ICCSA 2015, Part IV, LNCS 9158, pp. 204–216, 2015.
DOI: 10.1007/978-3-319-21410-8_16

time series spanning over a 25 year period (1985-2011) to detect and map vegetation anomalies. The Normalized Difference Vegetation Index (NDVI) was used as input to a selective Principal Component Analysis (PCA) procedure. The PCA was used as a first step of data transform to enhance regions of localized change in multi-temporal data sets [8], [5]. Results from PCA were further processed using Geospatial analysis to identify and map land degradation phenomenon.

NDVI is very effective for the identification of vegetation health being based on the normalized difference (see formula 1) between the infrared red band reflectance:

$$NDVI = (\rho_{RED} - \rho_{NIR})/(\rho_{RED} + \rho_{NIR}) \tag{1}$$

NDVI provides information about the spatial and temporal distribution of vegetation cover in term of types, amount (biomass) and conditions. It provides a reliable estimation of the amount and vigor of vegetation because it is strongly related to the photosynthetic activity (Fig. 1).

The success of the NDVI is due to its reliability in detecting vegetation as well as in its simplicity in terms of computation and interpretation. That is the main reason of its wider use and larger popularity compared to other satellite-based spectral vegetation index.

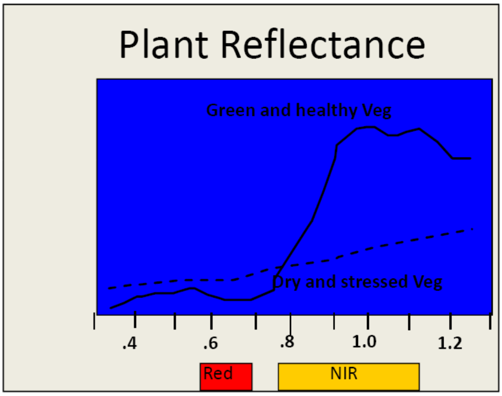

Fig. 1. Spectral response of vegetation in the red-nir spectrum regions

$$\operatorname{cov} k1, k2 = 1/nm \sum_{i=1}^{n} \sum_{J=1}^{m} \left(MVC_{i,j,k1} - \mu_{k2}\right)\left(MVC_{i,j,k2} - \mu_{k2}\right) \tag{2}$$

Where k1, k2 are two input time series dates, MVC_{ij} the annual maximum NDVI value in row i and column j , n the number of row, m the number of columns and μ is the mean of all pixel MVC values in the subscripted input dates.

The percent of total dataset variance explained by each component is obtained by formula 3:

$$\% i = 100 * \lambda_i / \sum_{i=1}^{k} \lambda_i \tag{3}$$

where λ_i are eigenvalues of S.

Finally, a series of new image layers (called eigenchannels or components) are computed (using formula 4) by multiplying, for each pixel, the eigenvector of S for the original value of a given pixel in the input bands

$$P_i = \sum_{i=1}^{n} P_k \times u_{k,i} \tag{4}$$

where P_i indicates a spectral channel in component i, u_{ki} eigenvector element for component i in input band k, P_k spectral value for channel k, number of input band.

A loading, or correlation R, of each component i with each input band k can be calculated by using formula 5.

$$R_{k,i} = u_{k,i} \times (\lambda_i)^{1/2} \times (\mathrm{var}_k)^{1/2} \tag{5}$$

where var_k is the variance of input date k (obtained by reading the kth diagonal of the covariance matrix).

PCA transformation produces new principal components (PC1, PC2, PC 3, …etc.), which are uncorrelated and ordered in terms of the amount of variance they represent with respect to the set of the original bands. These bands are often more interpretable than the source data.

PCA is widely used in detecting changes in time series data and has become one of the most popular techniques because of its simplicity and capability of enhancing even subtle changes.

In our investigations, PCA was used to emphasize the areas that present significant changes in multi-temporal data sets. This is a direct result of the (i) high correlation that exists among pixels related to regions that do not change significantly and the (ii) relatively low correlation associated with regions that change substantially.

The major portion of the variance in a multi-temporal image data set is associated with constant cover types and is represented in PC1, while the regions with localized change will be enhanced in later components. In particular, each successive component contains less of the total data set variance. In other words, the first component contains the major portion of the variance, whereas, later components contain a very low proportion of the total data set variance.

PC1 explains most of the variation in the NDVI integrals and represents the average spatial integrated NDVI pattern. PC1 shows the typical NDVI over the entire series.

PC2 explains the maximum remaining variation not explained by PC1 and subsequent components follow the same rationale.

Therefore the components subsequent to PC1 (especially PC2) tend to reflect specific events such as fires, drought periods and, in general, land degradation phenomena related to the vegetation, rather than to depict the general development during the period.

PCA is then used here as a tool for mapping areas that show a significant i.e measurable degree of inter-annual variability, able to discriminate unidirectional changes.

2 Dataset and Study Area

Data Set

The investigations were performed by using NDVI data derived from the Landsat TM images (Tab. 1) selected on the base of the data quality and low percentage of cloud cover in the months between June and September (period 1985-2011).

Table 1. Analytical categories

Thematic Mapper (TM)		
Landsat bands	**Wavelength (micrometers)**	**Resolution (meters)**
Band 1	0.45-0.52	30
Band 2	0.52-0.60	30
Band 3	0.63-0.69	30
Band 4	0.76-0.90	30
Band 5	1.55-1.75	30
Band 6	10.40-12.50	120
Band 7	2.08-2.35	30

The Landsat TM data were acquired free of charge from the United States Geological Survey (USGS) web site (Tab. 2).

Table 2. Landsat-TM Dataset

Year	Month	Day
1985	August	10
1986	August	13
1986	September	14
1987	June	13
1987	September	17
1993	July	15
2002	June	22
2003	June	25
2003	July	11
2003	July	27
2009	July	27
2010	September	16
2011	August	02
2011	August	18

Study Area

The analysis was performed in the Basilicata Region (see Fig. 2) that is characterized by typical Mediterranean climate with a pronounced bi-seasonality regime having hot/dry summers and cold/wet winters. Due to a combined effect of natural hazards (drought, wind and rain erosion, floods) and human activity (industry, fires, over tilling, land abandonment), this area recently increased its vulnerability [2].

The environmental equilibrium of the Basilicata is fragile and highly vulnerable to perturbations, as in other Mediterranean regions, and, therefore, it is expected that natural ecosystems, such as forest, shrubland and herbaceous cover, should be more sensitive to the changes that are presently affecting the whole Mediterranean basin.

Fig. 2. Study area

3 Methodology

Figure 3 shows a flow chart of the methodology that we adopted.

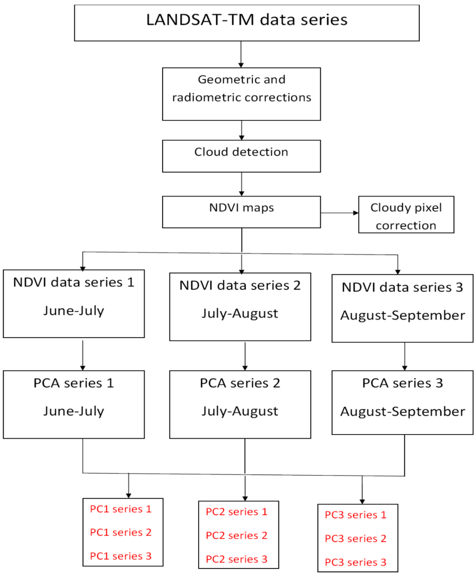

Fig. 3. Methodology flow chart

The Landsat-TM data series (summer images) were acquired on the basis of their free-of-charge availability, images quality and low percentage of cloud cover.

On each image we applied a geometric, radiometric and atmospheric correction, then a cloud detection if need.

Then we created three NDVI time series corresponding to June-July, July-August and August-September periods: this in order to avoid to apply PCA in the temporal windows too different on the phenological point of view.

For each NDVI series the PCA was applied and, finally, the first, the second and the third components of each series was combined and analyzed separately.

4 Results and Discussion

As mentioned before the most significant results in the evaluation of vegetation anomalies can derive from the second component of the PCA. The results shown in figure 4 concern the analysis of the second component of the PCA in some significant test sites selected in the Basilicata region.

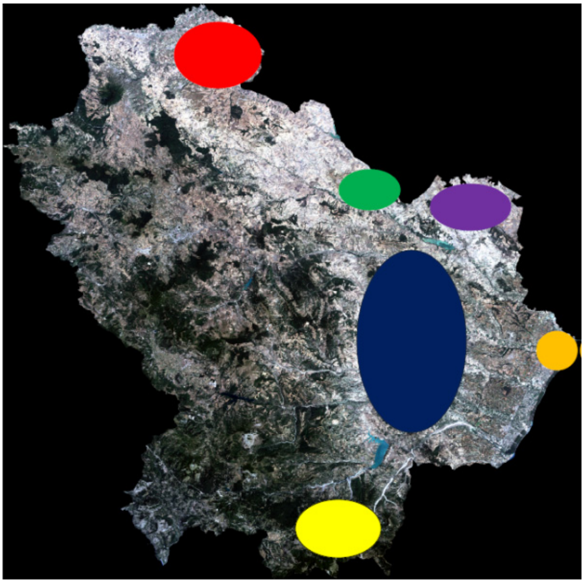

Fig. 4. Study areas. (purple=Matera; orange=Metaponto; yellow= Pollino; green=Irsina; red=North-East; Blue=Lucanian Badlands).

In figure 5 there is a comparison of the second components obtained for each NDVI data series for the Matera site. The graphics show the loadings of the second component of each series. The loadings representing the correlation of each component (in this case the second component) with each input date (NDVI). In this case the loadings show in general a decreasing trend in all three series.

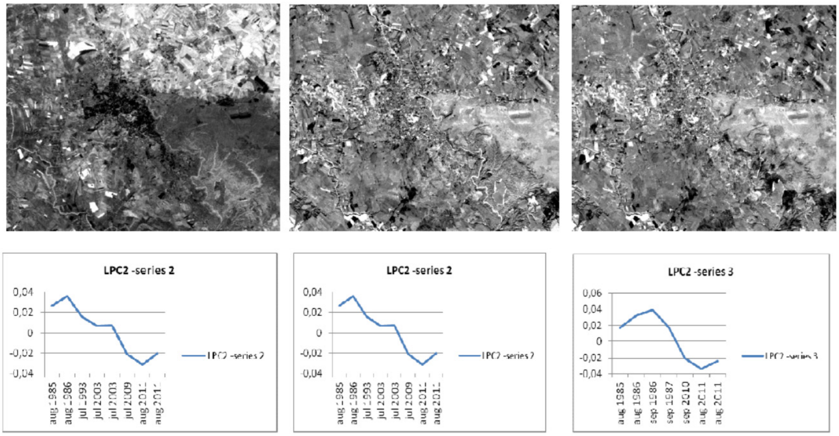

Fig. 5. Matera site. PC2 and loadings (3 series).

This trends justify the presence of negative anomalies (Fig. 6). This results concern only the anomalies detected in natural or semi-natural areas excluding, then, the agricultural areas in order to avoid to confuse the changes of agricultural land use with vegetation anomalies.

We distinguished highly negative anomalies (in red) and moderately negative anomalies (in orange). We defined the areas with highly negative anomalies as "critical areas" and the areas with moderately negative anomalies as "fragile areas".

The critical areas correspond to areas where negative anomalies are high in all three considered series, while the fragile areas correspond to areas where the negative anomalies are high only in two series. High negative values are the values below a set threshold.

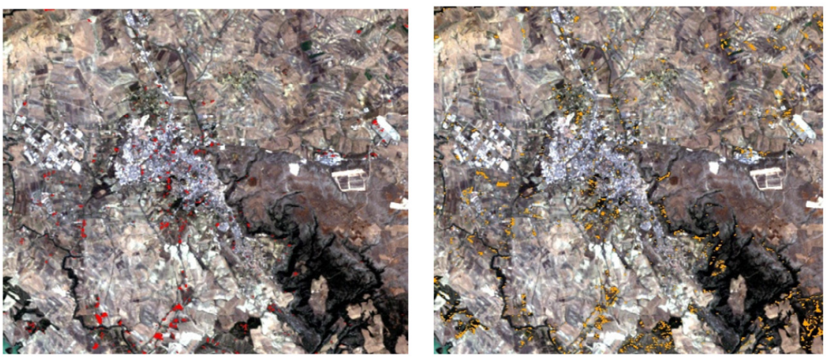

Fig. 6. Matera site. Negative anomalies (left: critical areas; right: fragile areas).

In the case of the Metaponto site, PC2 loadings trends (Fig. 7) are very different compared to the first case. Here, the loadings of series 1 show a decreasing trend until the 2003 and an increasing trend after; the loadings of series 2 tend to increase or (after 2003) they are stable; the loadings of series 3 show a decreasing trend more pronounced but in the last years they are stable.

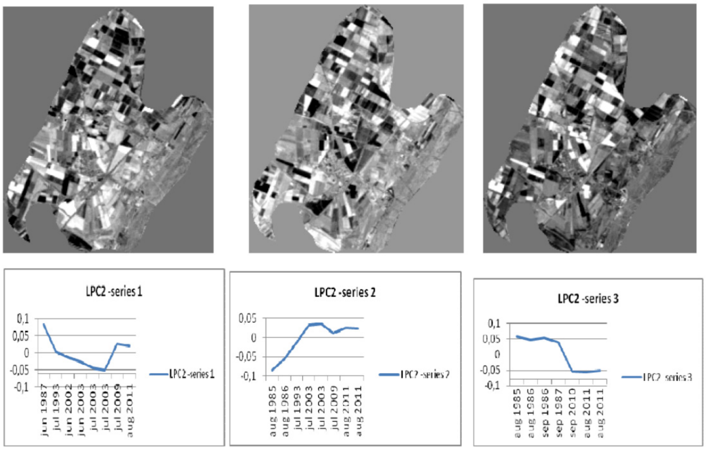

Fig. 7. Metaponto site. PC2 and loadings (3 series).

The map of anomalies reflects these trends because we did not detect any highly negative anomaly but only fragile areas (Fig. 8).

Fig. 8. Metaponto site. Negative anomalies (fragile areas).

Figure 9 shows the results obtained in the Pollino site. In this case PC2 loadings show an increasing trend in series 1 and 2 but a decreasing trend in series 3.

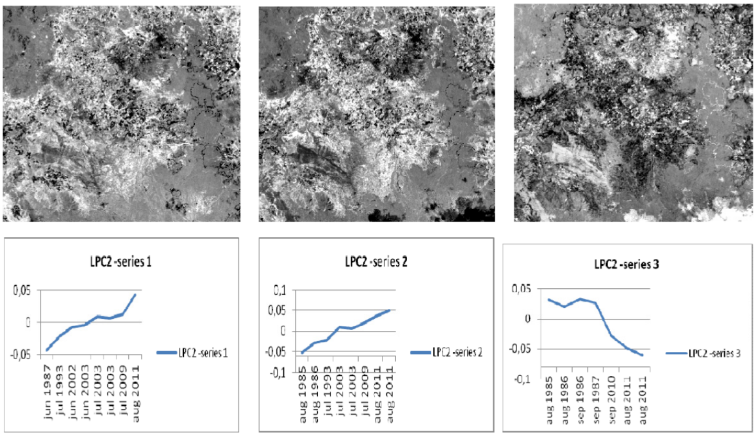

Fig. 9. Pollino site. PC2 and loadings (3 series).

These trends are confirmed in the map of anomalies where we observe some fragile areas (Fig. 10).

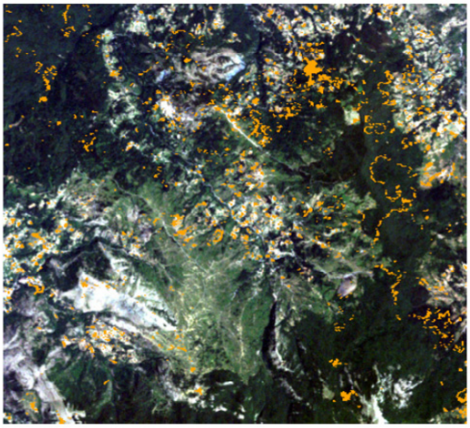

Fig. 10. Pollino site. Negative anomalies (fragile areas).

The same observations are suitable for the Irsina site (Figs. 11 and 12).

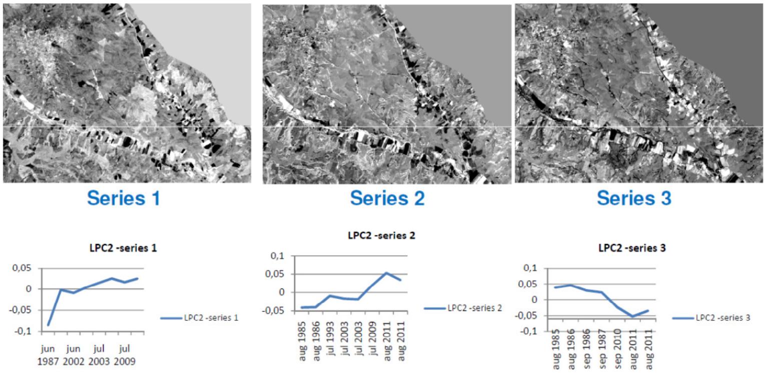

Fig. 11. Irsina site. PC2 and loadings (3 series).

Fig. 12. Irsina site. Negative anomalies (fragile areas).

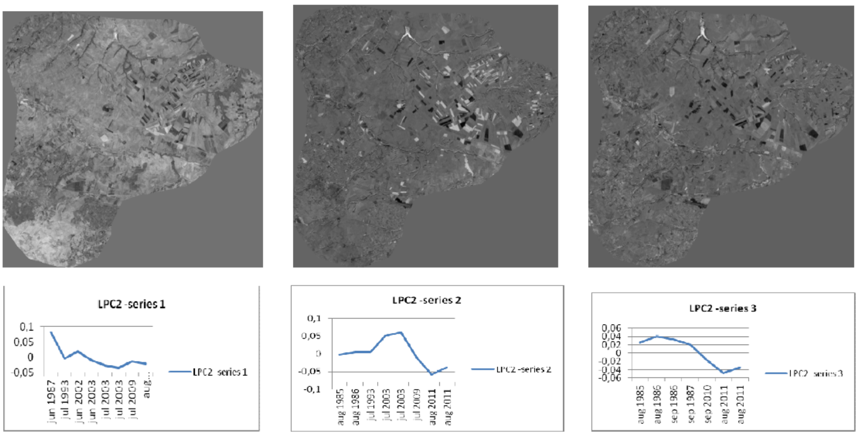

Fig. 13. North-East Basilicata site. PC2 and loadings (3 series).

In agreement with this analysis we found several highly and moderately negative anomalies (Fig. 14).

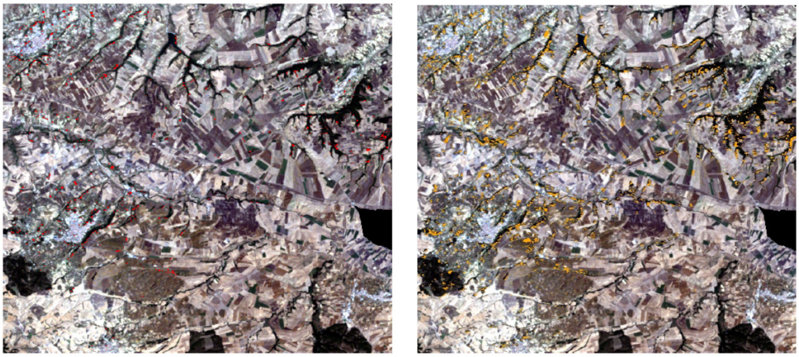

Fig. 14. North-East Basilicata site. Negative anomalies (left: critical areas; right: fragile areas).

The last site correspond to the Lucanian badlands area. In this case all loadings clearly show a decreasing trend (Fig. 15).

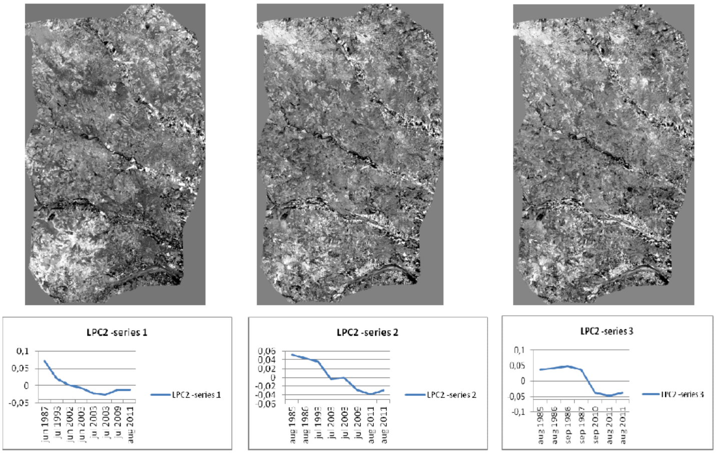

Fig. 15. Lucanian Badlands site. PC2 and loadings (3 series).

In fact we found many negative anomalies (critical and fragile areas) (Fig. 16).

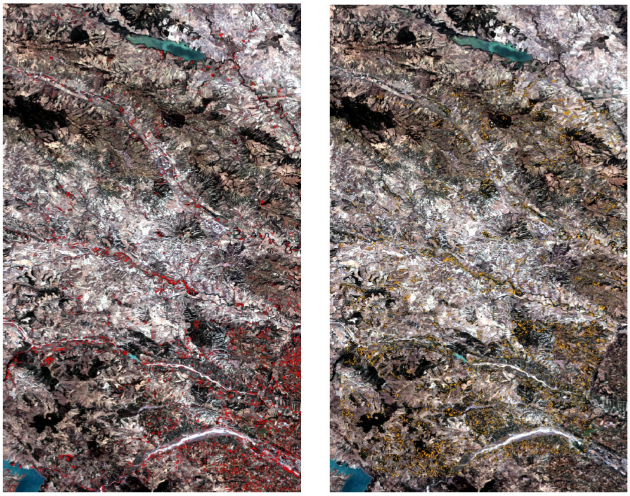

Fig. 16. Lucanian Badlands site. Negative anomalies (left: critical areas; right: fragile areas).

Finally, the results shown are consistent with the analysis performed with SPOT-VGT NDVI time series for land degradation monitoring in the Basilicata Region [6]. Moreover we found significant spatial correlations with the map of areas susceptible to desertification in Basilicata [4]. Thus confirming the reliability of spatial data and robust statistical analysis for operational risk monitoring as, for example, in [7,11,12, 16, 17,18].

5 Final Remarks

The technique and the methodology we adopted show a promising ability to identify and monitor vegetation anomalies also considering the relatively small number of Landsat images. The use of the PCA allowed us to monitor degradation phenomena in heterogeneous landscapes such those investigated for the Basilicata region at the medium-high scale provided by Landsat-TM data. The use of PCA can be an effective and low cost tool for extracting valuable information from NDVI temporal series regarding vegetation inter-annual variations. This study exemplifies the potential use of Landsat time series for environmental analyses performed on local scale and provides basic information applied in change detection analyses and in monitoring land degradation processes. The results accuracy could be increased by using other Landsat datasets series (ETM, Landsat 8), by improving data pre-processing and by integrating the results with ground surveys.

Acknowledgement. This research has be made in the context of Bilateral Italy-Argentina MAE project PGR00189 titled "Remote sensing technologies applied to the management of natural and cultural heritage in sites located in Italy and Argentina. Risk monitoring and mitigation strategies".

References

1. Alcantara, P.C., Radeloff, V.C., Prishchepov, A., Kuemmerle, T.: Mapping abandoned agriculture with multi-temporal MODIS satellite data. Remote Sensing of Environment **124**, 334–347 (2012)
2. APAT report 2006. La vulnerabilità alla desertificazione in Italia. Manuali e linee guida 40/2006
3. Dewan, A.M., Yamaguchi, Y.: Using remote sensing and GIS to detect and monitor land use and land cover change in Dhaka Metropolitan of Bangladesh during 1960-2005. Environ. Monit. Assess. **150**, 237–249 (2009). doi:10.1007/s10661-008-0226-5
4. Ferrara, A., Bellotti, A., Faretta, S., Mancino, G., Baffari, P., D'Ottavio, A.: Carta delle aree sensibili alla desertificazione della Regione Basilicata. Forest@-Journal of Silviculture and Forest Ecology **2**(1), 66 (2005)
5. Howarth, P., Piwowar, J., Millward, A.: Time-Series Analysis of Medium-Resolution, Multisensor Satellite Data for Identifying Landscape Change. Photogrammetric Engineering and Remote Sensing **72**(6), 653–663 (2006)
6. Lanorte, A., Aromando, A., De Santis F., Lasaponara, R.: Investigating satellite SPOT VEGETATION multitemporal NDVI maps for land degradation monitoring in the Basilicata Region: Preliminary Results from the MITRA project. In: Proceedings of the 33rd EARSel Symposium, Matera (Italy) (2013)
7. Lanorte, A., Danese, M., Lasaponara, R., Murgante, B.: Multiscale mapping of burn area and severity using multisensor satellite data and spatial autocorrelation analysis. International Journal of Applied Earth Observation and Geoinformation **20**, 42–51 (2013)
8. Lasaponara, R.: On the use of principal component analysis (PCA) for evaluating interannual vegetation anomalies from SPOT/VEGETATION NDVI temporal series. Ecol. Model. **194**, 429–434 (2006)
9. Lunetta, R.S., Johnson, D.M., Lyon, J.G., Crotwell, J.: Impacts of imagery temporal frequency on land-cover change detection monitoring. Remote Sensing of Environment **89**, 444–454 (2004)
10. Stefanov, W.L., Ramsey, M.S., Christensen, P.R.: Monitoring urban land cover change: An expert system approach to land cover classification of semiarid to arid urban centers. Remote Sensing of Environment, 173–185 (2001)
11. Telesca L., Lasaponara R.: Pre and post fire behavioral trends revealed in satellite NDVI time series. Geophysical Research Letters 33(14) (2006)
12. Tuia, D., Ratle, F., Lasaponara, R., Telesca, L., Kanevski, M.: Scan statistics analysis of forest fire clusters, pp. 1689–1694 (2008)
13. Tziztiki, J.G.M., Jean, F.M., Everett, A.H.: Land cover mapping applications with MODIS: a literature review. International Journal of Digital Earth **5**(1), 63–87 (2012)
14. Wilson, E.H., Sader, S.: Detection of forest harvest type using multiple dates of Landsat-TM imagery. Remote Sensing of Environment **80**, 385–396 (2002)
15. Yeh, A.G., Li, X.: An integrated remotes sensing and GIS approach in the monitoring and evaluation of rapid urban growth for sustainable development in the Pear River Delta, China. International Planning Studies **2**(2), 193–210 (1997)
16. Lanorte, A., Danese, M., Lasaponara, R., Murgante, B.: Multiscale mapping of burn area and severity using multisensor satellite data and spatial autocorrelation analysis. International Journal of Applied Earth Observation and Geoinformation **20**, 42–51 (2013)
17. Telesca, L., Lasaponara, R.: Pre and post fire behavioral trends revealed in satellite NDVI time series. Geophysical Research Letters 33(14) (2006)
18. Tuia, D., Ratle, F., Lasaponara, R., Telesca, L., Kanevski, M.: Scan statistics analysis of forest fire clusters, 1689–1694 (2008)