ENHANCED FILTERING OF MODIS TIME SERIES DATA FOR THE ANALYSIS OF DESERTIFICATION PROCESSES IN NORTHEAST BRAZIL

Stefan Erasmi [a, *], Mirko Bothe [a], Reinaldo A. Petta [b]

[a] University of Goettingen, Institute for Geography, Dept. of Cartography & Remote Sensing, Goldschmidtstr. 5, D-37077 Goettingen, Tel. ++49 551 39 8003, Fax: ++49 551 39 8006, EMail: (serasmi, mbothe)@uni-goettingen.de
[b] Universidad Federal do Rio Grande do Norte (UFRN), Dept. of Geology, Campus Universitário, Natal, Brazil, Email: petta@geologia.ufrn.br

KEY WORDS: land cover, change detection, algorithms, semi-automation, satellite, ecosystem

ABSTRACT

Long time series of satellite data provide key input for geographically explicit monitoring and modelling of the environmental condition in the context of land use and land cover change (LUCC) analysis. Within the present study, the quality of vegetation index (VI) time series data was analysed under the assumption that a post-processing filtering algorithm considerably reduces noise and enhances the usability for land cover degradation monitoring purposes. As a prerequisite for data analysis, a filtering workflow for VI time series data was developed that accounts for disturbances in the satellite data time series which are related to atmospheric effects. The filter algorithm includes linear interpolation of cloudy pixels, long-term change trend calculation, iterative filtering and fitting of the new time series based on the Savitzky-Golay filter algorithm. Missing or uncertain values are replaced by long-term trend values and fitted using a polynomial function fitting method The time series filtering results in a new data set of enhanced temporal trajectories of VI values. The filter method was applied to a four year time series of MODIS vegetation index data (MOD13Q1). The results indicate the high potential of the Savitzky-Golay based filtering workflow to minimize noise in lower frequencies while conserving natural peaks of VI values that are related to phenological and seasonal phenomena in vegetation physiology. As a consequence, the filtered time series data provide improved input for the assessment of climate related versus human induced land degradation in semi-humid regions.

1.   INTRODUCTION

The North-Eastern area of Brazil is strongly affected by land cover degradation and environmental problems related to global climate change. These climate factors are mainly characterized by long periods of drought. The drought enforces processes of land degradation and leads to substantial economical losses in agriculture. In addition to this, the region is severely threatened by man made processes of land degradation which are summarized as desertification. The desertification vulnerability is the consequence of a mix of the semi-humid climate in combination with an intensive anthropogenic activity of land use and land cover conversion.

The sustainable management of those areas demands a monitoring system that documents processes of land degradation and links land cover changes to climate and human induced factors. A prerequisite for this is the availability of spatially explicit information at high temporal frequency for the detection and separation of seasonal, phenological, annual as well as inter-annual variability in land cover. Satellite sensors provide such data of long time series starting with the launch of the first Advanced Very High Resolution Radiometer (AVHRR) system in the early 80-ies (Townshend, 1994). AVHRR time series have been used for many studies that aim at the documentation and analysis of land cover and its change at the global level ranging from qualitative land cover change studies (Hansen et al., 2000) to quantitative investigations of biophysical attributes of the land surface that are related to global environmental issues (Defries et al., 2000, Strugnell et al., 2001). Since the late 90-ies the availability of time series data increased substantially with the commissioning of advanced sensor systems such as Spot-VEGETATION, the Moderate Resolution Imaging Spectroradiometer (MODIS) and the Medium Resolution Imaging Spectrometer (MERIS).

The main challenge in time series data processing is, besides problems of hardware limitations, the elimination of noise due to atmospheric effects and variable sensor geometry and the generation of a consistent and reliable data set for environmental applications. Within this context, a variety of methods has been proposed to process time series of satellite image data. In general, the processing chain consists of three modules: 1) radiometric calibration (radiative transfer modelling); 2) temporal compositing and 3) time series smoothing of residual errors.

The present study deals with the third of the above mentioned aspects of time series processing and proposes a filter method for an enhanced smoothing of temporal profiles based on satellite data time series composites. The method is based on a previously developed concept by Chen et al. (2005) for Spot-VEGETATION data and is modified and tested here for a medium resolution MODIS vegetation index data set (MOD13Q1). The underlying hypothesis for this work is that clouds and other atmospheric constituents result in a decrease of vegetation index (VI) values over land. Time series compositing methods can only work if at least one image within a compositing period delivers an uncontaminated value for every pixel, which is not achievable in regions of at least one significant yearly rain period. The reduction of residual noise in satellite time series aims at the enhancement of data quality and accuracy of land cover degradation estimates and thus may deliver improved input for environmental modelling.

2. STUDY AREA AND DATA

The state of Rio Grande do Norte in the North-Eastern part of Brazil comprises one of the areas that are most vulnerable to be affected by desertification in South America (UNCD 1977). The area is geographically located within the tropical biome with a semi-humid climate and low annual mean precipitation. The semi-humid climate and the occurring droughts in combination with an intensive anthropogenic activity of land use and land cover conversion are the main driving forces for degradation and desertification.

Within the state of Rio Grande do Norte, the study area is located between parallel 6°15' and 6°50' S and 37°00' and 37°30' W, in the micro region of Seridó (Figure 1). It covers an approximate area of 3.320 km², constituted by the municipal districts of Serra Negra, São João Sabugi, Jardim de Piranhas and Caicó (Petta et al. 2005).

Within the present study MODIS VI products (MOD13) have been evaluated. The MOD13 products are composites of daily bidirectional reflectances and are provided at different re-sampled spatial and temporal resolutions. In this study the MOD13Q1 product (16-day, 250 m), has been used, covering a time period of four subsequent years from December 2000 to November 2004. The MOD13 products include two VIs as well as the nadir-adjusted spectral reflectance channels that are needed to compute the VIs and ancillary information about geometry and data quality (see table 1). The normalized difference vegetation index (NDVI) is mainly provided for data continuity purposes with AVHRR data. The enhanced vegetation index (EVI) offers improved sensitivity to differences in vegetation from sparse to dense vegetation conditions and hence accounts for the problem of signal saturation for high biomass values (Huete et al. 1999).

Table 1: Overview of image layers in MODIS-VI products (MOD13) (modified from Huete et al. 1999)

| Layer	Variable	Pixel coding (bits)	Unit	Scaling
1	NDVI	16		NDVI=DN/10000
2	EVI	16		EVI=DN/10000
3	NDVI quality	16u		
4	EVI quality	16u		
5	Red	16	$Wm^{-2}*sr^{-1}*\mu m^{-1}$	R_r=DN/10000
6	NIR	16	$Wm^{-2}*sr^{-1}*\mu m^{-1}$	R_{nir}=DN/10000
7	Blue	16	$Wm^{-2}*sr^{-1}*\mu m^{-1}$	R_b=DN/10000
8	MIR	16	$Wm^{-2}*sr^{-1}*\mu m^{-1}$	R_{mir}=DN/10000
9	Average view zenith angle	16	Degrees	θ = DN/100
10	Average sun zenith angle	16	Degrees	θ = DN/100
11	Average relative azimuth angle	16	Degrees	φ = DN/10
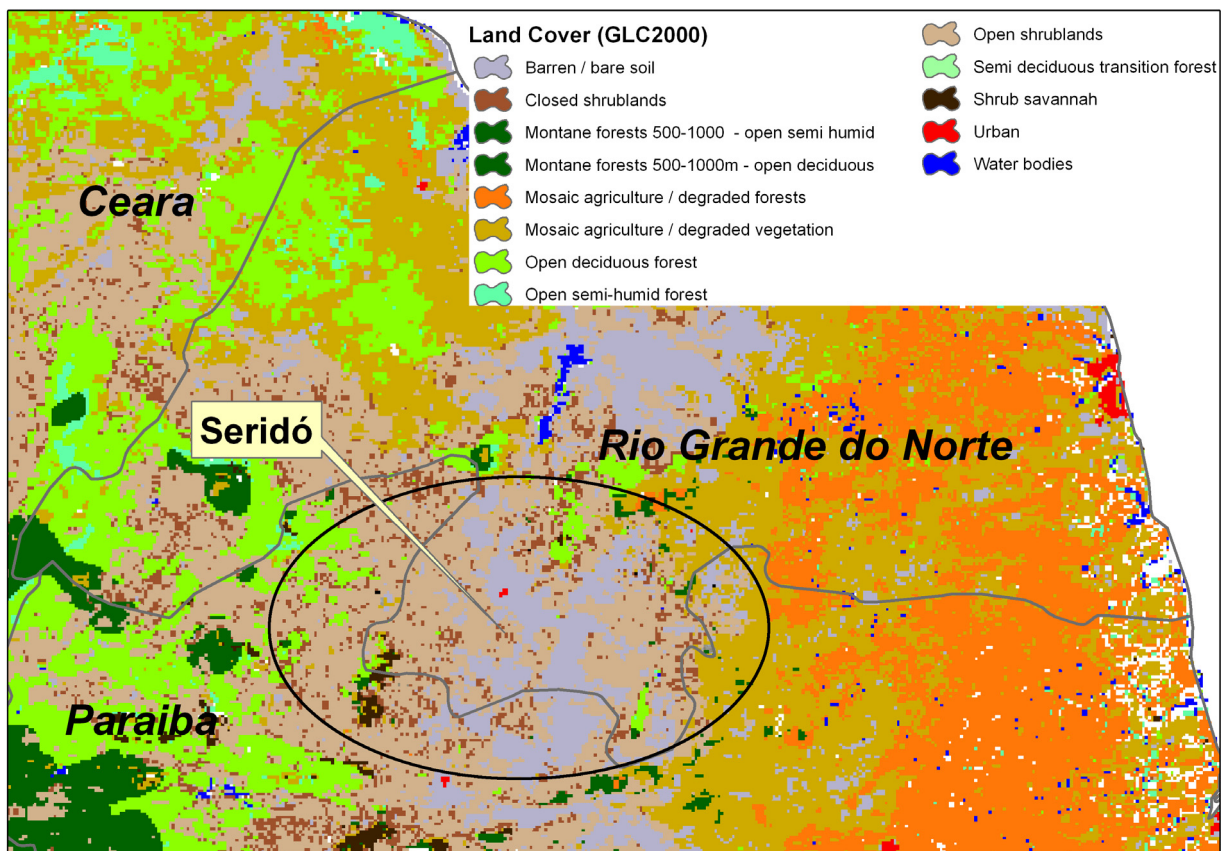
Figure 1: Overview of land cover types in the study area of Seridó (data source: Global Land Cover 2000 database. European Commission, Joint Research Centre, 2003, http://www-gem.jrc.it/glc2000)

3. METHODS

The development of a filtering algorithm for time series of satellite data for land surfaces has to account for the following requirements:
1. Missing or disturbed values have to be identified and have to be fitted to neighbouring values within a temporal trajectory
2. phenological (seasonal) phenomena have to be maintained
3. the algorithm can be applied to time series data at any spatial and temporal resolution

The problem of most smoothing algorithms operating in a simple moving window mode (e. g. mean, median, gauss filtering) is that they reduce noise and fill gaps in the frequency domain but at the same time modify undisturbed values that include important information. For this reason, a number of satellite time series filtering methods has been developed and tested such as threshold based methods (Viovy et al., 1992), Fourier-based fitting methods (Roerink et al., 2000) and asymmetric function fitting methods (Jonsson and Eklundh, 2002; Chen et al. 2005).
In this work, a polynomial function fitting method has been implemented and tested that is based on a weighted least-squares regression approach (Savitzky and Golay, 1964). Instead of calculating polynomial coefficients for every fitted interval at a position (x, y) within a gridded data set, the function uses weighted mean polynomial coefficients that minimize the mean square error between the polynomial function and the data values for the whole data set. The coefficients can be calculated for different polynomial degrees (2nd to 6th order) and filter intervals. The filter algorithm that uses these coefficients is summarized as follows:

$$Y_j^* = \frac{\sum_{i=-m}^{i=m} C_i \times Y_{j+i}}{N} \qquad (1)$$

and

$$N = 2 \times m + 1 \qquad (2)$$

where
Y_j^* = filtered value at position j
C_i = filter coefficient at position i
Y_{j+i} = data value at position j+i
m = filter interval
N = number of data points for calculation

The Savitzky-Golay filter constitutes the core part of a workflow that processes VI time series data according to the method that has been proposed by Chen et al. (2005) for Spot-VEGETATION NDVI time series.
In a first step, cloudy values within a time profile are determined based on the quality information of the MOD13Q1 dataset. These values are replaced by a linear interpolation of the nearest cloud free neighbouring values. In addition to the method of Chen et al. (2005), a pseudo-change threshold is applied to remove contaminated values that are not detected as cloudy. In the second step the Savitzky-Golay filter is applied to calculate a set of smoothed time series using different weighting options. The best result of this step is determined by

a least squares correlation and is used as the long term change trend. The third step calculates weights for the iterative filtering that are based on the long term change trend and the pre-processed time profile. The final filtering process then uses the initial curve, the change trend and the weights and calculates a new curve. In an iterative process, the result of the first fitting is used as input. The filtering process is terminated when the fitting effect index is at its minimum (see Chen et al. 2005 for a detailed explanation).

The edge effect of the filter algorithm has been accounted for by the assumption of a cyclic time series where the last points and the first points in a time series are regarded as neighbouring points. These points could be neglected in the current analysis in the way that the time series has been extended by 8 data points at the beginning and at the end of the data set, hence covering the period August 2000 to March 2005.

The filter method has been implemented into a graphical user interface environment using interactive data language (IDL). The complete module consists of a data import device tool, batch sub-setting, mosaicing, cloud masking and filtering in the temporal domain.

4. RESULTS AND DISCUSSION

Analysis of VI time series profiles has been performed using a sample of reference sites of the dominant land cover types within the investigation area as well as global statistics metrics. Additionally, annual spectral metrics have been evaluated for filtered and unfiltered data sets.

Analysis of annual VI metrics

The analysis of the annual performance of VI data has been carried out exploring mean EVI values that have been computed from annual time profiles for the years 2001 to 2004. A one-month-lagged annual mean has been calculated for the period December to November in order to account for the seasonality of vegetation dynamics that are mainly related to annual precipitation and drought periods. The statistical parameters that have been calculated from the annual mean images are summarized in Table 2.

In general, a shift of the EVI histograms (min, max, mean) to higher values can be observed. This is explained by the intrinsic structure of the filtering method that assumes that cloudy and atmospherically disturbed pixels generally show lower VI values and hence are adjusted to the neighbouring higher values within a temporal profile. The standard deviation and the Moran's I index (Moran, 1950), that have been investigated as measures of dissimilarity and homogeneity respectively, remain nearly unchanged. Hence, the filtering process does not affect the information content of the images. The mean residuals between filtered and non-filtered annual mean images range between 1.4 and 1.9 % but show substantial spatial variations within an annual data set (Figure 2). The residuals can be interpreted as a factor of spatial uncertainty and substantially affect change analysis at the annual level. The vegetation status change between the 2002 and 2003 annual mean images is shown in the lower part of Figure 2.

Table 2: Statistical results of the EVI annual mean images (based on MOD13Q1 data)

| Parameter	Data	annual mean 2001	annual mean 2002	annual mean 2003	annual mean 2004
min	EVI	-0.035	-0.015	-0.032	-0.014
	EVI filtered	-0.023	-0.007	-0.021	-0.102
max	EVI	0.529	0.607	0.589	0.507
	EVI filtered	0.588	0.630	0.625	0.548
mean	EVI	0.238	0.298	0.265	0.296
	EVI filtered	0.252	0.316	0.279	0.304
standard deviation	EVI	0.043	0.046	0.047	0.049
	EVI filtered	0.046	0.048	0.049	0.049
Moran's I	EVI	0.928	0.902	0.926	0.915
	EVI filtered	0.930	0.899	0.927	0.914
mean error (residuals)		0.014	0.019	0.015	0.015
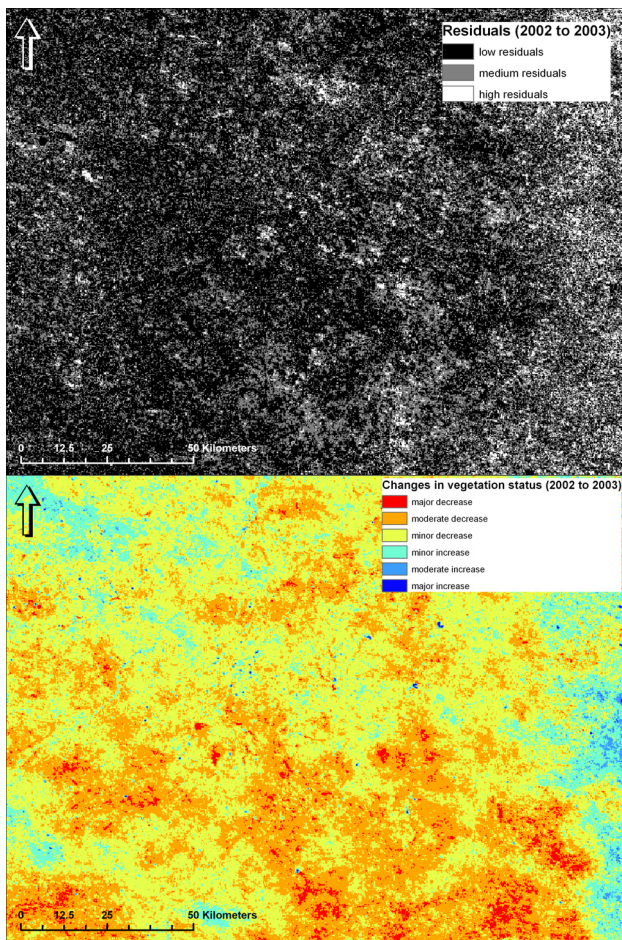
Figure 2: Residuals (filtered – non-filtered) of annual mean EVI images for 2002 and 2003 (a) and vegetation status change based on temporally smoothed annual mean EVI data (b)

Analysis of temporal profiles

The temporal profiles of EVI and NDVI data demonstrate the seasonality of vegetation growth within the study area. In general, the phenology of the vegetation follows the precipitation trend that comprises an extensive dry season from July to December and a wet season from January to June. Since ecosystems in semi-humid regions are well adapted to this

periodicity of seasons, temporal variations from long time mean climate such as extreme droughts are directly reflected in the temporal status of the vegetation. Against this background, the statistical performance of temporal VI profiles before and after time series smoothing has been analysed.

EVI and NDVI show high consistency regarding their temporal shape but the EVI comprises generally lower values (Figure 3). The analysis of the Moran's I index of spatial autocorrelation (Figure 4) reveals that the filtering well detects outliers in the annual trend whereas the EVI shows a higher variability in index values and generally lower index values for the dry season (lower VI values in Figure 3). This result documents the higher dynamic of the EVI and the improved sensitivity to vegetation status and changes in areas with sparse vegetation that is also reported in other studies (Fensholt and Sandholt, 2005).
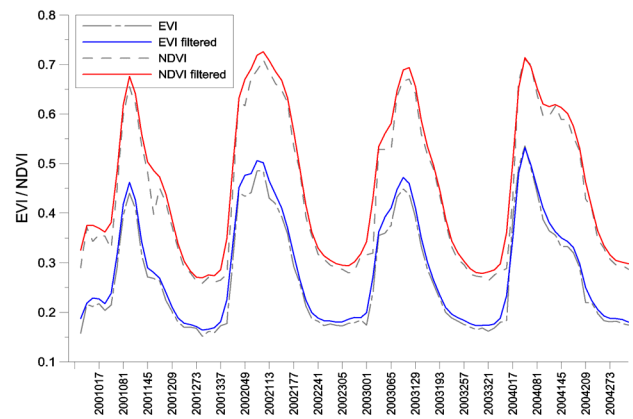

Figure 3: Time series profiles of EVI and NDVI raw and filtered data (image mean)
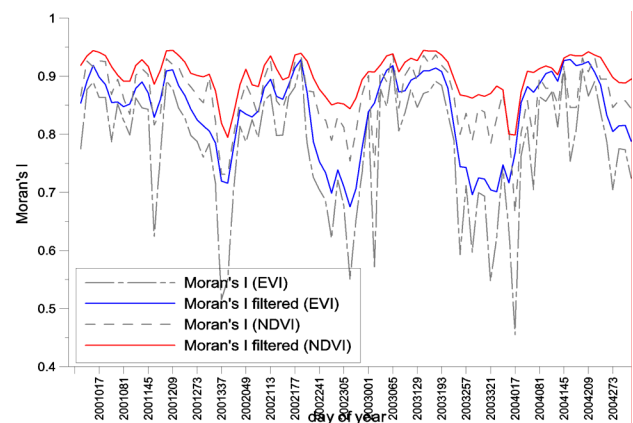

Figure 4: Time series profiles of Moran's I homogeneity index for EVI and NDVI raw and filtered data (image mean)

To test the spatially explicit effect of the filter method, a comparison of the time profiles for the predominating land cover types in the investigation are has been carried out. Profiles have been computed for enhanced and raw MODIS EVI and NDVI time series and for the classes bare land, open shrubland, closed shrubland and savannah shrubland (Figure 5). All profiles clearly document the positive effect of time series smoothing where outliers are fitted to the long-term curve trend and at the same time the shape and the positive local extremes of the curve are maintained. NDVI and EVI show similar trends
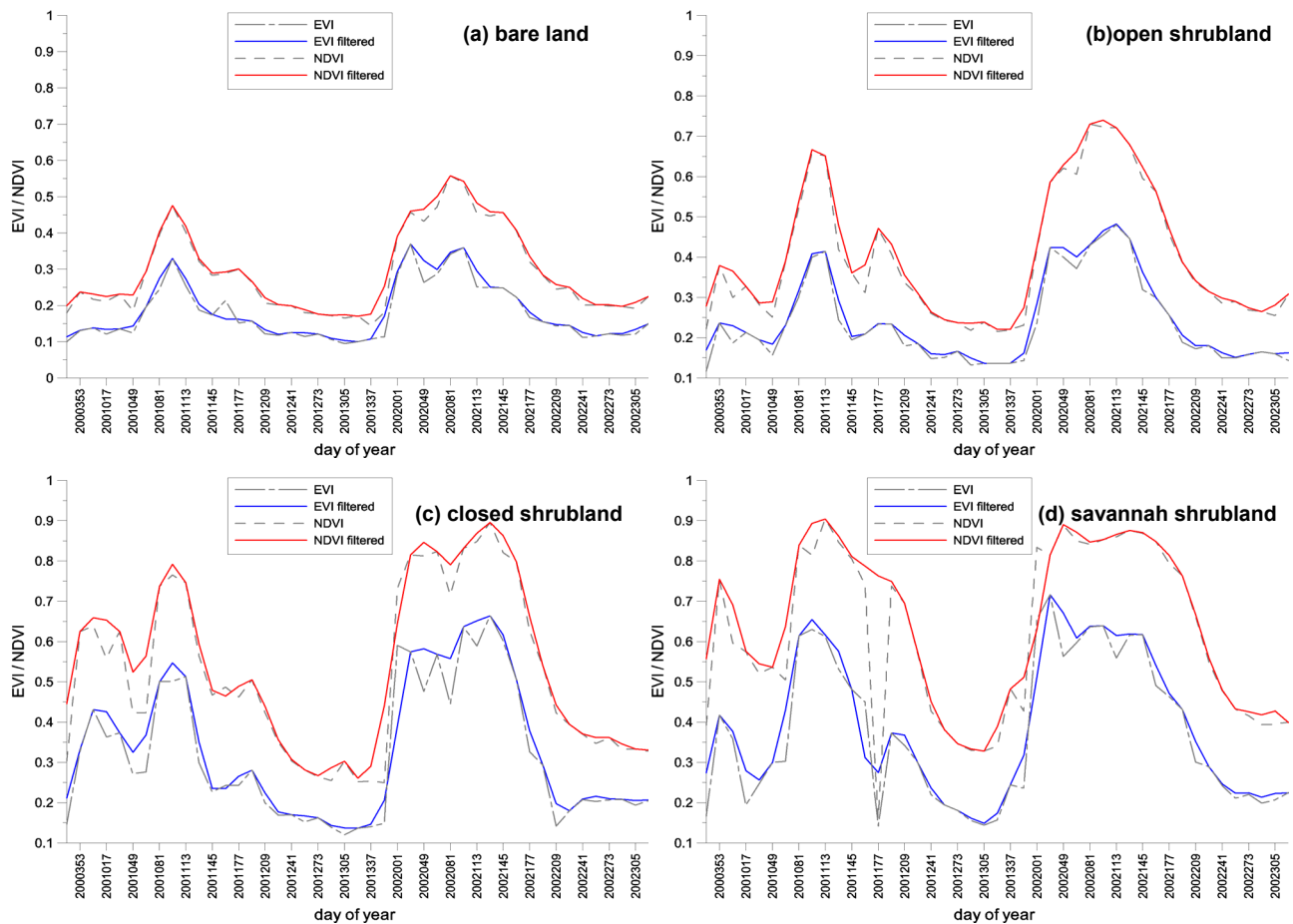
Figure 5: MODIS NDVI and EVI time series for selected land cover types in the Seridó region, Rio Grande do Norte, Brasil. The dashed lines show the 10-day composites of the MOD13Q1 product, the solid lines indicate the smoothed time series using the iterative Savitzky-Golay technique.

and dynamic ranges. Only the savannah shrubland plot indicates a higher resistance of the EVI to intra-seasonal fluctuations that can be seen from the short-term decrease of EVI values in the late growing season of 2001. Additionally, the NDVI nearly saturates in the savannah shrubland at the peak of the growing season. Generally, the results of the statistical analysis point out that the filtering method is independent of the VI that is processed. Both, the EVI and NDVI time series are substantially improved by the filtering method.

5. CONCLUSIONS

The hypothesis that has been investigated in the present study was, that time series smoothing considerably effects the spatial and temporal characteristics of pre-processed MODIS VI time series (MOD13Q1).

It has been shown that filtering in the temporal domain has only minor effects on the histogram statistics of annual mean images of EVI and NDVI. On the other hand, the analysis of the spatial patterns of such annual metrics reveals that temporal filtering noticeably influences the spatial variability of vegetation index values. Hence, the smoothing of VI temporal profiles contributes to enhanced LUCC analysis, even if such analysis is based on post-filtered temporal metric images.

The function analysis of the temporal profiles of EVI and NDVI data pointed to the superior usability of EVI time series for the detection of seasonal and inter-annual phenomena. The EVI better detects and preserves local temporal features (e.g. local minima) which may be correlated to phenological changes. This is of special relevance, if VI time series data are used in temporal correlation analysis with meteorological parameters in ecological studies (e.g. VI-rainfall correlation). Considering the investigation area in North-eastern Brazil, the enhanced MODIS VI time series data deliver essential input for the analysis of change and the separation of climate related change from human induced land degradation which is known as desertification.

REFERENCES

Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., Eklundh, L., 2004. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. *Remote Sensing of Environment*, 91, pp. 332-344.

Defries R. S.., Hansen, M. C., Townshend, J. R. G., Janetos, A. C., Loveland, T. R., 2000. A new global 1-km dataset of percentage tree cover derived from remote sensing. *Global Change Biology*, 6 (2), pp. 247-254.

Fensholt, R., Sandholt, I., 2005. Evaluation of MODIS and NOAA AVHRR vegetation indices with in situ measurements in a semi-arid environment. *International Journal of Remote Sensing*, 26 (12), pp. 2561-2594.

Griffith, D. A., 1987. Spatial Autocorrelation – A Primer. Association of American Geographers, Washington D.C.

Hansen, M. C., Defries, R. S., Townshend J. R. G., Sohlberg, R., 2000. Global land cover classification at 1km spatial resolution using a classification tree approach. *International Journal of Remote Sensing*, 21 (6-7), pp. 1331-1364.

Huete, A., Justice, C., van Leeuwen, W., 1999. MODIS vegetation index (MOD 13) algorithm theoretical basis document (ATBD), Version 3.0.

Jonsson, P., & Eklundh, L., 2002. Seasonality extraction by function fitting to time-series of satellite sensor data. *IEEE Transactions on Geoscience and Remote Sensing*, 40(8), pp. 1824– 1832.

Moran, P.A.P., 1950. Notes on continuous stochastic phenomena. *Biometrika* 37:17-23.

Petta, R. A., Ohara, T., Medeiros, C. N., 2005. Environmental GIS Database for Desertification Studies in the Brazilian North-Eastern Areas. In: *Remote Sensing & GIS for Environmental Studies: Applications in Geography* (Erasmi, S., Cyffka, B., Kappas, M. (eds.)), Göttinger Geographische Abhandlungen, 113, pp. 287-295.

Roerink, G. J., Menenti, M., Verhoef, W., 2000. Reconstructing cloudfree NDVI composites using Fourier analysis of time series. *International Journal of Remote Sensing*, 21(9), pp. 1911 –1917.

Savitzky, A., Golay, M. J. E., 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 36 (8), pp. 1627-1639.

Strugnell, N. C., Lucht, W., Schaaf, C., 2001. A global albedo data set derived from AVHRR data for use in climate simulations. *Geophysical Research Letters*, 28 (1), pp. 191-194.

Townshend, J. R. G., 1994. Global data sets for land applications from the Advanced Very High-Resolution Radiometer – an introduction. *International Journal of Remote Sensing*, 15, pp. 3319 1994.

UN Secretariat of the Conference on Desertification (UNCD), 1977. Desertification: An overview, In: *Desertification: Its Causes and Consequences.* New York, Pergamon Press.

Viovy, N., Arino, O., Belward, A. S., 1992. The Best Index Slope Extraction (BISE): A Method for reducing noise in NDVI time-series. *International Journal of Remote Sensing*, 13 (8), pp. 1585-1590.