Babatunde A. Osunmadewa, Christine Wessollek, Pierre Karrasch

Linear and segmented linear trend detection for vegetation cover using GIMMS normalized difference vegetation index data in semiarid regions of Nigeria

Erstveröffentlichung in / First published in:

Journal of Applied Remote Sensing. 2015 (9), S. 096029-1 – 096029-23 *[Zugriff am: 02.05.2019]*. SPIE Digital Library. ISSN 1931-3195.

DOI: https://doi.org/10.1117/1.JRS.9.096029

Journal of
Applied Remote Sensing

RemoteSensing.SPIEDigitalLibrary.org

Linear and segmented linear trend detection for vegetation cover using GIMMS normalized difference vegetation index data in semiarid regions of Nigeria

Babatunde A. Osunmadewa
Christine Wessollek
Pierre Karrasch

SPIE.

Linear and segmented linear trend detection for vegetation cover using GIMMS normalized difference vegetation index data in semiarid regions of Nigeria

Babatunde A. Osunmadewa,[a,*] Christine Wessollek,[a] and Pierre Karrasch[b]
[a]Technische Universität Dresden, Institute of Photogrammetry and Remote Sensing, Helmholtzstraße 10, Dresden 01069, Germany
[b]Technische Universität Dresden, Professorship of Geoinformation Systems, Helmholtzstraße 10, Dresden 01069, Germany

Abstract. Quantitative analysis of trends in vegetation cover, especially in Kogi state, Nigeria, where agriculture plays a major role in the region's economy, is very important for detecting long-term changes in the phenological behavior of vegetation over time. This study employs the use of normalized difference vegetation index (NDVI) [global inventory modeling and mapping studies 3g (GIMMS)] data from 1983 to 2011 with detailed methodological and statistical approach for analyzing trends within the NDVI time series for four selected locations in Kogi state. Based on the results of a comprehensive study of seasonalities in the time series, the original signals are decomposed. Different linear regression models are applied and compared. In order to detect structural changes over time a detailed breakpoint analysis is performed. The quality of linear modeling is evaluated by means of statistical analyses of the residuals. Standard deviations of the regressions are between 0.015 and 0.021 with R^2 of 0.22–0.64. Segmented linear regression modeling is performed for improvement and a decreasing standard deviation of 33%–40% (0.01–0.013) and R^2 up to 0.82 are obtained. The approach used in this study demonstrates the added value of long-term time series analyses of vegetation cover for the assessment of agricultural and rural development in the Guinea savannah region of Kogi state, Nigeria. © *2015 Society of Photo-Optical Instrumentation Engineers (SPIE)* [DOI: 10.1117/1.JRS.9.096029]

Keywords: trend analysis; advanced very high-resolution radiometer; GIMMS 3g normalized difference vegetation index; time series analysis; break point analysis; Kogi state Nigeria.

Paper 15061P received Jan. 23, 2015; accepted for publication Jun. 8, 2015; published online Jul. 8, 2015.

1 Introduction

Vegetation as an intrinsic part of the ecosystem, covering large areas of the earth's surface, plays enormous role in regulating various biogeochemical cycles; however, alteration of vegetation surface can affect its ecological cycle which can in turn influence (accelerate) global climate change.[1] However, proper analysis of vegetation productivity is important in order to determine the rate of vegetation degradation over time. Several studies of de Jong et al.[2] use advanced very high-resolution radiometer (AVHRR) normalized difference vegetation index (NDVI) data to analyze monotonic greening and browning (global), whereas other studies use NDVI and climatic data to examine vegetation trends.[3–5] NDVI generally provides a measure of the amount and vigor of vegetation on the land surface. The magnitude of NDVI, however, is related to the level of photosynthetic activity in the observed vegetation when seasonally integrated.[6]

In general, higher values of NDVI indicate greater vigor and amounts of vegetation. NDVI is generally calculated as a normalized ratio of red and near-infrared (NIR),[7] which is used to delineate vegetation productivity and can thus be used for the assessment of trends in vegetation.

*Address all correspondence to: Babatunde A. Osunmadewa, E-mail: Babatunde_Adeniyi.Osunmadewa@mailbox.tu-dresden.de

1931-3195/2015/$25.00 © 2015 SPIE

$$\text{NDVI} = \frac{\text{NIR} - R}{\text{NIR} + R}, \tag{1}$$

where NDVI is based on the principle that actively growing plants absorb radiation at the visible region of the electromagnetic spectrum due to chlorophyll absorption and reflect radiation at the NIR.[8]

In Nigeria, vast rural majority live in the rural communities where their major source of livelihood is completely dependent on the environment thereby altering the existing vegetation at the expense of food production in order to meet the demand of the growing population.[8] The vegetation of the Guinea savannah region of Nigeria has witnessed a tremendous change over the past decades due to anthropogenic pressure on the landscape; this has led to the alteration of vegetation composition and its density coupled with climatic variability (cycle) such as rainfall and temperature, which has a feedback on vegetation greenness.

Therefore, changes in the spatio-temporal pattern of vegetation greenness are important in the Guinea savannah of Nigeria, where anthropogenic activities induced by man have alter the natural vegetation over the last few decades. However, an assessment of vegetation trends using Earth observation (EO) data is essential as this gives an insight to the vegetation status of the study area since much has not been mentioned in the literature.

Time series trend analysis of NDVI have been using various instruments, which had yielded reliable results by several studies; such instruments include AVHRR, moderate resolution imaging spectrometer (MODIS), and satellite pour l'observation de la terre, amongst others.[8] Results of studies such like that from Usman et al.[9] using time series satellite data from National Oceanic and Atmospheric Administration (NOAA) (AVHRR and MODIS) show that there are both positive and negative long-term trends in vegetation cover of the northern part of Nigeria. Such result renders monitoring of changes in vegetation trend dynamics very important in other parts of Nigeria.

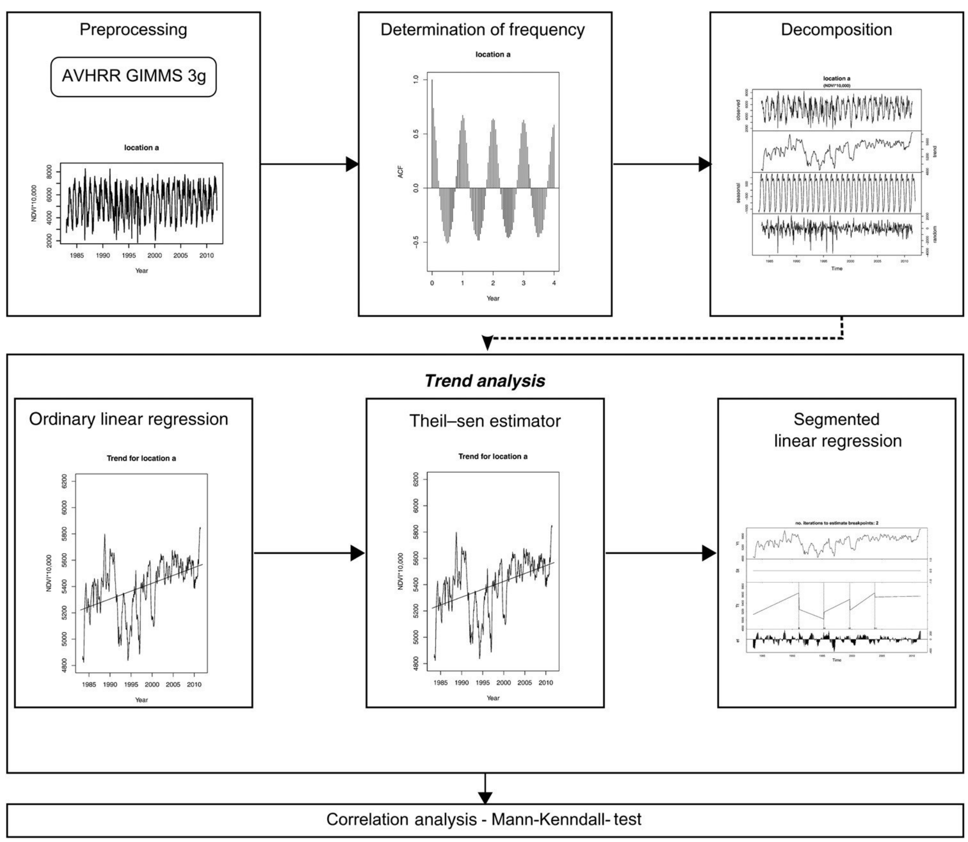


Fig. 1 Workflow of identification of long-term trends in vegetation dynamics in the Guinea savannah region of Nigeria.

It is obvious that a comprehensive knowledge of vegetation dynamics is mandatory in many fields of research. The basis of this knowledge can be different. One important source of data is long-term time series such as NDVI 3g data, available from the AVHHR satellite and distributed by the NOAA. Data of this nature promise much additional information. However, the basic requirement is the ability to extract the relevant information from the original time series signal regarding a particular research question.

Therefore, the objective of this study is to give recommendations or a short guide for land use change analyses using different methods of NDVI time series analyses. A 29-year-bimonthly time series (1983–2011) of global inventory modelling and mapping studies (GIMMS) AVHRR 3g NDVI is analyzed to answer the question whether a trend in vegetation dynamic is observable in the chosen study area of Kogi state in the Guinea savannah region of Nigeria.

The previously preprocessed NDVI data[10] are the origin of the analyses workflow displayed in Fig. 1.

The determination of the frequency of seasonality is the initial point for the decomposition of the time series into a trend, seasonal, and random component. For the trend analysis itself, different methods are tested. Besides the actual results, further test procedures are carried out for the assessment of quality of trend modeling. Based on this and other additional statistical parameters (e.g., standard deviation, coefficient of determination), the added value of the test procedures regarding the selection of a particular trend model becomes obvious. Additionally, correlation analysis for the different methods of trend modeling is performed. All analyses are exemplarily performed for four carefully selected locations (8×8 km) in Kogi state (Sec. 2.2). However, the presented workflow is also transferable to area-wide observations of vegetation dynamics.

At first, the data used in that study as well as the study area is introduced, followed by a detailed description of the used methodology in Sec. 3. The results of the analyses are shown in Sec. 4 and finally discussed in Sec. 5.

2 Data and Study Area

2.1 *GIMMS AVHHR NDVI 3g Data*

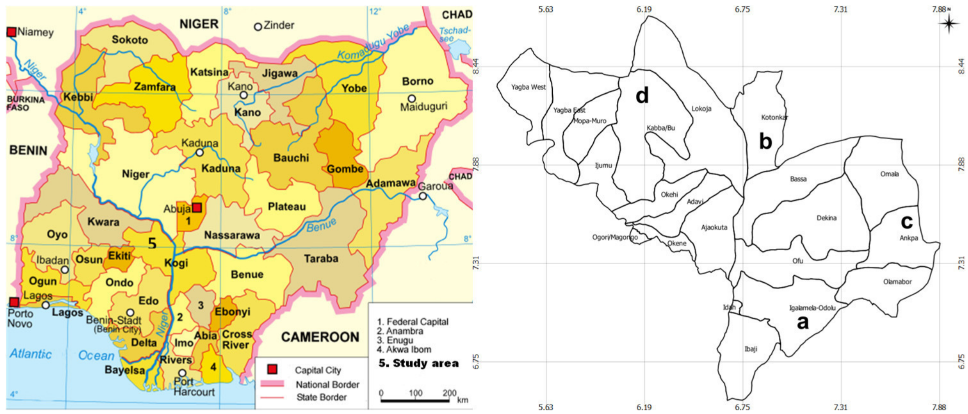
The new GIMMS AVHRR NDVI 3g with a spatial resolution of 8 km is used in this study to analyze vegetation trends. However, the GIMMS NDVI 3g is the reconstructed extension of the previous NDVI version and it comprises NDVI measurements derived from the seven NOAA instruments.[11] Because of the coarse resolution and inaccuracy of AVHRR, which might arise as a result of orbital drifts, it is therefore necessary to adopt a technique which overcomes the problems associated with using the AVHRR NDVI dataset for time series analysis.[12] In comparison with the older version of the AVHRR NDVI, the new NDVI 3g has been processed using adaptive empirical mode decomposition in order to remove artifacts such as solar zenith angle, orbital drifts, and/or volcanic aerosols which might cause systematic trends in NDVI time series (see Pinzon and Tucker[10] and Tucker et al.[12] for more details). Furthermore, various corrections such as stratospheric volcanic eruption for El Chichon and Mt. Pinatubo were applied to the new NDVI 3g; atmospheric correction (cloud cover scrutiny) was achieved using the maximum composite technique to improve its quality and minimize atmospheric and radiative geometry effects.[12] The spatial resolution of the GIMMS data is $8\ \text{km} \times 8\ \text{km}$ with a 15-day temporal resolution.[12]

Hence, the use of long-term NDVI 3g data is suitable not only at the global scale, but also at the regional and even local scale such as Kogi state, Nigeria, where technical resources are limited since analysis of vegetation trend over the last three decades is made possible.

The bimonthly NDVI dataset (GIMMS NDVI 3g) from 1983 through 2011 (29 years) used in this study was reprojected from the global projection to Universal transverse mercator and subset images of the study area were clipped out of the global data using the provincial boundary of Kogi state.

2.2 *Description of Study Area*

Kogi state (Nigeria) lies between latitude 6.6° and 8.7°N longitude 5.3° and 7.8°E (Fig. 2). It is bounded to the south by Edo state, to the north by Federal Capital Territory (Abuja), Niger,



(a) (b)

Fig. 2 (a) Map of Nigeria showing the study area designated as number 5 in the map BanyanTree; (b) administrative map of Kogi state showing the four selected locations [(a–d); data were extracted from the GADM database,[18] version 2.0, December 2011].

Nassarawa, to the east by Benue and it shares border with Ondo, Ekiti, and Kwara states.[13] The state is said to cover a land area of about 27,747 km^2.[14] The population of the study area (Kogi state) is mostly rural and had tremendously increased over the last decades from 661,515 (1952), 1,280,143 (1963),[15] 2,147,756 (1991), to 3,314,043 inhabitants according to the 2006 population census in Nigeria.[14] About 70% of the population lived in rural areas.[16] Agriculture is one of the major sources of economy in the state in which about 75% of the population are engaged in subsistence farming.[17] The climate of Kogi state is mainly tropical and it is characterized by two main seasons (dry and raining/wet season). The raining season starts towards the end of March and ends at the end of October with total annual rainfall which ranges from 804.5 mm to 1767.1 mm.[17] However, the start and end of rainfall considerably varied in some parts of the state which may influence vegetation productivity temporally as well as spatially. Another important attribute of the study area is the harmattan wind which is experienced during the dry months of December–January. The annual mean minimum temperature is 28°C while the annual mean maximum temperature is 35°C and relative humidity is considerably high.[16]

Kogi state, Nigeria, is found in the Guinea savannah region with the presence of gallery forest along water courses. The vegetation composition of the state is typically that of rain forest in the southern part and woody-derived, and Guinea savannah in the northern part.[16] The grass, weed, shrub, and leguminous component of the study region are very important to the grazing economy of the country.[19] The Guinea savannah provides dry-season grazing for long-distance Fulani nomads, but vegetation degradation through burning, over-cultivation, and fuel wood extraction and overgrazing had made studies on the vegetation trend dynamics in the study area very important.

Kogi state, Nigeria, is mostly referred to as the confluence state because the confluence of river Niger and Benue is at its capital. The region is rich in alluvial fertile soil which is very good for crop production.[17] Due to human perturbation, most of the natural vegetation is lost. For this study, four locations were selected for the analysis of changes in vegetation greenness in Kogi state, the Guinea savannah region of Nigeria. These four locations represent areas where vegetation transition had taken place over the last three decades.

Location [b] of the study area is composed of mixed leguminous wooded savannah, riparian forest, fresh water swamp, dominantly trees/woodlands, shrubs, and grasses in the late 1970s, but what were observed in this location today are rain fed agriculture and floodplain agriculture due to population upsurge and high level of poverty.

The vegetation type of location [a] particularly down south of the study area is composed of a mixture of vegetation of wooded savannah, rain, and mangrove forest, whereas locations [c] and [d], however, are composed of dominantly trees, wooded savannah, and shrubs with a subdominant grass. *Daniella Oliveri* and oil palm can also be found in the low-land forest and riparian forests are particularly dominated along the river valleys. Reconnaissance survey and observations during data collection show that most of the vegetation in the selected location had transited (altered by land-use changes) into both extensive and intensive (grazing, minor row crops) small holder rain-fed agriculture and increased urbanization; forested lands had also transited into shrub lands and in some cases grasses. In general, the natural vegetation of Kogi state had been greatly altered by land-use transformation due to anthropogenic activities which has influenced the vegetation types. Hence, the selection of the four study areas is made according to those criteria and based on very extensive local knowledge.

3 Methodology

3.1 *Methods of Determining the Frequency of Seasonality*

The existence of serial correlation is a major factor which complicates statistical inference of time series analysis.[20] In order to statistically test the general trends, autocorrelation was computed using the autocorrelation function (ACF) estimation.[21]

Partial autocorrelation was calculated as a further analysis to determine the existence of periodicity in the datasets. The reason for this is because of the slow decaying correlations as they occur in the ACF. Thus, another kind of auto correlation function gives consideration to intermediate values that are comparably found as an echo in the values of the ACF. This effect can be avoided with the partial autocorrelation function (PACF).[22]

Besides the ACF and PACF, other methods of determining frequencies in time series are available. For example, the Fourier transformation or more complex methods such as the wavelet transformation can confirm the results of the ACF/PACF.

Although it is obvious that a frequency of 1 year reflects the natural rhythm of vegetation cover over a year, it is therefore necessary to verify this assumption. The detailed analyses of the frequency of seasonality also allow identification of additional periodical phenomena. For the following process of decomposition of time series, it is of utmost significance to have detailed information about all possible periodical structures in the entire signal.

3.2 *Method of Time Series Decomposition*

Time series of environmental parameters primarily contain seasonality. In order to analyze the interannual trend, it is necessary to remove this seasonal component from the input data. In the first step, the time series (Y) must be separated into a trend component (T), a season component (S) and an additional component which contains the random noise (e) of the input signal.

The relationship between these three components can be described by an additive or multiplicative model. The additive model is described by

$$Y_t = T_t + S_t + e_t, \tag{2}$$

where t is time, T is trend, S represents the seasonal component of the signal, and e is the error or irregular component. The decomposition for the time series of the study area (Kogi state) was achieved using the "decompose" function implemented in the open-source software called R-programming language,[23] which was also used for the entire analyses presented in this paper. A moving average method was used to determine the trend:

$$\overset{t_{\max}-\frac{n}{2}}{\underset{t_{\min}+\frac{n}{2}}{T_t}} = \frac{\sum_{i=t-\frac{n}{2}}^{i=t+\frac{n}{2}} Y_i}{n+1}. \tag{3}$$

For the moving average, a window-size of frequency of seasonality $(n) + 1$ was used. After removing trend from the original time series, the seasonal figure (S_t) was computed by averaging, for each time unit, over all periods and centered in a second step. For one period (year), the seasonal component S_i is

$$\underset{i=1}{\overset{n}{S_i}} = \frac{\sum_{j=0}^{m-1} Y_{j \cdot n+i}}{m}, \tag{4}$$

where m is the number of years. If both, trend and seasonal figure are removed from the input data, the error component (random) remains.

$$e_t = Y_t - (T_t + S_t). \tag{5}$$

3.3 *Statistical Methods for Trend Analyses*

Several tests are available for the detection and estimation of trends. An important objective of vegetation dynamic studies is to detect changes or trends in NDVI over time. In this study, different statistical approaches were applied to assess long-term spatio-temporal change in vegetation greenness of the 29 years GIMMS NDVI dataset for Kogi state, Nigeria.

3.3.1 *Linear regression modeling: least-square method*

Temporal trends were examined by an ordinary least-square (OLS) linear regression model where time is the independent variable and the trend component of the NDVI signal is the dependent variable. The regression lines were calculated on the basis of the least-squares method for the values of the pixels selected in the study locations and their significances were tested at the $\alpha = 0.05$ level. OLS is a linear regression method which can be used for modeling time series of environmental variables such as vegetation, rainfall, and temperature. Equation (6) illustrates the OLS model as

$$\mathrm{NDVI} = a + b \cdot t, \tag{6}$$

where NDVI is the dependent variable which is a linear function of time (t).[24] The parameter b (slope) can be interpreted as the linear modeled trend of NDVI.

3.3.2 *Theil–Sen median slope*

Theil–Sen median trend estimator, a robust nonparametric statistical operator which is resistant to the presence of outliers, was used to analyze trends in NDVI data. This trend analysis is suitable for the assessment of the rate of change in time series analysis and the result is often similar to that of OLS regression when used for long time series analysis. The Theil–Sen median slope was determined by calculating the slope between all pairwise combinations and then assessing the median over time. The breakdown bound for the median is approximately 29%, which means the trends expressed in the image must have persisted for more than 29% of the length of the series (in time steps). Hence, the Theil–Sen's method for the analysis was used because of its ability to estimate trend slope even when data are missing. The slope of the Theil–Sen line will be significantly different from zero when Kendall's tau is significantly different from zero. Every N' data pairs ($i' > i$) were calculated by

$$Q = \frac{X_{i'} - X_i}{i' - i}, \tag{7}$$

where $X_{i'}$ and X_i are the data values at times (or during time period) i' and i, respectively. The median of these N' values of Q is represented as Sen's estimator of slope.[25,26]

3.3.3 *Mann-Kendall trend test*

Another approach used in this study was the Mann–Kendall's (MK's) monotonic test on trends. MK's test is a nonparametric test which is less sensitive to outliers, thus normal distribution of data is not required to carry out analysis of time series data using this test.[26] However, MK test was a preferred choice for this study because it takes problems such as autocorrelation into account. The test measures the degree to which a trend is consistently increasing or decreasing (-1 to $+1$).[27] A value of $+1$ thus indicates an increasing trend and a value of -1 means a decreasing trend, whereas a value of 0 is an indication of no trend. According to this test, the null hypothesis H_0 assumes that there is no trend and this is tested against the alternative hypothesis H_A, which assumes that there is a trend and was tested two-tailed at $\alpha = 0.05$.

The computational procedure for the MK test considers the time series of n data points and X_j and X_k as two subsets of data where $j = 1,2,3,\ldots,n-1$ and $k = i+1, i+2, i+3,\ldots,n$. The data values are evaluated as an ordered time series and each data value is compared with all subsequent data values (pair wise combination).[28] The result of all increasing and decreasing values gives the final value of S. The MK S statistic is computed as follows:[25]

$$\begin{aligned} \mathrm{sgn}(x_j - x_k) &= 1 \text{ if } x_j - x_k > 0 \\ &= 0 \text{ if } x_j - x_k = 0 \\ &= -1 \text{ if } x_j - x_k < 0, \end{aligned} \tag{8}$$

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \mathrm{sgn}(x_j - x_k), \tag{9}$$

where X_j and X_i are the annual values in years j and i, $j > i$, respectively.

The seasonal MK (sMK) test is also a nonparametric test which is appropriate when there is seasonality in time series data. This method was used for the time series analysis in this paper in order to remove seasonal cycle which might affect the significance of our result. The method involves computation of the MK test statistic S and its variance, VAR(S), separately for each month (season) with data collected over the year.[25]

3.3.4 *Segmented linear regression modeling*

Breaks for additive seasonal and trend (BFAST) is an approach which combines the iterative decomposition of a time series into seasonal, trend, and remainder with the detection of changes within the series. Breakpoints can be detected in the seasonal component as well as in the trend. As the decomposition was already proven, the detection of breakpoints was only performed for the trend components. In this case, the seasonal component is set to 0 and the input series is separated in trend and remainder. Within BFAST, the test for significant changes in the times series is done by the ordinary least-squares residual-based moving sum (OLS-MOSUM).[29] If changes are detected, the number and position of breakpoints are estimated. Additionally, robust regression based on M-estimation[21] was used to estimate the segment-specific slopes and intercepts. For further details about the used methods, the reader is referred to Verbesselt et. al,[30,31] Bai and Perron,[32] and Zeileis et. al.[33] For the detection of breakpoints, BFAST uses the R-package strucchange.[34]

For the results presented in Sec. 4.4.3, some boundary conditions define the character of regression lines. The most important is certainly the parameter h which represents the minimal segment size between two potential breakpoints which is expressed as the fraction of all available observations in the time series.[30,31] By means of several test procedures based on the use of different h-values, an optimal fraction of 0.15 is chosen, which means a minimum segment size of 4 years and 2.5 months. One criterion for this decision is the spread of the confidence interval of the breakpoints, which arises with decreasing h-value. The second one is the standard deviation of the residuals of the resulting segmented linear regression. These standard deviations arise with increasing h-value.

3.4 *Methods of Inferential Statistic*

3.4.1 *Stationarity Tests*

For the assessment of different mathematical methods (e.g., residuals of linear modeLing and random component of the decomposition), it Is necessary to perform a detailed residual analysis.

These residuals should meet two important conditions. At first, the mean value of residuals should be constant over time and the expectation value is 0. Second, the variance should be constant over the whole time series. Both conditions can be tested by means of tests of stationarity. In this study, three different tests for stationarity were conducted. The first two tests, i.e., the augmented Dickey–Fuller test (ADF) and Philipps–Perron test (PP), were used for testing unit root in time series. If there is a unit root, the time series is not stationary. Hence, both tests are tests on nonstationarity. For these, it is necessary to reject the null hypothesis to confirm the stationarity of the residuals or the random component of the decomposition. In addition to the ADF test, a further unit root test was performed. This will be necessary because the ADF test tends to reject the null hypothesis too often. Therefore, it could be happen that the tested random component of the time series will be deemed for nonstationary even through it is not. This limitation of the ADF test can be considered by the PP test.[35]

In contrast, the third test of Kwiatkowski–Phillips–Schmidt–Shin (KPSS) changed the null hypothesis and alternative hypothesis. Hence, the KPSS test is a direct test on stationarity.[35] Furthermore, the test is carried out with two different options. The first one allows to test the residuals and random component of the time series on stationarity concerning the remaining trend after decomposition (trend stationarity). The second type of KPSS test can be performed in a somewhat more strict manner. This option assumes that no further trend signal is in the random component. This method is used for the level stationarity test.[36]

3.4.2 *Tests on normal distribution*

Many methods of data analyses are based on the assumption of normal distributed measurements. Regarding the chosen method of correlation analyses, it is mandatory to test the distribution of the used NDVI time series. These results are the basis for the decision of the chosen correlation analysis method.

Therefore, two established tests are performed to confirm or reject the hypothesis of normal distribution (Shapiro–Wilks test, Anderson–Darling test). The test value W of the Shapiro–Wilks test is the ratio between the variance of an assumed normal distribution and the variance of the real distribution of measurements.[37] The null hypothesis of the Shapiro–Wilks test is that the measurements are normally distributed.

Whereas the Shapiro–Wilks test uses variances, the Anderson–Darling test is based on a comparison of frequency distributions. The test value A describes the difference between the transformed sample measurements and the normal distribution function.[38]

4 Results

4.1 *Visual Analysis of the Data*

Trends in vegetation parameters for the Guinea savannah region of Kogi state, Nigeria, from 1983 through 2011 are illustrated in Fig. 3. It can be easily seen that the AVHRR NDVI time series data for the four locations (study area) are dominated by the seasonal components of the signals. The observed signals are as a result of normal periodic changes in seasons over time. However, in Fig. 3, it is also obvious that long-term changes in signals are present in the NDVI data used for this study, especially in the data for location [c] (Fig. 3).

4.2 *Determination of NDVI Seasonality*

Further interpretations of the data can only be carried out by disaggregating the time series data into components of trend, seasonality, and noise (random) explained in Sec. 3.2. Therefore, it is necessary to know the frequency of the seasonality. For this reason, all available time series were subjected to autocorrelation analysis. Although it can be assumed that signals at all locations should have a strong period of 1 year, these analyses were done to verify that.

The variation of the autocorrelation functions in Fig. 4 confirmed the suspected course of the autocorrelation function in all study locations. Strong correlations are mainly observed by shifts in integer annual periods. Noticeable, however, is the autocorrelation result of location [c] where

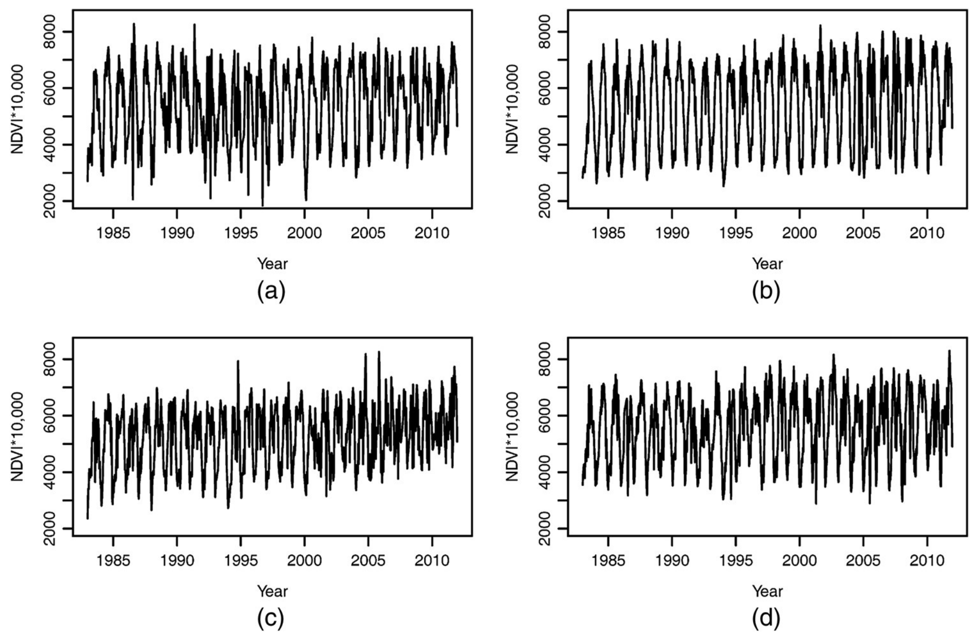


Fig. 3 Original time series of GIMMS AVHRR NDVI 3g data (1983–2011). (a) Location a, (b) location b, (c) location c, (d) location d.

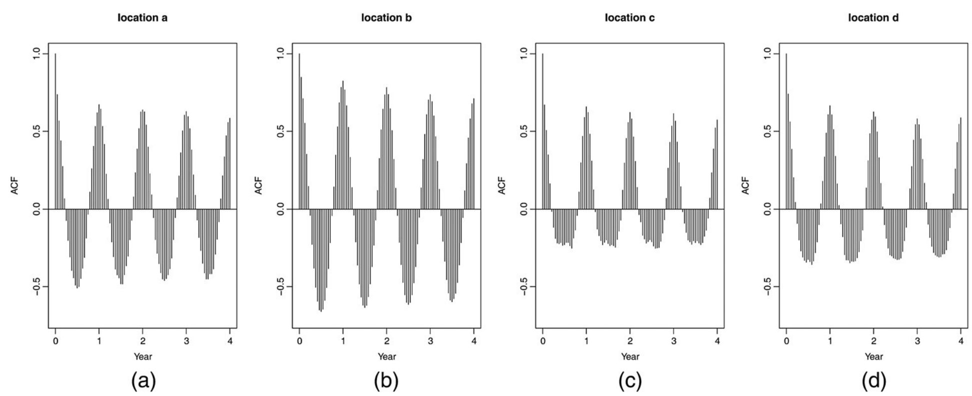


Fig. 4 Autocorrelation function (ACF) for the time series GIMMS AVHRR NDVI 3g data [1983–2011, (a)–(d) locations a–d].

strong negative correlations are significantly lower in the first periods and the decrease of its extremes is very low.

As a further analysis to determine the periodicity, partial autocorrelation was calculated. This is due to the slow abating correlations as they occur in the ACF. The reason for this is the consideration of the intermediate values for lags > 0, which have impact on the correlation values of the ACF. This effect can be avoided with the PACF.

It is clear that the correlations are significantly attenuated and faster than for the simple autocorrelation. It is evident from the result that a dominant period of 1 year is present. The initially incurred assumption is confirmed by the result. For this reason, in all following analyses, the (frequency of) seasonality was set to 24 (bimonthly). However, it cannot be ruled out that there is still another seasonality hidden in the signal (Fig. 5). As already mentioned in the methodology section, all results are also confirmed using a fast Fourier analyses.

4.3 *Decomposition of NDVI Time Series*

At first, the NDVI is translated into the time series, in which the specified frequency was 24 for NDVI because the data are bimonthly. Figure 6 shows the results of the decomposition of the

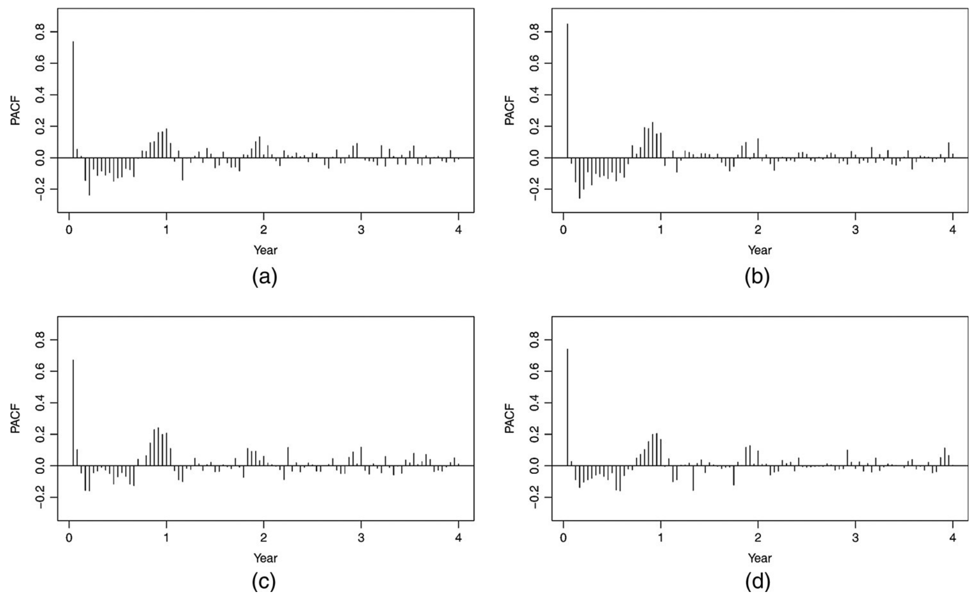


Fig. 5 Results of calculation of the partial autocorrelation function (PACF) for the time series GIMMS AVHHR NDVI 3g data [1983–2011, (a)–(d) locations a–d].

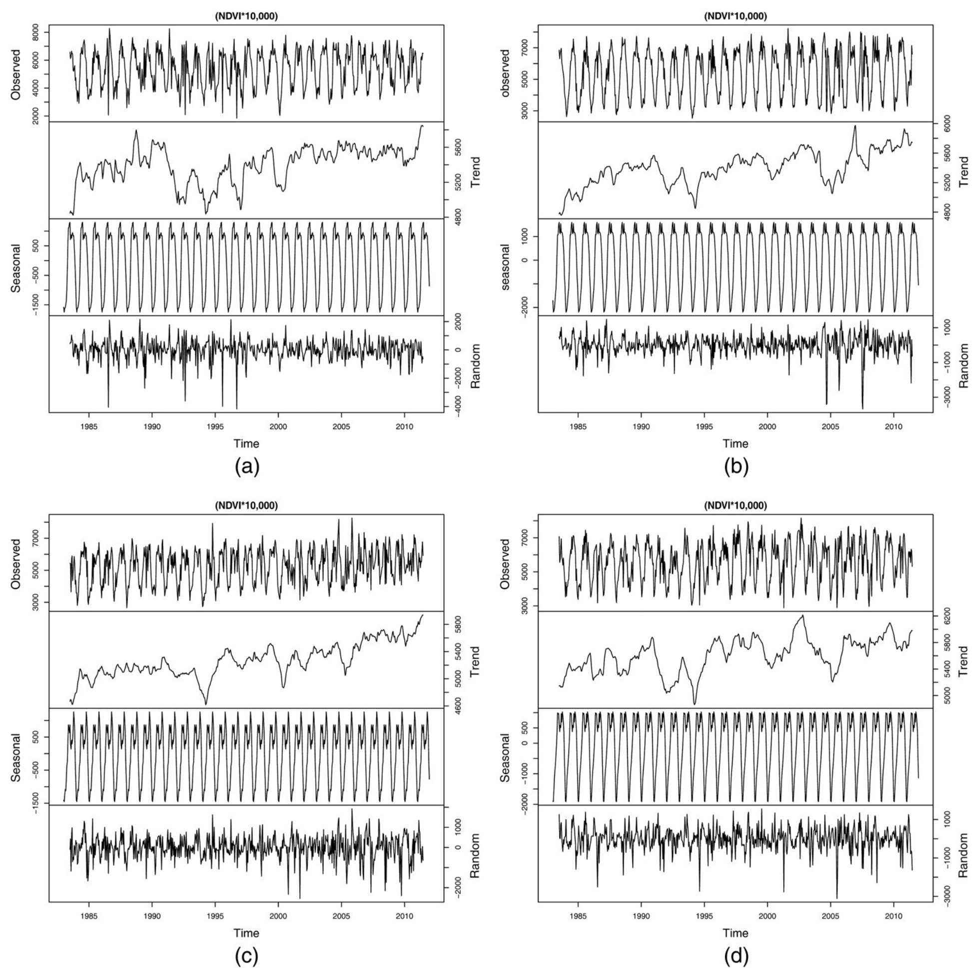


Fig. 6 Decomposition of the time series for GIMMS AVHHR NDVI 3g data [1983–2011, (a)–(d) locations a–d].

Table 1 Results of stationarity tests for the random component of the decomposition results normalized difference vegetation index (NDVI) for the four locations of interest. KPSS: Kwiatkowski–Phillips–Schmidt–Shin test on trend (*T*) and level (*L*) stationarity; ADF: augmented Dickey–Fuller test; PP: Philipps–Perron test (5% significance level).

	KPSS				ADF		PP	
Location	*T*	*p*	*L*	*p*	DF	*p*	Test value	*p*
a	0.0073	0.1	0.0197	0.1	−12.1430	0.1	−457.2789	0.01
b	0.0068	0.1	0.0135	0.1	−12.8923	0.1	−399.3910	0.01
c	0.0044	0.1	0.0048	0.1	−11.2990	0.1	−559.3890	0.01
d	0.0059	0.1	0.0166	0.1	−11.9175	0.1	−449.9169	0.01
Critical value [5%]	0.146		0.463		−3.41		−3.41	

NDVI time series into its trend, seasonal, and random component (noise). The calculations were made on the basis of the method of decomposition introduced in Sec. 3.2 and the identified frequencies from the previous section.

From the visual interpretation of the time series analysis of the four study areas, it can be concluded that there is a positive trend in NDVI (Fig. 6). At the same time, it can be recognized that the characteristic of the annual seasonality is different in the various locations. Thus, it should be noted that two growing seasons could be extracted from the original time series at all locations. In location [a], the second growing season has a lower intensity than the first one, whereas location [b] also has two vegetation periods which are very close together and of equal intensity. In contrast, the first of the two growing seasons in location [c] exhibits a lower intensity, whereas at location [d] in this regard no differences are noted.

Furthernore, the visual analyses of the random-component indicate that no other systematic (trend, seasonality) are in this part of the original signal. To prove this impression, further inferential statistics parameters were calculated. Based on the assumption that the aim of the chosen model of decomposition (Sec. 3.2) is to completely separate the components of trend and seasonality from the random part, a more detailed analysis of the random can help to test if the model fulfill these requirements. Therefore, different stationarity tests on the random component were performed to assess the quality of time series decomposition.

The results of the introduced tests (Sec. 3.4.1) for all four locations in the study area are displayed in Table 1.

The results of the several tests show in most cases a clear assessment concerning the stationarity of the examined data. The results show that the test values of ADF are lower than the critical values for a level of significance of 5%. For this reason, the null hypothesis (the random component is not stationary) must be rejected. Therefore, it can be stated that the random component of the original NDVI signal is stationary and the chosen model of decomposition of the time series is applicable. The same is also true for the results of the PP test. Just like the ADF test, the test values are clearly under the level of the critical value. Both tests conclude the stationarity of the random component. The same conclusion was reached with the KPSS stationarity test. Concerning the trend as well as the level, the test confirms the assumption of stationarity (null hypothesis).

Finally, it can be stated that the chosen model of decomposition fulfills the requirements of extracting the trend component of the original NDVI time series.

4.4 *Interannual NDVI Trend Analysis*

4.4.1 *Results of ordinary linear regression modeling*

The first approach of parameterization of the trend was realized using the OLS method, introduced in Sec. 3.3.1. The results of the linear regression of NDVI plotted against time for the study area are shown in Fig. 7.

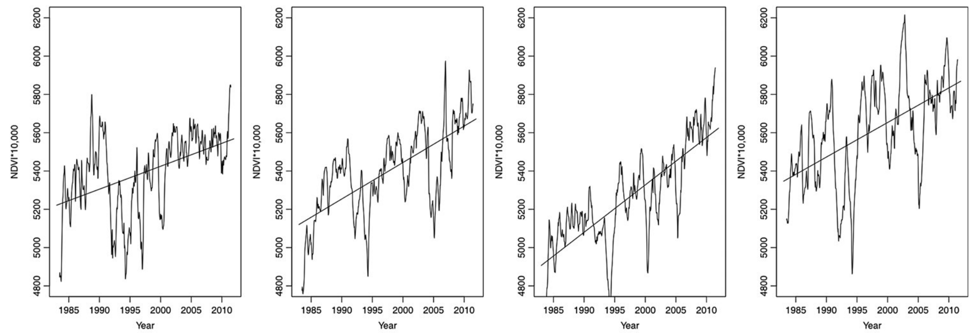


Fig. 7 Linear regression [ordinary least-square (OLS) method] for trend of GIMMS AVHHR NDVI 3g data [1983–2011, (a)–(d) locations a–d].

Table 2 Results of the linear regression [ordinary least-square (OLS) method].

Location	Slope	Residual standard error	*F*-statistic	R^2	*p*-value
a	11.97	181.1	191.9	0.222	$<2.2\times10^{-16}$
b	19.05	163.8	593.7	0.469	$<2.2\times10^{-16}$
c	24.75	149.5	1204.0	0.642	$<2.2\times10^{-16}$
d	18.05	212.5	316.6	0.320	$<2.2\times10^{-16}$

All results show a consistently positive trend in the NDVI signals. Nevertheless, it should be noted that the increase in the period of this study varies considerably in the four selected locations. This variation in trend complicates generalized statements about the trend in the entire study area (Kogi state). Additionally, some artifacts are detectable in the trend component of all locations. For example, the strong decrease of NDVI in the middle of the 1990s points out an impact which was not only present at one single location. This might be caused by a general change, e.g., in climate parameters. Table 2 shows the results of linear regression trend line including information about statistical parameters (F-statistic, R^2, and p-value).

4.4.2 *Results of Theil–Sen estimator*

Due to the high standard deviations (see Table 2), which can be as a result of high variability in NDVI trend signal, it seems to be more suitable to use a more robust method (Theil–Sen estimator) as explained in Sec. 3.3.2. The results of the Theil–Sen estimator are presented in Table 3. When compared to the results of the OLS method, two differences should be emphasized.

Although it is the goal of the calculation to find a better representation of the trend signal, it is observed that the residual standard error is higher than that of the OLS method. The second one is the change in the model parameters of the regression line as well as the slope and the intercept. While the slope of locations [a] and [b] increases, it decreases for the others. These different results between both methods could be an indication for outliers.

This assumption can be tested by analyzing the standardized residuals of all data values. The following Fig. 8 shows few outliers ($>\sqrt{(3)}\sim 1.73$):

Statistical outliers regarding the Theil–Sen estimation of the trend components were detected in all the study areas. However, further analyses using the Cook distance and the leverage effect[39] also showed that these outliers have no significant influence on the parameters of the regression. Clearly visible is also a systematic behavior of the residuals. For this systematic behavior, two causes can be considered. On one hand, there might still be seasonality in the trend data as already mentioned. Otherwise, the black line (Fig. 8) emphasizes that the chosen model of a linear regression (Theil–Sen estimation) could be insufficient.

Table 3 Results of the Theil–Sen estimator.

Location	Slope	Residual standard error	*p*-value
a	13.797	187.6	$<2 \times 10^{-16}$
b	20.333	168.7	$<2 \times 10^{-16}$
c	23.481	154.0	$<2 \times 10^{-16}$
d	17.200	214.4	$<2 \times 10^{-16}$

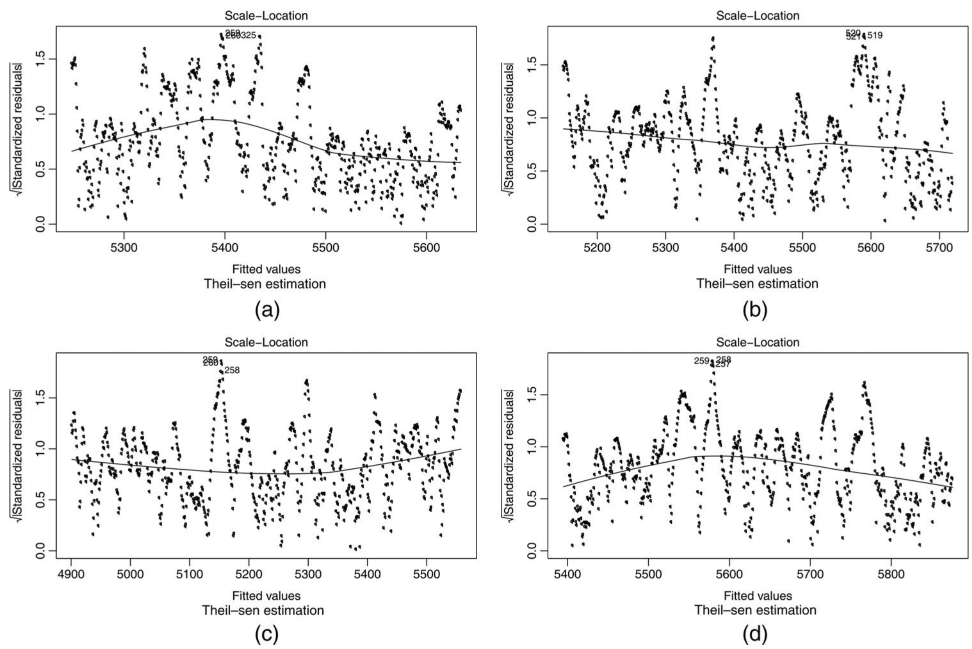


Fig. 8 (a)–(d) Standardized residuals for the Theil–Sen estimation of the trend signal.

Nevertheless, it is interesting to know whether the results of visual interpretation of the behavior of the standardized residuals (Fig. 8) are correct. Therefore, more detailed residual analyses of the OLS regression functions as well as the Theil–Sen estimations are needed. As already performed for the residual analysis of the time series decomposition, several stationarity tests can help to evaluate the quality of the chosen linear regression models. Table 4 shows the results of the mentioned methods of statistical testing. Due to the fact that both regression models follow a linear approach and the influence of outliers is comparatively limited, the results of the test statistic are identical.

Although the KPSS test on trend stationarity rejected the null hypothesis of stationarity, the same test on level stationarity confirms them for the residuals of locations [b] and [d]. For the other two locations, the assumption on stationarity was also rejected. However, the results of the ADF test reject the null hypothesis on nonstationarity and thereby they confirm the stationarity of the residuals for all locations even though the test values exceeds the critical value ($\alpha < 0.05$). The third test (PP test) also confirms the assumption of stationarity. The statements based on the test values for the different test procedures are quite ambivalent and justify the application of different test methods. Therefore, it should also be noted that a final evaluation of stationary is difficult.

The aim of evaluating linear trend modeling is the proof of stationarity of residuals. For this reason, results which confirm this hypothesis should be critically scrutinized in order to avoid type II errors (stationarity is assumed but false) regarding the overall hypothesis of stationarity, this applies particularly to the ADF test as well as PP test. Because both tests are in fact nonstationarity tests, the definition of null hypothesis and alternative hypothesis is reversed. Therefore, in the context of the ADF test and PP test, type I errors should be avoided. That means nonstationarity is rejected but true.

Table 4 Results of stationarity tests for the residuals of the ordinary least-square and Theil–Sen regression for the four locations of interest. KPSS: Kwiatkowski–Phillips–Schmidt–Shin test on trend (*T*) and level (*L*) stationarity; ADF: augmented Dickey–Fuller test; PP: Philipps–Perron test (5% significance level).

	KPSS				ADF		PP	
Location	*T*	*p*	*L*	*p*	DF	*p*	Test value	*p*
a	0.5750	<0.01	0.5750	0.02491	−3.7516	0.02142	−28.6133	0.01105
b	0.2572	<0.01	0.2572	>0.1	−3.8824	0.01488	−26.6031	0.01841
c	0.5032	<0.01	0.5032	0.04096	−4.1486	<0.01	−26.1860	0.01993
d	0.1856	0.02	0.1856	>0.1	−4.1708	<0.01	−22.2523	0.04430
Critical value [5%]	0.146		0.463		−3.41		−3.41	

To avoid these circumstances of different test results, it is decided to assume nonstationary residuals and consequently the necessity of better matching (linear) models arises. The reason could be the limited suitability of the chosen linear model. In its current form, the model cannot reflect sudden changes in the trend.

4.4.3 *Results of segmented linear regression*

As aforementioned, the chosen model of a simple linear regression over a time period of almost 30 years does not take into account that external impacts or events can have a decisive influence on the NDVI signal in the study areas. For this reason, it is necessary to analyze the development of NDVI for structural breaks with the aim of achieving breakpoints as limiters for a segmented linear modeling of NDVI time series. Therefore, a detailed analysis of breakpoints in the trend component of all the time series was performed using the R-package BFAST (Sec. 3.3.2).

Figure 9 displayed the results of the segmented linear regression for the four locations of interest. Because of the previous elimination of the seasonal part of the original time series, the analyses were performed without further estimation of periodical components.

A first visual interpretation of detected breakpoints provides a much better approximation of the segmented linear regression functions concerning the original trend signal than the results of the OLS method using the entire time series. Figure 10 gives a comprehensive overview of the resulting linear regression segments.

At all four locations of interest, four breakpoints and so five temporal segments for further linear regression analyses could be detected. In the following, a three-step assessment of breakpoint determination and segmented linear regression analysis is performed. These assessments should be more than only a calculation of several statistical parameters but also an interpretation of the results concerning the underlying database.

The first step of assessment covered the breakpoints itself. Every breakpoint is characterized by its position in the time series, its confidence interval, and the amount of break in the trend component of the NDVI signal. It can be seen in Fig. 10 that most of the breakpoints are in 1990s. In these years, the NDVI trend is very variable. This also applies to the confidence intervals on $\alpha = 0.05$ level. The results show different widths between 1 and 4 months. These confidence intervals can be interpreted in two ways. On one hand, it is an indicator for quality of the breakpoint on appropriate position. On the other hand, the width of the confidence interval can also be interpreted as a hint for the sharpness of change in the NDVI trend component. Thus, by choosing the parameter h the number of possible breakpoints and linear segments is limited by its number but also length of segments. As a result, in some cases, these procedure leads to the detection of smoother breakpoints with larger confidence intervals. Furthermore, every breakpoint is also characterized by its amount of break. The results clearly show strong breaks up to over 500. These sudden changes in the signal clarify the necessity of segmented linear regression modeling.

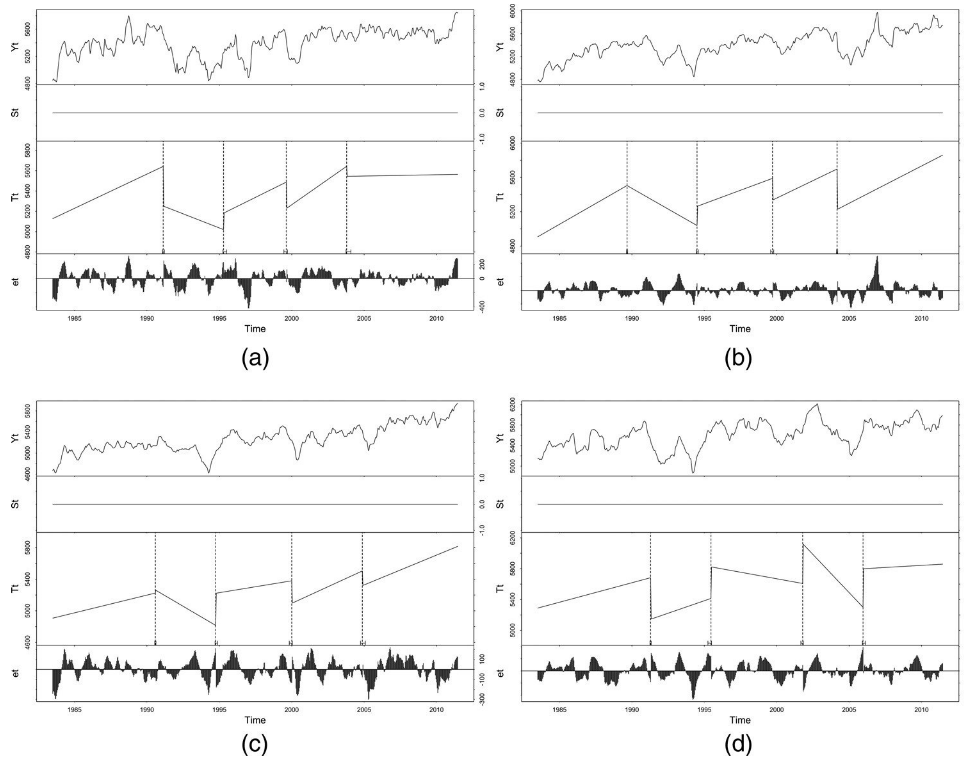


Fig. 9 (a)–(d) Results of the segmented ordinary least-square method ($h = 0.15$).

The second step of assessment involves the segmented linear regression lines which includes the slope values of the segments as well as R^2. The variability in slopes and particularly the changing algebraic sign shows the necessity of the segmented approaches. Also relevant is the evaluation of the coefficient of determination. In many cases, a comparison of the results presented in Table 2 (OLS) and Fig. 10 (segOLS) shows an improvement of the linear modeling or confirm the results. Only two exceptions should be taken into account. It is observed from the analysis that the linear regression for the fifth segment in locations [a] and [d] is not significant. Finally, as already performed for the OLS modeling, the results of different tests on stationarity applied on the residuals of the segmented linear regression are used as an additional indication for quality of modeling. Table 5 presents the results of these test procedures.

All tests are performed on the residuals of the entire time series for every location. The results impressively confirm the stationarity. Compared to the same tests on the OLS regression (Table 4), the results are much clearer. Especially, the values for the KPSS test are below the critical value and confirm the null hypotheses of stationarity. The results of the ADF and the PP test lead to the same conclusion. Also noticeable is that the test values when compared with each other are more homogeneous. This indicates that the distribution of the residuals between the locations is comparable with the interpretation of the results of the regression model.

In the final step of assessment, the standard deviation of the segmented linear regression should be used as a summary indicator concerning the quantitative improvement in the NDVI trend detection. Although it is not actually a model change, it is nevertheless obvious that the segmentation of the time series leads to an improvement in the resulting linear model and an increasing accuracy underlined by a reduction in the standard deviation of more than 30% (Fig. 10).

4.5 *Correlation Analyses NDVI Time Series*

Beside the already performed trend analyses, a correlation analysis should give information about the proportion of the trend which can be determined by time. Therefore, this information is also suitable for estimating the proportion of other unknown influencing variables.

Location a						
2,5%	**BP**	**97,5%**	**break**	**slope**	**R²**	**SD**
1991 (3)	1991 (4)	1991 (6)	-392.2	67.36	0.597	181.1 ↓ 119.0 [-34%]
				-55.60	0.185	
1995 (7)	1995 (8)	1995 (13)	164.5			
				70.52	0.265	
1999 (12)	1999 (16)	1999 (18)	-256.0			
				99.78	0.507	
2003 (19)	2003 (20)	2004 (3)	-98.8			
				2.59	-0.001	

Location b						
2,5%	**BP**	**97,5%**	**break**	**slope**	**R²**	**SD**
1989 (16)	1989 (17)	1989 (18)	-4.6	96.84	0.823	163,8 ↓ 98.2 [-40%]
				-96.36	0.596	
1994 (12)	1994 (13)	1994 (15)	226.3			
				62.56	0.751	
1999 (15)	1999 (18)	1999 (20)	-250.0			
				81.29	0.582	
2004 (4)	2004 (5)	2004 (6)	-467.6			
				86.81	0.666	

Location c						
2,5%	**BP**	**97,5%**	**break**	**slope**	**R²**	**SD**
1990 (14)	1990 (15)	1990 (16)	34.5	44.83	0.479	149.5 ↓ 100.4 [-33%]
				-107.60	0.647	
1994 (18)	1994 (19)	1994 (22)	407.8			
				30.07	0.182	
1999 (21)	2000 (1)	2000 (2)	-281.9			
				84.12	0.585	
2004 (19)	2004 (22)	2005 (3)	-183.2			
				75.55	0.617	

Location d						
2,5%	**BP**	**97,5%**	**break**	**slope**	**R²**	**SD**
1991 (7)	1991 (8)	1991 (9)	-536.0	50.53	0.432	212.5 ↓ 136.6 [-36%]
				64.76	0.145	
1995 (7)	1995 (12)	1995 (13)	408.2			
				-34.02	0.191	
2001 (17)	2001 (20)	2001 (21)	508.2			
				-197.93	0.729	
2005 (23)	2005 (24)	2006 (4)	499.7			
				10.865	0.019	

Fig. 10 Statistical summary for the segmented ordinary least-square method using BFAST (SD standard deviation; BP breakpoint).

4.5.1 *Selection of correlation method*

For the next step of analysis (calculation of correlation coefficients), it is necessary to think about the usability of different methods of correlation analyses. For one of the most common methods (Pearson correlation coefficient), the precondition of normal distributed measurements is mandatory. This means that for the selection of an appropriate correlation method, the distribution of measurements must be known. Therefore, with the Shapiro–Wilk test and the Anderson–Darling test (Sec. 3.4.2), two established tests for normal distribution were applied (Table 6).

The assessment of the normality tests by means of the critical values W (Shapiro–Wilk: $W < 0.996$,[37]) and A (Anderson–Darling: $A > 0.787$,[38]) indicates that the null hypothesis

Table 5 Results of stationarity tests for the residuals of the segmented ordinary least-square for the four locations of interest. KPSS: Kwiatkowski–Phillips–Schmidt–Shin test on trend (T) and level (L) stationarity; ADF: augmented Dickey–Fuller test; PP: Philipps–Perron test (5% significance level).

	KPSS				ADF		PP	
Location	T	p	L	p	DF	p	Test value	p
a	0.0285	>0.01	0.0296	>0.01	−6.3247	<0.01	−61.5932	<0.01
b	0.0400	>0.01	0.0258	>0.01	−6.2213	<0.01	−62.0524	<0.01
c	0.0491	>0.01	0.0360	>0.01	−6.1098	<0.01	−62.2395	<0.01
d	0.0299	>0.01	0.0233	>0.01	−5.9710	<0.01	−60.3478	<0.01
Critical value [5%]	0.146		0.463		−3.41		−3.41	

Table 6 Results of Shapiro–Wilk and Anderson–Darling normality test for $\alpha = 0.05$ (Critical values: $W_{\mathrm{critical}} = 0.996 / A_{\mathrm{critical}} = 0.787$.

	Shapiro–Wilk		Anderson–Darling	
Location	W	*p*-value	A	*p*-value
a	0.9547	1.635×10^{-13}	10.3497	$<2.2 \times 10^{-16}$
b	0.9929	0.00292	0.9972	0.01244
c	0.9855	3.29×10^{-6}	4.0906	3.498×10^{-10}
d	0.9875	1.708×10^{-5}	3.1219	7.794×10^{-8}

(there is a normal distribution) must be rejected. This decision is also confirmed by the probability value (p-value). Only for one location [b], the probability is comparatively high (1.24%). However, since the value is smaller than the chosen significance level ($\alpha = 0.05$), the result leads also to the rejection of the null hypothesis.[37,38]

This circumstance needs to be considered for the selection of an appropriate correlation method. For this reason, the validity of correlation coefficients such as Pearson correlation coefficient is limited. Therefore, the rank correlation coefficient of MK and the sMK correlation are is used for further analysis (Sec. 3.3.4).

4.5.2 *Results of the Mann-Kendall correlation analysis for the whole time series*

As already shown for the previous presented results, the MK correlation analyses are performed in a comparative way. This includes the sMK on the original time series and trend component as well as the MK correlation only on the trend component.

The result of the seasonal MK's rank correlation coefficient (sMK) for the original NDVI time series shows values that do not exceed 0.3 (Table 7). Although these correlations are relatively small, it shows that the relationship between time and NDVI trend is significant even though it is determined only partly by time. However, the associated p-values also show that this relationship is highly significant and in principle the null hypothesis cannot be rejected.

The result of the ordinary MK tau for the trend obtained from the decomposition of the NDVI time series shows that 35%–62% can be determined by time. The p-values for the analysis are highly significant at $\alpha = 0.05$ level. Thus, the use of the original trend component with the seasonal MK leads to a further improvement of the tau values. Although no remaining seasonality should be in the trend signal, the consideration of them leads to an improvement of the tau values. As aforementioned, these differences between the tau values of MK and sMK are indicators for the remaining seasonality in the trend signal.

Table 7 Results of (seasonal) Mann-Kendall (MK) correlation analyses for the original time series and the trend signal of the decomposition.

	Seasonal Mann–Kendall (sMK)				Mann–Kendall (MK)	
	Original TS		Trend		Trend	
Location	tau	*p*-value	tau	*p*-value	tau	*p*-value
a	0.117	1.3301×10^{-5}	0.360	2.22×10^{-16}	0.351	2.22×10^{-16}
b	0.252	2.22×10^{-16}	0.527	2.22×10^{-16}	0.509	2.22×10^{-16}
c	0.259	2.22×10^{-16}	0.638	$2.22e \times 10^{-16}$	0.624	2.22×10^{-16}
d	0.195	3.6859×10^{-16}	0.418	2.22×10^{-16}	0.403	2.22×10^{-16}

Table 8 Results of (seasonal) Mann–Kendall correlation analyses within the limits of the determined breakpoints.

	Location a				Location b			
	Mann–Kendall (MK)		Seasonal Mann–Kendall (sMK)		Mann–Kendall (MK)		Seasonal Mann–Kendall (sMK)	
Segment	tau	*p*-value	tau	*p*-value	tau	*p*-value	tau	*p*-value
1	0.573	2.22×10^{-16}	0.646	2.22×10^{-16}	0.735	2.22×10^{-16}	0.785	2.22×10^{-16}
2	−0.268	7.85×10^{-5}	−0.375	1.08×10^{-4}	−0.553	2.22×10^{-16}	−0.661	1.20×10^{-14}
3	0.408	2.22×10^{-16}	0.523	2.43×10^{-8}	0.647	2.22×10^{-16}	0.781	2.22×10^{-16}
4	0.435	2.22×10^{-16}	0.562	6.27×10^{-9}	0.550	2.22×10^{-16}	0.691	4.15×10^{-14}
5	−0.036	0.463	−0.049	0.423	0.625	2.22×10^{-16}	0.740	2.22×10^{-16}
	Location c				Location d			
	Mann–Kendall (MK)		Seasonal Mann–Kendall (sMK)		Mann–Kendall (MK)		Seasonal Mann–Kendall (sMK)	
Segment	tau	*p*-value	tau	*p*-value	tau	*p*-value	tau	*p*-value
1	0.490	2.22×10^{-16}	0.577	2.22×10^{-16}	0.454	2.22×10^{-16}	0.481	8.88×10^{-16}
2	−0.581	2.22×10^{-16}	−0.737	2.60×10^{-14}	0.283	3.10×10^{-5}	0.462	1.78×10^{-6}
3	0.272	6.58×10^{-16}	0.237	3.16×10^{-3}	−0.262	1.61×10^{-6}	−0.284	4.49×10^{-5}
4	0.578	2.22×10^{-16}	0.649	2.31×10^{-14}	−0.673	2.22×10^{-16}	−0.850	2.22×10^{-16}
5	0.570	2.22×10^{-16}	0.625	2.22×10^{-16}	0.076	0.20	0.04	0.61

4.5.3 *Results of the Mann–Kendall correlation analysis for the segmented time series*

The results presented in Sec. 4.4.3 indicate that a segmented linear regression better approximates the real change in trend. For this reason, the MK analysis was also performed within the limits of the determined breakpoints. Results of these correlation tests are displayed in Table 8.

It is obvious that depending on the relevant segment, the MK tau values differ when compared to the related correlation coefficient of the entire NDVI trend signal. According to the development of the NDVI within every segment, the MK tau value as well as the seasonal version of these correlation coefficients reflects the necessity of a segmented analysis. In many cases, the tau values them show a strong positive MK correlation interrupted only by the extracted breakpoints. Only in a few cases, a negative trend is observed over a comparable long period of more than about 5 years. Also, it should be noted that MK test is not significant for the fifth segment of locations [a] and [d].

5 Discussion

Our discussion focuses on two aspects. The first part of the discussion deals with the results of the statistical analyses and therefore underlined the methodical focus of the entire study. Nevertheless, all analyses were performed in the study area of Kogi state. Although the aim of the study is to establish a suitable methodology for detecting long-term trends in vegetation, it is also important to examine the effect of land-use and land-cover change on the results obtained from the trend analysis. The second part of the discussion tries to broaden the view to further usability of the presented methods as well as further statistical approaches.

5.1 *Discussion in the Context of Statistical Methods and the Study Area*

Trends in NDVI time series for the study period (1983–2011) were analyzed using linear regression models to delineate the coefficient of determination over time and a perfectly linear series. It is observed that the linear trends for the four locations vary considerably (Table 2) although they are statistically significant, which implies that there is positive trend in the vegetation productivity of the study area (greening trend). In order to investigate whether the trends (slopes) are not affected by artifacts or influence of outliers, a more robust slope estimator was used and the result of the p-value for the locations is significant, which implies that there is a greening trend in the study area over time.

The positive and significant trends of GIMMS 3g NDVI observed in the study areas correspond with results of other studies based on different AVHRR datasets across the globe including the study of Heumann et al.,[40] where trends are found in most parts of Nigeria. The findings from the MK trend analysis for the four locations showed an increasing NDVI trend after the decomposition (removal of seasonality) of the original data, which indicate vegetation greening over time (monotonic greening). The Kendall tau of NDVI against time for the four study locations was assessed and the results show that Kendall tau is greater than zero and the p-values were low at the significant level of $\alpha = 0.05$. This indicates that there is a monotonic trend in NDVI over time (that is, a greening trend when Kendall tau > 0) and this result corresponds with the monotonic assessment trend by de Jong et al.[2] and Boschetti et al.[3]

The visual interpretation of residuals (Fig. 9) as well as the detailed statistical residual analyses expects that an ordinary linear regression of NDVI time series can be an appropriate instrument of trend detection. However, both analyses also demonstrate that some parts (segments) of the analyzed time series may show quite different behavior. Therefore, it could be suggested that a decomposition of the time series into different temporal segments will lead to an improvement of linear trend modeling. The chosen method is evaluated using several criteria and underlined the significance and necessity of considering structural breaks from a purely statistical point of view. Furthermore, this evaluation can be completed by considering context-related changes in addition to statistical measures. This is particularly reflected in comparison of locations [a]–[d]. Changes in the phenological condition take place at approximately the same time and order. The changes itself are of varied nature and thus usable as indicators for the characteristic of the chosen locations. The amount of break as well as the direction of change (positive or negative slope) gives valuable information about the condition of vegetation and orientation of vegetation development.

There are several explanations for the shift we observed in the NDVI trend within the 1990s. One of the main reasons for the significant fluctuation in the trend pattern observed is the land-use change/transformation in the studied locations coupled with an increase in the population within this decade. However, several infrastructural developmental projects sprang up in almost all geopolitical zones in Nigeria in the 1990s due to the creation of new states and local government areas, which has a feedback on the natural vegetation in which Kogi state Nigeria is also included. Due to human perturbation and dependency on natural resources, the natural vegetation in Kogi state had witnessed tremendous change over the last decades (transition from derived savannah to Guinea savannah). Table 9 gives an overview of the land-use/land-cover changes of the selected locations.

It is clear especially in the 1990s that the land use in all four locations has strongly changed. Because of the coarse temporal resolution of the land-use and land-cover information, it becomes clear that only the use of the temporal high resolution time series allows a more detailed view on the processes of vegetation dynamics.

The results of our linear trend analyses are in conformity with other research, which shows a linear increase in vegetation.[41] Many other studies also reported a lack of clear trends in NDVI after the mid-1990s[41] due to the recovery from rain; this might also have an effect on the NDVI signal for the 1990s.[42] Other drivers of changes in vegetation trend (decreasing and increasing trends) as observed in the study area can be linked to population growth. This is in line with other studies where increasing population and human activities have an effect on the vegetation trend behavior.[3]

It should be noted that some of the areas, where an increase in NDVI trend are observed, are located in the surrounding of water bodies. In these locations, increase of vegetation productivity

Table 9 The assessment of land use change in the selected locations among 1976, 1995, 2000, and 2009 [Source: land use land cover map 1976 and 1995 FORMECU, Nigeria; global land cover 2000 and 2009 (European Space Agency)].

Location	LULC 1976/78	LULC 1995/96	LC 2000	LC 2009
a - Igbalamela	(1) Undisturbed forest (95.4%) (2) Intensive (row crops, minor grazing) small Holder rainfed agriculture (4.6%)	(1) Shrub/sedge graminoid fresh water marsh/ swamp (89.41%) (2) Grassland (10.59%)	(1) Mosaic forest/ croplands (20%) (2) Deciduous shrubland with sparse trees (40%) (3) Deciduous woodland (6.7%) (4) Irrigated crop land (33.3%)	(1) Mosaic grassland/ forest or shrubland (3.8%) (2) Closed to open (broadleaved or needleleaved, evergreen or deciduous) shrubland (58.8%) (3) Mosaic forest or shrubland/grassland (5.3%) (4) Open broadleaved deciduous forest/woodland (0.8%) (5) Mosaic cropland/ vegetation (grassland/ shrubland/forest) (31.3%)
b - Kotonkarfe	(1) Riparian forest (47.2%) (2) Dominantly trees/woodlands/shrubs with a subdominant grass component (38.9%) (3) Intensive (row crops, minor grazing) small Holder Rainfed Agriculture (13.9%)	(1) Dominantly trees/woodlands/ shrubs with a subdominant grass component (100%)	(1) Deciduous woodland (100%)	(1) Open broadleaved deciduous forest/ woodland (51.2%) (2) Mosaic forest or shrubland/grassland (24.88%) (3) Closed to open (broadleaved or needleleaved, evergreen or deciduous) shrubland (23.92%)
c - Ankpa	(1) Intensive (row crops, minor grazing) small Holder Rainfed Agriculture (100%)	(1) Intensive (row crops, minor grazing) small Holder Rainfed Agriculture (100%)	(1) Deciduous shrubland with sparse trees (54.4%) (2) Deciduous woodland (45.6%)	(1) Open broadleaved deciduous forest/ woodland (17.6%) (2) Closed to open (broadleaved or needleleaved, evergreen or deciduous) shrubland (82.4%)
d - Kabba Bunu	(1) Intensive (row crops, minor grazing) small Holder rainfed agriculture (26.4%) (2) Extensive (grazing, minor row crops) small holder rainfed agriculture (73.6%)	(1) Dominantly trees/woodlands/ shrubs with a subdominant grass component (100%)	(1) Deciduous woodland (100%)	(1) Open broadleaved deciduous forest/ woodland (74.9%) (2) Closed to open (broadleaved or needleleaved, evergreen or deciduous) shrubland (25.1%)

(NDVI) is related to expansion and/or intensification of irrigated agricultural activities by the rural communities and also land-cover conversion from forest into grassland as the study area supports dry-season grazing by nomads.

5.2 *Usability of Results in the Context of Further Research*

Further research in different areas is intended, such as an analysis of more locations, with the aim to find spatial dependences of NDVI in the entire Kogi state and beyond to examine potential discrepancies in the development of the Nigerian landscape. The presented results are limited to specially selected areas with the aim of proving the usability of the available methods, which should be understood as a method-oriented case-study. The inclusion of additional data for a better understanding of processes in land-use change is also important. Especially, the influence of rainfall is discussed in Osunmadewa et al.[42]

Furthermore, socioeconomic data such as the development of the population (population density) could provide more information about the anthropogenic impact and stress on vegetation distribution. For the validation of this present effect, other satellite data (e.g., LANDSAT) could increase the geometric resolution of NDVI information, although the data could not fulfill the requirements of a consistent time series. Existing data provided by the Nigerians authorities (land use, land cover, settlement development, population density, and so on) could help to downscale results of our analysis to recommendations on the level of municipalities.

Besides these data-driven perspectives, further analyses should also take into account other methods of trend detection. As outlined in this paper, simple linear regressions are well-suited. Nevertheless, other more robust regression algorithms besides the used Theil–Sen estimator will help to consider outliers (e.g., RANSAC[43]). Because of the limits of NDVI between -1 and $+1$, the chosen linear relation of NDVI and time is exclusively applicable for retrospective analyses. To predict further development of vegetation based on NDVI data, it is strictly necessary to consider these limits. One possibility of consideration is the use of further local regression methods such as LOESS/LOWESS (locally weighted scatter plot smoothing[44]) or other current approaches.[45] In this context, a further deep research of break points in time series gains in importance. This holds true for the determination of local regression functions as well as for the analyses of reasons that unperpin these points of discontinuity. The same applies for detailed analyses of outliers and their reasons.

A further aspect of future research that deals with the considerations of different seasonality is needed. As pointed out in this paper, the realized analyses only take into account an annual repetitive signal. However, the results also indicate the presence of other long-term periodical components. In this regard, it is necessary to think about the reasons as well as the methodology to be used. One way is to expand the part of residual analysis. The residuals displayed in Fig. 8 suggest the existence of a systematic and currently unknown driving force factor. In some cases, it might be possible to use this information as an indicator for further seasonality.[46] The same applies in respect to the random component of the time series decomposition. Finally, it should be pointed out that there are also approaches for time series analyses in the frequency domain, including spectral analyses and recently wavelet analyses.[47,48] Especially, the presence of different and changing seasonalities should be emphasized in further research.

6 Conclusion

There are two main objectives presented in this study. The first one refers to the analyses of vegetation cover in the Guinea savannah region of Nigeria, Kogi state. Based on the AVHRR GIMMS NDVI 3g data recorded between 1983 and 2011, four specially selected locations of the entire study area were analyzed regarding vegetation cover. Therefore, the NDVI product as one of the most important and common vegetation index used in the field of remote sensing was used to prove the presence of trends in vegetation cover as well as the identification of structural breaks in the long-term NDVI time series. The second most important focus is on the introduction of a basic workflow for statistical analyses of NDVI time series for the determination and modeling of trends over a period of 29 years with a bimonthly temporal resolution.

The results of the ordinary linear least-square modeling indicate only a few outliers in the trend components of NDVI signal and induce moderate changes of slopes in the trend signal. In turn, the analyses of residuals can show different results, and therefore the necessity of improvement in certain parts of the trend signal must be achieved. The results show that the consideration of structural breaks can resolve the apparent contradiction of residual analyses. At the same time, the aforementioned statistical values were considerably improved. For the shown examples, the residual standard error decreases up to 40% and for some temporal segments between two structural breaks, the coefficient of determination rises up to 0.82. Hence, it can be stated that the modified or segmented linear regression yields quite better results. This statement is also confirmed by the results of correlation analyses. The presented workflow can be used to decide on the suitability of different modeling approaches of different complexity depending on the used data and requirements of a particular application.

Thus, the use of remotely sensed data coupled with statistical methods is an important approach for assessing vegetation trend in regions such as Kogi state, Nigeria, where adequate information on vegetation productivity (greening) over the last decades has not been properly archived. In conclusion, it can be stated that the analyses of NDVI time series in Nigeria can assist the understanding and knowledge of natural- and anthropogenic-indicated processes.

References

1. X. Zhang et al., "Monitoring vegetation phenology using MODIS," *Remote Sens. Environ.* **84**, 471–475 (2003).
2. R. de Jong et al., "Analysis of monotonic greening and browning trends from global NDVI time-series," *Remote Sens. Environ.* **115**, 692–702 (2011).
3. M. Boschetti et al., "Identification of environmental anomaly hot spots in West Africa from time series of NDVI and rainfall," *ISPRS J. Photogramm. Remote Sens.* **78**, 26–40 (2013).
4. X. Zhang et al., "Monitoring the response of vegetation phenology to precipitation in Africa by coupling MODIS and TRIMM instruments," *J. Geophys. Res.* **110**, D12103 (2005).
5. R. Fensholt et al., "Greenness in semi-arid area across the globe 1981-2007- an Earth observing satellite based analysis of trends and drivers," *Remote Sens. Environ.* **121**, 144–158 (2012).
6. W. J. D. Van Leeuwen, "Monitoring the effects of forest restoration treatments on post-fire vegetation recovery with MODIS multitemporal data," *Sensors* **8**, 2017 (2008).
7. C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," *Remote Sens. Environ.* **8**, 127–150 (1979).
8. N.B. Eniolorunda and O.M. Bello, "Trends in vegetation cover over Niger-Benue trough from 1998–2007 using 1-km SPOT NDVI data," *Sokoto J. Soc. Sci.* **2** (2012).
9. U. Usman, S. A. Yelwa, and S. U. Gulumbe, "An assessment of vegetation cover changes across Northern Nigeria using trend line and principal component analysis," *J. Agric. Environ. Sci.* **1**(1), 01–18 (2012).
10. J. E. Pinzon and C. J. Tucker, "A non-stationary 1981-2012 AVHRR NDVI3g time series," *Remote Sens.* **6**(8), 6929–6960 (2014).
11. A. Anyamba et al., "Thirty-two years of Sahelian zone growing season non-stationary NDVI3g patterns and trends," *Remote Sens.* **6**(4), 3101–3122 (2014).
12. C. J. Tucker et al., "An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data," *Int. J. Remote Sens.* **26**(20), 4485–4498 (2005).
13. A. O. Onoja and O. Idoko, "Econometric analysis of factors influencing fuel wood demand in rural and peri-urban farm households of Kogi state," *J. Sustainable Dev.* **8**(1), 115–127 (2012).
14. National Bureau of Statistics (NBS), *Annual Abstract of Statistics*, Federal Republic of Nigeria (2010).
15. P. Ostien, "Percentages by religion of the 1952 and 1963 populations of Nigeria's present 36 states," in *NRN Backgroundpaper No 1*, Nigeria Research Network (2012).
16. S. J. Ibitoye, "Survey of the performance of agricultural cooperative societies in Kogi state, Nigeria," *Eur. Sci. J.* **8**(24), 857–7881 (2012).
17. E. B. Audu, "A descriptive analysis of rainfall for agricultural planning in Lokoja local government area of Kogi state, Nigeria," *Int. J. Sci. Technol.* **2**(12) (2012).
18. "GADM database of global administrative areas," www.gadm.org (2015).
19. J. S. Oguntoyinbo, O. O. Areola, and M. Filani, "A geography of Nigerian development,"in *Proc. Regional Conf. on International Geographical Union*, 2nd ed., Nigeria (1978).
20. Y. Hong, "Serial correlation and serial dependence," in *Proc. Wise Working Paper Series, WISEWP0601*, pp. 1–28, Xiamen University, China (2006).
21. W. N. Venables and B. D. Ripley, Eds., *Modern Applied Statistics with S*, 4th ed., Springer-Verlag, Berlin, Heidelberg (2002).
22. G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, Eds., *Time Series Analysis, Forecasting and Control*, 4th ed., John Wiley & Sons, Hoboken, New Jersey (2008).
23. M. Kendall and A. Stuart, *The Advanced Theory of Statistics*, Vol. **3**, pp. 410–414, Griffin, London (1983).

24. G. D. Hutcheson, "Ordinary least-squares regression," in *The SAGE Dictionary of Quantitative Management Research*, L. Moutinho and G. D. Hutcheson, Eds., pp. 224–228, SAGE Publications Ltd, London (2011).
25. R. O. Gilbert, *Statistical Methods for Environmental Pollution Monitoring*, Van Nostrand Reinhold Company Inc, New York (1987).
26. J. R. Eastman, "IDRISI Selva manual," in *Proc. Clark Labs for Cartographic Technology and Geographic Analysis*, Clark University, Worcester, Massachusetts (2012).
27. H. Tabari and P. H. Talaee, "Recent trends of mean maximum and minimum air temperatures in the western half of Iran," *Meteorol. Atmos. Phys.* **111**, 121–131 (2011).
28. A. Ramesh, *Response of Flood Events to Land Use and Climate Change, Analyzed by Hydrological and Statistical Modeling in Barcelonnette, France*, Springer Science +Business Media, Dordrecht (2013).
29. A. Zeileis, "A unified approach to structural change tests based on ML scores, F statistics, and OLS residuals," *Econom. Rev.* **24**(4), 445–466 (2005).
30. J. Verbesselt et al., "Detecting trend and seasonal changes in satellite image time series," *Remote Sens. Environ.* **114**(1), 106–115 (2010).
31. J. Verbesselt et al., "Phenological change detection while accounting for abrupt and gradual trends in satellite image time series," *Remote Sens. Environ.* **114**(12), 2970–2980 (2010).
32. J. Bai and P. Perron, "Computation and analysis of multiple structural change models," *J. Appl. Econom.* **18**(1), 1–22 (2003).
33. A. Zeileis et al., "Strucchange: an R package for testing for structural change in linear regression models," *J. Stat. Software* **7**(2), 1–38 (2002).
34. A. Zeileis et al., "*Package strucchange*," Technical Report (2015).
35. P. C. B. Phillips and P. Perron, "Testing for a unit root in time series regression," *Biometrika* **75**(2), 335–346 (1988).
36. D. Kwiatkowski et al., "Testing the null hypothesis of stationarity against the alternative of a unit root," *J. Econom.* **54**, 159–178 (1992).
37. M. Mahibbur Rahman and Z. Govindarajulu, "A modification of the test of Shapiro and Wilk for normality," *J. Appl. Stat.* **24**(2), 219–236 (1997).
38. M. A. Stephens, "EDF statistics for goodness of fit," *J. Am. Stat. Assoc.* **69**(347), 730 (1974).
39. S. Chatterjee and A. S. Hadi, Eds., *Regression Analysis by Example*, 5th ed., p. 424, John Wiley & Son (2012).
40. B. W. Heumann et al., "AVHRR derived phenological change in the Sahel and Soudan, Africa, 1982-2005," *Remote Sens. Environ.* **108**, 385–392 (2007).
41. S. Jamali et al., "Automated mapping of vegetation trends with polynomials using NDVI imagery over the Sahel," *Remote Sens. Environ.* **141**, 79–89 (2014).
42. B. A. Osunmadewa, C. Wessollek, and P. Karrasch, "Identification of long term trends in vegetation dynamics in the Guinea savannah region of Nigeria," *Proc. SPIE* **9239**, 92390F (2014).
43. M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM* **24**(6), 381–395 (1981).
44. W. S. Cleveland, "Robust locally weighted regression and smoothing scatterplots," *J. Am. Stat. Assoc.* **74**(368), 829–836 (1979).
45. L. Lin, Y. Song, and Z. Siu, "Local linear-additive estimation for multiple nonparametric regressions," *J. Multivar. Anal.* **123**, 252–269 (2014).
46. A. Schützenmeister and H.-P. Piepho, "Residual analysis of linear mixed models using a simulation approach," *Comput. Stat. Data Anal.* **56**(6), 1405–1416 (2012).
47. B. Boashash, *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference*, Elsevier, New York (2003).
48. E. A. Maharay and P. D'Urso, "Fuzzy clustering of time series in the frequency domain," *Inf. Sci.* **181**, 1187–1211 (2011).

Biographies of the authors are not available.