Journal of Arid Environments xxx (2017) 1–11

Contents lists available at ScienceDirect

Journal of Arid Environments

journal homepage: www.elsevier.com/locate/jaridenv

# Examining the NDVI-rainfall relationship in the semi-arid Sahel using geographically weighted regression

Stefanos Georganos [a, *], Abdulhakim M. Abdi [a], David E. Tenenbaum [a], Stamatis Kalogirou [b]

[a] Department of Physical Geography and Ecosystem Science, Lund University, Sölvegatan 12, 223 62 Lund, Sweden
[b] Department of Geography, Harokopio University of Athens, Greece

ARTICLE INFO

*Article history:*
Received 19 December 2016
Received in revised form
10 May 2017
Accepted 28 June 2017
Available online xxx

*Keywords:*
Geographically weighted regression
Non-stationarity
Scale dependency
Sahel
Earth observation
Drylands

ABSTRACT

The Sahel of Africa is an eco-sensitive zone with complex relations emerging between vegetation productivity and rainfall. These relationships are spatially non-stationary, non-linear, scale dependant and often fail to be successfully modelled by conventional regression models. In response, we apply a local modelling technique, Geographically Weighted Regression (GWR), which allows for relationships to vary in space. We applied the GWR using climatic data (Normalized Vegetation Difference Index and rainfall) on an annual basis during the growing seasons (June–September) for 2002–2012. The operating scale of the Sahelian NDVI–rainfall relationship was found to stabilize around 160 km. With the selection of an appropriate scale, the spatial pattern of the NDVI-rainfall relationship was significantly better explained by the GWR than the traditional Ordinary Least Squares (OLS) regression. GWR performed better in terms of predictive power, accuracy and reduced residual autocorrelation. Moreover, GWR formed spatial clusters with local regression coefficients significantly higher or lower than those that the global OLS model resulted in, highlighting local variations. Areas near wetlands and irrigated lands displayed weak correlations while humid areas such as the Sudanian region at southern Sahel produced higher and more significant correlations. Finally, the spatial relationship of rainfall and NDVI displayed temporal variations as there were significant differences in the spatial trends throughout the study period.

© 2017 Elsevier Ltd. All rights reserved.

## 1. Introduction

The normalized difference vegetation index (NDVI) is the most widely used surrogate for vegetation greenness in a wide range of studies spanning regional to global scales (Anyamba and Tucker, 2005; Vrieling et al., 2013). The variability of NDVI is a function of prevalent climatic conditions such as rainfall and temperature, and this relationship is well established at various spatial and temporal scales (Fabricante et al., 2009; Udelhoven et al., 2009; Wang et al., 2010). Rainfall is a particularly important predictor of vegetation distribution in the transition zone from humid to arid environments (Martiny et al., 2006; Huber et al., 2011). In most studies that characterize the relationship between vegetation and rainfall, NDVI is modelled as a function of rainfall using global linear models calibrated using ordinary least squares (OLS) regression methods. However, the NDVI-rainfall relationship varies spatially and temporally depending on land cover, soil type, vegetation composition and structure, microclimatic conditions and human impact (Propastin et al., 2008). As such, models that assume stationarity may fail to capture the true nature of the relationship between variables making the validity of their results questionable.

The Sahel region of Africa comprises various land cover categories and complex ecosystems, and is known to be sensitive to environmental change (Nicholson et al., 1990; Huber et al., 2011). The Sahel underwent a protracted drought from the mid-1960s through the mid-1980s in which there were several humanitarian crises. Eklundh and Olsson (2003) reported a recovery from this period and observed increases in satellite-derived NDVI from the mid-1980s onwards. This increase in landscape greenness was called the "greening of the Sahel" (Olsson et al., 2005), and was the result of increases in herbaceous and tree cover (Dardel et al., 2014; Brandt et al., 2015). The primary mechanism behind this observed greening is the increase in rainfall (Hickler et al., 2005), and, to a lesser extent, improved land use activities (Olsson et al., 2005; Lee et al., 2015).

* Corresponding author.
*E-mail address:* stefgeorganos@gmail.com (S. Georganos).

http://dx.doi.org/10.1016/j.jaridenv.2017.06.004
0140-1963/© 2017 Elsevier Ltd. All rights reserved.

ARTICLE IN PRESS



The majority of studies that examine the Sahelian NDVI-rainfall relationship are based on linear per-pixel time series analysis of NDVI and rainfall. However, the spatially variable relationship between these parameters has not been explored in depth. As such, this study attempts to model the complex relations between NDVI and rainfall by using a local non-parametric regression method known as geographically-weighted regression (GWR) (Fotheringham et al., 2003). GWR is commonly used in human geography (Fotheringham et al., 2001; Hu et al., 2012) and has recently become popular in ecology (Wang et al., 2005; Propastin et al., 2008; Gaughan and Waylen, 2012). GWR allows the relationships between dependent and explanatory variables to vary over space and directly deals with non-stationarity. The outputs of this method are useful for descriptive purposes and to detect areas of model misspecification or variability that would otherwise be lost in a global model. Thus, the objective of this study is to explore the NDVI – rainfall spatial relationship in the Sahel between 2002 and 2012, with a focus on vegetative growing season. The extremities of this period were chosen because of the large difference in total rainfall received in the region –2002 was a dry year and 2012 was a wet year (Fig. 1).

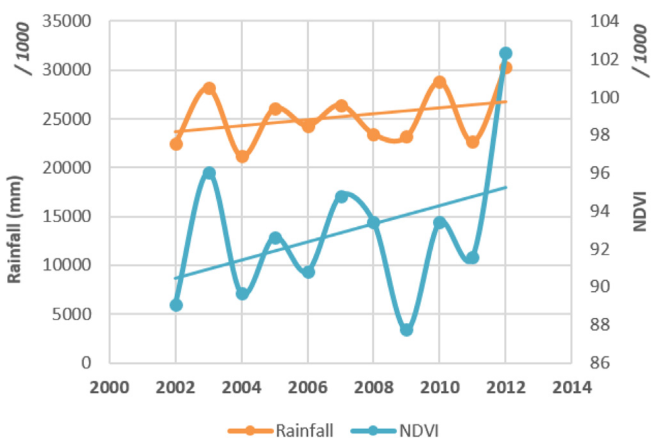


**Fig. 1.** Annual sums of monthly NDVI and monthly rainfall layers between 2002 and 2012 based on the GIMMS and CHRIPS datasets, respectively.

This enables the analysis of temporal variability in the spatial relationships, and the detection of areas where the Sahel is particularly sensitive to variations in rainfall by mapping the local regression results.

## 2. Study area

The Sahel is a 3.3 million km$^2$ region that separates the hyper-arid Sahara Desert in the north from the humid Sudano-Guinean zone in the south. The majority of the rainfall is distributed over 2–4 months during the summer growing season (June–September) whereas rainfall during the rest of the year is negligible (Brandt et al., 2015). The dominant land cover categories of the Sahelian belt based on the Global Land Cover (GLC) SHARE classification (Latham et al., 2014) are shown in Fig. 2. Most Sahelian plants have the $C_4$ photosynthetic pathway, and are acclimatized to warm, arid environments. These are primarily composed of herbaceous vegetation (Fig. 2). The canopy cover ranges between 3 and 10%, and is predominantly composed of trees that have the $C_3$ photosynthetic pathway.

## 3. Material and methods

### 3.1. Normalized difference vegetation index – NDVI

The independent variable in this study is the NDVI and is computed as:

$$NDVI = \frac{NIR - RED}{NIR + RED} \tag{1}$$

where, NIR and RED denote spectral reflectance in the near infrared (800–1000 nm) and red (620–750 nm) portions of the electromagnetic spectrum. The index ranges between −1 (water bodies) and 1 (dense vegetation). We utilized NDVI derived from the Advanced Very High Resolution Radiometer (AVHRR) instrument on board the National Oceanic and Atmospheric Administration (NOAA) polar orbiting satellites. The dataset is the third generation (3g) NDVI product developed by the Global Inventory Monitoring and Modelling System (GIMMS) project. The data are provided as

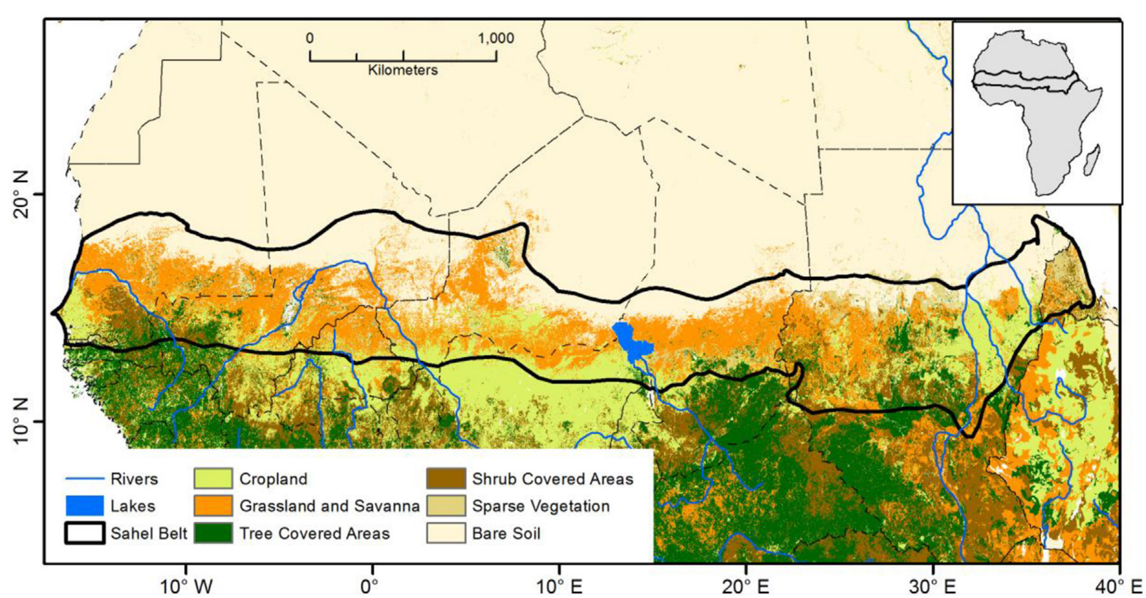


**Fig. 2.** Land cover of the Sahel based on the GLC-SHARE classification scheme from the Food and Agriculture Organization of the United Nations.

Please cite this article in press as: Georganos, S., et al., Examining the NDVI-rainfall relationship in the semi-arid Sahel using geographically weighted regression, Journal of Arid Environments (2017), http://dx.doi.org/10.1016/j.jaridenv.2017.06.004

ARTICLE IN PRESS



biweekly NDVI images at an 8 km spatial resolution and include corrections related to atmospheric change, satellite sensor decay and orbit drift (Pinzon and Tucker, 2014). A recent intercomparison of NDVI datasets by Tian et al. (2015) found that the GIMMS-3g possesses the highest temporal consistency and is appropriate for trend analysis. The biweekly NDVI data were converted into monthly values using the maximum value composite method to reduce cloud disturbance and increase the overall quality of the dataset (Fensholt and Proud, 2012). NDVI was integrated (iNDVI) over the growing season (June–September) for each year between 2002 and 2012 to provide an annual measure of growing season of vegetation productivity.

### 3.2. Climate hazard InfraRed precipitation with stations (CHIRPS)

The CHIRPS (v2.0) project provides rainfall datasets at a 0.05° (~5 km) resolution based on a conjunction geostationary infrared satellite rainfall estimates and rain gauge observations that are interpolated to produce robust precipitation grids (Funk et al., 2015). Monthly grids over the 2002–2012 study period were downloaded (http://chg.geog.ucsb.edu/data/chirps/) and averaged to an 8 km resolution to match the NDVI data. Then, growing season (June–September) was summed on an annual basis between 2002 and 2012 as the independent variable (Fig. 3).

## 4. Modelling methods

### 4.1. Geographically weighted regression

The Geographically Weighted Regression (GWR) is a local modelling technique appropriate for spatial data with some degree of spatial dependence (Kalogirou, 2003). The aim of GWR is to examine the existence of spatial non-stationarity in the relationship between a dependent variable and as set of independent variables. A complete presentation of the GWR method is available in the GWR book by Fotheringham et al. (2003). This section provides a short description of the method.

It can be argued that the GWR is a spatial disaggregation of the traditional regression model. In a simple traditional statistical model, we try to explain a phenomenon using one explanatory variable. Such a model can be formally written as follows:

$$y_i = a + bx_i + \varepsilon_i, \qquad i = 1:n \tag{2}$$

where $y$ is the dependent variable denoting the phenomenon, $x$ is the independent variable denoting the explanatory factor, $\varepsilon$ is the error term, $a$ and $b$ are the parameters to be estimated, and $n$ is the number of samples that correspond to spatial locations. For simplicity, we call this model “global”. When we calibrate the above

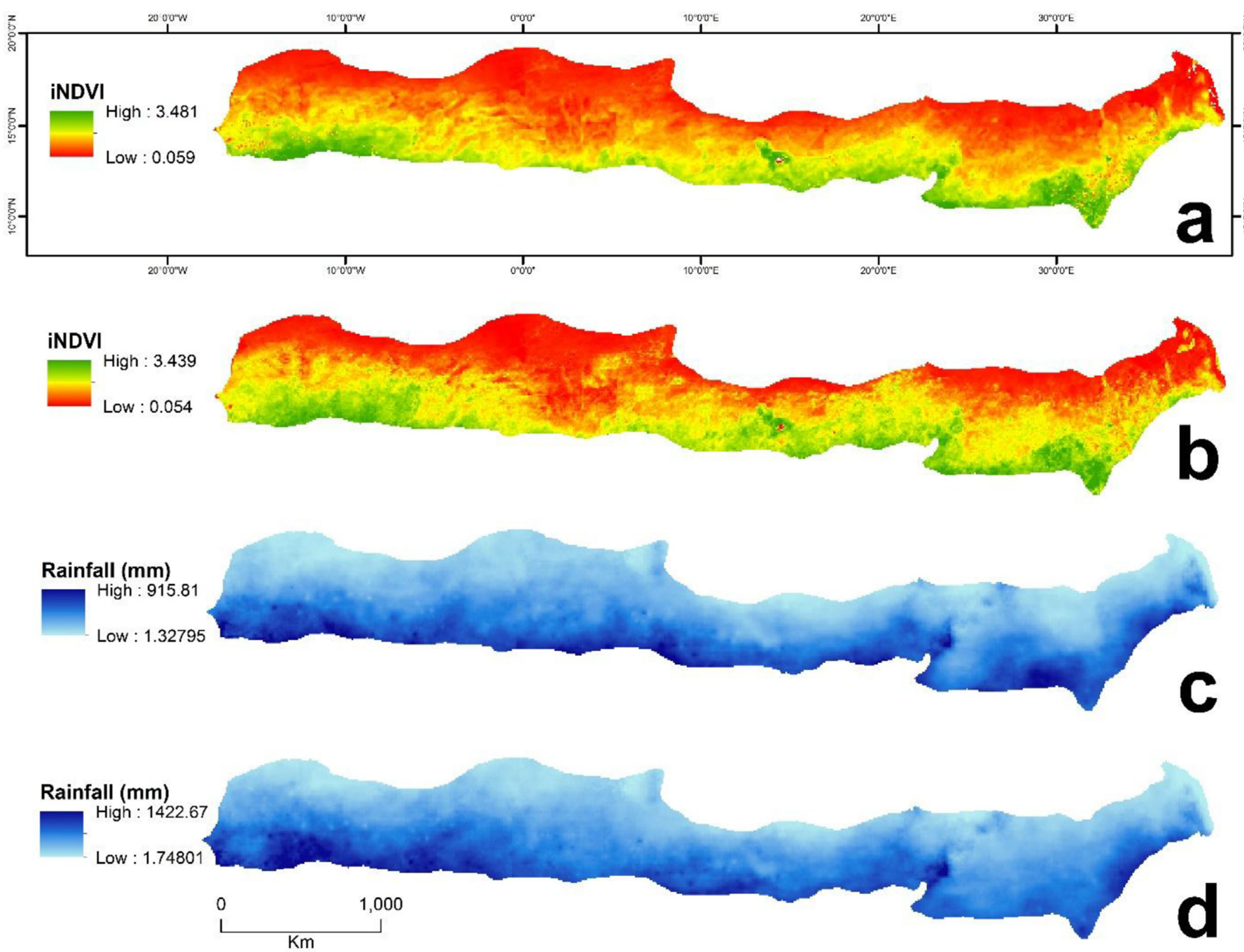


**Fig. 3.** Integrated NDVI (iNDVI) over the growing season for a) 2002, b) 2012, and cumulative rainfall over the growing season for c)2002, d) 2012.

model with the Ordinary Least Squared (OLS) regression, $a$ and $b$ are estimated in such a way that the sum of square of the model residuals are minimized (Brunsdon et al., 1998). Moreover, the residuals should satisfy the criteria of spatial independence and homoscedasticity. In OLS, the estimation of these parameters can be described as a set of matrix equations. For example, we can estimate the parameter $b$ as shown in Equation (3) below:

$$\widehat{b} = \left(X^T X\right)^{-1} X^T Y \tag{3}$$

where $\widehat{b}$ is the estimate of $b$; $X$ is the vector of the values of the independent variable; and $Y$ is the vector of the values of the dependent variable. The parameter estimate $\widehat{b}$, indicates the rate of change in the dependent variable with a unit change in the independent variable and is stationary across the study area. In the empirical example of this paper, the above parameter estimate refers to the relationship between the NDVI and rainfall.

GWR extends Equation (2) by allowing local parameters to be estimated. To achieve this, it takes into account the locations of the observations. The local model can be formally written as follows:

$$y_i = a(u_i, v_i) + b(u_i, v_i)x_i + \varepsilon_i, \qquad i = 1 : n \tag{4}$$

where $(u_i, v_i)$ refer to the coordinates of location $i$, and $a(u_i, v_i)$ and $b(u_i, v_i)$ are the local parameters to be estimated particularly for location $i$. To achieve this, a sub-model around each observation location is defined and fit taking into account a subset of the original observations. The study area of the sub-model is a neighbourhood defined by a weighting scheme in which nearby observations have a non-zero weight. Typically, the number of sub-models equal the number of observations. By computing a local parameter estimate for each observation of the study area, it is possible to examine the potential variability of the relationship between the dependent and independent variable. Consequently, Equation (3), can be rewritten as:

$$\widehat{\boldsymbol{b}}(u_i, v_i) = \left(\boldsymbol{X}^T \boldsymbol{W}(u_i, v_i) X\right)^{-1} \boldsymbol{X}^T \boldsymbol{W}(u_i, v_i) y \tag{5}$$

where, $\widehat{\boldsymbol{b}}$ is a vector of local estimates of $b(u_i, v_i)$ and $\boldsymbol{W}(u_i, v_i)$ is a weights matrix that denotes the weights of the observations for regression point $i$ (sub-model $i$). These weights are defined by a continues function of distance. Observations closer to the location of regression point $i$ have higher weights, while those further away have lower weights.

The type of weighting is important in GWR because it defines the neighbourhood. The most commonly used weighting functions are either Gaussian or Gaussian-like or bi-square. For example, a Gaussian-like weighting function proposed by Fotheringham et al. (2003) is:

$$w_{ij} = \exp\left[-1/2\left(d_{ij}/b\right)^2\right] \tag{6}$$

where $w_{ij}$ is the weight for observation $j$ that refer to the sub-model for location $i$; $d_{ij}$ denotes the Euclidian distance between $j$ and $i$; and $b$ is the size of the neighbourhood.

However, most GWR software uses the bi-square function that is written as follows:

$$\begin{aligned} w_{ij} &= \left[1 - \left(d_{ij}/b\right)^2\right]^2 \qquad \text{if } d_{ij} < b \\ &= 0 \text{ elsewhere} \end{aligned} \tag{7}$$

In the GWR terminology, the neighbourhood is called “kernel” and the maximum distance away from the regression location $i$ is called “bandwidth”. It is possible to have two types of kernels, a “fixed kernel” where the neighbourhood is defined by a circle whose radius is the bandwidth, and an “adaptive kernel” where the neighbourhood is defined by the number of nearest neighbours. The former is more appropriate for data that are evenly distributed across space (such as gridded data) and the latter is more appropriate for data with locations that are dense is some areas and sparse elsewhere (such as the centroids of administrative boundaries). Each type o kernel uses a different weighting scheme. Both weighting schemes were used in this study.

The benefit of a continuous weighting function is that it allows for the estimation of parameters at locations other than the locations of the observations. The bandwidth size describes the distance limit at which nearby observations are taken into account for fitting a sub-model. If the bandwidth is too large, GWR would turn into the global model since all observations would be used at each sub-model, thus hindering meaningful parameter variation and increasing model bias. On the contrary, if the bandwidth is too small the parameters will be less biased but will depend on only in a few observations and will have increased variance in their estimates and large standard errors due to low degrees of freedom (Propastin, 2009).

### 4.2. Model comparison

The Akaike Information Criterion corrected for small sample sizes (AICc) (Akaike, 1974) was used to compare the relative model performance while accounting for model complexity and differences in degrees of freedom. An F-test based on analysis of variance (ANOVA) was computed to assess the significance of the improvement:

$$F = \frac{RSS_{gwr}/DF_{gwr}}{RSS_{glm}/DF_{glm}} \tag{8}$$

where, *RSS_gwr* is the residual sum of squares for a GWR model, *RSS_glm* is the residual sum of squares for a global model, and *DF_gwr* and *DF_glm* are the degrees of freedom for GWR and the global model, respectively. The coefficient of determination ($R^2$) was calculated as a measure of the amount of variance explained by the model. Root mean square error (RMSE) was calculated to evaluate the performance of each model output. RMSE calculates the square root of the variance and smaller values denote better model performance. An RMSE of 0.0 indicates perfect simulation of the input data.

### 4.3. Measuring spatial non-stationarity and autocorrelation

The Stationarity Index (SI) proposed by (Osborne et al., 2007) was used as an approximation of spatial non-stationarity. The SI is calculated as:

$$\mathrm{SI} = \frac{iqr_{gwr}}{SE*2} \tag{9}$$

where *SI* is the stationarity index, *iqr_gwr* is the interquartile range of GWR coefficients, and *SE* is the standard error of a coefficient from a global model. Values less than 1 indicate stationarity at that scale, implying that the relationship has stabilized. The SI has often been used to study the scale dependency of non-stationarity between ecological variables. Calculating SI at incremental scales illustrates the difference between the observation and intrinsic scale-dependence of variables, thus aiding in the selection of a reliable bandwidth. Finally, Moran's I (Moran, 1948) was used to assess the degree of spatial autocorrelation of the data by constructing spatial correlograms at incremental spatial scales.

### 4.4. Data processing and software

The gridded NDVI and rainfall data were converted to vector format in order to use the GWR tool in ArcGIS 10.3, the *lctools* package (Kalogirou and Kalogirou, 2015) in the open source statistical software R 3.1.2 (R Core Team, 2016), and the GWR4 software package (Nakaya et al., 2009). Random sampling was performed to encapsulate 30% of the dataset corresponding to approximately 10,000 sample points for each year.

## 5. Results

### 5.1. Scale dependency of the relationship

The relationship between NDVI and rainfall in the Sahel during the growing season was scale dependent. The pattern was more homogenous as the bandwidth broadened and incorporated information from locations afar, thereby smoothing the regression coefficients and bringing them closer to those of a global model. On the contrary, with smaller bandwidths very detailed patterns were produced at the cost of increased standard errors. The SI for both 2002 and 2012 (Fig. 4) suggests that scale-dependence of non-stationarity was detectable by varying the scale of the analysis. In 2012, the SI was higher for small bandwidths than the values for 2002, but with earlier stabilization. Moreover, the index values were not stationary (SI < 1) in any of the examined spatial scales, which implies a strong non-stationary process in the data. The SI slopes declined abruptly with an increase in bandwidth that flattened out around 160 km, implying that this is the intrinsic scale of the Sahelian NDVI-rainfall relationship, i.e. the minimum geographical area with which a reliable relationship for can be built for the entire region. This could be interpreted as the size of a landscape unit that describes the natural arrangement of the phenomenon where variations of non-stationarity could be incorporated, while removing unnecessary bias and noise in the model.

### 5.2. Model comparison

The $R^2$ for GWR was higher than OLS for all years under examination (Fig. 5). OLS models are unstable in their predictions while

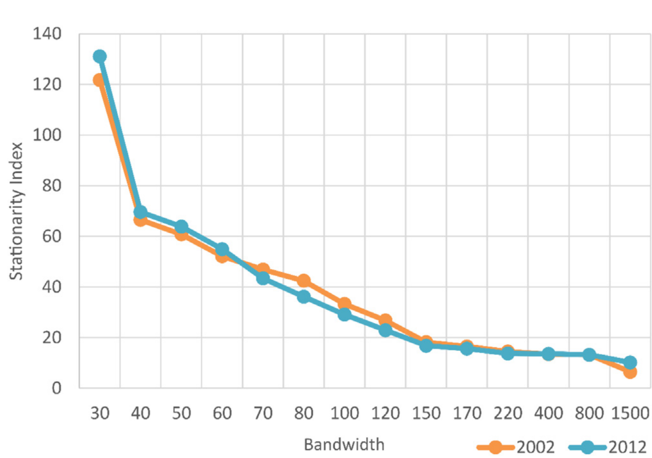


**Fig. 4.** Stationarity index (SI) for 2002 and 2012 at multiple bandwidths. The SI is that ratio of the interquartile range of standard errors for the GWR coefficients with twice the standard error of a spatially constant, global model (GLM). Values less than 1 on the y-axis indicate stationarity at the corresponding spatial scale.

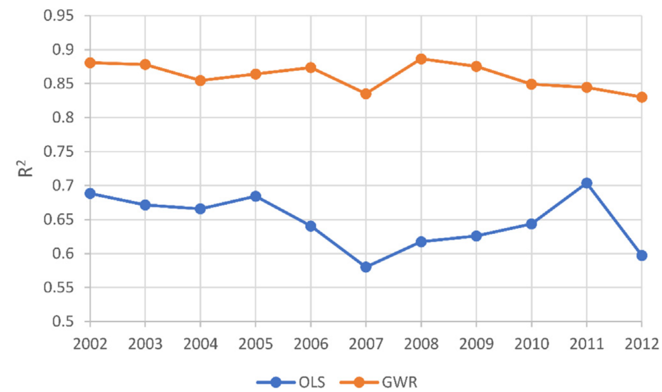


**Fig. 5.** Coefficient of determination ($R^2$) for OLS and GWR models between 2002 and 2012.

GWR deviates less over the study period. The $R^2$ values range between 0.60 and 0.70 for the OLS models, and between 0.84 and 0.88 for the GWR models. The ANOVA-based F-test shows that GWR produced statistically significant improvement over the OLS models in all years ($p < 0.01$). GWR explained more of variance and lowered AICc values, which accounts for changes in model complexity and degrees of freedom (Supplementary Table 2).

### 5.3. Spatial patterns of the NDVI – rainfall relationship

There is significant spatiotemporal variability in the strength of the NDVI-rainfall correlation throughout the Sahel (Fig. 6). The local fits are lower in wetlands around Lake Chad, and the Niger River near Mopti, Mali, as well as in the central and southwest parts of the Sahel. The influence of rainfall on NDVI variability in the growing season is low and perhaps other ecological or land use factors have stronger impact in these areas. High correlations ($R^2 > 0.5$) were found in the western Sahel (parts of Mauritania and Senegal) and the eastern Sahel (parts of Chad and most of Sudan) suggesting that rainfall is a very potent determinant in these areas. The local fits perform better in the dry year of 2002 than the relatively wet 2012, although the general clustering patterns are similar in both years.

The vast majority of the rainfall coefficients are positive, suggesting that an increase in rainfall relates to an increase in NDVI. However, the rate of increase differs significantly throughout the Sahelian belt (Fig. 7). The strength of the associations is higher in the growing season of 2002 than of 2012. In 2002, all local t-value estimates were statistically significant ($t = > 1.96$ or $t = < -1.96$). Comparing this to the land cover map, the significance of the t-values appears to be stronger in the transition zone between the bare soil and savanna. Although rainfall coefficients were lower in most of the Sahel in 2012 than 2002, a seemingly large cluster extending over Sudan had surprisingly high values that surpassed all the coefficient values of 2002. These findings suggest that NDVI was particularly sensitive to variations in rainfall over that area in that year. Instead of forming a spatially continuous sensitive geographical transitional zone, the Sahel forms clusters (Fig. 8) that operate as transitional passages between humid and hyper-arid environments, and their size being dependent on the amount of rain received.

Fig. 9 illustrates the linear temporal trends of rainfall coefficients between 2002 and 2012. The influence of rainfall as a local predictor of NDVI exhibits a negative trend in most of the western Sahel, while the trend is positive in the central regions of the region. The region of Kordofan in central Sudan displays a cluster of strong negative trends at its southern portion while high positive

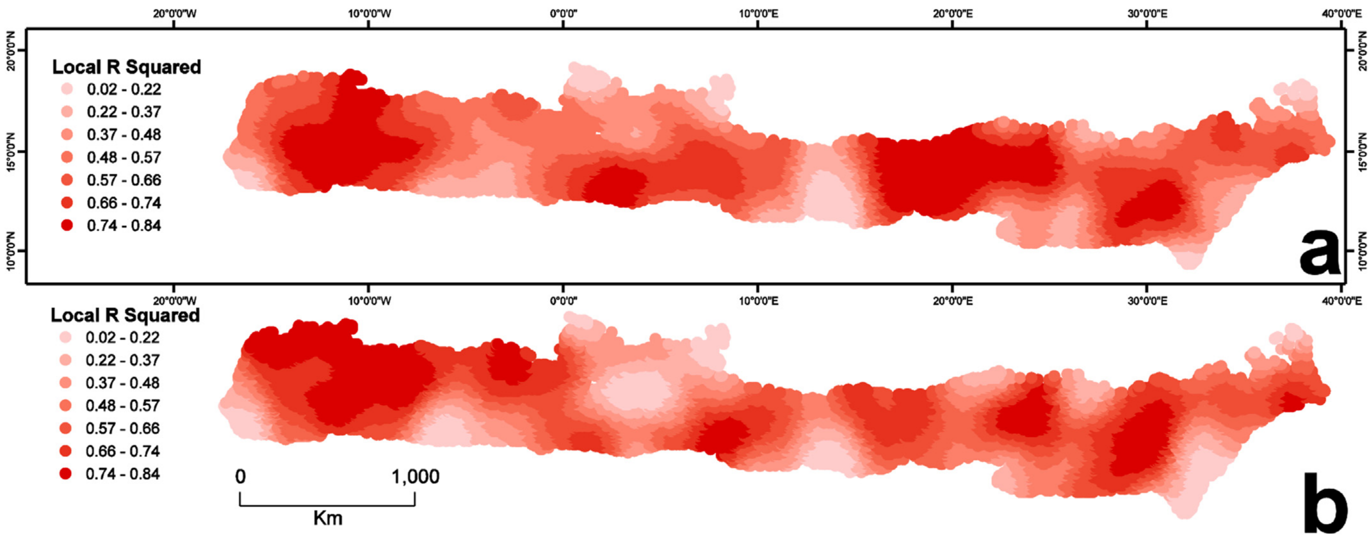


**Fig. 6.** Local $R^2$ patterns in 2002 and b) 2012. Low values indicate poor model performance and an indication for potentially missing variables in the model.

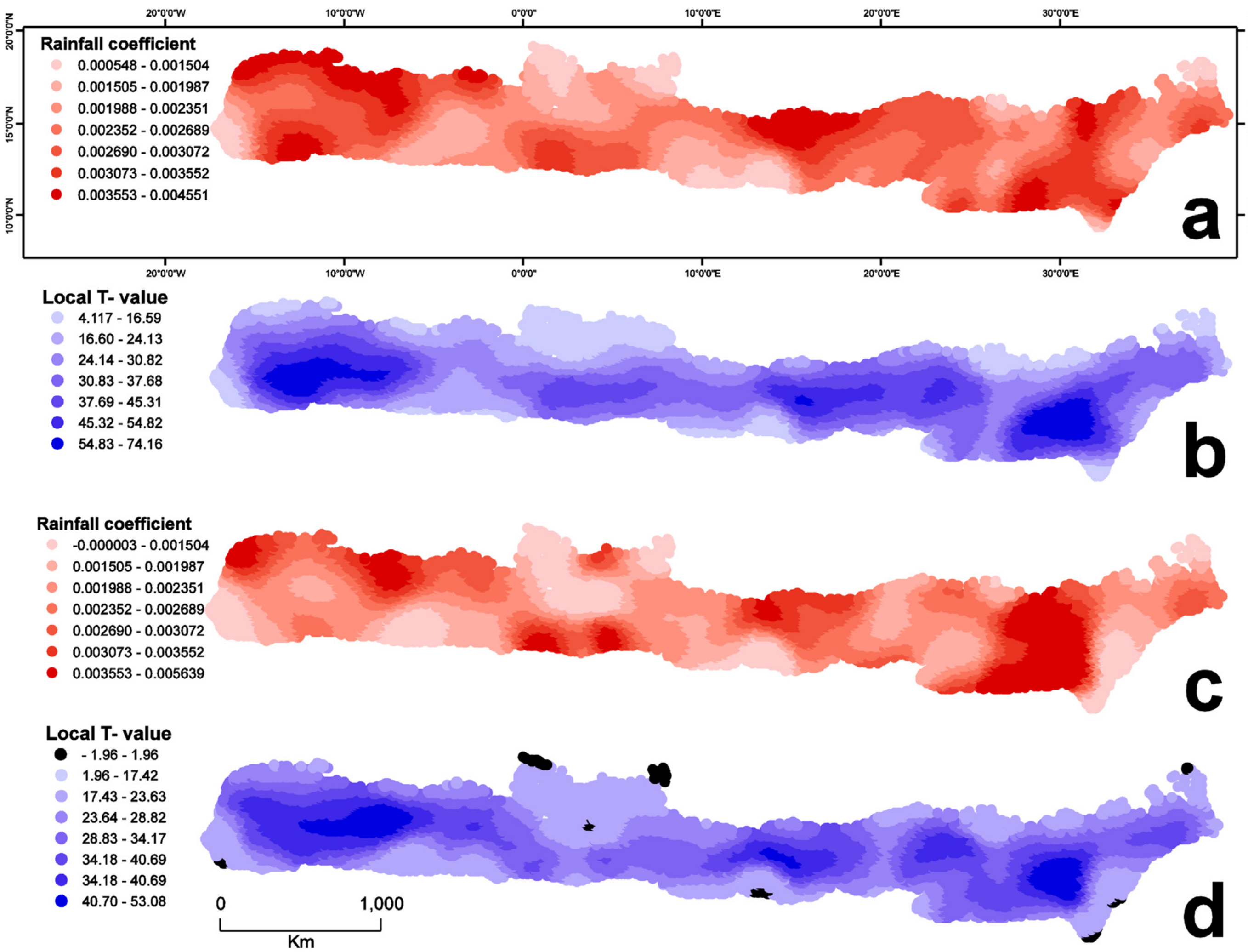


**Fig. 7.** Rainfall coefficients in the Sahel for a) 2002, local t-values of the rainfall coefficient for b) 2002, rainfall coefficients in the Sahel for c) 2012, local t-values of the rainfall coefficient for d) 2012. The rainfall coefficients depict the rate of change for a spatial unit of NDVI with an increase in a spatial unit of rainfall while the local t-values are computed as the ratio of a local estimate divided by its corresponding standard error. Values higher than 1.96 are statistically significant in the 95% confidence interval.

Please cite this article in press as: Georganos, S., et al., Examining the NDVI-rainfall relationship in the semi-arid Sahel using geographically weighted regression, Journal of Arid Environments (2017), http://dx.doi.org/10.1016/j.jaridenv.2017.06.004

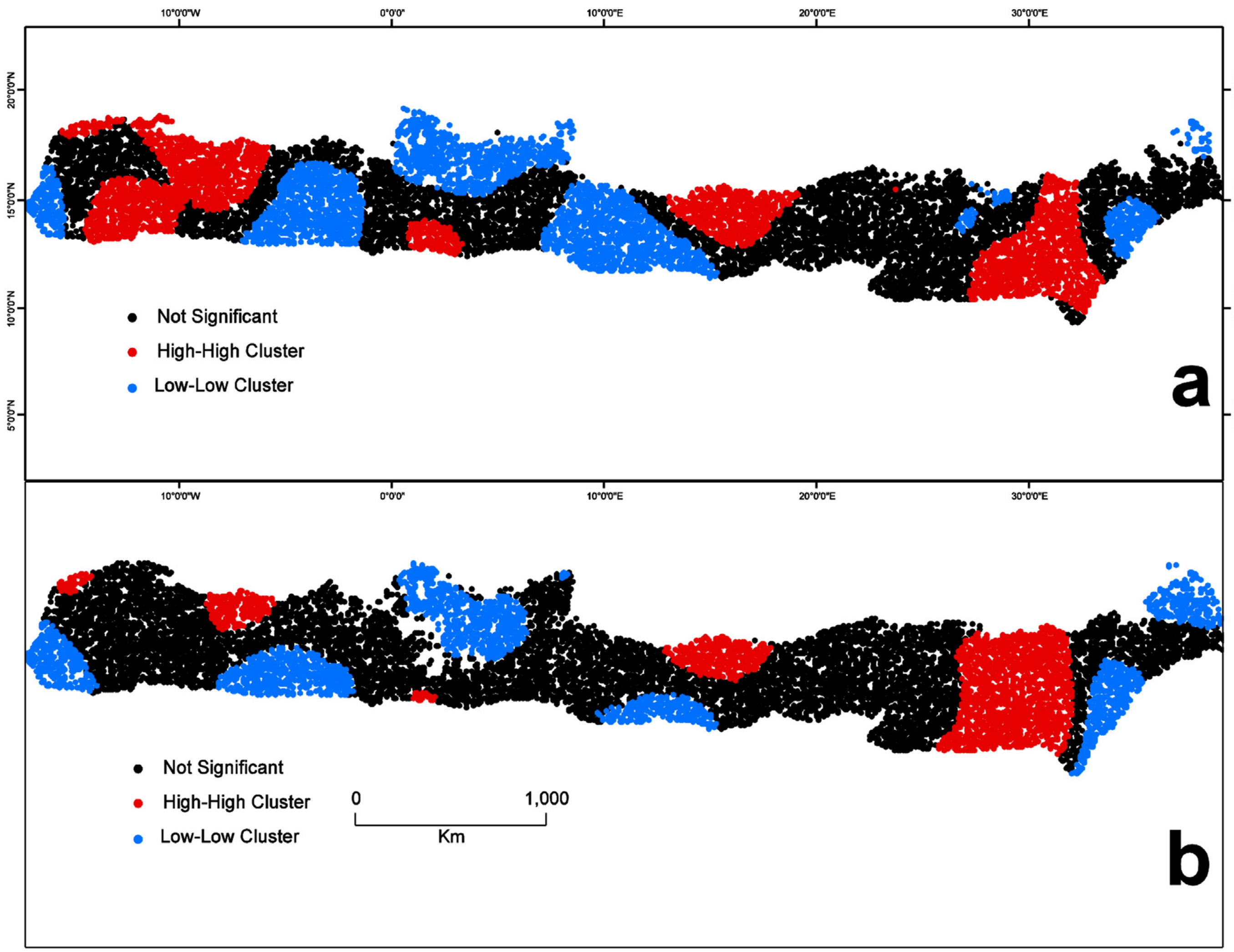


**Fig. 8.** Spatial clustering maps for the slope coefficient in the Sahel in a) 2002 and b) 2012, respectively. Red dots suggest significant clustering of similarly high values while blue dots significant clustering of low values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

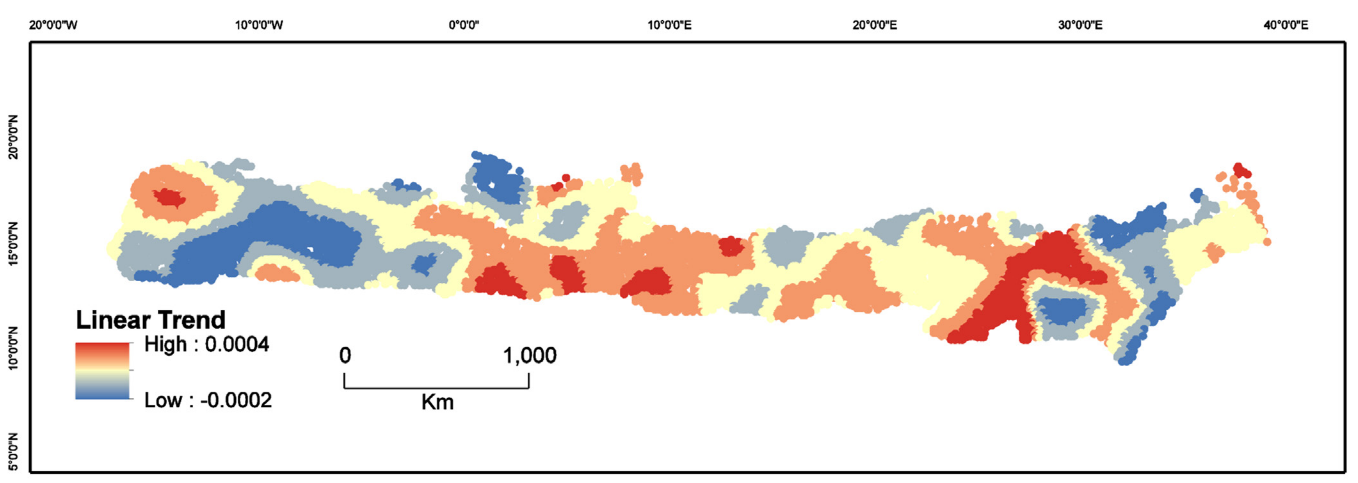


**Fig. 9.** Temporal trends of the rainfall coefficient between 2002 and 2012 over the growing season. The trends describe the temporal changes in the significance of rainfall as a local predictor of vegetation in the Sahel.

trends are found surrounding that region.

### 5.4. Predicted patterns

The scatterplots of the observed and predicted values are presented in Fig. 10. The global models were not consistent, underestimating high values of NDVI and overestimating low values. This is to be expected, since a constant set of parameters was used to capture a relationship over a very large area with high spatial heterogeneity. The local modelling approach managed to produce more accurate estimates by taking into account local characteristics and incorporating sample coordinates (Supplementary Table 2).

Figs. 11 and 12 show observed and predicted spatial patterns of NDVI for 2002 and 2012, respectively. The OLS model produced more generalized patterns that hindered local variability in the NDVI values across the Sahel (RMSE = 0.30 in 2002, RMSE = 0.39 in 2012). Similarly, predictions produced by combining OLS models in different land cover classes produced only slightly better results (RMSE = 0.27 in 2002, RMSE = 0.35 in 2012). Both models displayed the actual gradient of lower NDVI values in the north and higher values to the south. However, the GWR model had higher precision because it took into account the spatial variation in the NDVI—rainfall relationship and regional information (RMSE = 0.18 in 2002, RMSE = 0.24 in 2012).

## 6. Discussion

The results demonstrated that the NDVI—rainfall relationship is not stationary throughout the Sahel during the growing season in the years under examination. This non-stationarity was independent of land cover (Supplementary Table 1) as suggested by the comparison of standard deviations of global linear model with the interquartile range of GWR (Supplementary Tables 3 and 4). Our findings suggest that rainfall is a significantly more potent predictor of NDVI once spatial non-stationarity can be incorporated into the regression model. The relationship was positive across most of the region, however, negative or very weak relationships were observed in some locations. Moreover, the significance of the association varied dramatically through space. GWR functions as a continuously varying detector of geographical relationships by incorporating local information with significantly improved performance over the global models (Supplementary Table 2), explaining higher variance, and effectively reducing autocorrelation in the residuals (Supplementary Figs. 1 and 2). The underlying mechanism is that GWR can take into account variability within land cover classifications, accounting for species composition and distribution, and other factors that have a strong local component such as soil type or human disruption of ecological communities and locally unique climatic conditions (Propastin, 2009). These local variations demonstrated that the spatial patterns of NDVI are better correlated with rainfall in some parts of the Sahel than others. Herrmann et al. (2005) suggested that rainfall is the dominant determinant of vegetation growth in Sahel, but other factors such as human interference could also play a role. In this study, the southern parts of the Sahel generally exhibited weaker correlations (Fig. 7), for example the wetlands around Lake Chad.

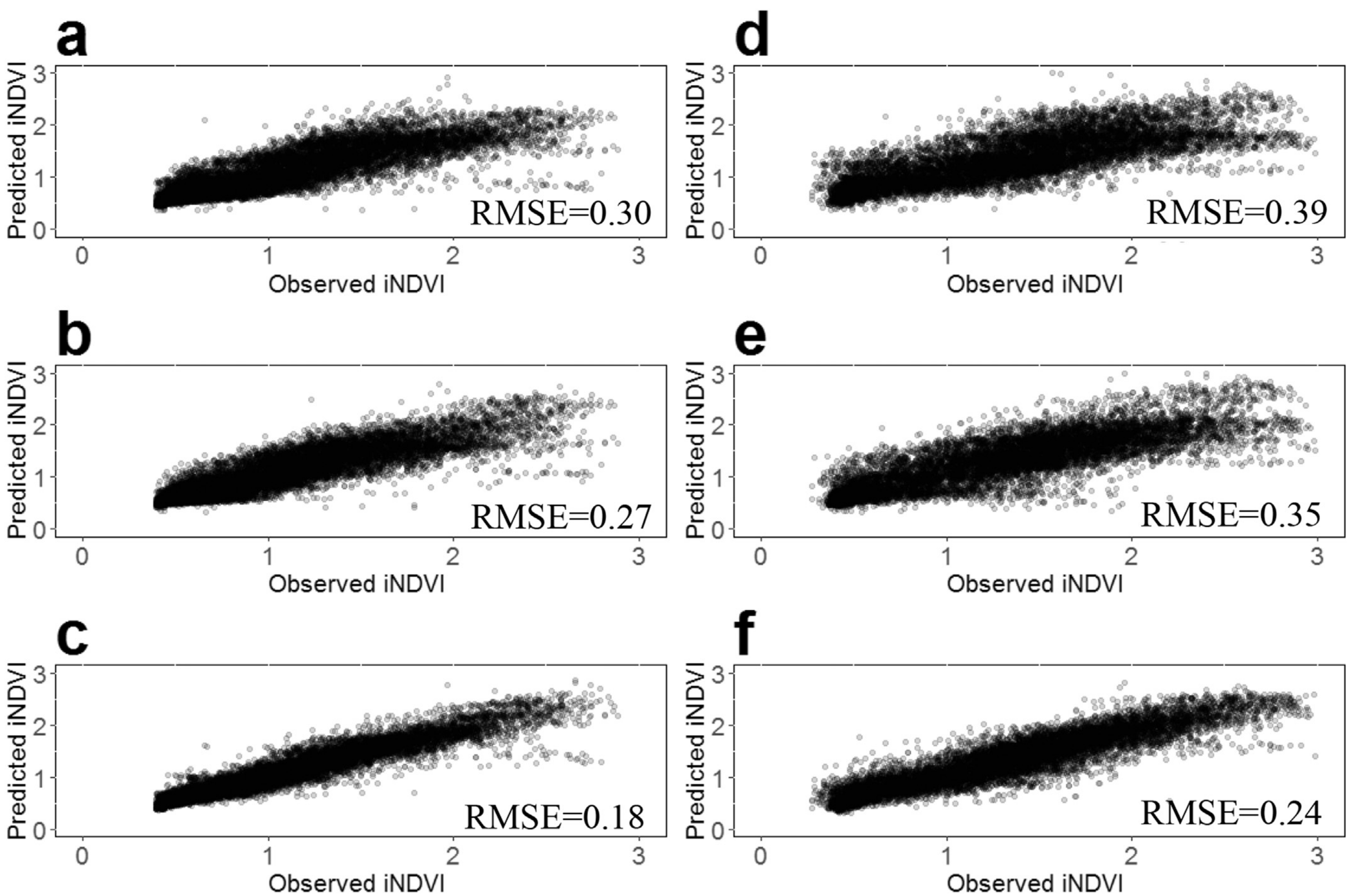


**Fig. 10.** Scatter plots of observed and simulated NDVI for the growing season. a) OLS model in 2002, b) Combined OLS models in separate land cover categories in 2002, c) GWR model in 2002, d) OLS model in 2012, e) Combined OLS models in separate land cover categories in 2012, f) GWR model in 2012. RMSE refers to the Root Mean Square Error.

Please cite this article in press as: Georganos, S., et al., Examining the NDVI-rainfall relationship in the semi-arid Sahel using geographically weighted regression, Journal of Arid Environments (2017), http://dx.doi.org/10.1016/j.jaridenv.2017.06.004

ARTICLE IN PRESS



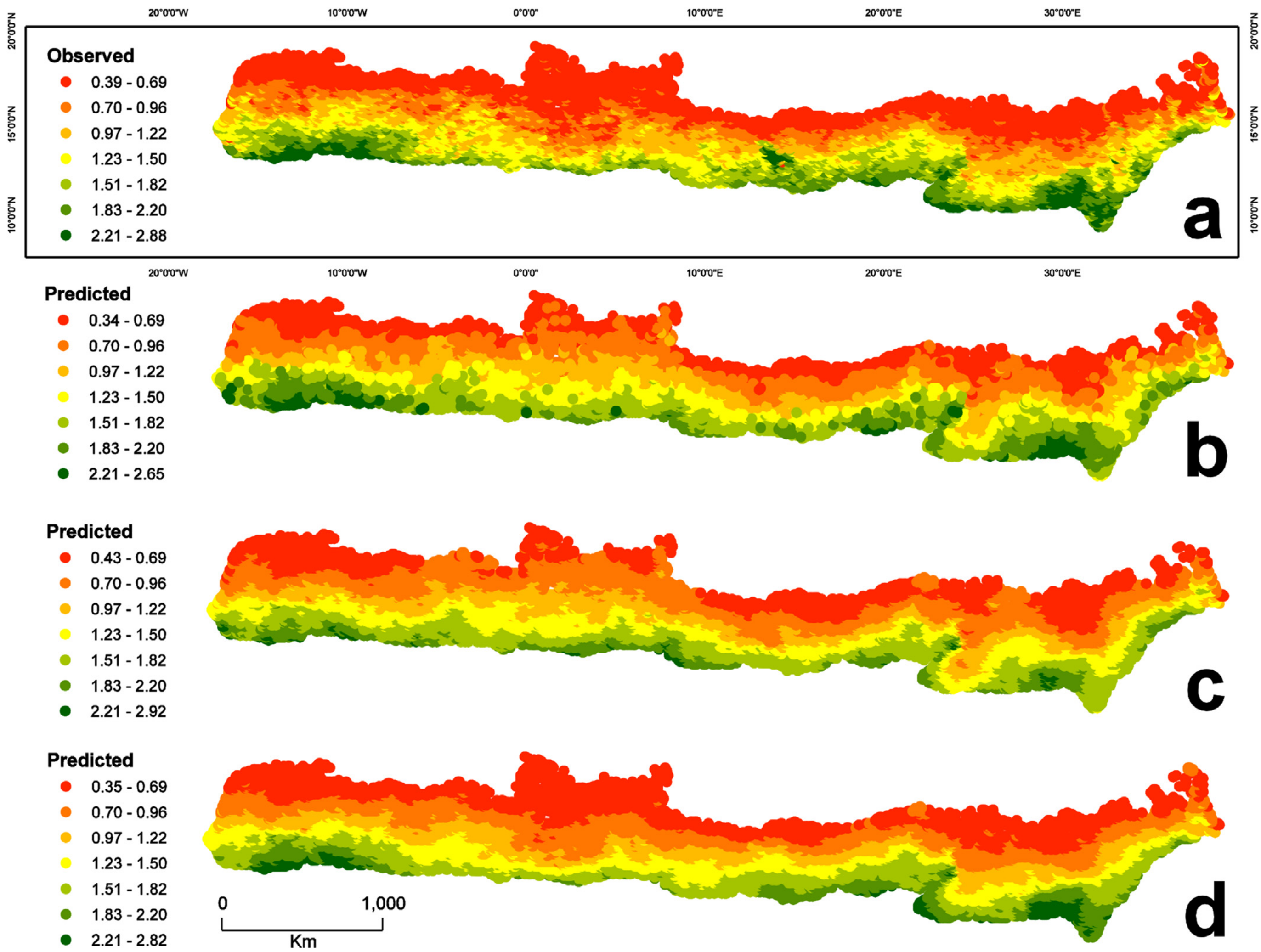


**Fig. 11.** Observed and Predicted spatial patterns of NDVI in 2002. a) Observed patterns, b) Combined OLS models in separate land cover categories, c) Global OLS model, d) GWR model.

This is probably due to the abundance of water and different irrigation practices (Fox and Rockström, 2003) coupled with increased human activity.

Hameed and Bannari (2016) demonstrated that the temporal correlations between NDVI and rainfall in semi-arid Sudan are highly correlated ($R^2 = 0.60–0.80$). These findings agree with the GWR results suggesting that spatially explicit modelling is an effective alternative to global models. The flexibility of the proposed method could be applied to modify rain use efficiency (RUE, Fensholt and Rasmussen (2011)) analysis by incorporating local characteristics. We suggest the use of the slope coefficients of GWR models as a spatial proxy for RUE, which will produce a localized perspective on ecosystem degradation rather single per-pixel analysis in a global model. The strong negative trends in the cluster located in Kordofan (Fig. 9) could be linked with soil degradation (Elgubshawi, 2008) or a change in grazing practices. The negative trends in most of the Western Sahel can be attributed to increases in woody cover (Brandt et al., 2016) as woody vegetation has been shown to be resistant to intra- and inter-annual variability of rainfall and soil moisture (Huber et al., 2011).

The relationship between vegetation and climatic variables such as rainfall has been shown to be scale dependent (Propastin et al., 2008; Zhao et al., 2015), and the selection of an appropriate bandwidth in GWR is crucial (Gao and Li, 2011). The bandwidth choice is a trade-off between variance in the local estimates and bias in the model. Usually, this decision is based on AICc, however, with large sample sizes AICc can suggest optimality in models with extremely small bandwidths, inflated $R^2$ values and large standard errors (Propastin et al., 2008). This hinders meaningful inferences and inserting a large amount of noise in the outputs. In this study, AICc was suitable for comparing OLS and GWR after the appropriate bandwidth was selected based on SI. However, it must be noted that in bivariate models with very large sample sizes ($n > 50,000$), it is very likely to encounter strong collinearity problems in the local explanatory terms. Although smaller bandwidths can be useful for exploratory purposes, it is strongly suggested that covariance matrices, local correlation coefficients, and condition indexes be applied before selecting a bandwidth, even when it is suggested by measures such as AICc. An alternative to AICc for large sample sizes is to employ stricter measures for bandwidth selection, such as Bayesian information criterion, which penalizes smaller bandwidths more and could be of potential use in further research (Nakaya, 2001).

The computation of SI at incrementing spatial scales provided important information with regard to the scale dependency of NDVI and rainfall in the Sahelian growing season. The two years that were selected for detailed examination showed that SI stabilized at roughly 160 km (Fig. 4) and the regressions constructed at that

Please cite this article in press as: Georganos, S., et al., Examining the NDVI-rainfall relationship in the semi-arid Sahel using geographically weighted regression, Journal of Arid Environments (2017), http://dx.doi.org/10.1016/j.jaridenv.2017.06.004

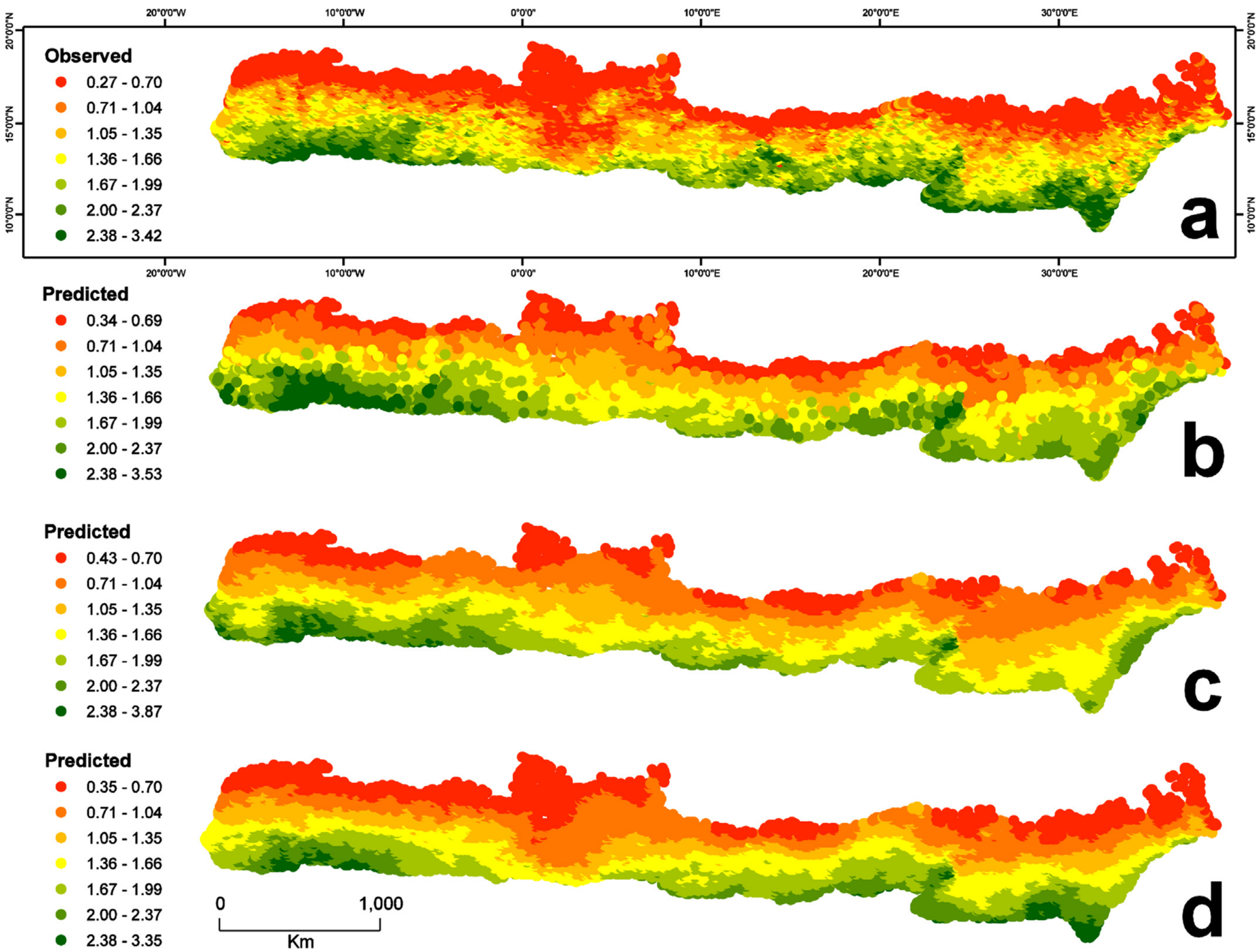


**Fig. 12.** Observed and Predicted spatial patterns of NDVI in 2012. a) Observed patterns, b) Combined OLS models in separate land cover categories, c) Global OLS model, d) GWR model.

scale were more reliable and stable. This is due to the fact that the condition indices and local correlation coefficients did not suggest any spurious local correlation in the coefficients even at the cost of introducing more bias than smaller bandwidths. Other studies in ecological transition zones similar to the Sahel found that the stability scales were found to be between 180 and 500 km (Gao et al., 2012; Zhao et al., 2015). It has to be noted that in those studies annual comparisons were performed, while in this study the focus was on the growing season. By analysing the data only in the months when most of the rainfall occurs and vegetation growth takes place, much of the unnecessary noise from the NDVI signal in the dry season can be eliminated.

## 7. Conclusion

GWR allows regression parameters to vary through space, producing different temporal and spatial patterns regarding the strength of the correlation and the significance of the relationship. By mapping the local diagnostics, the spatial heterogeneity and non-stationary of the NDVI-rainfall relationship are illustrated. Local coefficient values describe how well local models fit the observations and the nature of the relationships. This study examined the spatial patterns of NDVI and rainfall in the Sahel during the growing season between 2002 and 2012. The results validated our hypothesis that GWR is a viable alternative to OLS modelling in heterogeneous areas that are sensitive to environmental change. The results show that parts of the study area were particularly sensitive to variability in rainfall formed large clusters that connected humid and arid climatic zones. In these areas, rainfall appears to be the dominant determinant in understanding the distribution of vegetation. Moreover, regions mainly located around wetlands exhibited weak relationships with rainfall, indicating the need for the incorporation of additional variables to explain variations in NDVI. The GWR approach produced better predictions, lower autocorrelation in the residuals and highlighted interesting local variations. As such, GWR is strongly suggested as both an explanatory and exploratory method in spatiotemporal analysis and environmental modelling where spatial constancy in relations between variables is questionable.

## Acknowledgements

This work is a part of S.G.'s graduate thesis under the Master of Science in Geomatics program at Lund University. A.M.A. received support from the Swedish National Space Board (under contract No. 100/11) and the Lund University Centre for studies of Carbon Cycle and Climate Interactions (LUCCI). The authors thank the reviewers for their comments that helped improve the manuscript.

## Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jaridenv.2017.06.004.

## References


Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Contr 19, 716–723.

Anyamba, A., Tucker, C.J., 2005. Analysis of Sahelian vegetation dynamics using NOAA-AVHRR NDVI data from 1981–2003. J. Arid. Environ. 63, 596–614.

Brandt, M., Hiernaux, P., Tagesson, T., Verger, A., Rasmussen, K., Diouf, A.A., Mbow, C., Mougin, E., Fensholt, R., 2016. Woody plant cover estimation in drylands from Earth Observation based seasonal metrics. Remote Sens. Environ. 172, 28–38. http://dx.doi.org/10.1016/j.rse.2015.10.036.

Brandt, M., Mbow, C., Diouf, A.A., Verger, A., Samimi, C., Fensholt, R., 2015. Ground- and satellite-based evidence of the biophysical mechanisms behind the greening Sahel. Glob. Chang. Biol. 21, 1610–1620.

Brunsdon, C., Fotheringham, S., Charlton, M., 1998. Geographically weighted regression. J. R. Stat. Soc. Ser. D. (The Stat. 47, 431–443.

Dardel, C., Kergoat, L., Hiernaux, P., Mougin, E., Grippa, M., Tucker, C.J., 2014. Re-greening Sahel: 30 years of remote sensing data and field observations (Mali, Niger). Remote Sens. Environ. 140, 350–364. http://dx.doi.org/10.1016/j.rse.2013.09.011.

Eklundh, L., Olsson, L., 2003. Vegetation index trends for the african Sahel 1982–1999. Geophys. Res. Lett. 30.

Elgubshawi, A.A.I., 2008. Soil Degradation: its Aspects and Modelling a Case Study of Northeast Nuba Mountain South Kordofan State, Sudan. Institute of G. Ecology and E. protection Prof. Dr. R. Gloaguen TU Bergakademie, Institute of Geology.

Fabricante, I., Oesterheld, M., Paruelo, J.M., 2009. Annual and seasonal variation of NDVI explained by current and previous precipitation across Northern Patagonia. J. Arid. Environ. 73, 745–753. http://dx.doi.org/10.1016/j.jaridenv.2009.02.006.

Fensholt, R., Proud, S.R., 2012. Evaluation of earth observation based global long term vegetation trends—comparing GIMMS and MODIS global NDVI time series. Remote Sens. Environ. 119, 131–147.

Fensholt, R., Rasmussen, K., 2011. Analysis of trends in the Sahelian "rain-use efficiency"using GIMMS NDVI, RFE and GPCP rainfall data. Remote Sens. Environ. 115, 438–451.

Fotheringham, A.S., Brunsdon, C., Charlton, M., 2003. Geographically Weighted Regression: the Analysis of Spatially Varying Relationships. John Wiley & Sons.

Fotheringham, A.S., Charlton, M.E., Brunsdon, C., 2001. Spatial variations in school performance: a local analysis using geographically weighted regression. Geogr. Environ. Model 5, 43–66.

Fox, P., Rockström, J., 2003. Supplemental irrigation for dry-spell mitigation of rainfed agriculture in the Sahel. Agric. Water Manag. 61, 29–50. http://dx.doi.org/10.1016/S0378-3774(03)00008-8.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., et al., 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Sci. data 2.

Gao, J., Li, S., 2011. Detecting spatially non-stationary and scale-dependent relationships between urban landscape fragmentation and related factors using geographically weighted regression. Appl. Geogr. 31, 292–302.

Gao, Y., Huang, J., Li, S., Li, S., 2012. Spatial pattern of non-stationarity and scale-dependent relationships between NDVI and climatic factors—a case study in Qinghai-Tibet Plateau, China. Ecol. Indic. 20, 170–176.

Gaughan, A.E., Waylen, P.R., 2012. Spatial and temporal precipitation variability in the Okavango–Kwando–Zambezi catchment, southern Africa. J. Arid. Environ. 82, 19–30.

Hameed, N.A., Bannari, A., 2016. The relationship between vegetation and rainfall in Central Sudan. Int. J. Remote Sens. Appl. 6, 30–40. http://dx.doi.org/10.14355/ijrsa.2016.06.004.

Herrmann, S.M., Anyamba, A., Tucker, C.J., 2005. Recent trends in vegetation dynamics in the African Sahel and their relationship to climate. Glob. Environ. Chang. 15, 394–404. http://dx.doi.org/10.1016/j.gloenvcha.2005.08.004.

Hickler, T., Eklundh, L., Seaquist, J.W., Smith, B., Ardö, J., Olsson, L., Sykes, M.T., Sjöström, M., 2005. Precipitation controls Sahel greening trend. Geophys. Res. Lett. 32.

Hu, M., Li, Z., Wang, J., Jia, L., Liao, Y., Lai, S., Guo, Y., Zhao, D., Yang, W., 2012. Determinants of the incidence of hand, foot and mouth disease in China using geographically weighted regression models. PLoS One 7, e38978.

Huber, S., Fensholt, R., Rasmussen, K., 2011. Water availability as the driver of vegetation dynamics in the African Sahel from 1982 to 2007. Glob. Planet. Change 76, 186–195. http://dx.doi.org/10.1016/j.gloplacha.2011.01.006.

Kalogirou, S., 2003. The Statistical Analysis and Modelling of Internal Migration Flows within England and Wales.

Kalogirou, S., Kalogirou, M.S., 2015. Package "lctools".

Latham, J., Cumani, R., Rosati, I., Bloise, M., 2014. Global Land Cover Share (GLC-SHARE) Database Beta-release Version 1.0-2014. FAO, Rome, Italy.

Lee, E., He, Y., Zhou, M., Liang, J., 2015. Potential feedback of recent vegetation changes on summer rainfall in the Sahel. Phys. Geogr. 36, 449–470. http://dx.doi.org/10.1080/02723646.2015.1120139.

Martiny, N., Camberlin, P., Richard, Y., Philippon, N., 2006. Compared regimes of NDVI and rainfall in semi-arid regions of Africa. Int. J. Remote Sens. 27, 5201–5223. http://dx.doi.org/10.1080/01431160600567787.

Moran, P.A.P., 1948. The interpretation of statistical maps. J. R. Stat. Soc. Ser. B 10, 243–251.

Nakaya, T., 2001. Local spatial interaction modelling based on the geographically weighted regression approach. GeoJournal 53, 347–358.

Nakaya, T., Fotheringham, A.S., Charlton, M., Brunsdon, C., 2009. Semiparametric Geographically Weighted Generalised Linear Modelling in GWR 4.0.

Nicholson, S.E., Davenport, M.L., Malo, A.R., 1990. A comparison of the vegetation response to rainfall in the Sahel and East Africa, using normalized difference vegetation index from NOAA AVHRR. Clim. Change 17, 209–241.

Olsson, L., Eklundh, L., Ardö, J., 2005. A recent greening of the Sahel—trends, patterns and potential causes. J. Arid. Environ. 63, 556–566.

Osborne, P.E., Foody, G.M., Suárez-Seoane, S., 2007. Non-stationarity and local approaches to modelling the distributions of wildlife. Divers. Distrib. 13, 313–323. http://dx.doi.org/10.1111/j.1472-4642.2007.00344.x.

Pinzon, J.E., Tucker, C.J., 2014. A non-stationary 1981-2012 AVHRR NDVI3g time series. Remote Sens. 6, 6929–6960. http://dx.doi.org/10.3390/rs6086929.

Propastin, P.A., 2009. Spatial non-stationarity and scale-dependency of prediction accuracy in the remote estimation of LAI over a tropical rainforest in Sulawesi, Indonesia. Remote Sens. Environ. 113, 2234–2242.

Propastin, P., Kappas, M., Erasmi, S., 2008. Application of geographically weighted regression to investigate the impact of scale on prediction uncertainty by modelling relationship between vegetation and climate. Int. J. Spat. data infrastructures Res. 3, 73–94.

Tian, F., Fensholt, R., Verbesselt, J., Grogan, K., Horion, S., Wang, Y., 2015. Evaluating temporal consistency of long-term global NDVI datasets for trend analysis. Remote Sens. Environ. 163, 326–340. http://dx.doi.org/10.1016/j.rse.2015.03.031.

Udelhoven, T., Stellmes, M., del Barrio, G., Hill, J., 2009. Assessment of rainfall and NDVI anomalies in Spain (1989–1999) using distributed lag models. Int. J. Remote Sens. 30, 1961–1976. http://dx.doi.org/10.1080/01431160802546829.

Vrieling, A., De Leeuw, J., Said, M.Y., 2013. Length of growing period over africa: variability and trends from 30 years of NDVI time series. Remote Sens. 5, 982–1000. http://dx.doi.org/10.3390/rs5020982.

Wang, Q., Ni, J., Tenhunen, J., 2005. Application of a geographically-weighted regression analysis to estimate net primary production of Chinese forest ecosystems. Glob. Ecol. Biogeogr. 14, 379–393.

Wang, T., Kou, X., Xiong, Y., Mou, P., Wu, J., Ge, J., 2010. Temporal and spatial patterns of NDVI and their relationship to precipitation in the Loess Plateau of China. Int. J. Remote Sens. 31, 1943–1958. http://dx.doi.org/10.1080/01431160902929263.

Zhao, Z., Gao, J., Wang, Y., Liu, J., Li, S., 2015. Exploring spatially variable relationships between NDVI and climatic factors in a transition zone using geographically weighted regression. Theor. Appl. Climatol. 120, 507–519.