7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. Edited by M. Caetano and M. Painho.

Reducing uncertainty in analysis of relationship between vegetation patterns and precipitation

Pavel Propastin[1], Nadiya Muratova[2] and Martin Kappas[1]

1 Department of Geography, Georg-August-University
Goldschmidtstr. 3, 37077 Goettingen, Germany
Tel.: + 049 0551 389 487; Fax: + 049 0551
ppavel@gmx.de; mkappas@uni-goettingen.de

2 Institute for Space Research and Earth Observation
Shevchenko Str. 15, 480100, Almaty, Kazakhstan
Tel.: + 007 3272 494 274; Fax: + 007 3272 918 077
nmuratova@mail.com

Abstract

The spatial relationship between vegetation patterns and rainfall and its trend over the period 1985-2001 in desert, semi-desert and steppe grassland of the Middle Kazakhstan was investigated with Normalized Difference Vegetation Index (NDVI) images (1985-2001) derived from the Advanced Very High Resolution Radiometer (AVHRR), and rainfall data from weather stations. The growing season relationship was examined using conventional, global, ordinary least squares (OLS) regression technique, and a local regression technique known as geographically weighted regression (GWR). Regression models between NDVI and precipitation for every analysis year (1985-2001) were calculated using separately the both statistic approaches, the OLS and the GWR. The study found a presence of high spatial and temporal non-stationarity in the strength of relationships and regression parameters. The ordinary least squares regression model had been applied to the whole study area was superficially strong (R^2 *= 0.63), however it gave no local description of the relationship. Applying the OLS at the scale of the separate land cover classes revealed a different response of various vegetation types to rainfall within the study area. The strength of the relationship between NDVI and rainfall increased in order from desert (* R^2 *= 0.36), to semi-desert (* R^2 *= 0.52), and to steppe grassland (* R^2 *= 0.67). The approach of geographically weighted regression provided considerably stronger relationships from the same data sets (mean value of* R^2 *= 0.88), as well as highlighted local variations within the land cover classes. Relationships between vegetation patterns and rainfall amounts are generally assumed to be spatially and temporally stationary. This assumption was not satisfied in this study. The study found that the relationship varied significantly in space and time. In such circumstances the results provided by a global regression model were uncertain and incorrect presented the relationship between the both variables locally. The application of local regression techniques such as GWR, may reveal local patterns of relationship and significantly reduces the uncertainties of calculations.*

Keywords: NDVI, precipitation, regression modeling

1 Introduction

Climate is one of the most important factors affecting vegetation condition. Therefore, evaluation of the quantitative relationship between vegetation patterns and climate is an important object of applications of remote sensing at regional and global scales. The Normalized Difference Vegetation Index (NDVI) is established to be highly correlated to green-leaf density and can be viewed as a proxy for above-ground biomass (Tucker and Sellers, 1986). Spatial correlations between NDVI and climatic factors are investigated in many research works. Particularly well correlation in the arid regions show NDVI and rainfall, the relationship between NDVI and temperature are reported to be weaker but also significant (Yang *et al.,* 1998; Richard and Poccard, 1998; Ji and Petters, 2004; Li *et al.,* 2004; Wang *et al.,* 2001).

Regression and correlation techniques were the common empirical approaches used to quantify the relationships in most of these studies. However, the conventional statistical regression method (global OLS regression) assuming the relationship to be spatially stationary is usually not adequate for spatially differenced data, especially by quantifying relationships at regional or global scales. There is many cases that show non-stability of this relationship in space (Fotheringham *et al.,* 2002; Foody, 2003; Foody, 2004; Wang *et al.,* 2005). The causes of variance of relationship between NDVI and its explanatory variables are known to be spatial variations in properties such as vegetation type, soil type, soil moisture (Wang et al., 2001; Yang et al., 1998; Ji & Peters, 2004). The differences between regression models established at different locations can be large with both the magnitude and sign of the model parameters varying.

Local regression techniques, such as geographically weighted regression (GWR) help to overcome this problem and calculate the regression model parameters varying in space. This technique provides a more appropriate and accurate basis for modelling relationship between various spatial variables and significantly reduces uncertainty in model prediction (Brunsdon *et al.,* 2001; Fotheringham *et al.,* 2002). At the field of remote sensing there are only rare studies applying local regression techniques for analysis of spatial relationships between remotely sensing data and climatic variables (Foody, 2003; Foody, 2004; Wang *et al.,* 2005).

In this paper, we analysed relationship between NOAA/AVHRR-NDVI and rainfall amounts for every year from the study period 1985-2000. First, we fitted regression models between this both variables by using the conventional ordinary least squares method. Second, we investigated spatial and temporal non-stationarity in the NDVI-rainfall relationship through utilization of GWR. Third, we assessed accuracy in prediction of NDVI from rainfall amount achieved by application of various regression models, global and local and demonstrated that the prediction accuracy increases through taking into account local variability in the relationship between the two variables.

2 Study area

The study area is located in the middle part of Kazakhstan between 46 and 50° northern latitude and 72° and 75° eastern longitude. In terms of surface structure the study area is divided into two large regions: a plateau of rolling upland in the southern, western, and northern parts with average elevations between 300-700 meter; hills and low mountains in the central and north-eastern parts with elevations 700-1100 meter.

The climate of the region is dry, cold and high continental. Average annual precipitation is above 250-300 mm per year in the north of the study area, and below 150 mm in the south. The most part of precipitation falls during warm period from March to October. The temperature amplitude is relative high: average January temperature is below –12° C and average July temperature is about 26-28° C.

The south of the study region is vegetated by sagebrush and perennial saltwort associations (Figure 1c). The northern section of the study is occupied by steppe vegetation. The semi-desert vegetation occupying the mid of the study area.

3 Data and methods

3.1 NOAA-AVHRR NDVI

The NDVI images used in this research represent 10-day Maximum Value Composites covering the study area for the years 1985-2000. The data for period were calibrated for post-launch sensor degradation by using methods described by Los (1993). In addition to that, we removed noisy pixel areas characterized by exceptionally low NDVI values relatively to their pixel neighbourhood. The 10-day NDVI composites were integrated to mean monthly and then to mean growing season values for each of the analysis years.

3.2 Precipitation data

The climate data in the study consist of 10-day rainfall data collected and calculated by the National Hydro-meteorological Centre of Kazakhstan for 9 climate stations placed in the study area for the period April-October 1985-2001. For preparation of gridded maps of precipitation we summed 10-day rainfall data of each meteorological station for every year. Interpolation of data between stations was made using as gridding method known as the inverse distance to a power. After that all gridded maps were resized to pixel resolution of the NDVI data.

3.3 Regression models

The simple linear regression model, usually fitted by ordinary least squares methods (OLS), is:

$$y = \alpha + \beta * x + \varepsilon \qquad (1)$$

Where α is the intercept, β represents the slope coefficient for independent variable x, ε is random error.

In this model, the two variables to be related are y, the dependent variable (typically NDVI), and x, the independent variable (one of the environmental predictors, such as rainfall). The general assumption of the global regression model is that the relationship under study is spatially constant, and thus, the estimated parameters remain constant over space. However, in most cases, the relationship varies in space.

Geographically weighted regression is a local regression technique which overcomes the problem of non-stationarity through local disaggregating global statistics and calculates the relationship between NDVI and its predicting variables for every point. A detailed description of geographically weighted regression and its treatments is given in (Fotheringham et al., 2002). Here we provide only a simple illustration.

The regression model is calibrated on all data that lie within the region described around a regression point and the process is repeated for all regression points. The resulting local parameter estimates can then be mapped at the locations of the regression points to view possible non-stationarity in the relationship being examined. The size of the moving window (kernel) is less than the region size and can be varied from one point to another. With the GWR the relationship between the variables can be expressed as

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^{k} x_{ij}\beta_j(u_i, v_i) + \varepsilon_i \qquad (2)$$

where there are j = 1, k explanatory variables,

ε_i is a random error term and the location for each observation is defined by the coordinates (u_i, v_i),

$\beta_0 - \beta_k$ are the parameters of the model with $\beta_j(u_i, v_i)$ a realization of the continuous function $\beta_j(u, v)$ at location *i*.

GWR works in the way that each data point is weighted by its distance from the regression point. The geographical weighting is achieved through a spatial kernel. The weighting of an observation in the analysis is not constant, but a function of location. Data from observations close to point *i* are weighted more than data from observations father away. The matrix form of parameter estimation for *i* is expressed as

$$\hat{\alpha}(\theta), \hat{\beta}(\theta) = (X^T W(\theta) X)^{-1} X^T W(\theta) y \qquad (3)$$

where $\hat{\alpha}$ and $\hat{\beta}$ are intercept and slope parameter in location *i*,

$W(\theta)$ is weighting matrix whose diagonal elements represent the geographical weighting associated with each site at which measurements were made for location of *i*.

Spatial weighting function can be calculated by several various methods. For fixed kernel size, the weight of each point can be calculated by applying Gaussian function

$$w_{ij} = \exp[-1/2(d_{ij}/b)]^2 \qquad (4)$$

where d_{ij} is the distance between regression point *i* and data point *j*,

b is a bandwidth expressed in pixels.

To establish an appropriate bandwidth, *b*, we used the cross-validation approach (CV) described in (Fotheringham *et al.,* 2002) which determines *b* by minimisation of the sum of squared errors between predicted variables and those observed. According to the reference, the equation for the *cross-validation sum of squared errors* CVSS is statistically expressed as:

$$CVSS = \sum_{i=1}^{n} [y_i - \hat{y}_i(b)]^2 \qquad (5)$$

where y_i is the observed value,

$\hat{y}_i(b)$ is the fitted value of y_i for bandwidth *b*.

As general rule, the lower the CVSS, the closer the approximation of the model to reality. The best model is the one with the smallest CVSS. For our regression model, the bandwidth of 5 pixel was decided to be the most appropriate.

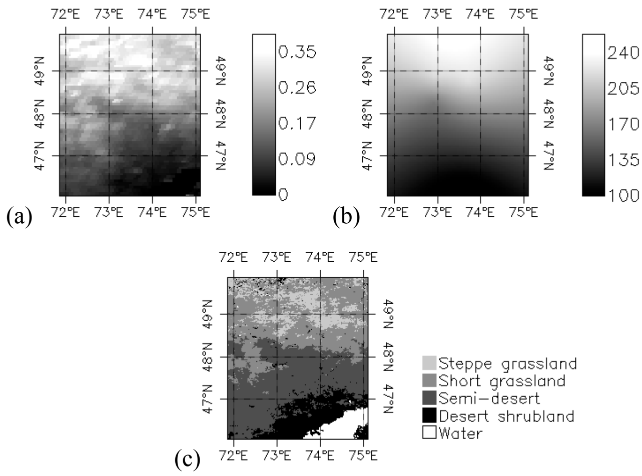


Figure 1 Average growing season NDVI (a), average rainfall amount in mm (b), and land cover types (c).

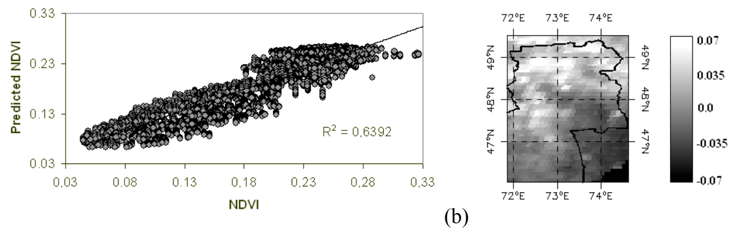
4 Results

4.1 Spatial patterns of NDVI and rainfall

Vegetation and rainfall spatial distribution in the study area display similar spatial patterns. Average precipitation increased markedly from south to north: from about 100 mm in the desert to over 240 mm in the steppe zone (Figure 1a). The 16-year average of NDVI ranges from less than 0.05 in the southern area of the study region to more than 0.30 in the steppe zone (Figure 1b). The spatial distribution of growing season NDVI roughly corresponds to that of rainfall.

4.2 OLS models

Regression analysis based on the conventional global OLS regression including all vegetated pixels in the study region revealed that there was a strong relationship between the spatial distribution of NDVI and precipitation. The estimated R^2 of the regression equations ranges from 0.41 to 0.83 over the period 1985-2000 and shows a mean value for all years of 0.64. Figure 2 (a) shows the scatter diagram between measured NDVI and predicted NDVI based on the regression parameters from Table 1. The regression model fitted between the 16-year averages of NDVI and rainfall factor explains about 64% of spatial variances in NDVI in the study area. The standard deviation of NDVI residuals was 0.038 or 26.1 % of mean NDVI value. A high degree of inter-annual variations in the regression parameters and the coefficient of determination of OLS was evident but the general nature of the relationship appeared to be relatively stable. The standard error of NDVI prediction through the global OLS regression equals 0.018.



(a) (b)

Figure 2 Results of the global OLS regression model for the data relating to the averages of NDVI and precipitation over the period 1985-2000. Scatter diagram between measured and OLS computed NDVI (a), and (b) map of residuals from the OLS regression analysis.

In order to investigate the influence of land cover type on the NDVI-rainfall relationship, the OLS regression analysis was performed with three individual land cover types, desert, semi-desert and steppe, presented in the study region. Table 1 shows the results of the OLS regression between rainfall and NDVI adapted for the land cover classes. With regard to vegetation type, the results indicate that R^2 increases from desert to semi-desert and to steppe, with a value of 0.36, 0.52, and 0.67 respectively. The using the regression models fitted for the individual land cover types enlarged the prediction power of the global OLS regression model over the study area. Thus, the standard deviation of the residuals was 0.027 or only 18 % of the mean NDVI value, and standard error was 0.017, 0.024 and 0.013 for desert, semi-desert and steppe, accordingly.

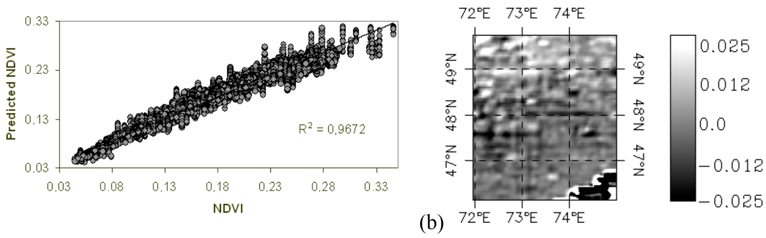
Table 1 Regression model parameters, coefficient of determination and standard error of prediction for the OLS regression analysis concerning all pixels in the study region and individual land cover types.

Model	α	β	R^2	Standard Error of model prediction
Global OLS version	-0.2039	0.002	0.64	0.018
OLS by land cover type:				
Desert shrubland	-0.0698	0.0011	0.36	0.017
Semi-desert	-0.2148	0.0021	0.52	0.024
Steppe	-0.0395	0.0012	0.67	0.013

4.3 GWR model

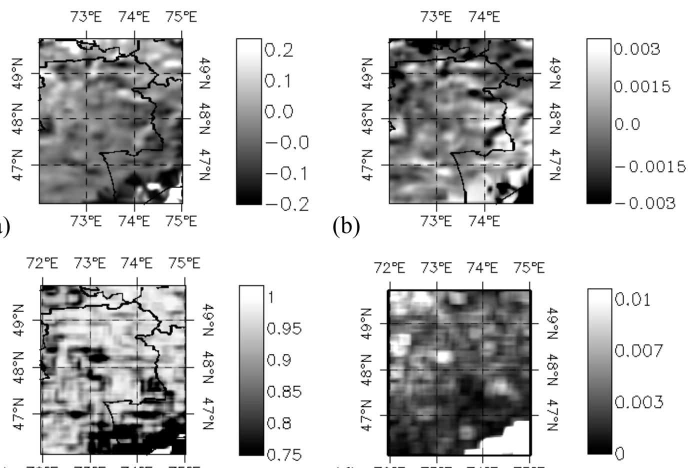
From the GWR modelling, it was evident that the relationship between NDVI and rainfall displays a high spatial non-stationarity. Figure 4 summarizes the results derived from the geographically weighted regression analysis between NDVI and rainfall for the data relating to the average values over the 1985-2000. The coefficient of determination, R^2, and the parameters of GWR varied over space. The strength of the relationship between the both variables increased remarkably, and, accordingly, the amount of variance unexplained is not so large as would be believed from the OLS analysis above. This suggest that the GWR model significantly improved prediction of NDVI by rainfall over the OLS model (Figure 3 a). The GWR model demonstrated a very good prediction power: the NDVI residuals ranged from –0.025 to 0.025 over the study area (Figure 3 b), the standard deviation of NDVI residuals was only 0.0127 or approximately 8 % of the mean NDVI value. In addition to estimating local parameters, we calculated local standard error, SE, in order to evaluate accuracy at local scale (Figure 4 d). SE calculated for every pixel ranges from 0-0.01 with a mean value of 0.008 and is much lower than that for the OLS analysis above.

Generally, the spatial patterns of the mapped intercept and slope parameters appear to correspond with some patterns in land cover. The intercept parameter increases in order from desert, to semi-desert, and to steppe, while the slope parameter decreases in the same direction. The coefficient of determination R^2 tends to display the highest values in the north of the study area where steppe vegetation dominates (R^2 = 0.92 - 0.96). For semi-desert vegetation formation, the R^2 ranges from 0.80 to 0.92 with the mean value of 0.90. Desert vegetation correlates much lower with the patterns of rainfall (R^2 = 0.75-0.85).



(a) (b)

Figure 3 Scatter diagram between measured NDVI and GWR computed NDVI (a), the GWR residuals (b).



(a) (b) (c) (d)

Figure 4 Summary results of the geographically weighted regression between NDVI and rainfall for the data relating to the mean values of the variables over the period 1985-2000. The images show the spatial variation in the local estimate of the intercept (a), the slope (b), the coefficient of determination R^2 (c), and spatial pattern of GWR standard error (d).

4.4 Analysis of spatial patterns in residuals

Spatial pattern of residuals are important indices to examine how accurate reveals a regression model the real relationship. Generally, the residuals of a linear regression model are required to be independently distributed over space with a mean of zero and constant variance. If the residuals exhibit some non-random patterns the model created is problematic. A diagnostic statistics indicating problems in regression modelling is the degree of spatial autocorrelation exhibited by the residuals from the model. The standard errors are usually underestimated when positive autocorrelation is present.

Spatial autocorrelation measures the similarity between samples for a given variable as a function of spatial distance. The Moran's *I* coefficient is the most commonly used coefficient in univariate autocorrelation analysis. For each model, we calculated the Moran's *I* of the residuals to examine the effect of calibrating the models locally by GWR rather than globally. It is proved that the local calibration removes much of the problems of spatially autocorrelated error terms included in traditional global OLS model (Wang et al., 2005; Fotheringham et al., 2003, pp. 112-117). We were interested in the comparison of the results from the global and local model.

Figure 5 shows the spatial correlograms for the OLS model residuals and the residuals from the GWR model. As expected, the error terms are most strongly autocorrelated for the OLS model. The OLS model residuals had significant spatial autocorrelation up to circa 130 km. In comparison, no significant positive spatial autocorrelation was found for the GWR model residuals. It suggests that the calibration of local model reduces the problem of spatially autocorrelated error terms. The GWR model displays the ability to deal with spatial non-stationary problems.

The reduction in the degree of spatial autocorrelation and in absolute values of the residuals through GWR can be seen if we compare maps of the residuals from the both models. In Figure 3 b, the residuals from the OLS model clearly exhibite positive spatial autocorrelation with an area of positive residuals grouped together (north) and an area of negative residuals groupped together (south). The absolute values of the residuals from the OLS model range between –0.07 and 0.07, while that of the GWR residuals between –0.025 and 0.025. The spatial autocorrelation in the residuals from the equivalent GWR model, shown in Figure 4 b, is no longer evident. There are no obvious patterns to the residuals which appear random over the region. The results suggest that GWR provides a better solution to the problem of spatially autocorrelated error terms in spatial modelling compared with the global regression modelling.

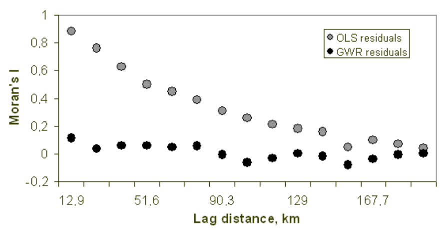


Figure 5 Spatial correlograms for residuals from OLS and GWR models.

4.5 Temporal drift of the NDVI-rainfall relationship

We also analysed and mapped the temporal non-stability of the NDVI-rainfall relationship on a pixel-by-pixel basis. Thus, we calculated coefficient of variation of R^2 and change of R^2 from the GWR for every pixel. Figure 6 shows the results of our calculations. Spatial patterns of the coefficient of variation of R^2 appear to correspond exactly with patterns in land cover: variability of R^2 decreases from desert, to semi-desert, and to steppe. These results indicate a high temporal changeability in the relationship between NDVI and rainfall in the study area. Figure 6b displays change in the coefficient of determination, R^2, for every pixel over the study period. Through an analysis of change in the NDVI-rainfall relationship the ability of the land surface to respond to rainfall over the time period can be implied. Thus, a decrease of R^2 over the study period would indicate a decreasing dependence of the vegetation cover on rainfall patterns and an increasing dependence on others factors such as temperature patterns or human influence. This negative trend would indicate an area with vegetation cover being damaged. On the contrary, an increase of R^2 over time may indicate a surface with an increasingly better response of the vegetation cover to rainfall and a decreasing role of other predictive factors.

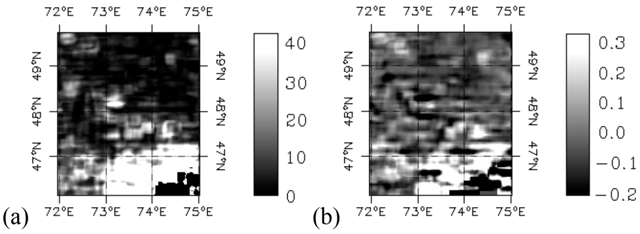


Figure 6 Variability in the R^2 (a) and linear change of R^2 (b) over the period 1985-2000.

5 Conclusion

In this paper we analyzed spatial relationship between NDVI and rainfall in an arid region of Kazakhstan. The analysis based on using two different regression techniques: the one is the global ordinary least squares regression, which assumes the relationship to be spatially stationary, and the other is the geographically weighted regression, which allows the regression parameters and the strength of the relationship vary over the space. The results of the GWR suggest that it provides more accurate predictions than the OLS regression model.

The coefficient of determination, R^2, was more higher for the GWR model. The amount of variance in NDVI unexplained was not as large as would be believed from the OLS analysis. The standard error (SE) was used as a guide to accuracy of the predictions. For the global OLS modeling, SE was 0.018. The SE calculated for global regressions fitted for the individual land cover types were much smaller for desert and steppe, while for semi-desert the both error indices were larger than for the whole region. Fitting the regression model at pixel scale which was achieved through application of the GWR significantly reduces error terms. As expected, the errors terms shown by the results of the GWR are mostly low with a mean value of 0.008 for SE.

The results suggest that the calibration of local rather than global models reduces the problem of spatially autocorrelated errors. The residuals from the global OLS model clearly exhibited positive spatial autocorrelation up to approximately 130 km. the residuals from the GWR model displayed no positive autocorrelation, suggesting the ability of GWR approach to deal with spatial non-stationary problems. The GWR provides a more directly interpretable solution to the problem of spatially autocorrelated errors in spatial modeling compared with the global forms of spatial regression modeling. In GWR, the spatial non-stationarity of the parameters is modeled directly, rather than allowing the non-stationarity to be reflected through the error terms in the global model. Similar results have been presented by Fotheringham *et al.* (2002) and Wang *et al.* (2005).

Our study proved the superiority of GWR over global OLS model in analysis the relationship between patterns in NDVI and precipitation. This superiority is mainly due to the consideration of the spatial variation of the relationship over the study region. Global regression techniques likes OLS may ignore local information and, therefore, indicate incorrectly that a large part of the variance in NDVI was unexplained. The non-stationary modeling based on GWR approach has the potential for greater prediction precision because the model is more tuned to local circumstances, although clearly a greater number of data is required to allow reliable local fitting.

Acknowledgements
We gratefully acknowledge Professor S. Fotheringham for providing the software GWR 3.0.

References
Brunsdon, C., McClatchey, J. and D. J. Unwin. 2001. Spatial variations in the average rainfall-altitude relationship in Great Britain: An approach using geographically weighted regression. *International Journal of Climatology,* 21: 455-466.

Foody G. M. 2003. Geographical weighting as a further refinement to regression modelling: an example focused on the NDVI-rainfall relationship. *Remote Sensing of Environment*, 88: 283-293.

Foody, G. M. 2004. Spatial non-stationary and scale-dependancy in the relationship between species richness and environmental determinants for the sub-Saharan endemic avifauna. *Global Ecol. Biogeogr.*, 13: 315-320.

Fotheringham, A. S., Brunsdon, C. and M. Charlton. 2002. *Geographically weighted regression: the analysis of spatially varying relationships*. Chichester: Willey.

Ji, L. and A. J. Peters. 2004. A Spatial Regression Procedure for Evaluating the Relationship between AVHRR-NDVI and Climate in the Nothern Great Plains. *Int. J. Remote Sensing,* 25: 297-311.

Los S. O. 1993. Calibration Adjustment of the NOAA AVHRR Normalized Difference Vegetational Index Without Resource to Component Channel 1 and 2 Data. *Int. J. Remote Sensing,* 14:1907-1917.

Richard Y. & I. Poccard. 1998. A statistical study of NDVI sensitivity to seasonal and interannual rainfall variations in southern Africa. *Int. J. Remote Sensing,* 19: 2907-2920.

Tucker, C. J. & P. J. Sellers. 1986. Satellite remote sensing of primary production. *Int. J. Remote Sensing,*7: 1396-1416.

Yang, L., Wylie, B., Tieszen, L. L., and B. C., Reed. 1998. An analysis of relationships among climate forcing and time-integrated NDVI of grasslands over the U.S. Northern and Central Great Plains. *Remote Sens. Environ.,* 65: 25–37.

Li, J., Lewis, J., Rowland, J., Tappan, G. and L. L. Tieszen. 2004. Evaluation of land performance in Senegal using multi-temporal NDVI and rainfall series. *Journal of Arid Environments*, 59: 463-480.

Wang, J., Price, K. P. and P. M. Rich. 2001. Spatial patterns of NDVI in response to precipitation and temperature in the central Great Plains. *Int. J. Remote Sensing*, 22: 3827-3844.

Wang, Q., Ni, J. and J. Tenhunen. 2005. Application of a geographically weighted regression analysis to estimate net primary production of Chinese forest ecosystem. *Global Ecol. Biogeogr.*, 14: 379-393.