INCREASING ACCURACY IN ANALYSIS NDVI-PRECIPITATION RELATIONSHIP THROUGH SCALING DOWN FROM REGIONAL TO LOCAL MODEL

P. A. Propastin, M. Kappas

Georg-August-University, Department of Geography, Goldschmidtstr., 3, 37073; Goettingen, Germany,
E-Mail ppropas@uni-goettingen.de, mkappas@uni-goettingen.de

KEY WORDS: NDVI-Rainfall Relationship, GWR Modelling, Uncertainty reducing, Kazakhstan

ABSTRACT

Spatial relationship between vegetation and rainfall in Central Kazakhstan has been modelled using Normalized Difference Vegetation Index (NDVI) and rainfall data from weather stations. The modelling based on application of two statistical approaches: conventional ordinary least squares (OLS) regression, and geographically weighted regression (GWR). The results support the assumption that the average impression provided by the OLS model may not accurately represent conditions locally. The OLS model applied to the whole study area was strong ($R^2 = 0.63$), however it gave no local description of the relationship. Application of the OLS at the scale of individual land cover classes resulted in a better prediction power of the model ($R^2 = 0.75$) and revealed that the response of vegetation to rainfall varies between the land cover classes. The GWR approach, dealing with spatial non-stationarity, significantly increases the model's accuracy and prediction power ($R^2 = 0.97$), as well as highlights local conditions within every land cover class. In order to compare the OLS and the GWR models in terms of prediction uncertainty, we calculated Moran's I for their residuals. Our results demonstrated that the GWR provides a better solution to the problem of spatially autocorrelated errors in spatial modelling compared to the OLS modelling.

1. INTRODUCTION

Climate is the most important factor affecting vegetation condition. Great research effort has been made to derive models that predict spatial variations in vegetation by climates which are based on utilizing remotely sensing obtained vegetation indices and climatic data. The Normalized Difference Vegetation Index (NDVI) is the most used multi-spectral vegetation index which is highly correlated to green-leaf density and can be considered as a proxy for above-ground biomass (Tucker & Sellers, 1986). Several previous studies already modelled relationships between spatial or temporal patterns of NDVI and that of climatic factors. Particularly strong relationships in the arid regions show NDVI and rainfall (Richard & Poccard, 1998; Wang et al, 2001), NDVI and temperature or growing-degree days (Yang et al, 1998; Li et al, 2002) as well as NDVI and evapotranspiration (Ji & Peters, 2004).

Modelling the spatial NDVI-climate relationship one should take into account that one hase to deal with a phenomenon of non-stationarity of this relationship in space. However, the conventional statistical regression method (global ordinary least squares regression, OLS) is stationary in a spatial sense. Stationarity means that a single model is fitted to all data and is applied equally over the whole geographic space of interest. This regression model and its coefficients are constant across space assuming the relationship to be also spatially constant. That is usually not adequate for spatially differenced data, especially by quantifying relationships at regional or global scales. The differences between regression models established at different locations can be large with both the magnitude and sign of the model parameters varying. The easiest method to improve the regression model and to reduce these differences is the fitting of an individual OLS model for each land-cover or vegetation type. On this way, the variance in regression parameters between land-cover types can be highlighted and the prediction power of the regression model increases significantly (Wang et al, 2001; Ji & Peters, 2004, Li et al, 2004). However, this method does not make up the local non-stationarity in the relationship within the land-cover type.

An interesting and efficient alternative is to allow the parameters of the model to vary with space. Such non-stationary modelling shows a greater prediction precision because the model being fitted locally is more attuned to local circumstances. Local regression techniques, such as localized OLS (moving window regression) or geographically weighted regression (GWR) help to overcome the problem of non-stationarity and calculate the regression model parameters varying in space. This techniques provide a more appropriate and accurate basis for modelling relationship between various spatial variables and significantly reduce uncertainty in model prediction. For example, the local regression techniques have been effectively used to quantify spatial relationships between different variables at the field of human and economical geography (Fotheringham et al, 1996; Pavlov, 2000: Fotheringham et al, 2002; McMillen, 1996), in soil science and climatology (Murray & Backer, 1991; Brunsdon et al, 2001). In the field of remote sensing there are only rare studies applying local regression techniques for the analysis of spatial relationships between remotely sensing data and climatic variables (Foody, 2003; Foody, 2005; Wang et al, 2005).

In the submitted paper, we analyse spatial relationships between NDVI obtained from the satellites NOAA AVHRR and rainfall amounts in the southern margin of the Kazakh low hills. The aim of the study was to derive a regression model with strong focus on the accuracy of model prediction at a local scale. In order to find a model with the best prediction power, we tested three different regression techniques, two OLS regression models, and a local model based on geographically weighted regression (GWR). We demonstrated that scaling down from regional to local relationships significantly increases the accuracy and the prediction power of a regression model.

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. 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. Most part of the 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. Dominating vegetation species here are *Artemisia terrae-albae, Artemisia pauciflora, Anabasis salsa, Salsola orientalis*. The northern section of the study region is occupied by steppe vegetation, where dominate short grassland species such as *Festuca sulcata, Stipa capillata* and *Stipa lessingiana*. The semi-desert vegetation complex occupying the mid of the study area represents a complex combination of real steppe turf grasses and semi-shrubs with halophytes. Distribution of land cover types in the study area is shown in Figure 1.

3. DATA USED IN THE STUDY

3.1 NDVI dataset

We used a 10-day 1-km NDVI data set from the global AVHRR archive for every growing season (April-October) during the years 1992, 1993, and 1995. The data set is generated using a maximum value composite (MVC) procedure, which selects the maximum NDVI value within 10-day period for every pixel (Holben, 1986). This procedure is used to reduce noise signal in NDVI data due to clouds or other atmospheric factors. In addition to that, we removed noisy pixels remained in the NDVI maps characterized by exceptionally high or low NDVI values relatively to their pixel neighbourhood. The method of the identification of noisy pixels used a window with a size of 3*3 pixels, which was moving over NDVI scenes and calculated a mean value of the surrounding pixels for every point. After subtracting the original pixel value from the mean value of surrounding pixels, differences of NDVI more than 0.12 were considered as noise. Then, pixels identified as noisy were replaced by the surrounding mean. From the AVHRR NDVI data set we computed a 3-year mean NDVI for every 10-day period beginning with April through October. At last, a NDVI data set accumulated over growing season, $NDVI_{accum}$, was produced by summing up the 10-day mean values derived. Several studies used the $NDVI_{accum}$ as a measure of the magnitude of greenness available through the growing season time which reflects the capacity of the land to support photosynthesis and net primary production for a growing season. A close relationship between $NDVI_{accum}$ and precipitation, especially in arid and semi-arid regions has well been established in the literature (Li et al, 2004; Budde et al, 2005).

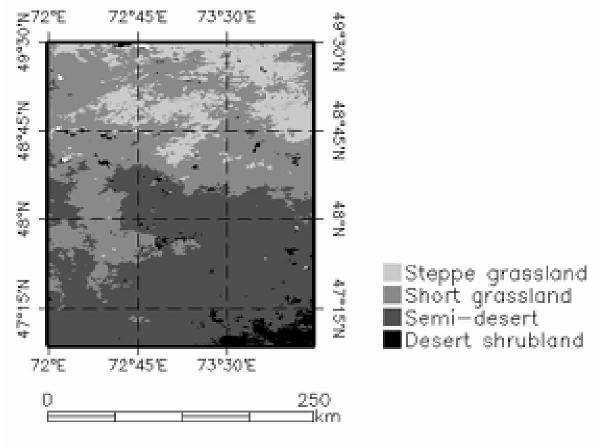

Figure 1. Map of land cover types.

3.2 Precipitation dataset

The precipitation 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 1985-2004. Using these data, we calculated a total average precipitation over the growing season for each climate station during the period 1985-2004.

The preparation of a gridded precipitation map was made by interpolating data between the stations. The interpolation method was kriging with an external drift. Use of a secondary variable, elevation, for the preparation of the gridded map was important because of a strong influence of relief on spatial pattern in precipitation in the study area (Figure 2, left).

4. METHODS

4.1 Regression models

Relationships between NDVI and precipitation were derived by using conventional ordinary least squares (OLS) and geographically weighted regression (GWR) analysis. The first one was fitted both to the whole study region (global OLS) and to each land-cover type (stratified OLS). The second one uses the location information for each observation and allows the model's parameters to vary in space. As OLS and GWR have been well documented, we just briefly describe the theoretical background, especially for the GWR method. A full description of geographic weighted regression and its treatments is provided by (Fotheringham et al, 2002).

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

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

Where a is the intercept of the line on the y axis (where x = 0), β represents the slope coefficient for independent variable x, and ε is the deviation of the point from the regression line.

Fitting the best-fit regression model incorporates the problem to find *a* and *β* so that the total error $\sum \varepsilon_i^2$ is minimized.

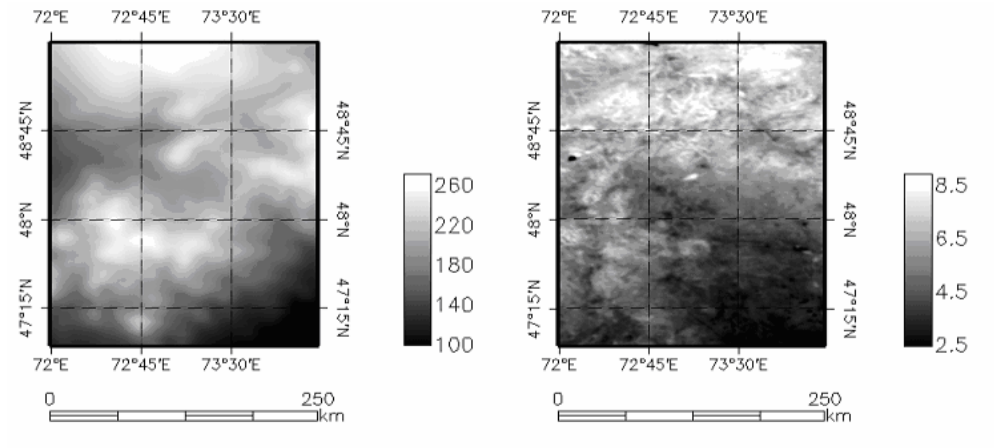

Figure 2. Growing season precipitation in mm (left) and NDVI accumulated over growing season (right).

In this model, the two variables to be related are *y*, the dependent variable (for this study - NDVI), and *x*, the independent variable (rainfall). The regression model's parameters *a* and *β* derived by the above approach are assumed to be stationary globally over the analysis space (the whole study region or the geographical space occupied by a land-cover type). In other words, applying the conventional global regression model to studying relationships between vegetation distribution and its conditions and environmental parameters, one bases his calculation on the assumption, that at each point of the study area this model is absolutely representative and the quantified relationship is constant.

Geographically weighted regression is a local regression technique that deals with the problem of non-stationarity through local disaggregating global statistics and calculates the relationship between NDVI and its explanatory variables for every point. In geographically weighted regression, the regression and its parameters in each point (pixel) of the study region is quantified separately and independently from other points. 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. GWR focused on deriving local parameters to be estimated. The above OLS model can be rewritten as:

$$y = \alpha(\Theta) + \beta(\Theta) * x + \varepsilon \qquad (2)$$

where Θ indicates that the parameters are to be estimated at a location for which the spatial coordinates are provided by the vector Θ.

GWR being a local technique is to be distinguished from other local regressions, as it works in the way that each data point is weighted by its distance from the regression point.

This means that a data point closer to the regression point is weighted more heavily in the local regression than are data points farther away. For a given regression point, the weight of a data point is at maximum when it has the same location as the regression point, and are more lightly when it has a location at a range of the moving window. In GWR an observation is weighted in accordance with its proximity to location *i* so that the weighting of an observation is no longer constant but varies with *i*. 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*; and $W(\theta)$ 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*, and *b* is referred to as a bandwidth.

An alternative way is the bi-square function

$$w_{ij} = [1-(d_{ij}/b)^2]^2, \qquad (5)$$

when d_{ij} < b and w_{ij} = 0 otherwise.

In the practice, for each variable from equation (2) its weighting value can be calculated by applying a weighting matrix W(Θ). The weighting matrix is an *n* by *n* matrix whose off-diagonal elements are zero and whose diagonal elements denote the geographical weighting of each of the n observed data for regression point *i*. After that, moving a kernel over the space can derive a local regression at each point in the analysis area.

Estimated parameters in geographically weighted regression depend on the weighting function of the kernel selected. As the bandwidth, *b*, becomes larger, the closer will be the model solution to that of global OLS. Conversely, as the bandwidth decreases, the parameter estimates will increasingly depend on observations in close proximity to regression point *i* and will have increased variance. The problem is therefore how to select an appropriate bandwidth in GWR. To establish an appropriate bandwidth, *b*, we used the cross-validation approach (CV) which determines *b* by minimisation of the sum of squared errors between predicted variables and those observed. According to (11), the equation for the *cross-validation sum of squared errors CVSS* is statistically expressed as:

$$CVSS = \sum_{i=1}^{n} [y_i - \hat{y}_i(b)]^2 \qquad (6)$$

where y_i is the observed value and $\hat{y}_i(b)$ 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 GWR model, the bandwidth of 9 pixels was decided to be the most appropriate.

4.2 Accuracy analysis

The results obtained using global OLS, stratified OLS and GWR were compared by the amount of $NDVI_{accum}$ variance explained by every regression model. A general rule is that the higher is R^2 the better is the understanding of the variables responsible for the variation in $NDVI_{accum}$ values observed. Generally, a prediction power of a regression model increases with the increase of R^2. We plotted the observed $NDVI_{accum}$ values against the predicted $NDVI_{accum}$ values to compare the prediction power of each regression model.

The standard error, SE, was used as a guide to the accuracy of the predictions. Using the standard error is based on central limit theorem, which says that for large sample size n the conditional distribution of the error should be approximately Gaussian (or normal). For a given sample the standard error can be calculated by the equation:

$$SE = \frac{s}{\sqrt{n}} \qquad (7)$$

where,

$$s = \sqrt{\frac{\sum_{i=1}^{n} (\bar{z}_i - \hat{z}_i)^2}{n-1}} \qquad (8)$$

where, *s* is the standard deviation of the variable *z* and n is the number of data used. In the GWR, $\bar{z}_i$ exposes the mean value of a given kernel. The number of data depends on the kernel size.

Regression residuals are deviations of the points from the regression line. They contain a very important information about the prediction accuracy of a regression model. We were interested not only in values of residuals, but also in the spatial information associated with error. The map of residuals might highlight areas of over-prediction (positive errors) and under-prediction (negative errors). An independent distribution of residuals over the analysis space is the sign for a non-problematic regression model. Spatial patterns of regression residuals containing positive autocorrelation indicate that a model created is problematic: the standard errors are underestimated and the correlation coefficient often indicates a significant relationship between variables when in fact there is not (Clifford et al, 1989). For each regression model, we calculated spatial autocorrelation of the residuals. We were interested in the comparison of the results from the global and the local models. In this study, the Moran's *I* coefficient was used as a measure of autocorrelation. It is the most commonly used coefficient in univariate autocorrelation analysis and is given as:

$$I = \left(\frac{n}{s}\right)\left[\frac{\sum_i \sum_j (y_i - \bar{y})(y_j - \bar{y}) w_{ij}}{\sum_i (y_i - \bar{y})}\right] \qquad (9)$$

where n is the number of samples, y_i and y_j are the data values in quadrats i and j, $\bar{y}$ is the average of y and w_{ij} is an element of the spatial weights matrix W. Under the null hypothesis of no spatial autocorrelation, Moran's *I* has an expected value near zero, with positive and negative values indicating positive and negative autocorrelation, respectively.

5. RESULTS AND DISCUSSION

Spatial distribution of $NDVI_{accum}$ roughly corresponds to that of rainfall (Figure 2). Regression analysis based on the applying of conventional global OLS regression revealed that there was a strong relationship between spatial distribution of the $NDVI_{accum}$ and precipitation.

The global OLS regression model between $NDVI_{accum}$ and precipitation fitted to all vegetated pixels explains about 64% of spatial variance in vegetation distribution and was expressed as:

$$NDVI_{accum} = 0.0854 + 0.0258 * P \qquad (10)$$

(R^2 = 0.64)

where *P* is precipitation.

The standard error used as a measure for prediction accuracy was 0.21 or about 5 % from the mean $NDVI_{accum}$ value. The relatively low value of the standard error might give assumption that the derived regression model provides an accurate description of the relationship between variables. However, the two regression variables, both $NDVI_{accum}$ and precipitation data contain positive autocorrelation (graphs not shown). Their Moran's *I* values up to a distance of ca. 100 km are significantly larger than the values expected under the null hypothesis of no positive autocorrelation. It is known that spatial autocorrelation is problematic for statistical analysis like OLS regression. When conventional OLS regression is applied to the analysis of data containing positive autocorrelation, there are two problems: (1) the standard error of the regression coefficient is underestimated, (2) the residual mean square may seriously underestimate the variance of the error term, hence the coefficient of determination (R^2) is overestimated (Clifford *et al*., 1989). Recent studies tried to overcome these problems by applying spatial regression technique that can adjust for spatial autocorrelation inherent in the regression model on the basis of a variogram function (Titelsdorf, 2000; Ji & Peters, 2004).

The result of the global OLS analysis encouraged us for a further work that aimed to increase the understanding of the relationship between variables. One of the ways to reduce the model uncertainty is to introduce other variables into the model specification. Another way may be an improving regression model by disaggregating (stratification) the global regression model into a separate model for each of land cover types.

We tried to reduce the amount of unexplained $NDVI_{accum}$ and the negative influence of spatial autocorrelation in that way that we performed the OLS regression analysis separately to the four main vegetation types represented in the study region. With regard to vegetation type, the results indicate that the coefficient of determination, R^2, increases from desert to semi-desert, to short grassland, and to steppe, with value of 0.36, 0.44, 0.52, and 0.67 respectively. The components of the regression equation vary in a wide range: there are notable differences in regression slope and intercept between the vegetation types. The stratification of the OLS model by land-cover types clearly illustrates presence of non-stationarity in the general relationship between $NDVI_{accum}$ and precipitation, which may now be written as:

$$NDVI_{accum} = (0.6668 - 6.3178) + (0.0017 - 0.0205) * P \qquad (11)$$

($R^2 = 0.75$)

Values for the range in both intercept and slope parameters are given in brackets. We concluded that the global OLS regression can not possibly be considered stationary. There is spatial variation in intercept and slope parameters as well as in the coefficient of determination, R^2, between the land-cover categories. These results assume a different response of vegetation to precipitation by various land cover categories. That agrees with the results of other research works about dry regions which also deal with the influence of NDVI-rainfall relationships by land cover type (Yang *et al*., 1998; Li *et al*., 2002; Wang *et al*., 2001).

Disaggregating of global OLS regression model into four stratified OLS models has significantly improved the modelling certainty and quality. Although, the amount of variance in $NDVI_{accum}$ unexplained remains relatively high, but the accuracy of the prediction is increased: there is a significant decrease of standard error in comparison with the global OLS model. The smallest SE was 0.10 for steppe grassland, while the largest SE of prediction is equal to 0.17 and was calculated for short grassland.

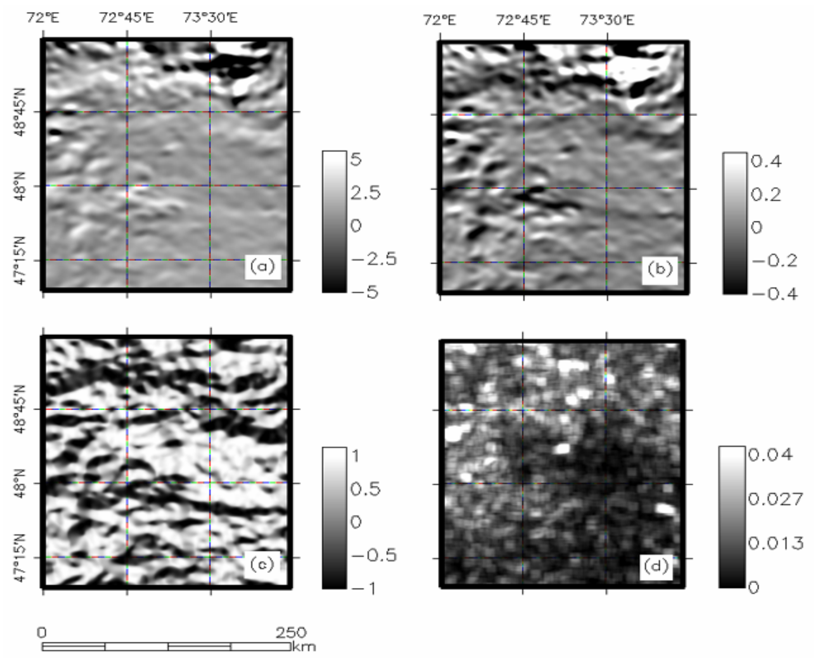

Figure 3. Spatial variations in regression outputs from the GWR analysis of $NDVI_{accum}$ against precipitation: (a) model intercept; (b) model slope parameter; (c) local coefficient of determination, R; (d) standard error of the model prediction.

The GWR method has also been applied for modelling the relationship between $NDVI_{accum}$ and precipitation. To obtain localized results, a 9 by 9 pixels window was placed over each pixel which provides 81 data points for the model calibration at each pixel location. This was the window size which was determined by minimisation of the *cross validation square sum*, CVSS. The GWR model allows the regression parameters to vary in space and establishes considerably stronger relationship between the two variables. The general regression equation may be given as:

$$NDVI_{accum} = (-4.98 - 5.03) + (-2.36 - 1.97) * P \qquad (12)$$

($R^2 = 0.97$)

In the brackets we have written range values for regression intercept and slope parameters.

Figure 3 summarizes the results derived from the geographically weighted regression analysis between $NDVI_{accum}$ and rainfall. Panel a shows spatial distribution of the intercept that had a mean of –0.32 and a range of –4.98 to 5.03. Large positive values are distributed mainly in the north of the region where short grassland and steppe grassland dominate while low values are mainly in the mid and in the south. Here dominate semi-desert and desert vegetation.

Panel b shows spatial variation in the slope parameter. This parameter had a mean of 0.0418 with a range of –2.36 to 1.97 and a standard deviation of 0.22. Negative values of the slope parameter indicate that in some locations $NDVI_{accum}$ decreases when precipitation increases. Negative values are mainly in the northern and western parts of the study region where crop fields/grassland mosaics dominate. The valley bottoms in the northeast also exhibit negative values of the slope parameter.

Panel c displays the spatial variation in the strength of the relationship. The goodness-of-fit, measured by the coefficient of determination, R^2, varied in the space and ranged from 0.016 to 0.99, with $R^2 > 0.75$ for two-thirds of the study region. Low values of R^2 are mainly distributed in the west and over a swath of land from the east to the northwest in upper part of the map. The entire model performance was significantly improved both for standard error of prediction accuracy and for the prediction power.

Panel d in Figure 3 indicates standard error term which has been used as a guide to prediction accuracy. The standard error estimated for the GWR ranged from –0.0012 to 0.04. Values of standard error are several times smaller than that estimated for the global OLS (SE = 0.21) and the stratified OLS models (SE = 0.10-0.17). The GWR model enables to map the standard error for every pixel. The spatial patterns in the standard error reveal the danger of using the single estimate for SE derived from a global OLS locally, they vary in magnitude from pixel to pixel. The spatial patterns of SE clearly correspond to that of land-cover categories. This suggests that the GWR model significantly improved prediction of $NDVI_{accum}$ by rainfall over the OLS model.

The spatial patterns in regression residuals are important indices to examine how accurate the regression model reveals the real relationship. The validity of the regression statistics depends on the distribution of the residuals. There are three conditions which have to be fulfilled by the residuals: (a) the residuals must be normally distributed; (b) the residuals must be homoscedastic; (c) the residuals must not be autocorrelated (Tiefelsdorf, 2000). 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. Visual interpretations of the residual maps shown in Figure 4 give us a good impression that there is a clear separation of the residuals from the global OLS in the space. In the northern part of the study area, the residuals tend to exhibit positive values, while in the south

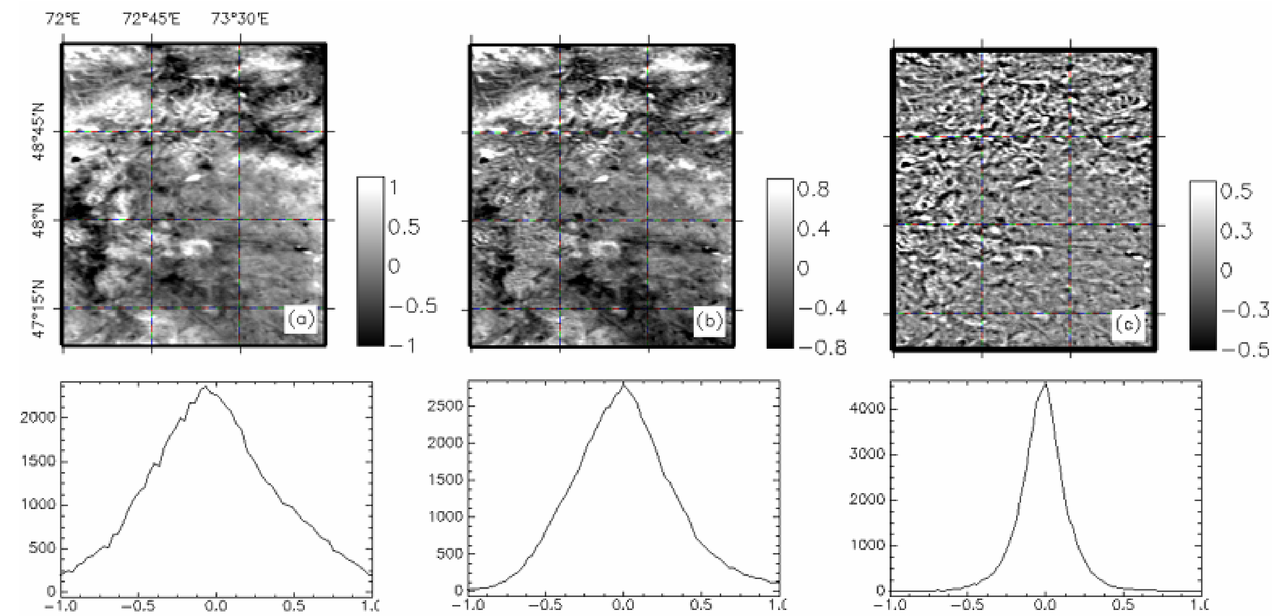

Figure 4. Spatial patterns of regression residuals (upper panels) and corresponding residuals histograms (lower panels) for the global OLS model (a), the OLS model based on stratification by land-cover types (b), the GWR model (c).

the residuals values are mainly negative (Figure 4, a). The global OLS model underestimates when $NDVI_{accum}$ is high and overestimates when NDVI is low. Patterns in the mapped residual values appear to correspond clearly with patterns in land-cover. The positive deviations are associated with the dry steppe vegetation cover, while the negative deviations are mainly observed in the desert zone. The spatial patterns of residuals from the OLS model stratified by land-cover types are not so clear as those for the global OLS model, but the separation in the space also remains (Figure 4, b). Only the GWR model allowed destructing the spatial dependence of the regression residuals (Figure 4, c). The GWR residuals display no clear spatial patterns and their distribution over the study area seems to be close to random.

Spatial autocorrelation measures the similarity between samples for a given variable as a function of spatial distance. For the global OLS model and the GWR 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 (Fotheringham et al, 2002, pp. 112-117). We were interested in the comparison of the results from the global and local models.

Figure 5 shows the spatial autocorrelograms for the global OLS model residuals and the residuals from the GWR model. As expected, the error terms are most strongly autocorrelated for the global OLS model. The OLS model residuals had significant spatial autocorrelation up to circa 50 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 demonstrates the ability to deal with problems of spatial non-stationary.

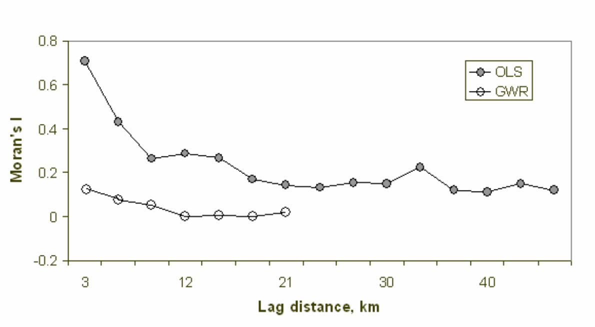

Figure 5. Spatial autocorrelograms for OLS residuals and residuals from the GWR model. The autocorrelogram of the OLS residuals indicates that Moran's I values up to a lag distance of more than 40 km are significantly larger than the value expected under the null hypothesis of no positive autocorrelation. The autocorrelogram of the GWR residuals displays no significant positive autocorrelation.

6. CONCLUSION

In this paper we modelled the spatial relationship between NDVI and rainfall in a semi-arid region of Kazakhstan. We demonstrated how the model accuracy and model prediction power increase through scaling down from regional to local analysis. The analysis based on the use of two different regression techniques: one is the global ordinary least squares regression, OLS, which assumes the relationship to be stationary in space, and the other is the geographically weighted regression, which allows the regression parameters to vary over space. The results of the GWR suggest that it provides more accurate predictions than the OLS regression model.

The study found a high spatial non-stationarity in the strength of relationship and regression parameters both between the land-cover types and within each land-cover type itself. The ordinary least squares regression model has been applied to the whole study area was superficially strong (R^2 = 0.63), however it delivered no local description of the relationship. Applying the OLS at the scale of the separate land cover classes reduced significantly the amount of unexplained variance in $NDVI_{accum}$ (for the whole model R^2 = 0.75) and revealed a different response of various vegetation types to rainfall. The strength of the relationship between NDVI and rainfall increased from desert (R^2 = 0.36), to semi-desert (R^2 = 0.44), to short grassland (R^2 = 0.52), and to steppe grassland (R^2 = 0.67) respectively. The coefficient of determination, R^2, was higher for the GWR model. The approach of geographically weighted regression provided considerably stronger relationships from the same data sets (R^2 value for the general regression = 0.97), as well as highlighted local variations within the land cover classes. The amount of variance in NDVI unexplained was not as large as had been anticipated 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 0.21. The SE calculated through the stratified OLS model for the land-cover types were a few smaller than for the whole region. Fitting the regression model into a pixel scale, what was achieved through application of the GWR, significantly reduces error terms. As expected, the errors terms shown by the results of the GWR are several times lower ranging from 0.0012 to 0.04.

Applying GWR method for dealing with spatial relationship significantly reduces both the degree of autocorrelation and absolute values of the regression residuals. Figure 4 (upper right panel) displays that the residuals from the global OLS model clearly exhibit positive spatial autocorrelation with an area of positive residuals grouped together (in the north) and also an area of negative residuals grouped together (in the south). The spatial autocorrelation in the residuals from the equivalent GWR model, shown in Figure 4 (lower right panel), is no longer evident. There are no obvious patterns to the residuals which appear randomly over the region. The results suggest that GWR provides a better solution to the problem of spatially autocorrelated error terms in spatial modelling compared to the global regression modelling.

The results also 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 50 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 modelling. In GWR, the spatial non-stationarity of the parameters is modelled directly, rather than allowing the non-stationarity to be reflected through the error terms in the global model. This agrees with the results that have been discussed by (Fotheringham et al, 2002; 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 modelling 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 Unwin, D. J. 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.

Brunsdon, C. F., Fotheringham, A. S. and Charlton, M. E. 1996. Geographically weighted regression: a method for exploring spatial non-stationarity. *Geographical Analysis*, 28: 281-298.

Budde, M. E., Tappan, G., Rowland, J. Lewis, J. and Tieszen, L. L. 2004. Assessing land cover performance in Senegal, West Africa using 1-km integrated NDVI and local variance analysis. *Journal of Arid Environments*, 59: 481-498.

Clifford, P., Richardson, S. and Hemon, D. 1989. Assessing the significance of the correlation between two spatial processes. *Biometrics,* 45: 123-134.

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. 2005. Spatial nonstationary and scale-dependancy in the relationship between species richness and environmental determinants for the sub-Saharan endemic avifauna. *Global Ecology & Biogeography*, 13: 315-320.

Fotheringham, A. S., Brunsdon, C. and Charlton, M. 2002. *Geographically weighted regression: the analysis of spatially varying relationships*. (Chichester, Willey) 232 pp.

Fotheringham, A. S., Charlton, M. E. and Brundson, C. 1996. The geography of parameter space: and investigation into spatial non-stationarity. *International Journal of GIS*, 10: 605-627.

Holben, B. N. 1986. Characteristics of maximum-value composite images from temporal AVHRR data. *International Journal of Remote Sensing,* 7:1417-1434.

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. *International Journal of Remote Sensing*, 25: 297-311.

Li, B., Tao, S. and Dawson, R. W. 2002. Relation between AVHRR NDVI and ecoclimatic parameters in China. *International Journal of Remote Sensing*, 23: 989-999.

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

McMillen, D. P. 1996. One hundred fifty years of land values in Chicago: a non-parametric approach. *Journal of Urban Economics,* 40: 100-124.

Murray, M. R. and Baker, D. E. 1991. MWINDOW: an interactive FORTRAN-77 program for calculating moving-window statistics. *Computers and Geoscience*, 17: 423-430.

Pavlov, A. D. 2000. Space-varying regression coefficients: a semi-parametric approach applied to real estate markets. *Real Estate Economics*, 28: 249-283.

Richard Y. and Poccard I. 1998. A statistical study of NDVI sensitivity to seasonal and inter-annual rainfall variations in southern Africa. *International Journal of Remote Sensing,* 19: 2907-2920.

Tiefelsdorf, M. 2000. *Modelling spatial processes: the identification and analysis of spatial relationships in regression residuals by means of Moran's I.* (Berlin, Springer-Verlag) 345 pp.

Tucker C. J. and P. J. Sellers. 1986. Satellite remote sensing of primary vegetation. *International Journal of Remote Sensing,* 7: 1395-1416.

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. *International Journal of 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 Ecology & Biogeography*, 14: 379-393.

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