Available online at www.sciencedirect.com
SCIENCE DIRECT®

Remote Sensing of Environment 88 (2003) 283–293

Remote Sensing of Environment

www.elsevier.com/locate/rse

Geographical weighting as a further refinement to regression modelling: An example focused on the NDVI–rainfall relationship

G.M. Foody

School of Geography, University of Southampton, Highfield, Southampton, SO17 1BJ, UK

Received 13 May 2003; received in revised form 8 August 2003; accepted 18 August 2003

Abstract

The regression analyses undertaken commonly in remote sensing are aspatial, ignoring the locational information associated with each sample site at which the variables under study were measured. Typically, basic ordinary least squares regression analysis is used to derive a relationship that is believed to be uniformly applicable across the study area. Although such global analyses may appear satisfactory, often with large coefficients of determination derived, they may provide an inappropriate description of the relationship between the variables under study. In particular, a global regression analysis may miss local detail that can be significant if the relationship is spatially non-stationary. Local statistical approaches, such as geographically weighted regression, include the spatial coordinates of the sample sites in the analysis and may provide a more appropriate basis for the investigation of the relationship between variables. The potential value of geographically weighted regression to the remote sensing community is illustrated with reference to the relationship between the normalised difference vegetation index (NDVI) and rainfall over north Africa and the Middle East over an 8-year period. For each year, spatial non-stationarity was evident, particularly with regard to the slope parameter of the regression model. Moreover, the conventional ordinary least squares regression models, while superficially strong (minimum $R^2 = 0.67$), were relatively poor local descriptors of the relationship. Relative to this, the geographically weighted approach to regression provided considerably stronger relationships from the same data sets (minimum $R^2 = 0.96$) as well as highlighting areas of local variation. The implications of the difference in the outputs from the two types of regression analysis are illustrated with reference to the use of the derived NDVI–rainfall relationships in mapping desert extent. For example, with the data relating to 1987 the southern limit of the Sahara was generally estimated to lie at a more southerly position when the relationship derived from OLS rather than geographically weighted regression was used.
© 2003 Published by Elsevier B.V.

Keywords: Normalised difference vegetation index; Rainfall; Regression modelling

1. Introduction

Regression techniques have been used widely in remote sensing. Frequently, regression has been used to describe the relationship between an environmental variable measured at the Earth's surface (e.g. biomass) and some measure of its associated remotely sensed response (e.g. a vegetation index). Often, the regression analysis is undertaken with the aim of using the model formed to make predictions of the environmental variable at other sites from their remotely sensed response. Although a variety of approaches to regression modelling exist (Curran & Hay, 1986), the remote sensing community has tended to use uncritically conventional ordinary least squared (OLS) regression analysis (Cohen, Maiersperger, Gower, & Turner, 2003). Since OLS regression has important limitations, its use may not always be appropriate and alternatives should be evaluated (Cohen et al., 2003; Curran & Hay, 1986).

Cohen et al. (2003) present an improved strategy for regression modelling in remote sensing. Recognizing the need to critically assess the techniques used commonly in research and considering the merits of alternative methods, they illustrate some of the different options to OLS regression that may be of immense value to the remote sensing community. One aspect that is infrequently addressed is that the regression analyses commonly used in remote sensing are global techniques, with a single set of model parameters taken to apply uniformly in space. Such analyses are based implicitly on an assumption that the relationship is spatially stationary. The assumption of spatial stationarity in a relationship may often be untenable, particularly when consid-

E-mail address: g.m.foody@soton.ac.uk (G.M. Foody).

0034-4257/$ - see front matter © 2003 Published by Elsevier B.V.
doi:10.1016/j.rse.2003.08.004

ering the large area of coverage provided by remote sensing, and a local technique may be more appropriate (Maselli, 2002). The aim of this article is to outline to the remote sensing community a recent refinement to regression modelling, geographical weighting, that has attracted interest within the geographical research community (Fotheringham, Brunsdon, & Charlton, 2002) and which has attractive features for use with remotely sensed data. In particular, geographically weighted regression is a local technique that allows the regression model parameters to vary in space. This paper will briefly introduce the salient features of geographically weighted regression and then illustrate its application in comparison to standard OLS regression with regard to the widely used relationship between the normalised difference vegetation index (NDVI) and rainfall.

2. Geographically weighted regression

The basic linear regression model that has been used widely in remote sensing may be expressed in the form

$$y = \alpha + \beta x + \varepsilon. \tag{1}$$

In this model, the two variables to be related are y, the dependent variable, and x, the independent variable. Typically in remote sensing studies, y is a remotely sensed variable and x the environmental variable of interest. The remaining parts of the model are its parameters, α which represents the intercept and β which expresses the slope of the relationship between the two variables, and an error term, ε. If there is more than one independent variable, the regression model is typically expressed as

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n + \varepsilon \tag{2}$$

in which β_0 is the intercept and β_1–β_n represent the slope coefficients for the independent variables x_1–x_n, respectively. This type of model is aspatial, with the location of the sites at which the variables were measured deemed irrelevant to the analysis.

The regression model parameters derived following the above approach are assumed to apply globally over the region from which the measurements used in the analysis were derived. Such use of regression models is based implicitly upon the assumption of spatial stationarity in the relationship between the variables under study. Unfortunately, relationships are often not stable in space. The remote sensing literature contains many examples of relationships that appear to vary spatially (e.g. Foody, Boyd, & Cutler, 2003; Grist, Nicholson, & Mpolokang, 1997; Li, Tao, & Dawson, 2002). Moreover, the differences between regression models established at different locations can be large with both the magnitude and sign of the model parameters varying. Many reasons may be put forward to account for such a situation. These range from the view that the conventional global regression model has not been adequately specified through to the existence of intrinsic differences in the relationship over space (Fotheringham et al., 2002). However, as a relationship established at one location may differ greatly from those at others, spatial non-stationarity may be a major limitation to the role of remote sensing as a source of environmental information if conventional global statistical techniques continue to be used inappropriately.

Because of spatial non-stationarity, it may be more appropriate to undertake a local rather than global analysis since the relationship may be a function of location (Fotheringham et al., 2002). If non-stationarity is evident the global model, even if apparently strong (e.g. high R^2), will provide only an average impression of the relationship between the variables under study. It would be quite possible for the parameters from the global model to differ markedly from those derived from an analysis undertaken locally and, indeed, for the global model parameters to not represent the true relationship anywhere within the region of study. Non-stationarity in a relationship, therefore, has a great impact on the use of regression for both descriptive and predictive purposes. If a relationship is spatially non-stationary, a regression model developed at one small region may not have strong predictive power when applied to another site and one developed globally, over a large region, may have weak local predictive power within that region (Osborne & Suarez-Seoane, 2002).

Geographically weighted regression is a technique that expands standard regression for use with spatial data. Since remote sensing is a major source of spatial data on the environment, geographically weighted regression may be a valuable technique to add to the remote sensing toolbox. A key feature of the technique is that it allows the parameter estimates to vary locally. With geographically weighted regression, the relationship between the variables may be expressed as

$$y(\theta) = \alpha(\theta) + \beta(\theta)x + \varepsilon \tag{3}$$

for the simple bivariate situation or

$$y(\theta) = \beta_0(\theta) + \beta_1(\theta)x_1 + \ldots + \beta_n(\theta)x_n + \varepsilon \tag{4}$$

for the multivariate version, where θ indicates that the parameters are to be estimated at a location for which the spatial coordinates are provided by the vector θ. The derived parameter estimates may be mapped to show the nature of their variation in space. This can help reveal spatial variations in the relationship between the variables that would pass unnoticed in a global analysis, in which the estimated parameters are assumed to be constant over space.

In geographically weighted regression, the parameter estimates are made using an approach in which the contribution of an observational site to the analysis is weighted in accordance to its spatial proximity to the specific location under consideration. The weighting of an observation in the analysis is, therefore, not constant but a function of location.

The weighting associated with an observation declines the further the observation is from the location for which predictions and parameter estimates are required. Specifically, in matrix form, the parameter estimates are obtained from

$$\hat{\boldsymbol{\beta}}(\theta) = (\mathbf{X}^{\mathrm{T}}\mathbf{W}(\theta)\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}(\theta)\boldsymbol{y} \quad (5)$$

in which $\mathbf{W}(\theta)$ is a weighting matrix. The off-diagonal elements of the matrix $\mathbf{W}(\theta)$ are set to zero and its diagonal elements represent the geographical weighting associated with each site at which measurements were made for the specific location under consideration; note that the model parameters can be predicted for locations other than those at which the variables have been measured (Fotheringham et al., 2002). The weights act to ensure that observations near to the location at which the parameter estimates are to be made have more influence on the analysis than those further away. This is implemented with a spatial kernel function, the bandwidth of which defines the distance decay in the weighting function and is specified by the analyst. Often, a fixed kernel size with a Gaussian function has been used. However, the specification of the bandwidth of this kernel function and assumption of its constancy across the study area are problematic issues. Indeed the use of this approach may exaggerate the degree of non-stationarity present (Paez, Uchida, & Miyamoto, 2002a, Paez, Uchida, & Miyamoto, 2002b). An alternative approach is to use a kernel with variable bandwidth and various methods to produce a spatially varying kernel are available (Fotheringham et al., 2002). With this approach a bi-square function is used commonly to specify the weights such that the weight of the jth observation at the specific location of interest, regression point i, is

$$w_{ij} = [1 - (d_{ij}/b)^2]^2 \text{ when } d_{ij} \le b$$

$$w_{ij} = 0 \text{ when } d_{ij} > b$$

where d_{ij} is the Euclidean distance between the locations of the sites and b the bandwidth. This bandwidth is adaptive in size and acts to ensure that the same number of non-zero weights are used for each regression point in the analysis. In addition to the estimation of model parameters at each location of interest, a local estimate of the R^2 may be derived. These regression model parameters and diagnostic statistics are interpreted in the same way as with a conventional global regression, although the local estimates of the R^2 should not be interpreted with the same confidence as that from a global model (Fotheringham et al., 2002).

Parameter estimation is highly dependent on the weighting function and kernel used. For example, as the bandwidth increases the parameter estimates will tend towards the estimate from a global model and their spatial pattern will appear increasingly generalised if mapped. The selection of an appropriate bandwidth and weighting function, therefore, requires care and in some instances may benefit from a measure of how well the model fits the data, accommodating for model complexity, such as the Akaike Information Criterion (Fotheringham et al., 2002).

In summary, when using a global regression technique, it is assumed that the relationship under study is spatially stationary. The model's parameters are therefore taken to be constants, applying uniformly over space. However, the relationship may be spatially varying. In such circumstances, the parameters of the global model may not represent conditions locally within the study area. A local technique such as geographically weighted regression allows the model's parameters to vary in space. Although the technique is inappropriate for extrapolating a relationship beyond the region in which the model was established, it does allow the parameters to vary locally within the study area and may provide a more appropriate and accurate basis for descriptive and predictive purposes. A fuller description of geographically weighted regression is provided by Fotheringham et al. (2002) and examples of its application include those presented by Brunsdon, McClatchey, and Unwin (2001) and Paez et al. (2002a).

3. Example: the NDVI–rainfall relationship

Vegetation amount and condition are a function of environmental variables such as rainfall. Consequently, a strong relationship, involving a brief time–lag in the vegetation response to rainfall, would be expected between vegetation indices, such as the NDVI[(infrared reflectance (IR) – red reflectance (R))/(IR + R)] and rainfall (Li et al., 2002; Potter & Brooks, 1998; Richard & Poccard, 1998).

Many studies have focused on the relationship between the NDVI and rainfall. These studies have been undertaken over a range of spatial and temporal scales and in a variety of environments (e.g. Grist et al., 1997; Li et al., 2002; Santos & Negri, 1997; Wang, Price, & Rich, 2003). Although they have often differed in detail, notably in terms of how the variables are expressed (e.g. maximum monthly NDVI, mean annual NDVI, total annual rainfall, etc.), each has essentially sought to link the variation in some measure of the NDVI of a region to a measure of the rainfall incident upon it. Commonly, in studies of a large areas, composite NDVI images generated over periods of days (e.g. du Plessis, 1999; Eklundh, 1998), months or annually (e.g. Kawabata, Ichii, & Yamaguchi, 2001; Richard & Poccard, 1998; Wang, Price, & Rich, 2001) have been used and related typically to rainfall in the immediate few months prior to and including that of image acquisition or on an annual basis (e.g. Grist et al., 1997; Richard & Poccard, 1998; Santos & Negri, 1997). Although the studies reported in the literature differ in detail, a strong positive relationship between rainfall and NDVI has often been observed (e.g. Grist et al., 1997; Wang et al., 2003). Since rainfall data are difficult to acquire over large areas the NDVI has often been used to indicate rainfall. For example, the NDVI has often been used for applications ranging from the mapping and monitoring of desert extent and dynamics (e.g. Foody, 2001; Tucker, Dregne, & Newcomb, 1991) through to the estimation of rainfall itself (Grist et al., 1997).

The relationship between NDVI and rainfall is known to vary spatially, notably due to the effects of variation in properties such as vegetation type and soil background (du Plessis, 1999; Li et al., 2002; Nicholson & Farrar, 1994), with the sensitivity of NDVI values to fluctuations in rainfall, therefore, varying regionally (Richard & Poccard, 1998). Additionally, the problems of satisfying the underlying assumptions and concerns, such as non-stationarity, in analyses of the relationship between NDVI and rainfall have been discussed explicitly (e.g. Eklundh, 1998). However, it is common for basic regression analysis to be used in studies based on the relationship (e.g. du Plessis, 1999; Grist et al., 1997; Potter & Brooks, 1998). That is, a basic ordinary least squares regression analysis is commonly used to study the relationship between the variables globally over the area of study. Even though the relationships observed have generally been strong (e.g. Grist et al., 1997; Wang et al., 2003), the global regression may provide a distorted view of the relationship. In particular, the global relationship may provide a poor description of the relationship locally and may result in important detail being missed. To illustrate the limitation of global techniques such as OLS regression and the potential of geographically weighted regression, the relationship between the NDVI and rainfall over a large area, spanning a region centered on north Africa and the Middle East, will be assessed as an example.

3.1. Data

The NDVI and rainfall data sets used were extracted from the Climatology Interdisciplinary Data Collection generated by the NASA Goddard Distributed Active Archive Center. Specifically, NDVI data generated from

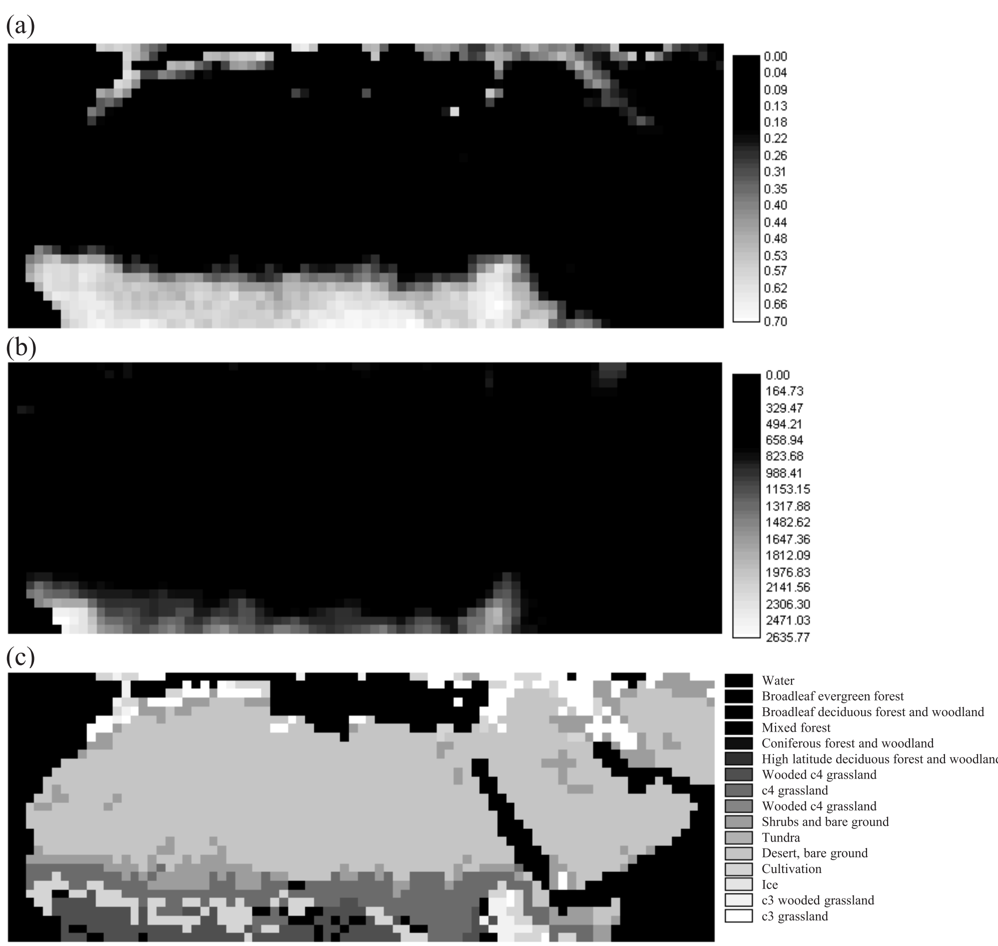

Fig. 1. Examples of the data used. Shown are data relating to 1987. (a) NDVI data, specifically the maximum value composite for the year determined from monthly composite images, (b) rainfall, specifically the total annual rainfall for the year (mm), and (c) land cover (ISLSCP global vegetation and land cover map; some class names have been abbreviated). In this and following figures, regions outside the study area are masked-out in black.

Table 1
Summary of the results

Year	OLS regression			Geographically weighted regression		
	α	β	R^2		Non-stationarity	
				R^2	α	β
1986	0.049895	0.0004940	0.6995	0.9721	0.01	0.01
1987	0.061118	0.0004993	0.7095	0.9691	0.04	0.00
1988	0.062425	0.0004544	0.7205	0.9722	0.05	0.00
1989	0.063133	0.0004608	0.7528	0.9718	0.07	0.00
1990	0.071436	0.0005074	0.6683	0.9693	0.07	0.00
1991	0.064343	0.0004375	0.7151	0.9697	0.00	0.00
1992	0.078344	0.0004448	0.6880	0.9633	0.00	0.00
1993	0.076150	0.0004706	0.7292	0.9738	0.05	0.00

The table shows the regression model parameters and coefficient of determination for the OLS regression analyses together with the coefficient of determination and significance of non-stationarity for the model parameters derived from the geographically weighted regression analyses. The final columns indicate the significance of the spatial variation in the local parameter estimates derived from a Monte Carlo test.

NOAA AVHRR imagery and rainfall data from the Global Precipitation Gauge Analysis for each month in the 8-year period 1986–1993 inclusive were used. Both data sets have a 1° spatial resolution, the same as data sets used in many studies of the NDVI–rainfall relationship (e.g. Potter & Brooks, 1998; Richard & Poccard, 1998). Using the ISLSCP map depicting land cover of 1987, an area centered on north Africa and the Middle East covered by the classes of desert, shrubs and bare ground, grassland (c3 and c4), and wooded grassland only were selected for the analysis (Fig. 1c). For the purposes of this illustrative example, all 1667 1° pixels representing the land area of the region were used.

For each year covered by the data set, a maximum value composite NDVI image was generated and the total annual rainfall calculated (Fig. 1). The relationship between the derived NDVI (dependent variable) and rainfall (independent variable) products was then established for each year using both conventional OLS and geographically weighted regression analysis. This allowed an assessment

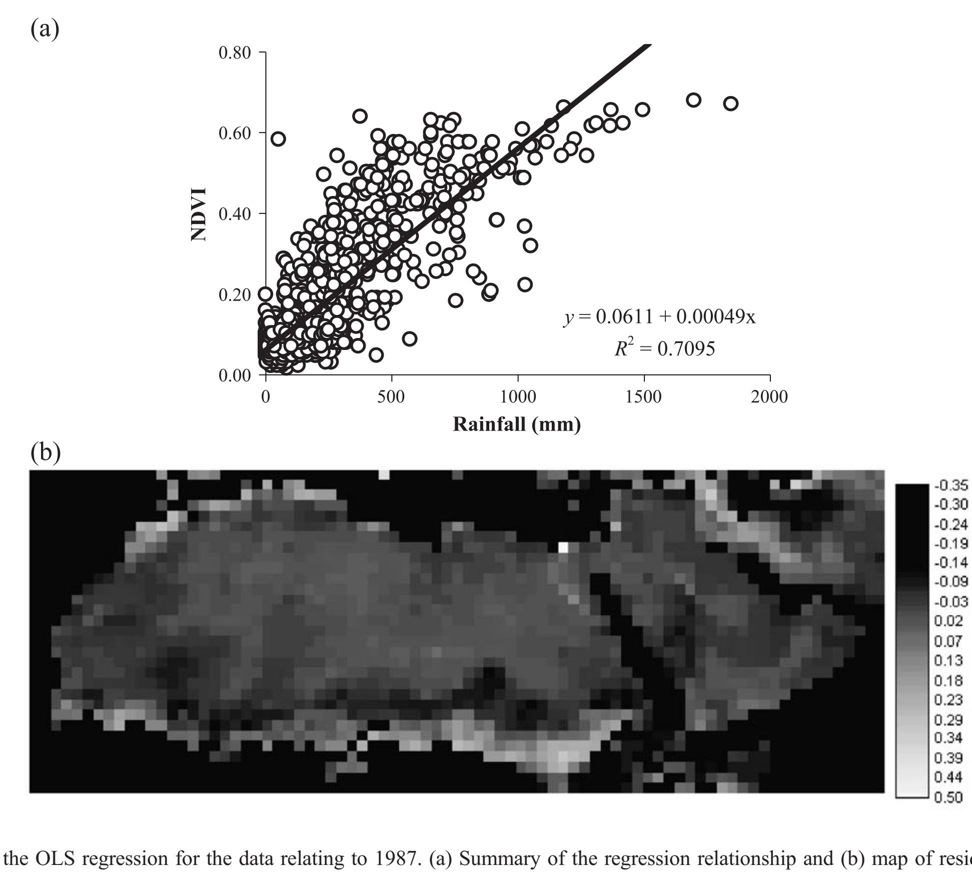

Fig. 2. Results of the OLS regression for the data relating to 1987. (a) Summary of the regression relationship and (b) map of residuals from the regression analysis.

of the spatial and temporal stationarity of the relationship between NDVI and rainfall over the region. The geographically weighted regression analyses were undertaken using an adaptively defined kernel with a bi-square function in which the bandwidth was determined by minimisation of the Akaike Information Criterion (Fotheringham et al., 2002). A Monte Carlo significance testing approach was used to determine if the model parameters displayed significant non-stationarity (Fotheringham et al., 2002; Hope, 1968).

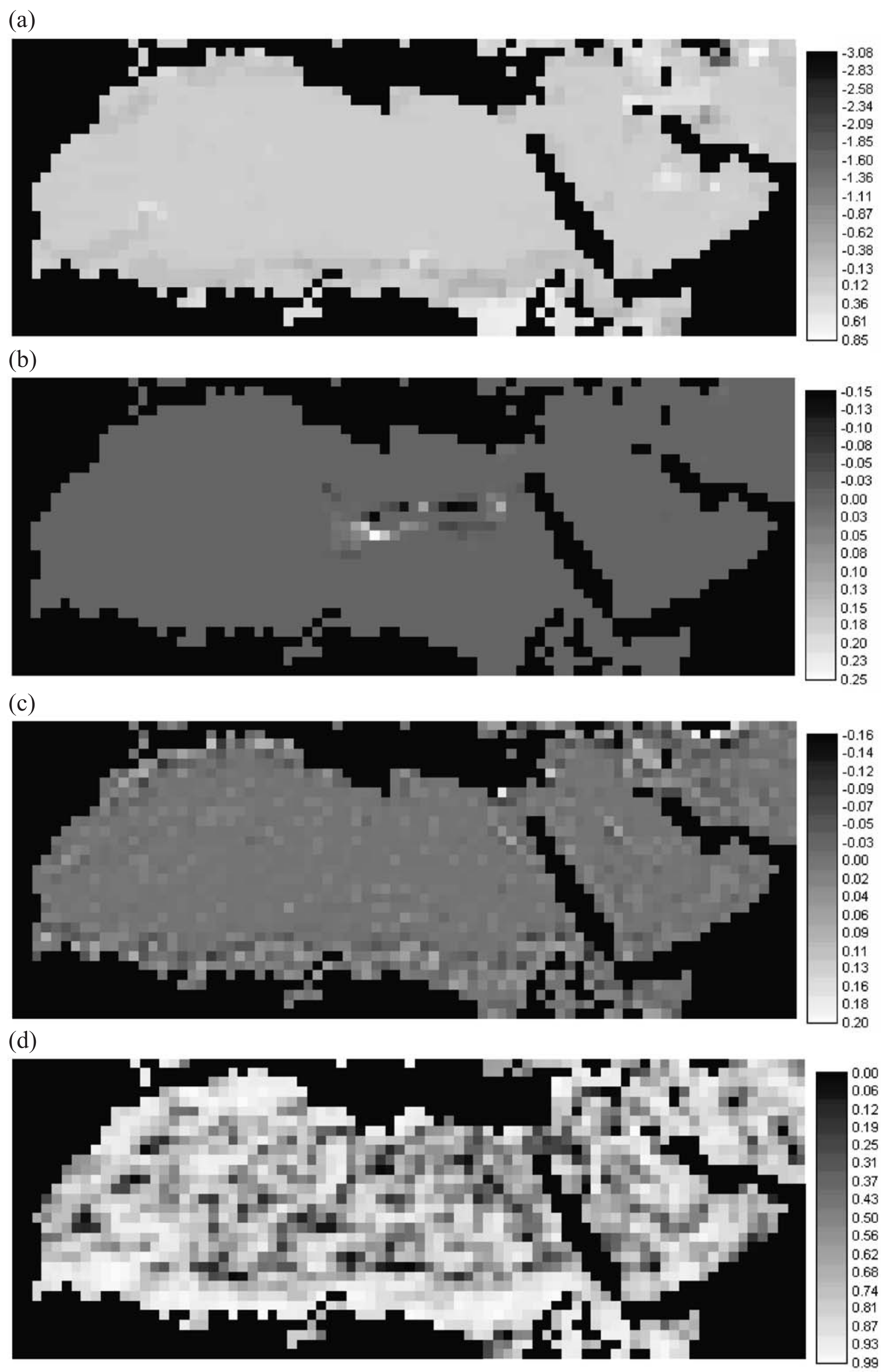

Fig. 3. Results of the geographically weighted regression for the data relating to 1987. The images show the spatial variation in (a) the intercept, α, (b) the slope parameter, β, (c) the residuals, and (d) the local estimate of the coefficient of determination (R^2).

3.2. Results

For the data sets acquired for each of the 8 years, the OLS regression analysis revealed a strong and highly significant (99% level of confidence) relationship between NDVI and rainfall (Table 1). A degree of inter-annual variation in the parameters of the regression equations was evident but the general nature of the relationship appeared relatively stable. As an example, Fig. 2 summarizes the main results derived from the data for 1987, the year to which the ancillary data on land cover relates. With a minimum correlation of $r = 0.817$ ($R^2 = 0.6683$) observed, it may be tempting to assume that the regression relation derived for each year provided an accurate description of the relationship for that period and one that may be applied uniformly over the entire region. Furthermore, the results of each OLS regression analysis indicate a large amount of unexplained variance, and this may drive further work that aims to increase the understanding of the variables responsible for the variation in NDVI values observed (Potter & Brooks, 1998).

From the geographically weighted regression analyses, however, it was apparent that the relationship between NDVI and rainfall was spatially non-stationary. By adopting the geographically weighted regression approach, the strength of the relationship between NDVI and rainfall increased markedly, with a minimum of 96.3% of the variation in the NDVI values explained by that in rainfall ($R^2 = 0.9633$; Table 1). Thus, the amount of variance unexplained may not be as large as would be believed from the OLS analysis. Fig. 3 summarizes the results derived from the geographically weighted regression analysis for the data relating to 1987.

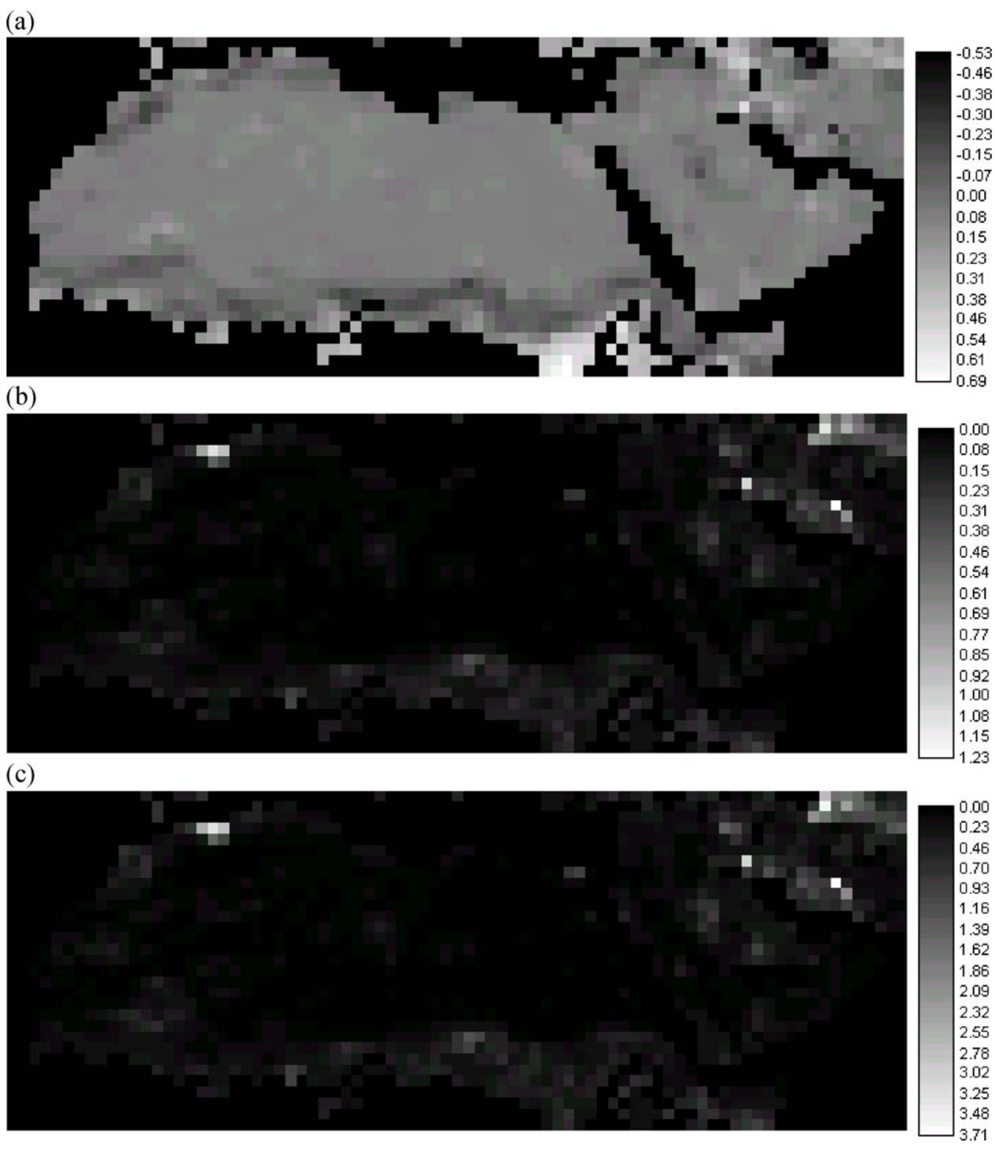

Fig. 4. Summary of the estimates for the intercept parameter, α, over the period 1986–1993. (a) Mean value, (b) standard deviation, and (c) range (maximum–minimum).

3.3. Discussion

Superficially, the OLS regression analysis results appeared to show a strong and statistically significant relationship between NDVI and rainfall (Table 1). However, it is evident that this relationship was non-stationary and that the variation in rainfall could explain a larger amount of the variation in NDVI than was apparent from the basic OLS regression analysis. Although, both regression model parameters appeared to vary significantly in space (Figs. 4 and 5) the variation was most apparent with the slope parameter. In particular, there was evidence for significant spatial non-stationarity in the slope parameter for each year (Table 1, Fig. 6). This was most apparent in the central part of the eastern Sahara, which contained a marked clustering of locations with relatively high or low slope parameters and was associated with considerable variation in parameter values in time (Figs. 5 and 6). These results indicate a major degree of spatial and temporal variation in the relationship between rainfall and the NDVI in this region. There may be many reasons for this, relating, for example, to variations in soil type and land cover as well as issues connected with the generation and accuracy of the data sets used (e.g. Foody, 2001; Ringrose, Matheson, Matlala, O'Neill, & Werner, 1994). For example, patterns in the mapped intercept values (Fig. 4) appear to correspond with some patterns in soil and, in particular, land cover distribution (Fig. 1c), and there are also concerns relating to the analyses undertaken (e.g. in effectively treating the data as if relating to specific points, kernel and bandwidth selection as well as the effect of Earth's curvature on the measures of distance used, etc.). The key issue here, however, is that the

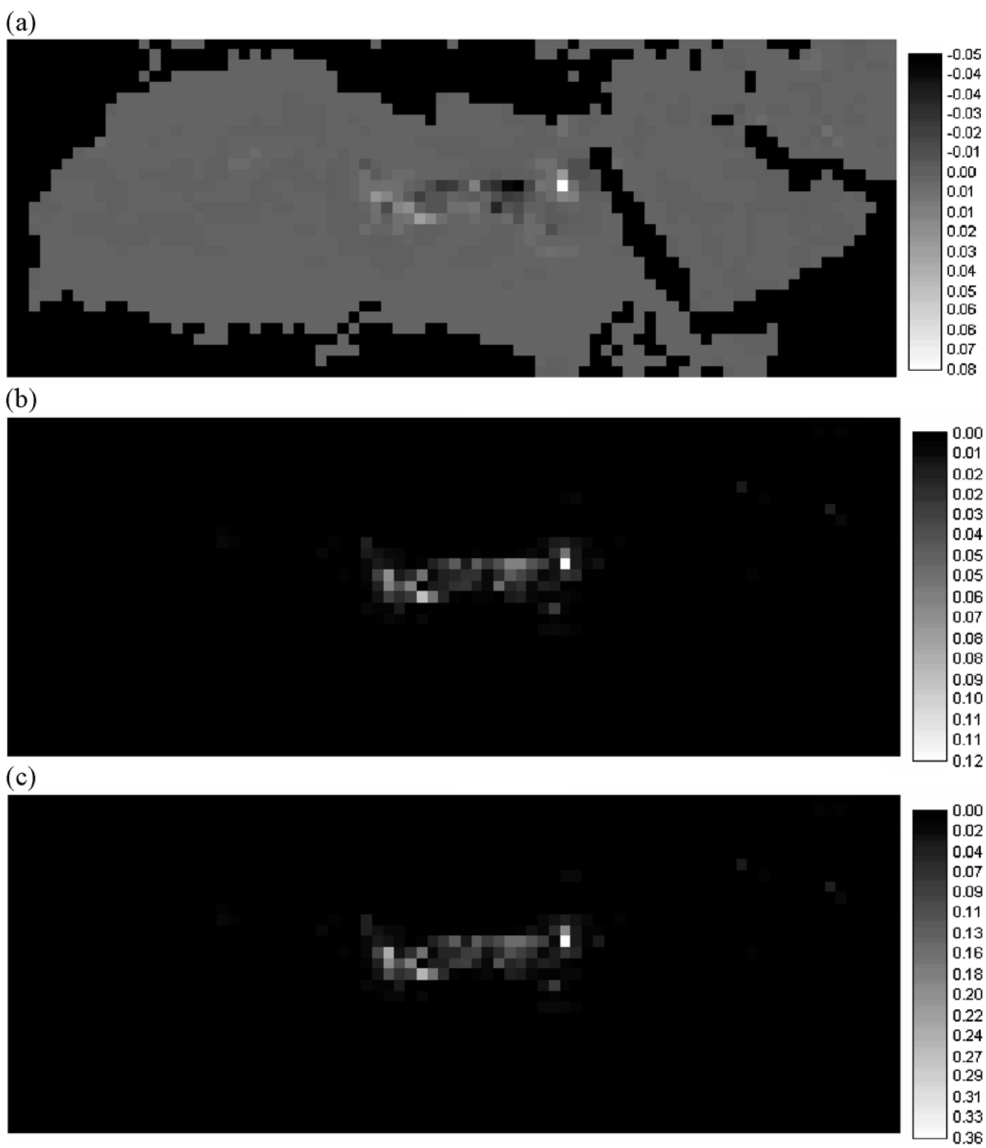

Fig. 5. Summary of the estimates for the slope parameter, β, over the period 1986–1993. (a) Mean value, (b) standard deviation, and (c) range (maximum – minimum).

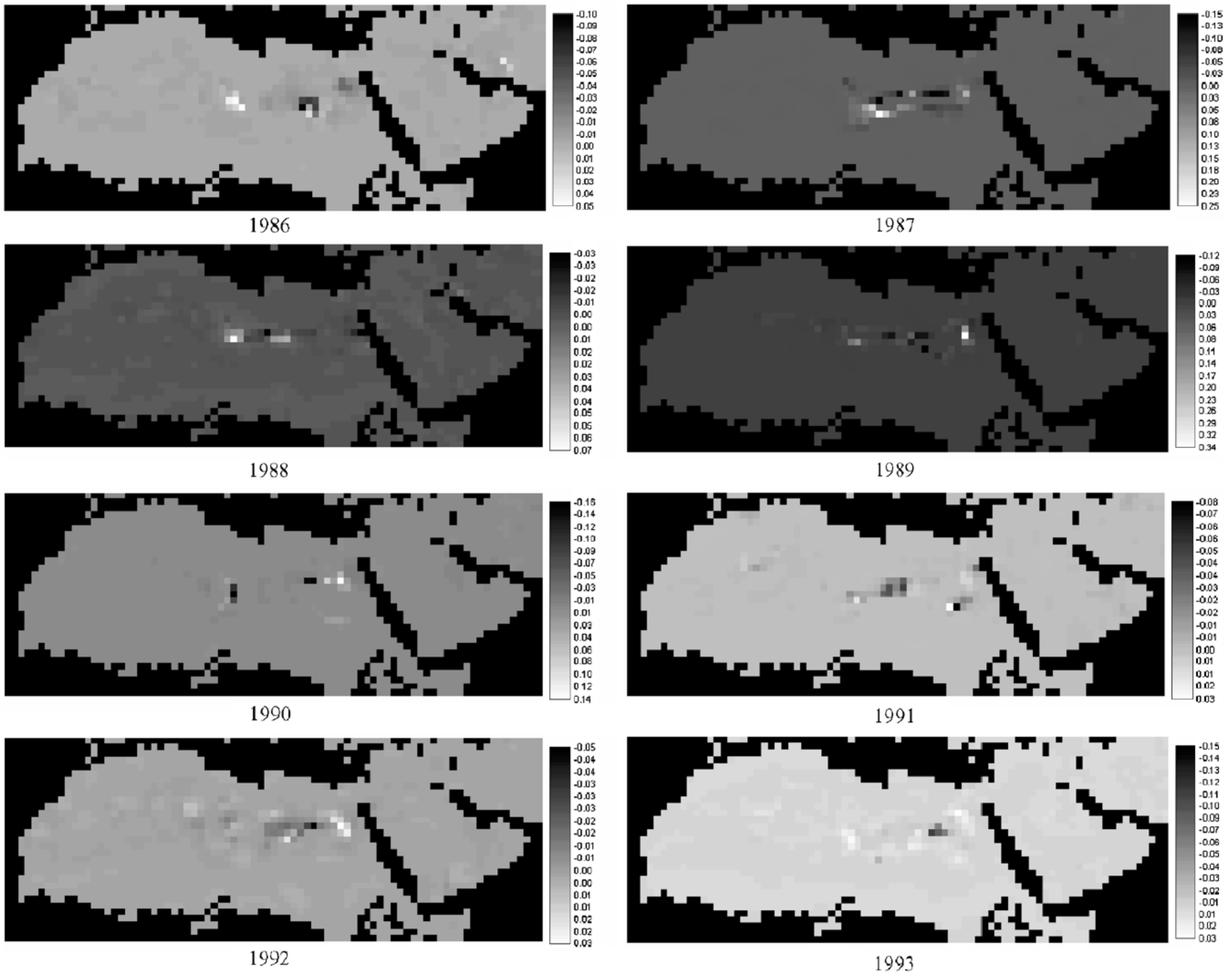

Fig. 6. Spatial distribution of the estimate of the slope parameter, β, for each year.

local variation in the relationship between the NDVI and rainfall would have gone unnoticed in a conventional, global, OLS regression analysis. Moreover, the strength of the relationship derived from the OLS regression analysis, while very strong, could lead to inappropriate and unproductive effort being directed towards accounting for the apparently large unexplained variance in the relationship. The outputs of the geographically weighted regression may provide a more appropriate starting point for efforts directed at increasing the ability to explain the variation in the NDVI values observed.

The dissimilarities in the models derived from OLS and geographically weighted regression may impact markedly on predictions derived from them. As a crude example, the relationships between the NDVI and rainfall established for 1987 were used to plot the southern boundary of the Sahara desert. Using each regression model, the annual rainfall for the study area was predicted from the NDVI data, running the models inversely simply to illustrate the issue. A contour representing the 200-mm annual rainfall isoline was fitted to the predicted rainfall data to represent the desert boundary (Tucker et al., 1991). The location of the predicted boundary of the Sahara desert differed depending on which regression model was used (Fig. 7). In particular, it was evident that the use of the relationship between NDVI and rainfall established with the OLS regression resulted in the boundary being plotted at a more southerly position than if that from the geographically weighted regression analysis had been used. Without an independent data set on which to evaluate the accuracy of the boundary predictions, it is not possible to assess their relative accuracy. However, the important observation is that the two approaches yielded markedly different predictions from the same data.

Geographically weighted regression and related approaches have considerable potential in remote sensing as the data generally used have an explicit spatial character. The technique is not problem-free, with concerns over issues such as kernel and bandwidth selection together with some shared with conventional regression analysis. However, geographically weighted regression can be more appropriate than global techniques such as OLS regression analysis and a suite of other related techniques, including

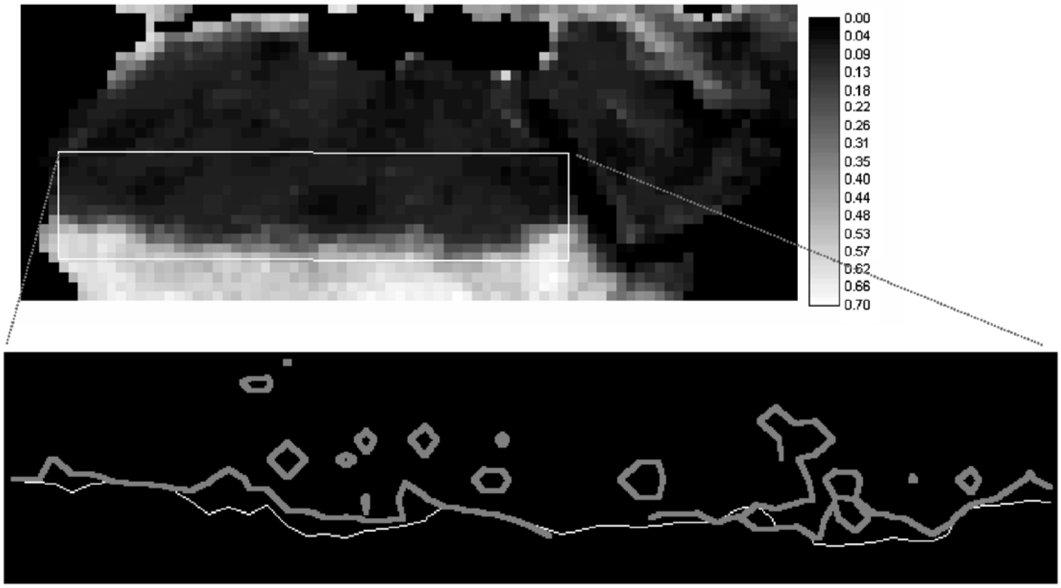

Fig. 7. Predictions of the location of the southern boundary of the Sahara from NDVI data. The image shows the NDVI data for 1987 and highlights an extract shown below that illustrates the boundary position derived using the relationships between the NDVI and rainfall established for 1987 using OLS (thin white line) and geographically weighted regression (thick grey line). Note the break in the contour for the boundary derived using the geographically weighted regression arose as a result of the presence of a small area of a land cover class that had been excluded from all analyses and for which, therefore, parameter estimates had not been made.

some spatial regression models, if relationship is spatially non-stationary (Fotheringham et al., 2002). It is hoped that this article will raise awareness of the approach and lead to further developments that will help in the full realization of the potential of remote sensing as a source of environmental data.

4. Summary and conclusion

Regression analysis is used widely in remote sensing. From the range of regression techniques available, OLS regression is generally used unquestioningly. OLS regression, however, may not always be appropriate and other approaches have been suggested for use by the remote sensing research community (Cohen et al., 2003; Curran & Hay, 1986). One further refinement that could be added to the suggestions made by previous authors is the use of geographically weighted regression (Fotheringham et al., 2002). With particular regard to the example focused on the widely used relationship between the NDVI and rainfall the following key conclusions may be made:

(1) Many relationships are spatially non-stationary (and possibly temporally non-stationary). While an OLS regression may show a strong relationship between the variables, this may be a poor summary of the relationship in reality. Although the outputs from a global analysis are typically assumed to apply equally to all parts of the region of study, they may actually apply to no single location within it. That is, the global relationship derived from an OLS regression analysis may deviate considerably from that observed locally, and may in fact never provide a true description of the relationship at any site but rather some average impression of the relationship over a region.

(2) For the NDVI–rainfall relationship, OLS regression indicated very strong and significant relationships in each year (minimum $R^2 = 0.6683$). However, the geographically weighted regression analysis highlighted local variation in both of the regression model parameters and explained substantially more of the variance (minimum $R^2 = 0.9633$). By revealing local variation that may be missed in a global analysis, the geographically weighted regression approach may be considered to act as a spatial microscope, identifying interesting and unusual locations which may beneficially direct future work (Fotheringham et al., 2002). Thus, rather than the aspatial picture of uniformity derived from a conventional, global, OLS regression the geographically weighted regression analysis may be used to reveal local patterns of relationships. Consequently, the outputs of the geographically weighted regression may form a more appropriate base than those from the OLS regression for studies that aim to further the understanding of the variables impacting on the observed NDVI values, especially in focusing on local issues.

(3) The predictions derived from the OLS and geographically weighted regression approaches differ as a consequence of their dissimilarities in parameter values. The importance of this issue was illustrated with

reference to the use of the relationship between the NDVI and rainfall in mapping of the southern boundary of the Sahara. With the mapping based on the relationship derived from the OLS regression analysis the boundary was generally located at a more southerly position than when based upon the relationship derived from the geographically weighted regression.

Since spatial non-stationarity may be common and remotely sensed data have a strong spatial character, locational information should perhaps be included more commonly in analyses than is currently the practice. Geographically weighted regression is one technique that allows locational information to be usefully included in a type of analysis that is common in remote sensing and it is hoped that it will be of value to the research community.

Acknowledgements

I am grateful for the data sets used that are available in the public domain from major activities funded through the NASA Mission to Planet Earth program. Specifically, the NDVI and rainfall data were extracted from the Climatology Interdisciplinary Data Collection (CIDC), a Goddard Space Flight Center activity sponsored by the NASA Earth Science Enterprise, and the ancillary data on soils and land cover were extracted from the Distributed Active Archive Center (Code 902.2) at Goddard Space Flight Center. The geographically weighted regression analyses were undertaken with the GWR2.0 package developed by Brunsdon, Charlton, and Fotheringham at Newcastle University. I am also grateful to the two referees for their constructive comments on the original manuscript.

References

Brunsdon, C., McClatchey, J., & 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.

Cohen, W. B., Maiersperger, T. K., Gower, S. T., & Turner, D. P. (2003). An improved strategy for regression of biophysical variables and Landsat ETM+ data. *Remote Sensing of Environment*, *84*, 561–571.

Curran, P. J., & Hay, A. (1986). The importance of measurement error for certain procedures in remote sensing at optical wavelengths. *Photogrammetric Engineering and Remote Sensing*, *52*, 229–241.

du Plessis, W. P. (1999). Linear regression relationships between NDVI, vegetation and rainfall in Etosha National Park, Namibia. *Journal of Arid Environments*, *42*, 235–260.

Eklundh, L. (1998). Estimating relations between AVHRR NDVI and rainfall in East Africa at 10-day and monthly time scales. *International Journal of Remote Sensing*, *19*, 563–568.

Foody, G. M. (2001). Monitoring the magnitude of land-cover change around the southern limits of the Sahara. *Photogrammetric Engineering and Remote Sensing*, *67*, 841–847.

Foody, G. M., Boyd, D. S., & Cutler, M. E. J. (2003). Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. *Remote Sensing of Environment*, *85*, 463–474.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester: Wiley.

Grist, J., Nicholson, S. E., & Mpolokang, A. (1997). On the use of NDVI for estimating rainfall fields in the Kalahari of Botswana. *Journal of Arid Environments*, *35*, 195–214.

Hope, A. C. A. (1968). A simplified Monte Carlo significance test procedure. *Journal of The Royal Statistical Society, B*, *30*, 582–598.

Kawabata, A., Ichii, K., & Yamaguchi, Y. (2001). Global monitoring of interannual changes in vegetation activities using NDVI and its relationships to temperature and precipitation. *International Journal of Remote Sensing*, *22*, 1377–1382.

Li, B., Tao, S., & Dawson, R. W. (2002). Relations between AVHRR NDVI and ecoclimatic parameters in China. *International Journal of Remote Sensing*, *23*, 989–999.

Maselli, F. (2002). Improved estimation of environmental parameters through locally calibrated multivariate regression analyses. *Photogrammetric Engineering and Remote Sensing*, *68*, 1163–1171.

Nicholson, S. E., & Farrar, T. J. (1994). The influence of soil type on the relationships between NDVI, rainfall and soil moisture in semiarid Botswanna: I. NDVI response to rainfall. *Remote Sensing of Environment*, *50*, 107–120.

Osborne, P. E., & Suarez-Seoane, S. (2002). Should data be partitioned spatially before building large-scale distribution models? *Ecological Modelling*, *157*, 249–259.

Paez, A., Uchida, T., & Miyamoto, K. (2002a). A general framework for estimation and inference of geographically weighted regression models: 1. Location-specific kernel bandwidths and a test for local heterogeneity. *Environment and Planning A*, *34*, 733–754.

Paez, A., Uchida, T., & Miyamoto, K. (2002b). A general framework for estimation and inference of geographically weighted regression models: 2. Spatial association and model specification tests. *Environment and Planning A*, *34*, 883–904.

Potter, C. S., & Brooks, V. (1998). Global analysis of empirical relations between annual climate and seasonality of NDVI. *International Journal of Remote Sensing*, *19*, 2921–2948.

Richard, Y., & Poccard, I. (1998). A statistical study of NDVI sensitivity to seasonal and interannual rainfall variations in Southern Africa. *International Journal of Remote Sensing*, *19*, 2907–2920.

Ringrose, S., Matheson, W., Matlala, C.J.S.S., O'Neill, T., & Werner, P. A. (1994). Vegetation spectral reflectance along a north-south vegetation gradient in Northern Australia. *Journal of Biogeography*, *21*, 33–47.

Santos, P., & Negri, A. J. (1997). A comparison of the normalised difference vegetation index and rainfall for the Amazon and Northeastern Brazil. *Journal of Applied Meteorology*, *36*, 958–965.

Tucker, C. J., Dregne, H. D., & Newcomb, W. W. (1991). Expansion and contraction of the Sahara desert from 1980 to 1990. *Science*, *253*, 299–301.

Wang, J., Price, K. P., & Rich, P. M. (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, J., Rich, P. M., & Price, K. P. (2003). Temporal responses of NDVI to precipitation and temperature in the central Great Plains, USA. *International Journal of Remote Sensing*, *24*, 2345–2364.