Mathematical Theory and Modeling www.iiste.org
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.5, No.8, 2015

Study of the Geographically Weighted Regression Application on Climate Data

U. Usman[1], Aliyu M. L.[2] and Aminu M. K.[2*]
1. Department of Mathematics, Usmanu Danfodiyo University, Sokoto, Nigeria.
2. Department of Mathematics and Statistics, Sokoto State Polytechnic, Sokoto, Nigeria.

Abstract

This study used Geographical Weighted Regression (GWR) technique to find spatial relationship between Elevation and climate (Rainfall, Temperature) in Northern Nigeria using climate (Rainfall, Temperature) data from weather stations from 1980 – 2010 obtained from Nigerian Meteorological Agency (Nimet). From the results of the analysis it was shown that there is significant relationship between the elevation and climate variables (Rainfall, Tmax and Tmin). The study also shows that GWR has smaller residual sum of square than OLS in analysing the relationship between Elevation and Climate data. This may be due to the consideration of the spatial variation of the relationship over the study region. When mapping the results of GWR model it was observed that the effect of Elevation on climate variables appears to vary geographically

Keyword: Geographical Weighted Regression (GWR), Ordinary Least square (OLS),

Introduction

Climate is one of the most important factors affecting ecological, societal systems and vegetation condition. Therefore, evaluation of the quantitative relationship between vegetation patterns and climate is an important object of applications of Geostatistics at regional- and global scales. Relationship between vegetation and its spatial predictors appears to vary as a function of geographical region and a number of the underlying environmental factors such as vegetation type, soil type and land use (Wang *et al.*, 2001; Yang *et al.*, 1997; Pavel *et al.*, 2007, Ji and Peters, 2004). Moreover, the NDVI-climate relationship is also not the same within one landcover type. There are many cases that show a non-stability of this relationship in space within the same land cover or vegetation type (Fotheringham *et al.*, 1996; Foody, 2003; Foody, 2004; Wang *et al.*, 2005; Propastin and Kappas, 2008). According to these studies, when modelling the spatial vegetation-climate relationship one should take into account that one has to deal with a phenomenon of non-stationarity of this relationship across space. Nonstationarity means that the relationship between variables under study varies from one location to another depending on physical factors of the environment which are spatially autocorrelated. Local regression techniques, such as geographically weighted regression (GWR) help to overcome the problem of nonstationarity and calculate the regression model parameters varying in space (Fotheringham *et al.*, 2002). Because of spatial non-stationarity, the parameters of the model describing the relationship may actually vary greatly in space producing a mosaic that reflects distribution of interaction between the response variable and the predictor factor. This mosaic, however, might demonstrate different patterns at each scale, because different results may be obtained from an analysis by varying its spatial resolution (Openshaw, 1984). Obviously, that the scale-dependent results may be expected with a change in the spatial resolution if a relationship is spatially non-stationary. Spatial variation in the relationship between variables both at and between spatial scales is reported in the recent literature for studies with spatially distributed environmental data. The study by Foody (2003 and 2004), Propastin and Kappas (2008) showed that the predictive power as well as the rank order of explanatory variables in spatial models between remotely sensed data and climatic parameters is a function of scale.

Pavel *et al.*, (2007), Study Application of Geographically Weighted Regression to Investigate the Impact of Scale on Prediction Uncertainty by Modelling Relationship between Vegetation and Climate. The analysis revealed the presence of spatial non-stationarity for the NDVI-precipitation relationship. The results support the assumption that dealing with spatial non-stationarity and scaling down from regional to local modelling significantly improves the model's accuracy and prediction power. The local approach also provides a better solution to the problem of spatially autocorrelated errors in spatial modelling.

Foody (2003) meant by the term "scale effect" the influence of scale on the outputs of a model (strength of the relationship, parameter values and direction, prediction accuracy, etc.) and suggested that the scale effect is a consequence of the relationship between the variables varying in space. Observations of scale dependent results can indicate that the explanatory processes and variables operate at different spatial scales. Concerning the spatial distribution of vegetation, the scale effect may be used (1) to analyse variations of microclimate and their effect to vegetation, (2) to determine the minimal size of landscape units reacting to climate factors as a homogeny area, and (3) to find a model with the best prediction power.

In particular, Dye and Tucker (2003) studied the seasonality and trends of snow-cover, vegetation index, and temperature in northern Eurasia. Nicholson and Farrar (1994) examined the variability of NDVI over semiarid Botswana during the period 1982-1987. Their study demonstrated a linear relationship between precipitation and NDVI when precipitation was less than approximately 500 mm/yr or 50-100 mm/month. Similar results were also found by Wang et al. (2003), who examined the temporal responses of NDVI to precipitation and temperature in the central Great Plains, USA in Kansas and concluded relationship between precipitation and NDVI is strong and predictable when viewed at the appropriate spatial scale. There are also a number of different studies that have analyzed the influence of precipitation, temperature, atmospheric circulation on vegetation dynamics and biomass at high latitudes.

Nicholson *et al.* (1990) compared the vegetation response to precipitation in Sahel and East Africa during 1982 to 1985 and found out that the spatial patterns of annually-integrated NDVI closely reflected mean annual precipitation.

Rodríguez-Lado *et al.* (2007) Carried out a study on spatial modelling of air temperature (maximum, mean and minimum) of the State of São Paulo (Brazil) using multiple regression analysis and ordinary kriging. Climatic data (mean values of five or more years) were obtained from 256 meteorological stations distributed uniformly over the State. It was found that the correlation between the climatic dependent variables, with latitude and altitude as independent variables was significant and could explain most of the spatial variability. The coefficients of determination ($P < 0.05$) varied in the range of 0.924 and 0.953, showing that multiple regression analysis gave an accurate method for the modelling of air temperature for the State of São Paulo. Finally, these regression equations were used together with the kriged maps of the residual errors to build 15 digital maps of air temperature using a 0.5 km^2 Digital Elevation Model in a Geographic Information System. Pavel *et al.*, (2006), assessed a human-induced dryland degradation in the cachment basin of the Balkhash Lake in the Middle Kazakhstan based on time series of rainfall data and normalized difference vegetation index (NDVI) for the period 1985-2000. They developed a method to remove the climatic signal from the change in vegetation activity over the study period. By applying a local regression technique known as geographically weighted regression (GWR), relationship between spatial patterns of the growing season NDVI and the growing season rainfall were estimated for every pixel and every year. The relationship between NDVI and the explanatory variable was found to vary spatially and temporally. At local scales, the regression models indicated that over 90% of spatial variations in NDVI are accounted for by the climatic predictor. Deviations in NDVI from this relationship, expressed in regression residuals, were calculated for each year of the study period 1985-2000. Residuals, laying out of the “Standard Error of the Estimate” are regarded as outliers and interpreted as human-induced. The results of the modelling were validated by comparison of the remote sensing data of high spatial resolution (Landsat TM and ETM) and the data from field trips to degrading areas.

Materials and Methods

The meteorological observation stations in the Federal Republic of Nigeria are administrated by the Nigerian Meteorological Agency (Nimet). For this study, climate stations in the study area and the nearest to that were used. Climate data contain monthly records of 12 climate stations placed in the study area for growing seasons (June- September) during the period of 1981-2010. Rainfall, minimum Temperature, Maximum Temperature variables were used for the analysis.

Model Specification

Geographically weighted regression was first explored by Fotheringham (1997), Brunsdon *et al.,* (1998), Fotheringham and Brunsdon (1999), and Fotheringham (2000). Fotheringham *et al*., (2002) discussed in detail of geographically weighted regression.

For the value $z(s_0)$ at a given location s_0 , it can be estimated using its neighbors with the set of values $z = z(z(s_1), z(s_2), ..., z(s_n))$. Considering *k* predictors of *q*, The GWR model can be written as Eq. (3.7).

$$\hat{z}(s_0) = \sum_{k=0}^{p} \hat{\beta}_k . q_k(s_0) + \varepsilon \qquad (2)$$

Where ε is the residuals, and other notes as above. The objective of GWR is to obtain non-parametric estimates for each predictor q_i and at the location s_0. This can be processed using neighboring data of the location s_0. The basic process using GWR for spatial prediction can be summarized below: (1) determine the samples, (2) determine the unsampled location s_0, (3) design and compute a weight matrix (*W*) based on this location (i.e., Eq. 8), (4) compute the model coefficients using weighted least-squares regression using Eq. 9, and (5) estimate the values of an interesting property at the given locations using the fitted GWR model.

$$W_{s_0} = \begin{pmatrix} w_{01}\, 0 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 \;\; 0 & \cdots & w_{0n} \end{pmatrix} \qquad (3)$$

$$\hat{\beta}_{s_0} = (Q^T . W_{s_0} . Q)^{-1} . Q^T . W_{s_0} . z \qquad (4)$$

A number of weighting functions can be used. Gaussian function is given as an example as follows, and the weight at the location s_0 is calculated as:

$$w_{s_0} = \exp(-0.5(d/\tau)^2) \qquad (5)$$

Where *d* is the Euclidean distance between the location s_0 and its neighbors, τ is the bandwidth of the kernel. Detailed discussion of bandwidth and weight matrix can be found in the software of GWR (Fotheringham *et al*, 2002). Once the *w* for each unsampled location has been calculated, the coefficient matrix can be computed by repeated application of Eq. (4). Therefore, without specifying a function of the spatial variation a set of estimates of spatially varying parameters can be obtained at the unsampled locations. In the process of interpolation, each regression coefficient is predicted to characterize each predictor at a given location, and the GWR "lets the data speak for themselves" (Brunsdon *et al.*, 1998).

Using GWR, the parameters are estimated using Eq. (6). For a given unsampled location s_0, the estimated value is calculated using Eq. (7), where $q_{s_0}^T$ are the s_0 row of the *Q*, and $\hat{\beta}_{s_0}$ are the estimated parameter vector at the location s_0.

$$\hat{z}(s_0) = q_{s_0}^T . \hat{\beta}_{s_0} = q_{s_0} (Q^T . W_{s_0} . Q)^{-1} . Q^T . W_{s_0} . z \qquad (7)$$

Results and Discussion

Geographical Weighted Regression Estimates

The advantage of the GWR is its local approach to analysing relationship between spatial variables. This enables the use of the non-stationarity in the relationship for better prediction. Also GWR approach disaggregates spatial patterns in the model residuals and reduces the spatial autocorrelation of the residuals.

Table 1: Summary of GWR Coefficient Estimates for Climate Data:

	Min.	1st Qu.	Median	3rd Qu.	Max.	Global
X.Intercept	-2276.0	3199.000	3503.000	3935.000	5792.000	4686.9281
Rainfall	-11.880	-1.853	-1.432	-1.276	-0.681	-2.9972
Tmax	26.260	144.100	265.300	344.200	619.100	148.2521
Temp	-835.100	-498.500	-337.400	-263.200	-132.700	-310.0528

R^2: 0.9899725, F = 3.5562, df1 = 8.000, df2 = 5.448, p-value = 0.07963

Table 2: Analysis of Variance Table (OLS and GWR)

	Df	Sum Sq	Mean Sq	F value
OLS Residuals	4.0000	129796		
GWR Improvement	3.6077	93297	25860.4	3.112
GWR Residuals	4.3923	36499	8309.8	

Regression analysis based on applying conventional global OLS regression in Table 1 shows that there is significant relationship between the elevation and climate variables (Rainfall, Tmax and Tmin). The estimated R^2 of the regression equations was found to be 0.99. The GWR model allows the regression parameters to vary in space and from (Table 2) shows that GWR has smaller residual sum of square than OLS.

Applying the GWR method for dealing with spatial relationship significantly reduces both the degree of autocorrelation and absolute values of the regression residuals. 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 modeling.

Conclusion

The degree to which GWR shows higher accuracy is a function of the relationship between climate variables and elevation locally at a given time. From the results of the analysis it shows that there is significant relationship between the elevation and climate variables (Rainfall, Tmax and Tmin). Our study shows that GWR has smaller residual sum of square than OLS in analysing the relationship between Elevation and Climate data, that means GWR is better than OLS in analysing climate date. This may be due to the consideration of the spatial variation of the relationship over the study region. Global regression techniques like OLS may ignore local information and, therefore, indicate incorrectly that a large part of the variance in Climate data was unexplained Pavel *et al*., (2006). The non-stationary modelling based on the GWR approach has the potential for a more reliable prediction because the model is more aligned to local circumstances, although definitely a greater number of data is required to allow for a reliable local fitting.

Acknowledgments

The author thanks Nigerian Meteorological Agency (NIMET) for providing the rainfall and temperature data.

REFERENCES

Brunsdon C, Fotheringham A.S. and Charlton, M (1998). Geographically weighted regresson-modeling spatial non-stationarity. *The Statistician* 47:431-443.

Dye, D. G., & C. J. Tucker, (2003), Seasonality and trends of snow-cover, vegetation index, and temperature in northern Eurasia. *Geophysical Research Letters*, 30 (7): Art. No. 1405.

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

Foody, G. M. (2004). Spatial nonstationary and scale-dependancy in the relationship between species richness and environmental determinants for the sub-Saharan endemic avifauna. *Global Ecology and Biogeography*, 13: 315-320.

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

Fotheringham AS (1997). Trends in quantitative methods I: stressing the local. *Progress in Human Geography* 21: 88-96.

Fotheringham AS (2000). Context-dependent spatial analysis: A role for GIS. *Journal of Geographical Systems* 2:71-76.

Fotheringham AS, and Brunsdon C (1999). Local forms of spatial analysis. *Geographical Analysis* 12 31:340-358.

Fotheringham, A. S., Brunsdon, C., and M. E., Charlton, (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, Chichester: Wiley, pp282.

IGBP (1992). J. R. G. Townshend (Ed.), Improved global data for land applications. IGBP Global Change Report, vol. 20. Stockholm, Sweden: International Geosphere– Biosphere Programme.

Ji, L. & 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.

Nicholson, S. E., M. L. Davenport, and A. R. Malo, (1990). A comparison of the vegetation response to precipitation in the Sahel and East Africa, using normalized difference vegetation index from NOAA AVHRR. *Climatic Change*, 17(2-): 209-241.

Nicholson, S. E. and T. J. Farrar, (1994). The influence of soil type on the relationships between NDVI, precipitation, and soil moisture in semiarid Botswana. I. NDVI response to precipitation. *Remote Sensing of Environment*, 50(2): 107-120.

Openshaw, S. (1984). *The modifiable areal unit problem*. CAT-MOG, 38. Geo Abstracts, Norwich.

Pavel A. P., Martin K. and Nadiya R. M., (2006). *Application of Geographically Weighted Regression Analysis to Assess Human-Induced Land Degradation in a Dry Region of Kazakhstan.* Shaping the Change XXIII FIG Congress Munich, Germany, October 8-13, 2006.

Pavel Propastin, Martin Kappas & Stefan Erasmi (2007). Application of Geographically Weighted Regression to Investigate the Impact of Scale on Prediction Uncertainty by Modelling Relationship between Vegetation and Climate. *International Journal of Spatial Data Infrastructures Research, 2008, Vol. 3, 73-94*

Propastin, P. & M. Kappas. (2008). Reducing uncertainty in modelling NDVI precipitation relationship: a comparative study using global and local regression techniques. *GIScience and Remote Sensing*, 45: 1-25.

Rodríguez-Lado Luis, Sparovek Gerd , Vidal-Torrado Pablo , Dourado-Neto Durval and Felipe Macías-Vázquez (2007). Spatial modelling of air temperature (maximum, mean and minimum) of the State of São Paulo (Brazil). *Sci. Agric. (Piracicaba, Braz.), 64,(5),* 460-467.

Vicente-Serrano, S. M., Grippa, M., Delbart, N., Toan, T. L., & L. Kergoat, (2006), Influence of seasonal pressure patterns on temporal variability of vegetation activity in central Siberia. *International Journal of Climatology*, 26: 303-321.

Wang, J., Price, K. P. & 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, J., P. M. Rich, & K. P. Price, (2003), Temporal responses of NDVI to precipitation and temperature in the central Great Plains, USA. *International Journal of Remote Sensing*, 24(11): 2345-2364.

Wang, Q., Ni, J. & 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.