Journal of Mathematical Sciences and Applications, 2013, Vol. 1, No. 2, 24-28
Available online at http://pubs.sciepub.com/jmsa/1/2/2

DOI:10.12691/jmsa-1-2-2

Modelling Relationship between NDVI and Climatic Variables Using Geographically Weighted Regression

U. Usman[1], S. A. Yelwa[2,*], S.U. Gulumbe[1], A. Danbaba[1]

[1]Department of Mathematics, Usmanu Danfodiyo University, Sokoto, Nigeria
[2]Department of Environmental Sciences, Federal University Dutse, Jigawa State, Nigeria
*Corresponding author: bubakar9@yahoo.co.uk

Received June 08, 2013; Revised August 05, 2013; Accepted August 10, 2013

Abstract 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 the type of vegetation, soil and land use. However, NDVI-climate relationship also varies within one landcover type because there are many cases that show a non-stability of this relationship in space within the same land cover or vegetation type. The purpose of this study is to investigate the applicability of Geographically Weighted Regression (GWR) with the objective of finding the spatial relationship between Normalized Difference Vegetation Index (NDVI) derived from NOAA/AVHRR and Aqua/ Moderate Resolution Imaging Spectroradiometer (AQUA/MODIS) as well as climatic variables (Rainfall, Temperature) data obtained from some weather stations across Northern Nigeria from 1980 - 2010.The results of this study show that there is significant relationship between the NDVI and climate variables (Rainfall, Tmax and Tmin). The study proved the superiority of the local approach provided by GWR over the global Ordinary Least Square (OLS) approach in analysing the relationship between patterns of NDVI and precipitation. This superiority however, was mainly due to spatial variation of the relationship over the local study area because global regression techniques like OLS tend to ignore local information and, therefore, indicate incorrectly that a large part of the variance in NDVI was unexplained. The non-stationary modelling based on the GWR approach therefore, has the potential for a more reliable prediction because the model is more aligned to local circumstances, although more time-series data is needed to allow a more reliable local fitting.

Keywords: *geographical weighted regression, Ordinary Least Square, Normalized Difference Vegetation Index, rainfall, temperature*

Cite This Article: U. Usman, S. A. Yelwa, S.U. Gulumbe, and A. Danbaba, "Modelling Relationship between NDVI and Climatic Variables Using Geographically Weighted Regression." *Journal of Mathematical Sciences and Applications* 1, no. 2 (2013): 24-28. doi: 10.12691/jmsa-1-2-2.

1. Introduction

Time-series NDVI data derived from National Oceanic Atmospheric Adminitration/Advance Very High Resolution Radiometer (NOAA/AVHRR) and Aqua/Moderate Resolution Imaging Spectroradiometer (AQUA/MODIS) have been used for detecting long-term land-use/cover changes and for modeling terrestrial ecosystems on the global, continental, and regional scales. From various studies [1-9] NDVI has valuable information regarding land-surface characteristics. Furthermore, NDVI provides a measure of the amount and vigor of vegetation at the land surface. The magnitude of NDVI is related to the level of photosynthetic activity in the observed vegetation. Most often, higher values of NDVI indicate greater vigor and amounts of vegetation while low values indicate otherwise except bare surface and water bodies which indicate minus or close to zero. Other research that utilized NDVI from these type of sensors were able to obtain useful and reliable results which are closely related to percent cover, leaf area Index (LAI), and plant canopy [10,11,12,13]. The NDVI approach is based on the fact that healthy vegetation has a low reflectance in the visible portion of the Electromagnetic Spectrum (EMS) due to chlorophyll and other pigment absorption and has high reflectance in the NIR because of the internal reflectance by the mesophyll spongy tissue o green leaf [14]. NDVI can be calculated as a ratio of red and NIR bands of a sensor system and is represented by the following equation:

$$NDVI = \frac{NIR - R}{NIR + R} \qquad (1)$$

NDVI values range from -1 to +1. Because of high reflectance in NIR portion of the EMS, healthy vegetation is represented by NDVI values between 0.1 and 1. Conversely, non-vegetated surfaces such as water bodies yield negative values of NDVI because of the electromagnetic absorption quality of water. Bare soil areas represent NDVI values which are closest to 0 due to high reflectance in both visible and NIR portions of the EMS [15]. NDVI has also been shown to be related to photosynthetically active radiation (PAR) and basically measures the capability of leaves, which is related to vegetative canopy resistance and water vapour transfer [16].

2. Theoretical Background

The vegetative surface cover has an important function in the earth system [17] which is linked via several feedback mechanisms in hydrological and climatological processes. Identifying and quantifying these linkages delivers important insight for environmental modelling, management and informed decision making. Satellite earth observation data with high temporal repeat intervals such as NOAA/AVHRR and AQUA/MODIS deliver spatially significant information about the earth surface and are well-suited for monitoring continental scale surface processes.

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 [18,19,20,21]. However, the NDVI-climate relationship is not the same within one landcover type because there are many cases where this relationship shows a non-stability in space within the same land cover or vegetation type [22,23,24,25,26]. Accordingly, because these studies were dealing with the modelling of spatial vegetation-climate relationship they took into account the non-stationarity of the relationship between the phenomena across space. Nonstationarity here refers to the relationship between variables under study which varies from one location to another depending on physical factors of the environment that are spatially autocorrelated. Local regression techniques, such as geographically weighted regression (GWR) help to overcome the problem of non-stationarity and calculate the regression model parameters varying in space [27]. Because of spatial non-stationarity, the parameters of the model describing the relationship may actually vary greatly in space thereby 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 [28]. Thus, it is most likely 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. Studies conducted [23,24,26] 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. However, the study conducted [20] on the application of GWR to investigate the impact of scale on prediction of uncertainty by modelling relationship between vegetation and climate revealed that spatial non-stationarity for NDVI-precipitation relationship exists. 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.

On the other hand, [23] referred the influence of scale on the outputs of a model (strength of the relationship, parameter values and direction, prediction accuracy, etc.) as "scale effect" and further suggested that the scale effect is a consequence of the relationship between the variables varying in space. A review of the scale dependent results therefore may infer that the explanatory processes and variables operate at different spatial scales. With regards to spatial distribution of vegetation, the scale effect may be used firstly, to analyse variations of microclimate and their effect to vegetation. Secondly, to determine the minimal size of landscape units reacting to climate factors as a homogenous area, and thirdly to find a model with the best predictive power. In this regards therefore, [29] studied the seasonality and trends of snow-cover, vegetation index, and temperature in northern Eurasia; while [30] studied the spatial distribution of the inter-annual variability of vegetation activity in central Siberia and its relationship with atmospheric circulation variability. The strongest relationships between the atmospheric circulation variability, climate and the NDVI variability were reported to exist in areas where the climatic characteristics are more limiting for the vegetation development (in this case, the northern hemisphere).

On the other hand, [31] 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 [32] who examined the temporal responses of NDVI to precipitation and temperature in the central Great Plains, USA in Kansas and concluded that the relationship between precipitation and NDVI is strong and predictable when viewed at the appropriate spatial scale. Furthermore, [33] 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.

3. Materials and Methods

Two Global datasets utilized in this study comprised of monthly NDVI and Enhanced Vegetation Index (EVI) maximum value composite images of 1981-2000, (with the exception of September-December 1994 due to cloud contermination). Each of the datasets were subjected to radiometric and atmospheric corrections and projected to a 0.05 degree climatic modeling grid (CMD) in the latitude/ longitude reference system in IDRISI format. The original values were however, not altered. Climatic data in the form of monthly records such as rainfall, minimum and maximum temperatures obtained from the Nigerian Meteorological Agency (NIMET's) 12 climate stations covering the study area were used as input variables in the analysis for growing season months (June- September) between 1981-2010.

4. Model Specification

GWR was first explored by [22] and discussed in detail in [27]. For the value $z(s_0)$ at a given location s_0 , it can be estimated using its neighbors with the set of values

$z = z\left(z\left(s_1\right), z\left(s_2\right), \ldots, z\left(s_n\right)\right)$ Considering k predictors of q , The GWR model can be written as

$$\hat{z}\left(s_0\right) = \sum_{k=0}^{p} \hat{\beta}_k \cdot q_k\left(s_0\right) \quad (2)$$

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: firstly, to determine the samples, secondly to determine the unsampled location s_0 thirdly to design and compute a weight matrix (*W*) based on this location, fourthly, to compute the model coefficients using weighted least-squares regression, and fifthly to estimate the values of an interesting property at the given locations using the fitted GWR model.

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

$$\hat{\beta}_{s_0} = \left(Q^T \cdot W_{s_0} \cdot Q\right)^{-1} \cdot Q^T \cdot W_{s_0} \cdot z \quad (4)$$

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

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

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 [27]. Once the w for each unsampled location is determined, the coefficient matrix is computed by a repeated computation. Hence, without even specifying a function of the spatial variation a set of estimates of spatially varying parameters will be obtained at the unsampled locations. In the process of interpolation, each regression coefficient was then predicted to characterize each predictor at a given location and the GWR. Thus, for a given unsampled location s_0 the estimated value was then calculated using Equation (6) where $q_{s_0}^T$ are the s_0 row of the Q , and β_{s_0} are the estimated parameter vector at the location s_0 .

$$\hat{z}\left(s_0\right) = q_{s_0}^T \cdot \hat{\beta}_{s_0} = q_{s_0}\left(Q^T \cdot W_{s_0} \cdot Q\right)^{-1} \cdot Q^T \cdot W_{s_0} \cdot z \quad (6)$$

5. Results and Discussion

The advantage of the GWR is its local approach to analysing relationship between spatial variables. This enables a non-stationarity in the relationship to be utilized for better prediction. GWR approach also disaggregates spatial patterns in the model residuals and reduces the spatial autocorrelation of the residuals as shown in Table 1.

Table 1: Summary of GWR coefficient estimates

	Min.	1st Qu	Median	3rd Qu	Max.	Global	Pr(>|t|)
X.Intercept	32.5700	83.7100	114.7000	129.7000	141.1000	78.3297	0.2232
Rainfall	0.3895	0.4511	0.5810	0.7673	1.1290	0.8114	0.0264

R^2 : 0.7516789,Pearson Correlation Matrix: 0.636

Table 1 summarize the GWR coefficient estimates indicating that the estimated R^2 of the regression equations which was 0.75 for the period 1981-2010. The standard error used as a measure for prediction accuracy was 0.38. Hence, this model explains about 75 % of spatial variance, and was expressed as:

$$NDVI = 78.3 + 0.81 * Rainfall \qquad \left(R^2 :\ 0.75\right) \quad (7)$$

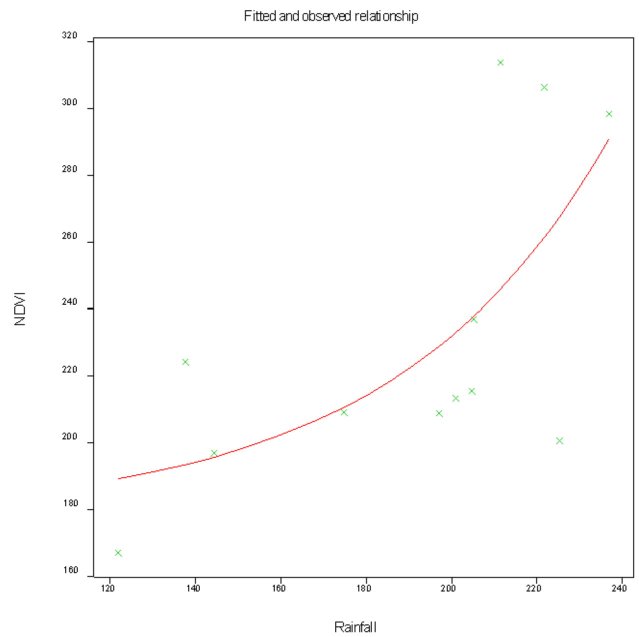

Figure 1. Scatter plot for NDVI and Precipitation

Table 2. Summary of GWR coefficient estimates

	Min.	1st Qu.	Median	3rd Qu.	Max.	Global
X.Intercept	162.40000	222.60000	307.30000	532.70000	900.80000	1016.0888
Rainfall	0.20690	0.25610	0.29120	0.29750	0.34650	0.0914
Tmax	99.91000	179.80000	216.10000	231.90000	308.20000	259.4890
Tmin	136.20000	293.00000	365.90000	384.50000	416.50000	312.8872
Temp	-704.8000	-585.4000	-564.3000	-464.3000	-254.7000	-592.2378
Elev	-0.04459	0.04045	0.09032	0.14650	0.20890	-0.0347

Residual sum of squares: 288.8093; Quasi-global R2: 0.97 ; F = 15.3298, df1 = 6.00, df2 = 1.55, p-value = 0.1012.

When GWR method was applied in the modeling of relationship between vegetation NDVI and precipitation, the results assumed a different response of vegetation to precipitation by various land cover categories. This agrees with the research conducted in the dry regions with regards to NDVI-rainfall relationships on different land cover types [18,19].

Figure 1 shows the scatter plot between measured NDVI and Precipitation. Based on analyses of correlation coefficients values, the spatial relations between NDVI and precipitation indicates positive correlation at every decade of the growing season suggesting that there is positive correlation between NDVI and Precipitation.

Table 2 summarize the results of the GWR model between NDVI, precipitation amounts and temperatures comprising all vegetated pixels in the study area.

Table 3. Analysis of Variance Table (OLS and GWR)

	Df	Sum Sq	Mean Sq	F value
OLS Residuals	6.00000	4427.4		
GWR Improvement	5.18828	4138.6	797.68	
GWR Residuals	0.81172	288.8	355.80	2.2419

The regression analysis based on applying GWR shows that there is significant relationship between the NDVI and climate variables (Rainfall, Tmax and Tmin). The estimated R^2 of the regression equations based OLS was found to be 0.7986. The GWR model allows the regression parameters to vary in space, the goodness-of-fit, measured by the coefficient of determination was higher for the GWR model with $R^2 = 0.97$ as well as the highlighted local variations within dataset and Table 3 shows that GWR had lower residual sum of squares than OLS.

By applying GWR method when dealing with spatial relationship this significantly reduces both the degree of autocorrelation and absolute values of the regression residuals. The results thus, suggest that GWR provides a better solution to the problem of spatially auto-correlated error terms in spatial modeling within a local environment compared with the global regression modelling.

The residual from the annual GWR models shown in Table 3 further represents the remained anthropogenic noise in the NDVI time series dataset after removing the climatic signal. Thus, in order to further detect areas experiencing human-induced change in vegetation cover, the time-trend of the residuals was computed for every pixel. The residuals with positive trend were shown to be widely distributed in the southern part of the study areas falling in the guinea savannah. The concluded model suggests that the remained noise in the inter-annual NDVI time series dataset represent human induced signal while the climatic impact is likely to be the driving force for the changes in vegetation cover between 1980 to 2010.

6. Conclusion

This study shows that the relationship between NDVI and climatic variables appeared to vary as a function of geographical region and other environmental factors. However, when GWR was used in modelling the relationship between rainfall and other climatic variables in the semi arid areas of northern Nigeria the results indicated a better prediction spatially as compared with a derived lower coefficient of determination when a global model was utilised. This agrees with the results pointed out by [25,27]. Thus, this study shows that there is significant relationship between the NDVI and climate variables (Rainfall, Tmax and Tmin). The study therefore, proved the superiority of the local approach provided by GWR over the global OLS approach in analysing the relationship between patterns of NDVI and precipitation because global regression techniques like OLS tend to ignore local information and, therefore, indicate incorrectly that a large part of the variance in NDVI was unexplained. The non-stationary modelling based on the GWR approach on the other hand has the potential for a more reliable prediction because the model is more aligned to local circumstances, although a large quantity of similar dataset utilised in this study is required as input in order to derive a more reliable local fitting.

Acknowledgement

The Researchers are grateful to Clarks Lab. for sourcing the monthly global dataset provided by NASA – MODIS/NDVI/EVI/CMD used in this study. They are also grateful to NIMET for providing the rainfall and temperature data.

List of Abbreviations

CMD	Climate Modelling Grid
EVI	Enhanced Vegetation Index
GWR	Geographically Weighted Regression
IGPB	International Geosphere Biosphere Programme
NASA	National Aeronautics Space Administration
NDVI	Normalised Difference Vegetation Index
NIMET	Nigerian Meteorological Agency
MODIS	Moderate Resolution Imaging Spectrometer
OLS	Ordinary Least Square
Tmax	Maximum Temperature
Tmin	Minimum Temperature

References

[1] 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.

[2] Justice, C. O., Townshend, J. R. G., Holben, B. N., & Tucker, C. J. (1985). Analysis of the phenology of global vegetation using meteorological satellite data. International Journal of Remote Sensing, 6(8), 1271-1318.

[3] Myneni, R. B., Keeling, C. D., Tucker, C. J., Asrar, G., & Nemani, R. R. (1997). Increased plant growth in the northern high latitudes from 1981 to 1991. *Nature*, 386(6626), 695-702.

[4] Potter, C. S., Randerson, J. T., Field, C. B., Matson, P. A., Vitousek, P. M., Mooney, H. A., & Klooster, S. A. (1993). Terrestrial ecosystem production: A process model based on global satellite and surface data. Global Biogeochemical Cycles, 7(4), 811-841.

[5] Prince, S. D. (1991). A model of regional primary production for use with coarse-resolution satellite data. International Journal of Remote Sensing, 12(6), 1313-1330.

[6] Reed, B. C., Brown, J. F., VanderZee, D., Loveland, T. R., Merchant, J. W., & Ohlen, D. O. (1994). Measuring phenological variability from satellite imagery. Journal of Vegetation Science, 5(5), 703-714.

[7] Running, S. W., & Nemani, R. R. (1988). Relating seasonal patterns of the AVHRR vegetation index to simulated photosynthesis and transpiration of forests in different climates. Remote Sensing of Environment, 24, 347-367.

[8] Tucker, C. J., & Sellers, P. J. (1986). Satellite remote sensing of primary production. *International Journal of Remote Sensing*, 7(11), 1395-1416.

[9] Tucker, C.J., Townshend, J.R.G., & Goff, T.E.,1985. African Land cover Classification using Satellite data. *Science*, 227, 369-375.

[10] Malingreau, J.P., 1986. Global Vegetation Dynamics: Satellite observations over Asia.

[11] Marsh, S.E., Walsh, J.H., Lee, C.T., Beck, L.R & Hutchison, C.F., 1992. Comparison of multi-temporal NOAA/AVHRR and SPOT-XS Satellite data for Mapping Land cover dynamics in West African Sahel. *International Journal of Remote sensing*, 13: 2997-3016.

[12] Di, L., Rundquist, D.C & Han, L., 1994. Molelling Relationships between NDVI and Precipitation during Vegetation Growth Cycles. International Journal of Remote Sensing 15(10): 2121-2136.

[13] John, G., Yuan D., Lunetta R.S &Elvidge C.D., 1998. A Change Detection experiment using Vegetation Indices. Photogrammetry Engineering Intepretation. New York: John Wiley.

[14] Campbell, J. B., 1987. Introduction of Remote sensing. New York: the Guilford press.

[15] Lillesand TM, Kiefer RW (2005). Remote Sensing and Image Interpretation, 4th ed, John Wiley, and sons, inc. USA. ISBN: 0471255157, London, UK.

[16] Malo, A.R. & Nicholson S.E.,1990. A study of Rainfall dynamics in African Sahel using Normalized Difference Vegetation Index. Journal of Arid Environment 19:1-24.

[17] Steffen W., and Tyson P., 2001, Global Change and the Earth System: A planet under Pressure.

[18] 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.

[19] 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.

[20] 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*.

[21] Ji, L. & A. J. Peters. (2004). A Spatial Regression Procedure for Evaluating the Relationship between AVHRR-NDVI and Climate in the Northern Great Plains. *International Journal of Remote Sensing*, 25: 297-311.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] Fotheringham, A. S., Brunsdon, C. and Charlton, M. (2002). *Geographically weighted egression: the analysis of spatially varying relationships.* Chichester, Willey.

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

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.