Scale-Dependency of Prediction Uncertainty by Modelling Relationship between Vegetation and Precipitation Patterns in Central Sulawesi, Indonesia

Pavel Propastin, Martin Kappas, Stefan Erasmi

Department of Geography, Georg-August-University, Göttingen, Germany
{ppropas, mkappas, serasmi}@uni-goettingen.de

Abstract. Scale-dependence of spatial relationship between vegetation and rainfall in Central Sulavesi has been modelled using Normalized Difference Vegetation Index (NDVI) and rainfall data from weather stations. The modelling is based on application of two statistical approaches: conventional ordinary least squares (OLS) regression, and geographically weighted regression (GWR). The analysis scales ranged from the entire study region to spatial unities with a size of 5*5 pixels (1250*1250m). 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.

1 INTRODUCTION

Studies of the geographical patterns of vegetation have often been based upon relationships between characteristics of the vegetation activity such as biomass, vegetation cover fraction, leaf area index etc. versus a set of perceived explanatory variables. Commonly, these studies revealed a measure of any indicator for the vegetation activity against a set of environmental determinants, in the main including climatic factors such as precipitation, temperature or growing-degree days, evaporation, soil moisture or others. The Normalized Difference Vegetation Index (NDVI) derived from multispectral satellite data is the most used surrogate of vegetation activity and vegetation characteristics in these studies (e.g. Richard and Poccard, 1998; Yang et al., 1998; Li et al., 2002; Wang et al., 2001; Ji and Petters, 2004).

One commonly noted feature is that the relationship between vegetation and its spatial predictors appear 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; Ji and Peters, 2004). Moreover, the NDVI-climate relationship is also not the same within one land-cover type. There are many cases that show a non-

Florian Probst and Carsten Keßler (Eds.): GI-Days 2007 - Young Researchers Forum.
IfGIprints 30. ISBN: 978-3-936616-48-4

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, 2007). 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. Non-stationarity 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 themselves. 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 (Fotheringham et al., 2002; Foody, 2003). 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. 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 reported in the recent literature for a study on bird's diversity in sub-Saharan Africa. This study also revealed an increase of model accuracy by scaling down to local analysis (Foody, 2004).

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 the submitted paper, we analyse scale-dependency of spatial relationships between NDVI and rainfall amounts in Central Sulawesi, Indonesia. The aim of the study was to determine the spatial scale at which the NDVI-precipitation modelling achieves the best prediction power and the best prediction accuracy. In order to find this model, we tested seven different scales (ranging from the entire study area to local) using two regression techniques - the conventional global OLS regression, and a local regression based on geographically weighted regression (GWR). The results produced in the study should be used in further research for investigation of the primary production and its dependence on the environmental factors in the Lore-Lindu National Park.

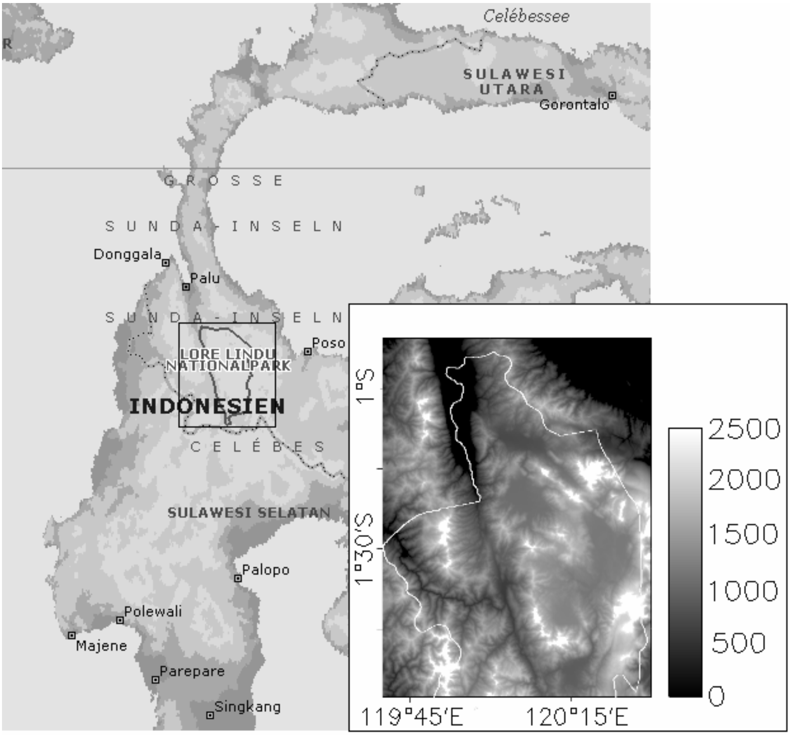


Figure 1: Maps present the location of the study area in Sulawesi (left) and its relief (elevation above sea level in m). The white line on the right map shows the border of the Lore-Lindu National Park.

2 MATERIALS AND METHODS

2.1 Study Area

The analysis area is located in Central Sulawesi, Indonesia (Latitude 0°55’-01°54’ South, Longitude 119°40’- 120°29’ East) and comprises the region of the Lore-Lindu National Park together with the bordering areas (Figure 1). The area has a very complicated relief with elevations from zero in the north to more than 2300 m above sea level in the middle part and is cut by four river valleys: the Palolo to the north, Napu to the east, Bada to the south and Kulawi to the west. The highest peaks are Mt. Nokilalaki (2355 m) and Mt. Rorekatimbu (2610 m).

In terms of the climate type, the study area belongs to the belt of equatorial humid climate. The annual rainfall ranges from about 2000 mm in the north to more than 3000 mm in the south. It falls throughout the year and the heaviest period is during the northern monsoon which lasts from November to April. There is no pronounced wet and dry season. The daytime temperature in lowland areas of the region ranges from 26-28°C throughout the year. However, due to the complex terrain and the diverse geomorphological setting the climate is characterised by large spatial variations. For instance, the main valley of the Palu River receives only 600-800 mm precipitation, while mountain slopes east and west of the valley may have up to 2500-3000 mm of annual precipitation. The spatial distribution of mean daily temperature is also depending strongly on elevation and falls in the mountainous areas to 15-16° C.

The natural vegetation is generally classified into two major vegetation types based on altitudinal distribution with lowland rainforest below 1000 m and mountain rainforest above 1000 m. Most areas of the river valleys are completely deforested and used for production of paddy rice; the most common upland cropping systems in the research area are maize and perennial agroforestry systems with cocoa and/or coffee.

2.2 NDVI Dataset

The most recent studies on spatial and temporal relationships between vegetation and climate at global or regional scales have been based on the using of the satellite derived Normalized Difference Vegetation Index (NDVI). The NDVI is established to be highly correlated to green-leaf density, absorbed fraction of photosynthetically active radiation and aboveground biomass and can be viewed as a major surrogate for vegetation activity (Tucker and Sellers, 1986). The vegetation absorbs a great part of incoming radiation in the visible portion of the spectrum (VIS=380-730 nm) and reaches maximum reflectance in the near-infrared channel (NIR=730-1100 nm). The NDVI, defined as ratio (NIR-VIS)/(NIR+VIS), represents the absorption of photosynthetic active radiation and hence is a measurement of the photosynthetic capacity of the canopy. Negative NDVI values indicate non-vegetated areas such as snow, ice, and water. Positive NDVI values indicate green, vegetated surfaces, and higher values indicate increase in green vegetation.

Within the present study NDVI data products with the spatial resolution of 250 m obtained from the Moderate Resolution Imaging Spectroradiometer (MODIS) have been used. The used dataset covers a time period of four subsequent years from January 2002 to December 2005 and was composed as maximum 16-day values. Although the use of maximum values significantly reduces noise due to atmospheric effects, particularly, amount of clouds in the dataset (Holben, 1986), the MODIS data over the study region comprised many areas whose NDVI values were contaminated by clouds. The removal of the remained clouds from the 16-day NDVI time series was achieved by using a filtering algorithm based on a weighted least-squares regression approach described by Savitzky and Golay (1964). This algorithm has been successfully used for Spot-VEGETATION data by Chen et al. (2004) and improved for MODIS data by Erasmi et al. (2006).

From the filtered MODIS NDVI 16-day data sets, we computed a mean NDVI for the whole period.

2.3 Precipitation Dataset

The climate data in the study consist of daily rainfall data collected from a number of automatic climate stations placed throughout the study area. The daily rainfall data were summed to monthly values and a gridded map of precipitation amount for each month from 2002 was obtained by interpolating data between the stations using the method known as kriging with external drift. After that, we calculated the annual precipitation amount over the 2002.

2.4 Land Cover

The land cover data in the study area were taken from a digital land-cover map derived from a combination of Landsat ETM and Envisat-ASAR data by Twele et al. (2006). The map reveals 12 land cover types in the study area which could be generalized to 5 mean land cover categories used as stratification unities in this study: evergreen broadleaf forest, broadleaf cropland (cacao and coffee areas), irrigated cropland (rice areas), grassland and shrubland.

2.5 Regression Models

Relationships between NDVI and precipitation were modelled 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 category (stratified OLS). The second one uses the location information for each observation and allows the model's parameters to vary in space. The GWR was performed with 5 different kernel sizes (from 50*50 pixels to 5*5 pixels). The land area represented by each pixel was 62500 m^2 (0.625 ha).

As OLS analysis has been well documented, we just briefly describe the theoretical background for the GWR method. A full description of the geographically weighted regression and its treatments is provided by Fotheringham et al. (2002) and Paez et al. (2002 a, 2002 b).

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

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

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.

In this model, the two variables to be related to are y, the dependent variable (for this study - NDVI), and x, the independent variable (rainfall). The regression model parameters a and β derived by the above approach are assumed to be stationary 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 study the relationships between vegetation distribution and its conditions and environmental parameters, our calculation is based on the assumption, that at each point of the study area this model is absolutely representative and the quantified relationship is constant.

As our introduction shows, in the case of land cover types with different response of vegetation to climate and a diverse orographic environment it is incorrect to hold that the same linear relationship is appropriate in all places. Retaining the same linear model, we can allow its two parameters, the intercept and slope to change over space. That is, if Θ represents coordinates of a location on a Cartesian projection, the simple linear model can be expanded and rewritten as:

$$y=\alpha(\Theta)+\beta(\Theta)^{*}x+\varepsilon \tag{2}$$

This regression equation orders the regression parameters to be estimated at a location for which the spatial coordinates are provided by the vector Θ. The revised model is a local regression technique and overcomes the problem of non-stationarity through local disaggregating global statistics and calculates the relationship between NDVI and its predicting variables for each point. The regression parameters may be thought of as a three-dimensional surface over the geographical area rather than a single, fixed real number obtained by the global OLS model. In this model, 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 estimates of the local parameter 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.

Geographically weighted regression distinguishes from the local model described above, as it works in the way that each data point is weighted by its distance from the regression point. The closer is a data point to the regression point, the more weight it reveals. This means, that a data point closer to the regression point is more profound in the local regression than are data points farther away. The matrix form of the parameter estimation for a data point *i* is expressed as:

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

where $\hat{\alpha}$ and $\hat{\beta}$ are intercept and slope parameters in location *i*; and $W(\theta)$ is a 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:

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

where d_{ij} is the distance between regression point *i* and data point *j*, and *b* is referred to as a bandwidth.

In practice, its weighting value can be calculated for each variable from equation (2) 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, a local regression at each point in the analysis area can be derived by moving a kernel over the space.

Estimated parameters in geographically weighted regression depend on the weighting function of the kernel selected. As the bandwidth *b* becomes larger, the closer the model solution to that of global OLS will be. Conversely, as the bandwidth decreases, the parameter estimates will increasingly depend on observations in close proximity to regression point *i* and have increased variance. 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.

The equation for the *cross-validation sum of squared errors CVSS* is statistically expressed as:

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

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

As a general rule, the lower the CVSS is, the closer is the approximation of the model to reality. The best model is the one with the smallest CVSS. In this work we demonstrate a dependency of the model's approximation on the bandwidth of the used kernel. In our modelling the bandwidth ranged from 3 to 25 pixels giving different values of the CVSS and, as consequence, different accuracy of the model prediction.

2.6 Uncertainty Assessment

The results obtained at each spatial scale were compared by the amount of NDVI variance explained by the corresponding regression model. A general rule is that the higher is R^2, the deeper is the understanding of the variables responsible for the variation in NDVI values observed. A prediction power of a regression model increases with the increase of R^2. The cross-validated sum of squared errors, CVSS (Equation 5), was used as a guide to the accuracy of the predictions.

Regression residuals contain the very important information about the prediction correctness of a regression model. As the source data demonstrate a strong spatial autocorrelation, a regression modelling with these data is problematic and requires a careful treatment of this phenomenon. To consider the spatial autocorrelation in NDVI-rainfall analysis is of ecological significance can lead to nearby sites in space tending to have more similar values than would be expected by chance. Spatial autocorrelation of the source data makes an application of classical statistical tests like OLS regression for violating the assumption of independently distributed errors problematic. In this case spatial distribution of the regression residuals serve as a significant indicator for the model's uncertainty. 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 created model is problematic: the standard errors are underestimated and the correlation coefficient often indicates a significant relationship between variables when in fact there is none (Clifford et al., 1989). In this study, the Moran's *I* coefficient was used as a measure of autocorrelation for the regression residuals. 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. We computed and compared Moran's *I* autocorrelation for residuals from each regression model, the lower the autocorrelation of the residuals, the better is the model.

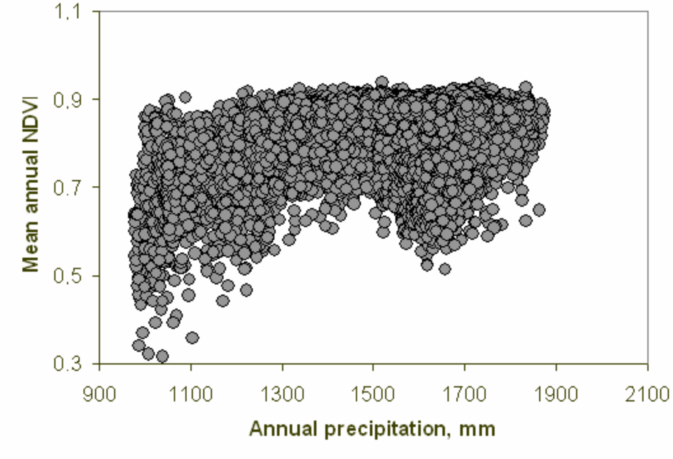


Figure 2: Scatter plot of the mean annual NDVI versus precipitation amount.

3 RESULTS AND DISCUSSION

Conventional OLS model fitted to all vegetated pixels of the study area (study area scale), stratified OLS model fitted to individual land cover categories and GWR model with different bandwidth (from 3 to 25 pixels) have been adapted to analyse NDVI in relationship to precipitation pattern. Correlation analysis of NDVI with precipitation revealed that NDVI had significant correlation ($p < 0.05$) with precipitation within all the above models, but the strength of this correlation, prediction power as well as prediction uncertainty of the models showed high variation between the modelling scales. Table 1 shows the derived characteristics for each model. The best model determined from our analysis was the GWR model which had a bandwidth of 7 pixels (1750 m). The worst model was the OLS regression fitted at the scale of the study area.

Table 1. Summary of the fitting characteristics for the regression models analysed in the study.

Regression model	**R²**	**CVSS**	**Moran's *I* autocorrelation of residuals (distance, pixel/m)**
OLS			
Global	0.18	19,306.43	60 (15000)
Stratified	0.41	16,345.75	46 (11500)
GWR with a bandwidth, *b*, of pixel (m)*			
25 (6250)	0.64	3,216.21	10 (2500)
15 (3750)	0.73	316.26	5 (1250)
11 (2750)	0.77	123.28	4 (1000)
7 (1750)	0.82	34.40	3 (750)
3 (750)	0.92	298.93	5 (1250)

* Distance in meter is given in the brackets

3.1 Study Area Scale

Figure 2 shows the scatter plot between NDVI and precipitation amounts comprising all vegetated pixels in the study area. Overall, there is a weak linear relationship with a determination coefficient of just 0.18. This regression model was expressed as:

$$NDVI = 0.0002 + 0.555 * P \quad (R^2 = 0.18) \tag{6}$$

where P is precipitation.

It is clear that much of the variance remains unexplained. However, the graph reveals that the relationship is linear only in a limited range of rainfall conditions, approximately up to 1500-1600 mm. Above this limit, NDVI shows no increase or rather a decrease with rainfall. Results derived from a polynomial regression model supported this suggestion. The polynomial model of second order exhibits a better goodness of fit statistics ($R^2 = 0.28$) and is written as:

$$NDVI = -5 * 10^{-7} * P^2 + 0.0018 * P - 0.5995 \; (R^2 = 0.28) \tag{7}$$

The results of both global regression models suggest that across the study region NDVI is positively related to precipitation, but the huge amount of the variance in NDVI remains unexplained. This fact encouraged us to undertake additional investigations which aim to increase the understanding of the relationship between the variables. The further analysis suggests a link between the strength of this relationship and the physical conditions of the underlying factors, particularly vegetation type and composition of vegetation communities.

3.2 Stratification of the NDVI-Precipitation Model by Land Cover Category

We tried to reduce the amount of NDVI which is unexplained in that way that we performed the OLS regression analysis separately within the main land cover types represented in the study region. With regard to land cover category, the results indicate that the coefficient of determination, R^2, increases from evergreen broadleaf forest to shrubland, to broadleaf cropland, and to grassland, with a value of -0.065, 0.090, 0.13, and 0.27 respectively. The components of the regression equation vary in a wide range: there are notable differences in the regression slope and the intercept between the land cover types. The stratification of the OLS model by the land-cover types clearly illustrates the presence of non-stationarity in the general relationship between NDVI and precipitation which may now be written as:

$$NDVI = (0.4424 - 0.902) + (-0.00003 - 0.0002) * P \; (R^2{=}0.41) \qquad (8)$$

In the brackets we have given values for the range in both intercept and slope parameters. These results assume a different response of vegetation to precipitation by various land cover categories: three land cover classes correlate positively with precipitation, while for rainforest we attained a negative correlation. The low or negative response of forested areas and the high response of grasslands to precipitation have been good documented in the recent literature (Yang et al, 1998; Li et al, 2002; Wang et al, 2001). The results of this work support the prior investigations and underline the fact that the global regression model represents only average values of the parameter estimates which hide interesting spatial variations. It is likely that, although a sub-division of the global model into a number of models significantly enlarged the prediction power of the global OLS ($R^2 = 0.41$) and improved our understanding of the relationship, most of the variance in NDVI remains still unexplained. We suggested that this relationship might vary over space within the sub-divided regions too, because of a high diversity of relief and hydrological conditions in the study area. Therefore, the modelling scale of the individual land-cover categories did not imply to be the most appropriate to obtain the highest certainty. The use of geographical weighted regression realized the incorporation of non-stationarity into the stratified model and enabled the analysis of the NDVI-precipitation relationship at local scale.

3.3 GWR Model

By accommodating spatial non-stationarity into the model, the GWR analysis allowed the parameters of the models to vary in space and showed considerably stronger relationships with NDVI than from the corresponding conventional global and stratified regression analysis. It was apparent, that the explanatory power of the models varied spatially, with local estimates of R^2 varying from 0.11 to 0.98 depending on the used bandwidth. A plot of bandwidth against CVSS (Table 1) suggests an optimal value for bandwidth of 7 pixels. A supplementary decrease of the bandwidth can give a higher value of R^2 (the finest scale studied was 3 pixels) but it also results in an increase of the CVSS. The non-stationary nature of the relationship between NDVI and precipitation can be expressed in an equation for every bandwidth. We give here the GWR equation for the model used bandwidth of 7 pixels (1750 m):

$$NDVI = (-1.98 - 2.03) + (-0.002 - 0.001) * P \qquad (R^2 = 0.83) \qquad (9)$$

In the brackets we have written range values for regression intercept and slope parameters. Figure 3 shows the scatter plot between measured NDVI and NDVI predicted using the Equation 9. Figure 4 demonstrates distribution of the parameters from Equation 9 over the space. Panel a displays the spatial variation in the strength of the relationship. The correlation coefficient varies in the space and ranges from -0.92 to 0.99. Panel b shows spatial distribution of the intercept which had a mean of –0.32 and a range of –1.98 to 2.03. Panel c shows spatial variation in the slope parameter. This parameter ranges from –0.002 to 0.001.

The GWR model exposed the presence of non-stationarity not only between different land-cover categories but also within each of these categories. GWR works in the way that it blows out the arbitrary boundaries between the land-cover categories and represents the NDVI-rainfall relationship as a continuous geographical process. However, GWR does not distract the general nature of this relationship, preserving the general differences in the vegetation response to precipitation between individual land-cover categories proved by the stratified OLS model.

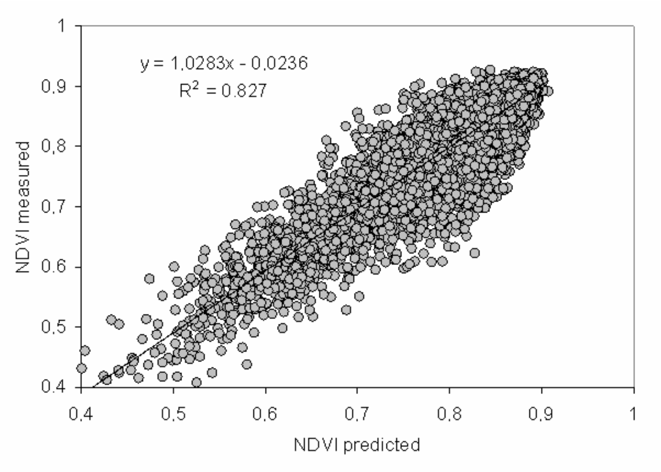


Figure 3: Scatter plot of measured NDVI versus NDVI predicted by the GWR model with the bandwidth of 7 pixels.

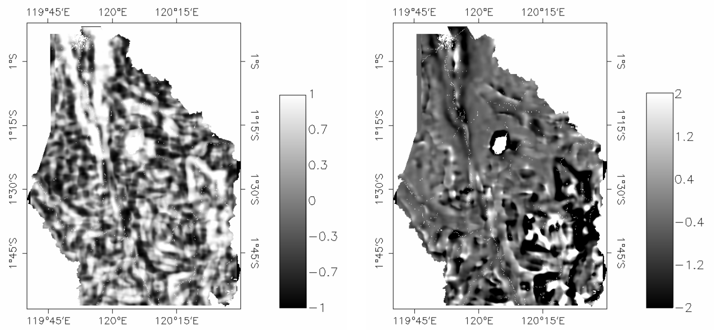
When observing the distribution of the R^2 across the study region (Figure 4) one can recognize the general pattern which agrees with the land-cover pattern. But in comparison to the stratified model the GWR model also exposes a mosaic of variance in R^2 within this general pattern scaling down to the individual locations. It means that the general nature of the relationship appears relatively stable according to the response of different vegetation types to rainfall. Nevertheless, the local variances in this response caused by the variance in underlying physical factors are also included in the model.

3.4 Effect of Scale Variation on Results of the GWR Model

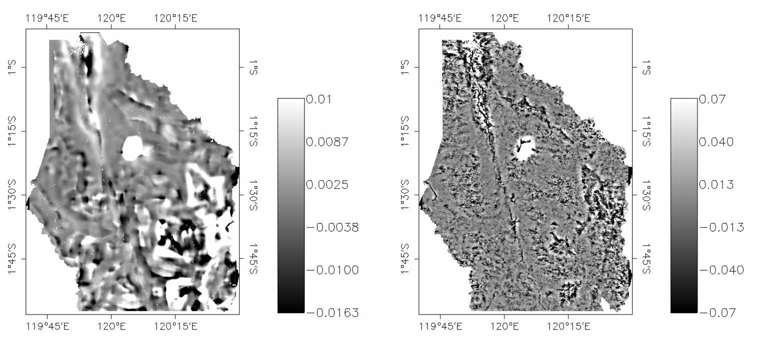
Obviously, the magnitude of local variance in the underlying physical factors should be sensitive to the observation scale. As a consequence, the explanatory power and accuracy of the model between NDVI and precipitation depends on the bandwidth used. Foody (2004) reported about a rapid increase of the explanatory power of the GWR model by declining the scale of the analysis. In our study such a decline would indicate that the spatial dependency exists over relatively small distances and that there is high frequency spatial variation that reflects the confined range characteristics of the underlying environmental factors.

In this work the effect of variation in spatial scale of the regression model's prediction was explored by varying the bandwidth used in the GWR model. As the bandwidth declines, the analysis becomes increasingly local, revealing greater geographical detail. With a very small bandwidth (*b* = 3 pixels), the relationship between NDVI and precipitation was very strong, $R^2 = 0.92$. With an increasing of the bandwidth the spatial patterns in the local estimates of the model parameters became more generalized and the magnitude of the estimated parameters tended towards the global model estimate (Figure 5).

Large windows improve the *t*-value of the model parameters (in sampling terms, increasing the size of the sample) by borrowing large amount of local information but at the expense of introducing bias because information is being borrowed from areas, further away, that may be different. Small windows reduce the risk of bias in the statistics but because little information is being borrowed the precision is not much improved. The effectiveness of local borrowing depends on the local homogeneity of the spatial data which depends on the size of spatial units in relation to the true scale of spatial variation. If adjacent areas are very different in nature then borrowing information locally may introduce bias that distorts the underlying patterns through inappropriate bandwidth. The GWR analysis showed that the most appropriate bandwidth is 7 pixels (a window of 13*13 pixels or 3500*3500 m in the reality). This dimension may be considered to reflect the “normal” size of homogeny landscape units in the study region. The presence of the scale effect in the strength of the NDVI-precipitation relationship and the prediction uncertainty of the model indicates that non-stationarity plays an important role in the ecological modelling and that the geography matters and location should be considered as a variable.

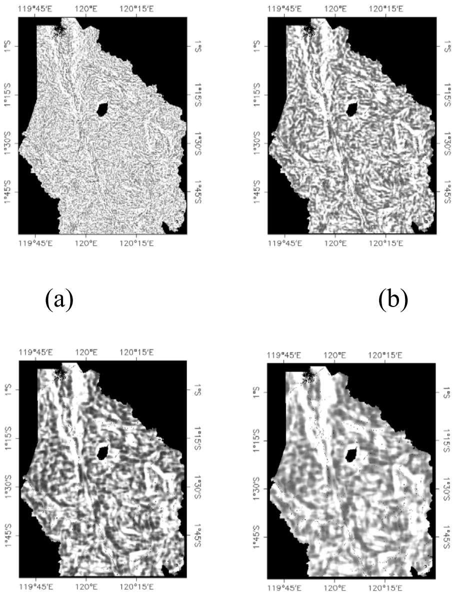


(a) (b)



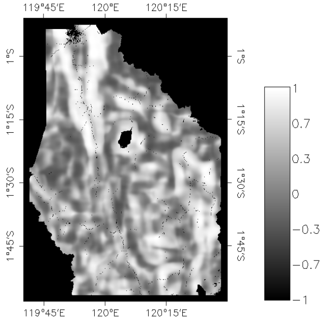
(c) (d)

Figure 4: Spatial variation in regression outputs from the GWR analysis of NDVI against precipitation for the bandwidth of 7 pixels. (a) Correlation coefficient, r; (b) model intercept; (c) slope parameter; (d) residuals.



(a) (b)

(c) (d)



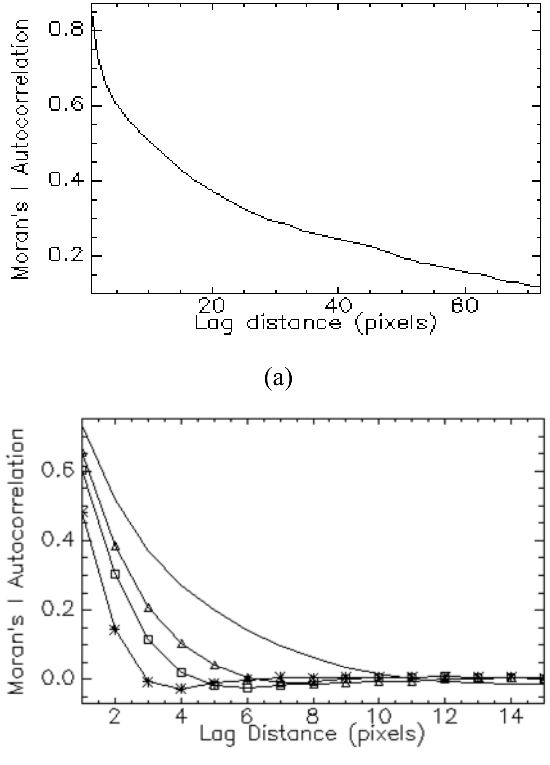
(e)

Figure 5. Spatial variation in the correlation coefficient between NDVI and precipitation at five bandwidths: (a) *b* = 3, (b) *b* = 7; (c) *b* = 11; (d) *b* = 15 and (e) *b* = 25. Spatial detail increases with a decrease in bandwidth, *b*.

3.5 Autocorrelation of Regression Residuals

Geographical phenomena tend to vary smoothly. That means values of locations close together tend to be more similar than values at distant locations. Spatial autocorrelation of the source variables represents a problematic aspect that has to be treated by modelling the relationship between these variables. The importance of considering spatial autocorrelation in studies on geographical ecology has been addressed by a number of recent papers (Griffith, 2003; Ji & Petters, 2004; Wang et al, 2005). The problematic of a regression model reflects in its residuals which spatial distribution is considered to be an effective indicator for the analysis uncertainty. The spatial patterns of residuals are important indices to examine how accurate the regression model reveals the real relationship. The residuals of a linear regression model are required to be independently distributed over space with a mean of zero and a constant variance. If the residuals exhibit some non-random patterns the model created is problematic. A diagnostic statistics indicating problems in regression modelling is the degree of spatial autocorrelation exhibited by the residuals from the model. The standard errors are usually underestimated when positive autocorrelation is present.

For each regression model we calculated the Moran's *I* autocorrelation of the residuals to examine the effect of calibrating the model between NDVI and precipitation at different spatial scales. As it has been proved, the local calibration solves much of the problems of spatially autocorrelated error terms included in the traditional global OLS model (Wang et al., 2005; Fotheringham et al., 2003). We were interested in the comparison of the results from the global and local models. Figure 6 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 60 pixels. In comparison, no significant positive spatial autocorrelation was found for the GWR model residuals at the distance more than 5 - 8 pixels. This suggests that the calibration of a local model reduces the problem of spatially autocorrelated error terms.



(a)

(b)

Figure 6. Moran's I autocorrelation of regression residuals from the global OLS model (a) and from the GWR model with different bandwidth (b): solid line – 25 pixels, line with triangles – 15 pixels, line with squares – 11 pixels, and line with asterisk – 7 pixels. Obviously, that the residuals from the GWR model exhibit no significant autocorrelation at the distance more

that 3-10 pixels, while the residuals from the OLS model are autocorrelated to a distance up to 60 pixels.

4 CONCLUSIONS

In this paper we revealed spatial variation in the spatial relationship between normalized difference vegetation index (NDVI) and rainfall in a subequatorial region of Central Sulawesi, Indonesia. The study investigated this variation both at and between spatial scales. The analysis based on the use of two different regression techniques: one is the global ordinary least squares regression, OLS, and the other is the geographically weighted regression, a relatively new local regression technique which allows the regression parameters and the strength of the relationship to vary over space. The analysis proofed the presence of non-stationarity in the NDVI-precipitation relationship both between the main land-cover types and between locations. It means that the modelling of this relationship with the global or stratified OLS regression attains results with high amount of uncertainty. The variance in the relationship across the space of the study region is explained by the variance in the underlying environmental factors such as vegetation composition, soil type, hydrology, land use etc. caused by the diversity of terrain. That agrees with the results of recent studies on vegetation-climate relationships from other regions (Yang et al, 1998; Ji & Petters, 2004). However, in spite of the spatial non-stationarity of the NDVI-rainfall relationship the general nature of the vegetation response to precipitation is considered to remain relatively stable: forested areas respond to precipitation only weak (from $R^2 = -0.3$ to $R^2 = 0.3$), whereas areas covered by shrubland or grassland show high response to precipitation ($R^2 = 0.3 – 0.98$).

Spatial non-stationarity of the relationship between NDVI and precipitation contributes essentially to scale-dependency in the results of the analysis (Foody, 2005). The GWR model enabled to use kernel bandwidth with different size (750 - 6250 m) working like some sort of a spatial microscope and scaling the modelling relationship from sub-regional to local scale and helping to determine the most appropriate scale. The results have shown that the regression parameters, the predictive power as well as the rank of the explanatory variable in the model of vegetation patterns is considered to represent a function of spatial scale. The results suggest that the explanatory power of the analysis increased very significantly with a diminishing of the scale. The NDVI-precipitation modelling provides the most accurate prediction by the use of the GWR model with a bandwidth of 7 pixels. This model explains about 83 % of all variance in NDVI over the study area. Further decreasing of the analysis scale results in enlarge of CVSS of the model and Moran's *I* autocorrelation of its residuals.

The results suggest that the calibration of local rather than global models reduces the problem of spatially autocorrelated errors. The residuals from the global OLS model clearly exhibited positive spatial autocorrelation up to approximately 60 pixels. In comparison to that, the residuals from the GWR model showed positive autocorrelation at the distance at least 10 times shorter, 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) and Wang et al. (2005).

Our study 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. This superiority is mainly 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 NDVI was unexplained. 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 a reliable local fitting.

ACKNOWLEDGEMENTS

This research work has been carried out within the special research project of SFB 552 STORMA, part project D6. STORMA is an acronym for Stability of Rainforest Margins in Sulawesi, Indonesia. This project is funded by the Deutsche Forschungsgemeinschaft (DFG).

REFERENCES

Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., Eklundh, L. (2004). “A simple method for reconstructing a high quality NDVI time series data set based on the Savitzky-Golay filter”. Rem. Sens. Environment, 91: 332-344.

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

Erasmi, S., Bothe, M., Petta, R. A. (2006). “Enhanced filtering of MODIS time series data for the analysis of desertification processes in northern Brazil”. The international Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 34, Part XXX.

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

Foody, G. M. (2004). “Spatial 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., Brunsdon, C. and Charlton, M. (2002). Geographically weighted regression: the analysis of spatially varying relationships. Chichester, Willey.

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.

Griffith, D. A. (2003). „Spatial autocorrelation and Spatial Filtering”. Berlin, Springer-Verlag.

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.

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

Paez, A., Uchida, T. and Miyanmoto, 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.

Propastin, P. and M. Kappas. (2007). “Reducing uncertainty in modelling NDVI-precipitation relationship: a comparative study using global and local regression techniques”. (accepted for publication in “GIScience and Remote Sensing).

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.

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

Twele, A., Holecz, F., Erasmi, S. (2006). “Synergy of optical and radar satellite data for enhanced land cover mapping in tropical regions – a case study from Central Sulawesi, Indonesia”. (submitted to Int. Journal Rem. Sensing).

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.