Use of Geographically Weighted Regression to Determine Natural Rubber Productivity and Their Driving Forces: A Case Study in the Kalutara District of Sri Lanka

Sankalpa, J. K. S.*[1], Wijesuriya, W.[1], Karunaratne, S.[2] and Ishani, P. G. N.[1]

[1] *Rubber Research Institute of Sri Lanka, Dartonfield, Agalawatta, 12200, Sri Lanka*

[2] *University of Western Sydney NSW, 2006, Australia*

**Corresponding Author*:
Tel: (94)766947839, Email:ssankalpa2@gmail.com

ABSTRACT

The goal of this study was to analyze the productivity variation in smallholder rubber lands in Kalutara district located in the wet zone of Sri Lanka and spatial relationship of key drivers to the productivity variation. Low productivity has been a major challenge in rubber plantations in the country in recent years. In this study spatial modelling tools available in geographic information science were used to explore the spatial variability of the rubber productivity and explored the key drivers of it in spatial context. Geostatistical kriging analysis is a simple type of prediction method which includes the cross validation of prediction and error terms in forecasting techniques. The productivity of smallholder rubber lands in Kalutara district varied from 777 to 1463 kg/ha/year, while the highest average productivity was recorded in the Divisional Secretariat (DS) divisions; Palindanuwara, Beruwala and Kalutara. Low productivity was recorded in Matugama and in a few areas in Ingiriya and Bandaragama divisions. Local variation of driving forces behind the average productivity was explored using Geographically Weighted Regression (GWR) method. GWR explored the spatial variability of the relationship between productivity and fertilizer usage, weeding, soil conservation, number of tappable trees and age of trees under tapping. All the variables were found to present significant spatial variabilities. Apart from generating global significant value, the model resulted local variation of each parameter estimates with respect to the projected coordinates of the area. Emerge of sign change of local parameters observed in some areas cannot be observed globally. It is necessary to understand the significance level of local coefficient subject to the multicolinearity and spatial auto correlation.

KEYWORDS: Geographically weighted regression; Kriging geostatistical analyst; Spatial auto correlation

Introduction

Rubber sector plays a vital role in the Sri Lankan economy, in terms of export earnings valued about US$ 0.8 billion in year 2016 (Sri Lanka Customs Department, 2016). The Sri Lankan rubber sector comprises of smallholders (land extent less than 20 ha) and large estates (land extent equal or more than 20 ha). The average land productivity of rubber plantations was recorded as 819 kg ha^{-1} yr^{-1} in 2016 (ANRPC, 2017). However, this is far below the potential productivity which is about 2500 kg ha^{-1} yr^{-1} with the new recommended clones.

The average productivity has decreased in the country due to various reasons; *viz*. land fragmentation, low level of soil fertility and conservation practices, low quality of tapping and several rubber plant diseases recorded in the country. Among the major rubber producing countries, highest average productivity has been recorded as 1680 kg ha^{-1} yr^{-1} in Vietnam while Philippine has recorded the lowest productivity as 694 kg ha^{-1} yr^{-1} among the members of the Association for Natural Rubber Producing Countries in year 2016 (ANRPC, 2017). Further, average productivity varies with the geographic location within the country due to different environmental and management practices. Awareness on recommendations which is dependent on the effectiveness of extension services is generally related to adoption of management practices. However, finding a solution to low productivity issue is problematic due to unavailability of farm level information. Use of recommended agronomic practices and the way of land management by farmer vary from one land to another. Investigation of all possible reasons isn't a practical application. Hence, quantification of prevailing management practices by land is an essential requirement before applying any solution. Effect of agronomic and land management conditions and productivity has a spatial variability. Use of Spatial analysis tools in forecasting of unknown locations for the agronomic practices as well as the average yield is possible, and similar techniques are successfully used by many scientists in different occasions. Therefore, agricultural extension workers and policy decision makers have capability in deciding the level of improvement of agro-management practices to achieve desired productivity in the field. This is especially important in devising demand driven extension programs in the rubber sector.

With the advancement of Information Science, the Geographic Information System (GIS) has been used extensively to demonstrate the spatial variability of interested attributes for scientific and policy level decision making. GIS is a computer based information system which integrates hardware, software and data for capturing, managing, analyzing and displaying all forms of geographically referenced information. GIS also helps to answer questions and solve problems by looking at available data in a way that is quickly understandable and easily shared (Anon, 2012). Use of GIS in socio-economic applications including different policy sectors, has become an important tool in scientific research (Higgs *et al.,* 2003). Geostatistics is one of the most important spatial analyst tools that can be used for analysis of point data and combinations with various GIS layers. One of the common uses of Geostatistics is spatial interpolation or prediction which is used to predict values of a sample variable over the whole area of interest.

In order to model and understand the relationship and effect of different factors and rubber productivity at a given point of spatial location, Geographically Weighted Regression (GWR) can be applied as a method of quantitative analysis which facilitated global and local understanding of the processes. GWR is a method of analyzing spatially varying relationships. It is increasingly being used in geography and other disciplines that deal with spatial data. Recently, this technique has extensively applied in the disciplines of economics, urban studies and environment. It has also been used to study spatial variability in areas such as industry and nutritional poverty (Jaimes *et al.,* 2010). The behavior of variables at local level on deforestation of the state of Mexico has been analysed in that study. Several authors have found that socio-economic variables; land ownership, availability of credits, biophysical variables and other proximity variables have spatially varying relationships. Also some proximity variables of wood processing industry, *viz.* distance to highways and land allocation to households have been found as drivers of afforestation and there have been a positive and negative correlations with significant spatial variability in the Northern Vietnam (Clement *et al.,* 2009). Spatial heterogeneity is evident in most geographical phenomena. Partridge *et al.,* 2008 has observed the spatial heterogeneity of non-metropolitan growth mechanism including employment growth of US using GWR and has compared them with global estimates. It reported that influence of different variables have spatial variation but their predictive capacity depends on the location. In certain cases, global regression suggested that there is no average effect on growth, while such variables have positive and negative effects for some regions. GWR allows individuals to explore and understand the spatial distribution of variables or processes by fitting a regression equation to any point in space. For this study, GWR is used because unlike the "classic" methods it considers the location of the phenomenon studied. This paper does not seek any causal depth in its models, but it rather concerns about providing an approach to an explanation of relationships that occur spatially.

Land productivity is among the key indicators which drives the sustainability of the rubber sector. Analyzing this indicator in the spatial domain is important in extension planning and various policy implications. Currently, it is being realized that GIS provides easy access to spatial information for policy makers and administrators. This reflects in the growing interest in the concept of Spatial Data Infrastructure (SDI) at national and global levels. Since SDI helps to provide geographic information to decision makers, it offers the prospect of better decision-making in the management and development of resources and, hence, improves socio-economic growth.

Kalutara district, which was considered in this study, has a rubber land extent of 19,053 ha according to the rubber land census conducted by Rubber Development Department (RDD, 2011). The extent of land in Kalutara district is the second highest when the country's total rubber extent, *viz.* 134,000 ha, recorded in year 2014 is concerned (MPI, 2015). The study is focused on the smallholder rubber sector in Kalutara district with the objectives of estimating the land productivity and developing maps employing GIS tools for the purpose of efficient decision making especially in extension activities.

Methodology

Study Area

The study area is located between 6^0 46' and 6^0 25' (Northern latitude) and 80^0 09' and 80^0 18' (Eastern longitude). Rubber is found in all 14 Divisional Secretariat (DS) divisions (Several Grama Niladari regions form a DS division) in Kalutara district belonging to agro ecological regions, WL_1 and WL_4.

Data Collection

Data were collected from a primary survey carried out in Kalutara district. Based on the number of smallholders in Kalutara DS division, the sample was selected by using stratified proportionate random sampling technique. The sample consisted of 250 smallholder farmers covering the rubber growing areas of Kalutara district. Sampling frame was decided based on the rubber smallholdings with mature rubber lands. The sample included farmers from all DS divisions based on the proportion of rubber lands in each division.

Preparation of Spatial Data Layers

Average productivity of smallholder farmers was calculated using production (kg) and mature extent of the land. ArcGIS version 10.2 was used for Geostatistical prediction to estimate the Productivity in unmeasured locations in Kalutara district. Ordinary kriging was practiced as the type of kriging while using optimum semivariogram model selection. The model for simple kriging explained in this study is given in equation 1. The spatial analysis framework which involved kriging is depicted in the Figure 1. All prediction were masked using smallholder natural rubber area map of Kalutara District.

$$Z(s) = \mu + \varepsilon(s) \qquad (1)$$

Where, μ is a known constant (mean) and ε is the error term. Simple kriging uses the semivariogram or covariance model. Cross validation was carried out in the final steps of kriging methodology. Cross-validation and validation help to make an informed decision as to which model provides the best predictions. Average standardized error and root mean standardized error were calculated to assess the prediction performance. If the prediction errors are unbiased, the mean prediction error should be around zero.

Geographically Weighted Regression (GWR)

Geographically weighted regression analysis was done using GWR tool developed by Geoda center for geospatial analysis and computation, Arizona State University US. (Brunsdon *et al.,* 1998; Fotheringham *et al.,* 2002). Gaussian type of model and geographically variability test was selected as model settings. Regressions were carried out in the localized points within Kalutara DS area. The regression model can be expressed as follows (eq.2) (Fotheringham *et al.,* 2002).

$$Y(u,v) = b_0(u,v) + b_1(u,v)x_1 + b_2(u,v)x_2 + \ldots + b_n(u,v)x_n + e(u,v) \qquad (2)$$

Coefficients $b_1,..b_n$ represent the magnitude of effects of the exploratory variables $x_1,..x_n$ and variable Y is the yield per hectare. The coefficient, b_0 is the constant term. Error term is described by e while u and v represent different influences on each specific location. GWR was carried out using management practices; fertilizer application, weeding and soil conservation as independent variables. Further, the tapping stand (tappable trees/ha) and age of tappable trees were used as independent variables to represent cultivation information. Exploratory variables in unmeasured survey locations were interpolated using kriging spatial interpolation method which is similar to the procedure described in Figure 1. While ordinary kriging was practiced for the tappable trees and age of trees, indicator kriging was applied for management practices; fertilizer application, weeding and soil conservation. Probability of receiving those management practices were calculated for the non-sample points.

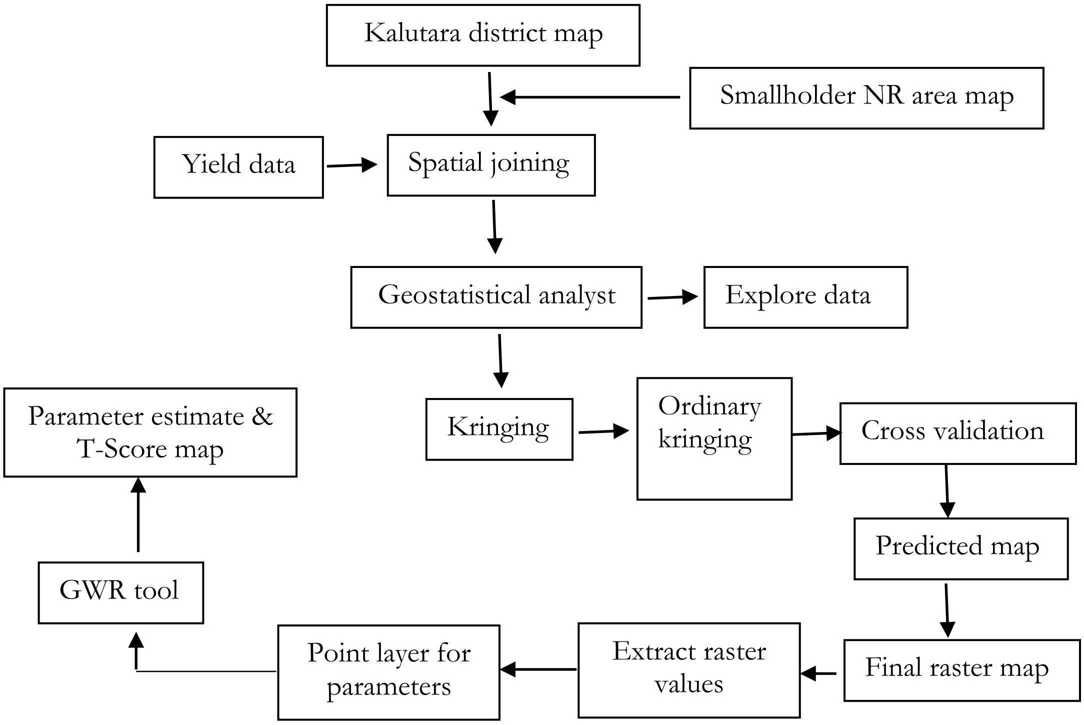


Figure 1. Flow Diagram of Kriging Spatial Analysis

Different forms of kriging, namely ordinary kriging for continuous variables and indicator kriging for categorical variables were used to create the surface maps. Summary of the methods deployed together with kriging model is depicted in Table 1.

Table 1. Different Forms of Kriging Methods Applied to Create Continuous Surfaces

Parameter	**Kriging interpolation type**
Tapped trees	Ordinary

Age of plant	Ordinary
Fertilizer application	Indicator
Soil Conservation	Indicator
Weeding	Indicator

Probability values of Non-sample points were extracted using centroid of the each Grama Niladari division in Kalutara district. It was observed that the GWR model parameters best explain the driving forces of productivity change in Kalutara district. One of the important GWR model result was the spatial variation of the model fit. In this case local R^2 values for the fitted model was calculated.

Results and Discussion

Productivity Variation in Kalutara District

Geostatistical analysis generated a continuous map that depicts the average productivity in Kalutara district (Figure 2.). Higher, Medium and Lower productivity values have shown in the Figure 2. Predicted average productivity was high in Palindanuwara (1208 kg ha^{-1} yr^{-1}), Beruwala (1202 kg ha^{-1} yr^{-1}) and Kalutara (1196 kg ha^{-1} yr^{-1}) DS divisions followed by Dodangoda DS (1176 kg ha^{-1} yr^{-1}). According to the statistics released by Rubber Development Department, highest extent of rubber lands are available in Bulathsinghala DS division followed by Palindanuwara (2443 ha) and Horana DS (1300 ha). About 540 ha of rubber lands are available in Beruwala DS and only 225 ha are available in Kalutara DS division. While the maximum productivity recorded in the Walallawita and Kalutara DS divisions, the minimum was predicted in Ingiriya and Millaniya DS divisions. Descriptive statistics of Predicted Productivity are summarized in the Table 2.

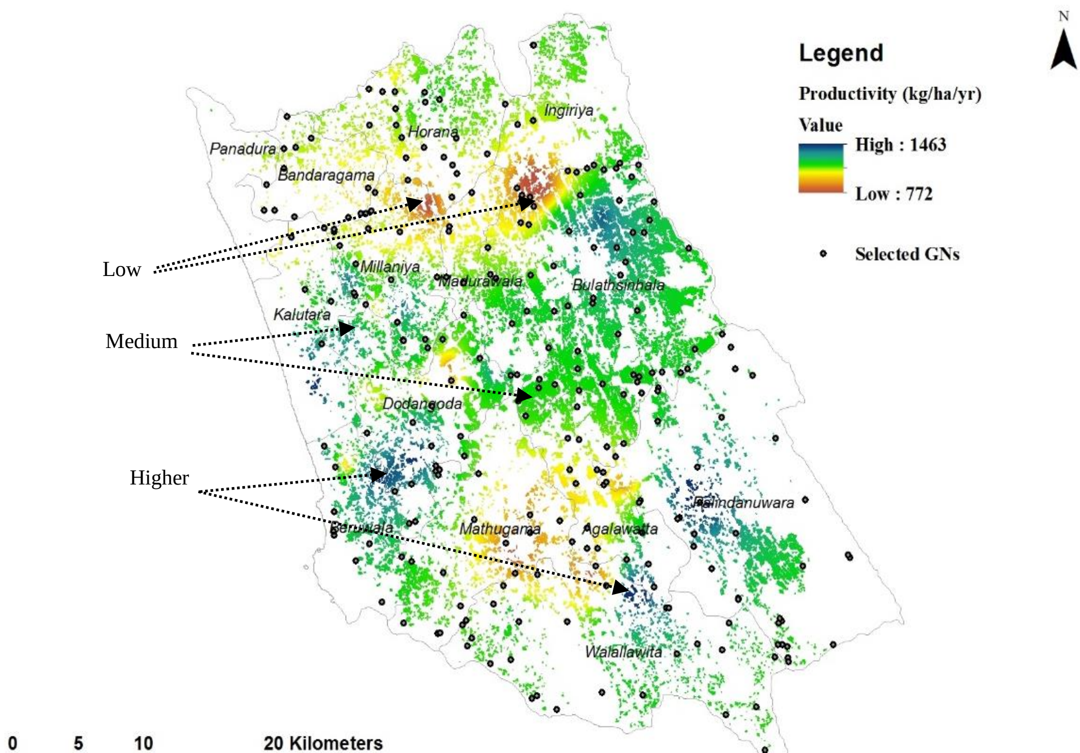


Figure 2. Predicted Productivity Raster Map for Kalutara District

Table 2. Summary Statistics of Productivity Values (kg ha^{-1} yr^{-1}) of Different DS Divisions in Kalutara District

DS Division	Mean	Max	Minimum	Standard Deviation
Agalawatta	1124	1331	982	49
Bandaragama	1082	1122	1033	18
Beruwala	1202	1351	1017	49
Bulathsinghala	1173	1310	958	48
Dodangoda	1176	1373	946	61
Horana	1102	1178	905	43
Ingiriya	1075	1169	772	96
Kalutara	1196	1384	1051	70
Madurawala	1120	1202	951	57
Matugama	1075	1243	872	67
Millaniya	1108	1307	849	82
Palindanuwara	1206	1374	971	60
Panadura	1090	1143	988	24
Walallawita	1160	1463	894	78

Validation of the Geostatistical Model for Productivity

The root mean square standardized error which reflects the model accuracy reported as 1.03. This should be close to one if the prediction standard errors are valid. If it is greater than one, the model is underestimating the variability in predictions. The reported value suggesting that predicted productivity values at non-sample points were close to the actual values against the collected sample points (250 sample points) which were used for validation.

The Spatial Drivers of the Productivity

Summary of the fitted GWR model is depicted in the Table 3. Global t-value explains the direction of relationship in the area and the statistical significance level change in the GWR. The Global regression estimates or parameters are given in column 2 of Table 3. Variables selected in the analysis were significant at 95% confidence interval in the global model and there were spatial variability in the area. However, some variables indicate weak significant spatial variability in the area. Variables which were significant in the GWR can either give a positive or negative spatial variability. Mapping of local-pseudo t-values was done to draw more attention on such cases and to distinguish different variables.

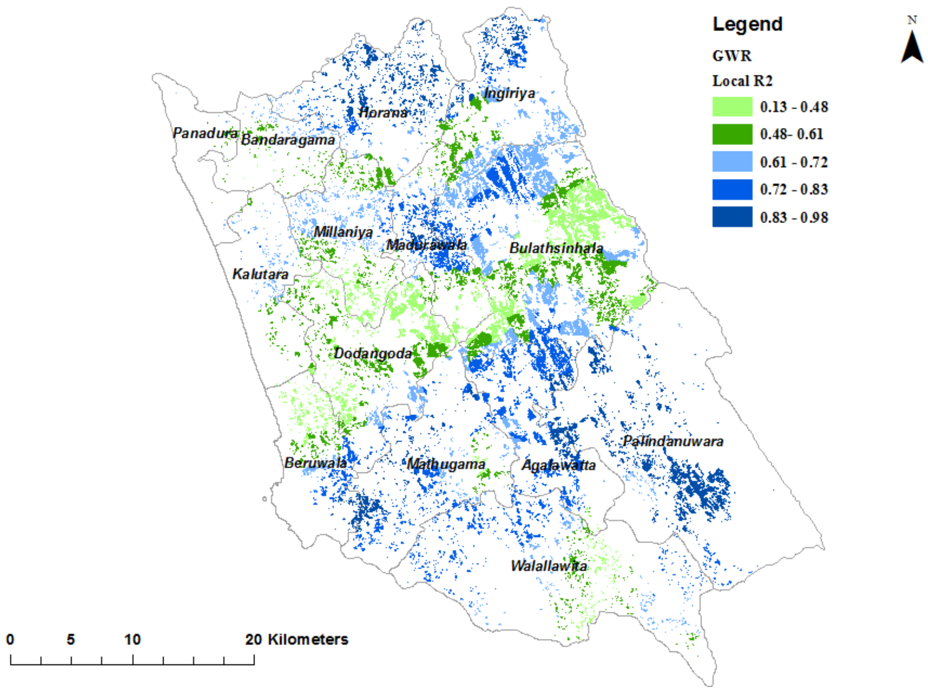


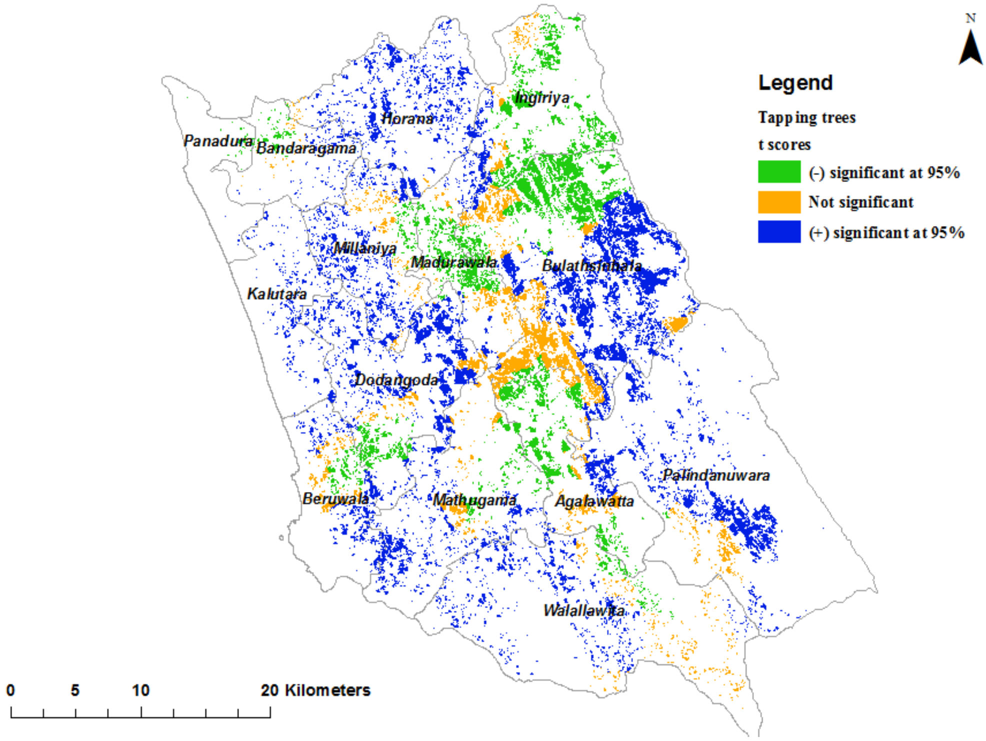
Figure 3. Local R^2 Value of GWR

Table 3. Results of Geographically Weighted Regression (GWR)

Variables	Global t-value	p-value*	GR estimate	GR SE	DIFF C
Intercept	624.54	0.00001	1146.84	1.83	-1229.68
Fertilizer application	2.66	0.004	5.43	2.03	-129.49
Weeding	4.48	0.00001	10.77	2.40	-165.05
Soil conservation	11.02	0.00001	26.41	2.39	-73.86
Age of trees	-9.21	0.00001	-18.89	2.05	-139.88
Tappable trees	5.55	0.00001	11.53	2.07	-122.50
N	1337				
AICGR	15,051				
AICGWR	12,934				
GR adjusted R^2	0.23				

GWR- adjusted R^2	0.85
GWR F value	49.71

*Note: GR: Global regression; GWR: Geographically weighted regression, AIC: Akaike information criterion, DIFF C: Difference of criterion, N= number of samples selected. GR SE: Global regression standard error of estimate. *All variables were significant at 95% confidence interval*



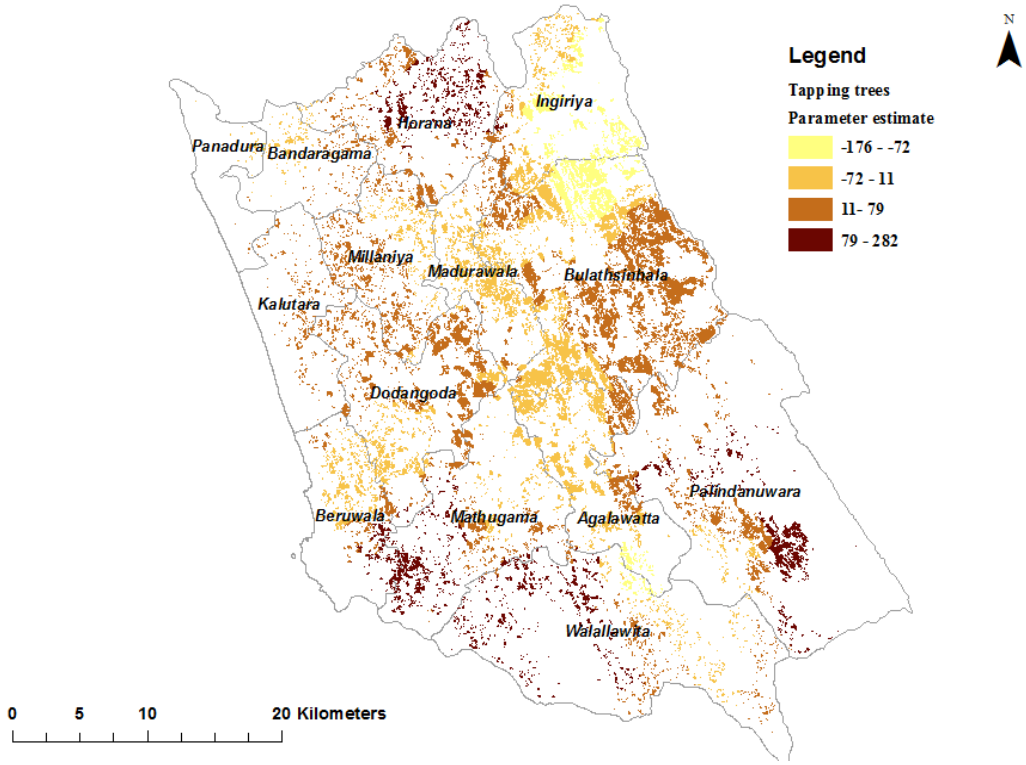


Figure 4a. Spatial Variability of Parameter Estimate and t-values of the Variable Tapped Tree

An important spatial distribution obtained from the GWR analysis was the spatial variation of the model fit. Spatial variation of the local adjusted R^2 values for the fitted model is shown in Figure 3. Local adjusted R^2 values were higher than the global model which varied from 0.13 to 0.98. It is important to note that compared to OLS or GLM where regression coefficients are fixed, GWR are variable regression coefficients across space. This enables to model the local variability of the rubber productivity more meaningfully when compared to the global model.

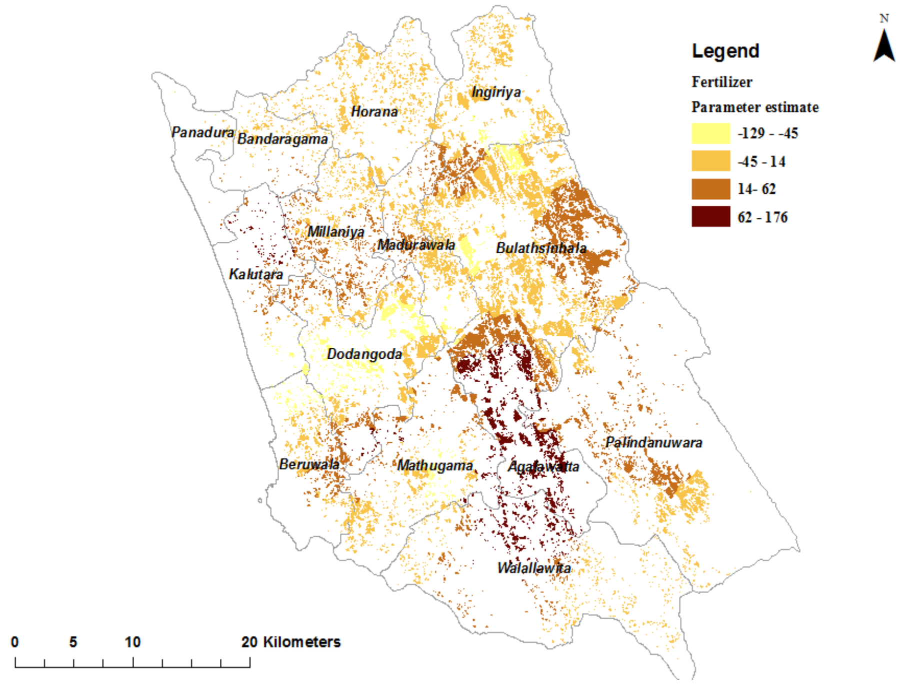
Moreover, the increase in adjusted R^2 confirms that GWR-adjusted model explains the variance of the data better than the global model. Higher R^2 values, permits that the correlation of the selected variables in the model and productivity are captured better by the GWR model.

When GWR results were mapped, estimated parameters and t values changed throughout the area. Furthermore, 95% confidence level was considered to map the t-values change throughout the region. Model results show that variable related to tapped trees have significant positive and negative spatial variability (Figure 4a.). The parameter estimates for tapping tress were positive for most divisions except for Ingiriya, Dodangoda and Agalawatta areas.

It was identified that the spatial distribution of the use of fertilizer is considered as an important driver of rubber productivity (Table 3). Estimated regression parameters for the fertilizer use were positive in most of the areas except for few areas in Ingiriya, Dodangoda, Horana and Matugama divisions (Figure 4b.). The Sri Lankan government has continued the fertilizer subsidy for small holder farmers for the last consecutive three years. However, most of the farmers haven’t applied fertilizer during the last two years in Mathugama area.

As for the soil conservation variable in the model, it shows a significant variability while impact of relationship was positive and negative in areas of Kalutara district (Figure 4c.). The values for the estimates showed that Millaniya, Bulathsinghala and Agalawatta areas reported positive relationships (for the estimated parameters), while the Dodangoda, Horana and Ingiriya areas reported a significant negative relationship. No responses were received from most of the farmers in Dodangoda and Beruwala area regarding any soil conservation methods practiced during the cultivation since most of them haven’t applied considerable measures for soil conservation in their lands.

Improvement of soil conservation practices maintaining the soil cover and minimizing the soil disturbance are important land management practices. Weeding is another important land management practice. Proper weeding management enhances the nutrient absorption by natural rubber without wasting fertilizer. The estimated parameters for weeding practices in the field shows significant variability in Kalutara district, and they were significantly negative in Mathugama, Beruwala, Ingiriya and Madurawala areas (Figure 4d.). Proper weeding practices were mostly recorded in Bulathsinhala, Agalawatta and Walallawita areas while weeding practices didn't produce significant positive variability in most of the areas in other divisions.



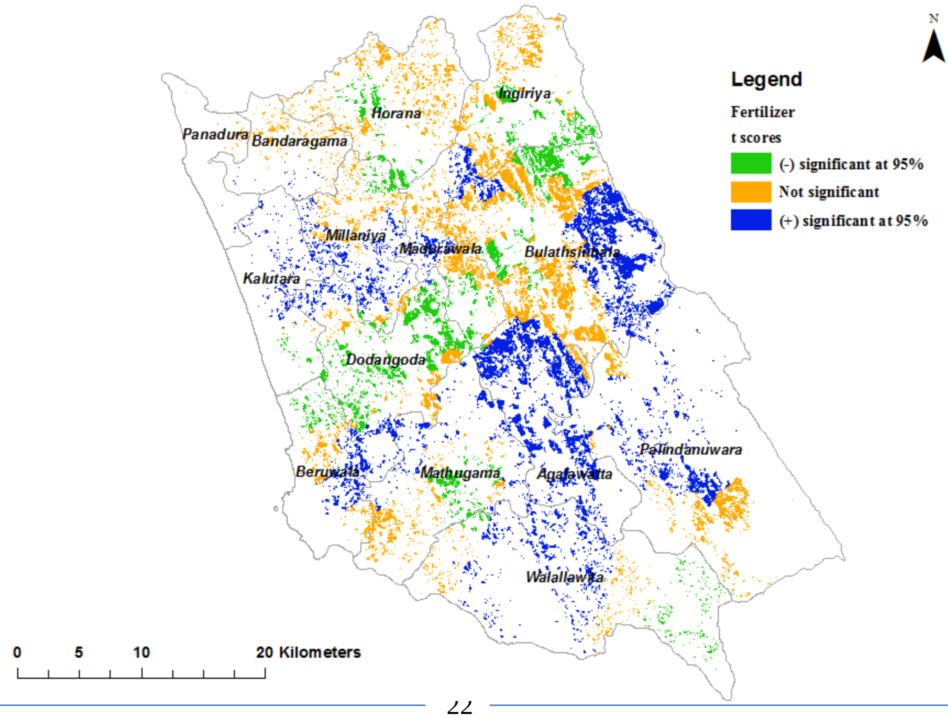
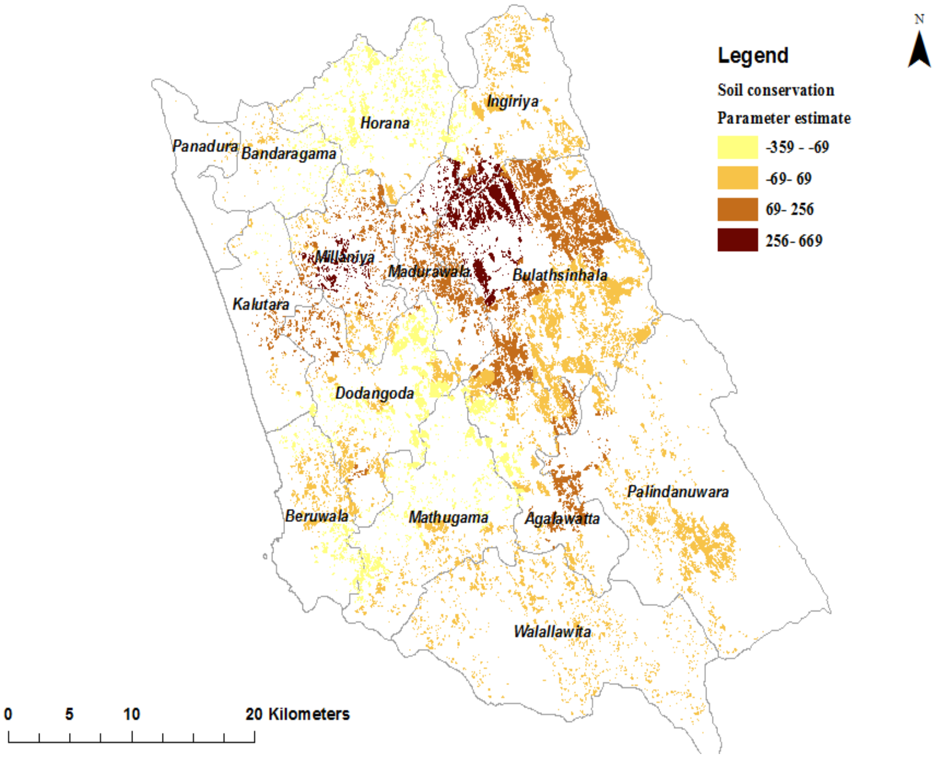


Figure 4b. Spatial Variability of Parameter Estimate and t-values of the Variable Fertilizer



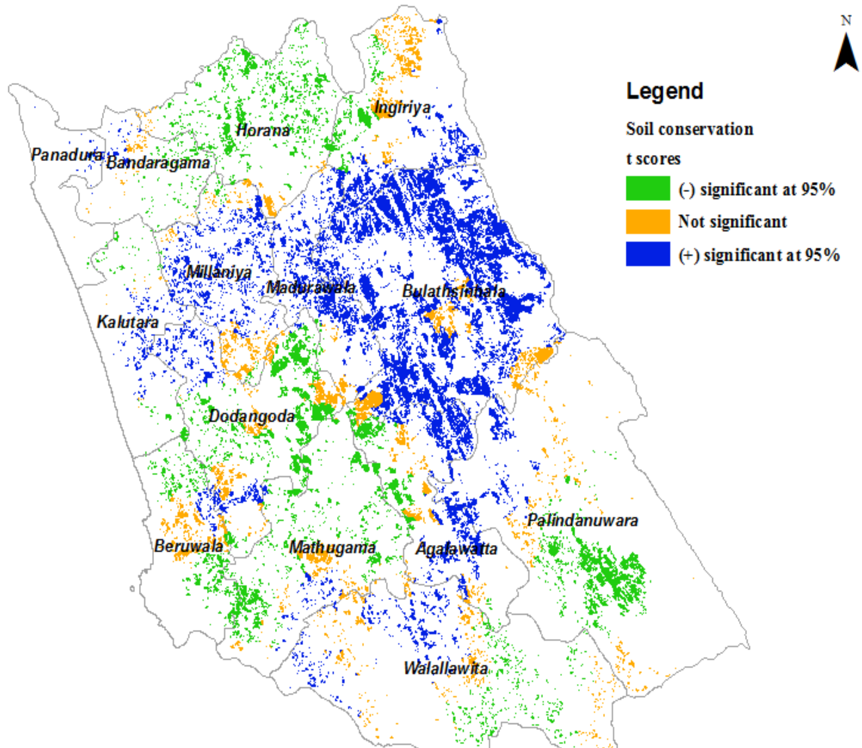


Figure 4c. Spatial Variability of Parameter Estimate and t-values of the Variable Soil Conservation

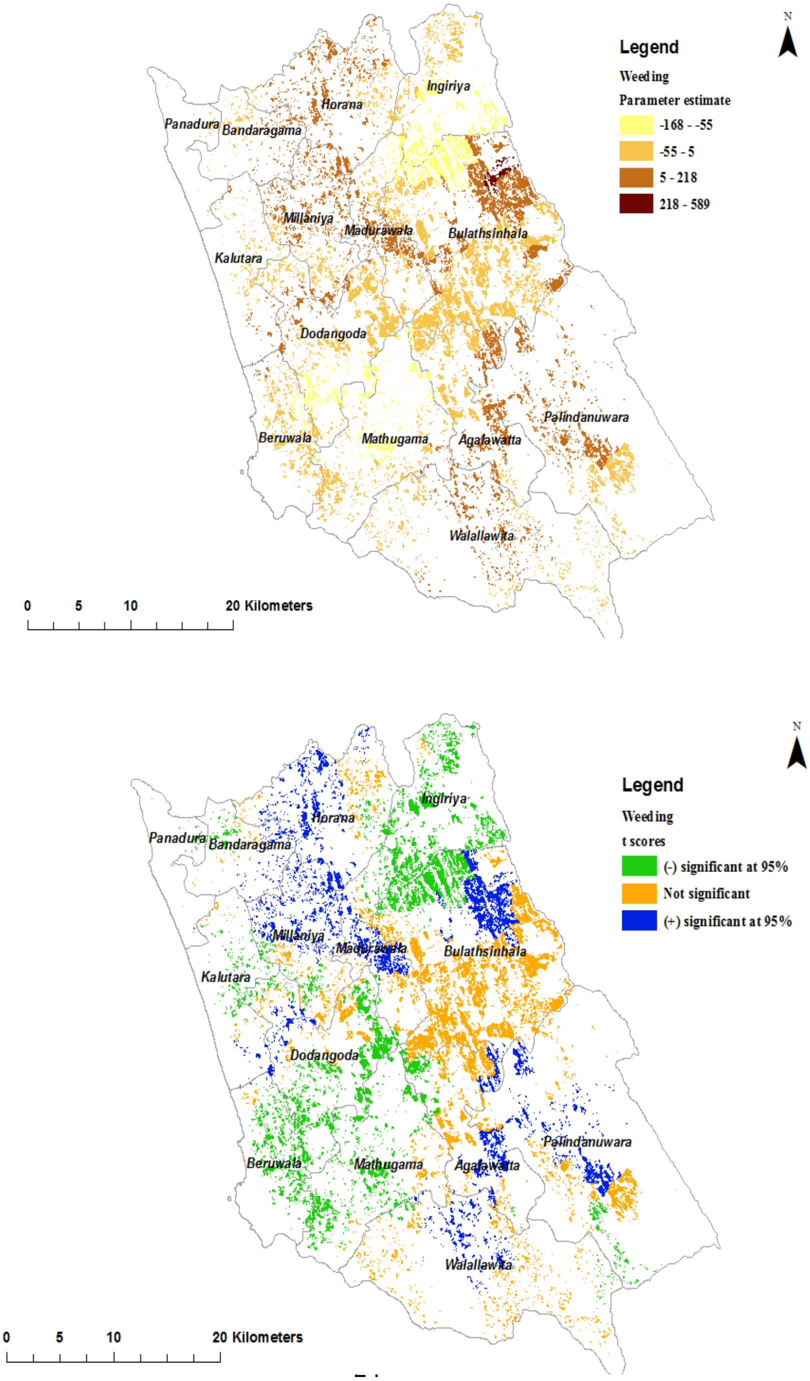
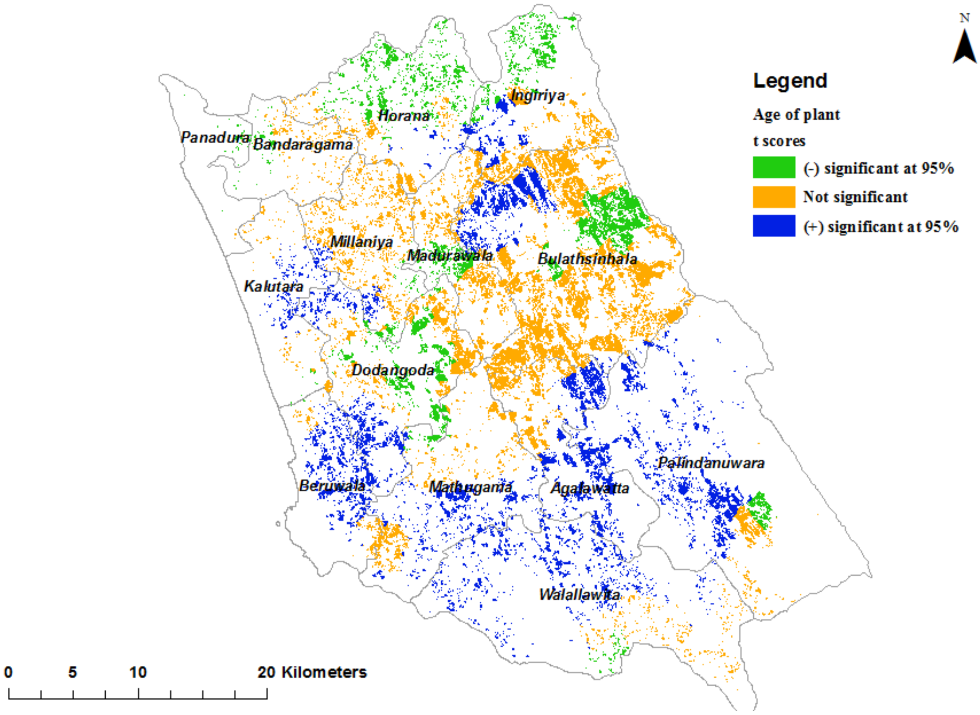


Figure 4d. Spatial Variability of Parameter Estimate and t-values of the Variable Weeding



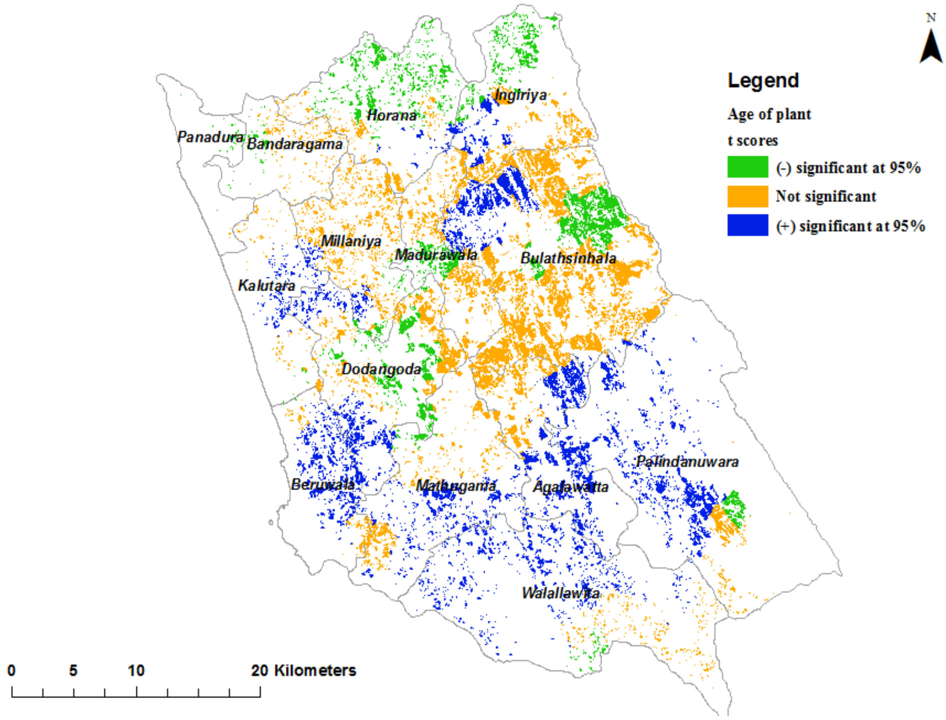


Figure 4e. Spatial Variability of Parameter Estimate and t-values of the Variable Age of Tappable Trees

Significance of the estimates for the age of rubber plantation shows considerable variability and it depends on the age profile of surveyed sample. The estimates for the age of rubber presented here display a significant positive variation for Agalawatta, Walallawita and Beruwala areas (Figure 4e.). According to the yield curve of natural rubber, yield increases up to age of 19 years. Thereafter, yield shows a negative relationship with the age. (Munasinghe *et al.,* 2008; Wijesuriya *et al.,* 2012). Mature extent of rubber lands in Kalutara district was different within the district. Highest extent of mature areas was recorded in the Bulathsinhala divisional secretariat followed by Palindanuwara and Walallawita. Lowest mature extent was recorded in Panadura area. With respect to age class, Bulathsinghala and Palindanuwara followed by Agalawatta and Walallawita areas were having higher land extents in age between 10 to 19 years than mature lands in other categories.

Conclusions

GWR explored the behavior of variables in the model at a local level and revealed significance of their spatial variability. This is further enhancement of understanding of a local analysis, rather than obtaining global average for the entire district. One of the important contributions of this study is that it enables the spatial variability of results to be studied. This has facilitated the identification of those parameters which exercise influence throughout Kalutara district. Results obtained from the GWR demonstrated that explanatory variables such as fertilizer, weeding, soil conservation, age of plant and tapped trees do not have constant homogenous parameters across the study area. This study has thus identified spatial variability of parameters.

Acknowledgements

The authors wish to acknowledge all the staff members of Rubber Research Institute of Sri Lanka and Land Use Policy Planning Department Sri Lanka for their support given during the study.

References

Anon. (2012). *Geographically Weighted Regression, the landscape toolbox, tools and methods for effective land health monitoring*, Retrieved on http://wiki.landscapetoolbox.org_ analysis _methods: geographically weighted regression.

Association for Natural Rubber Producing Countries (ANRPC) (2017). *Natural Rubber Trends and Statistics, Statistical Annexure Tables-June-2017*, Bangunan Getah Asli, 148 Jalan Ampang, 50450 Kuala Lumpur, Malaysia.

Brunsdon, C., Fotheringham, S., & Charlton, M. (1998). Geographically Weighted Regression: A Natural Evolution of the Expansion Method for Spatial Data Analysis, *Journal of Environmental and planning A: Economy and space*, 30(6), 957-973.

Clement, F., Orange, D. Williams, M., Mulley, C. & Epprecht, M. (2009). Drivers of afforestation in Northern Vietnam: Assessing local variations using geographically weighted regression. *Journal of Applied Geography*. 29. 561-576.

Fotheringham, A., Brunsdon, C., & Charlton, M. (2002). *Geographically weighted regression: The analysis of spatially varying relationships*. West Sussex, England: John Wiley & Sons

Higgs, G., Kidner, D. B. & White, S. D. (2003). *Socio Economic Applications of Geographical Information Science*, New Fetter Lane, London: Taylor and Francis.

Jaimes, N. B. P., Sendra, J. B., Delgado, M. G., & Plata, R. F. (2010). Exploring the driving forces behind deforestation in the state of Mexico (Mexico) using geographically weighted regression. Applied Geography, 30(4), 576-591

Ministry of Plantation Industries (MPI) (2015). *Plantation Sector Statistical Pocket Book*, Ministry of Plantation Industries, 11th Floor, Sethsiripaya, Battaramulla, Sri Lanka.

Munasinghe, E .S., Rodrigo, V. H. L., Gunawardena, U. A. D. P. & Karunatilake, P. K. W. (2008). Quantification of growth and economically important yield components of rubber plantations in wet and intermediate zones of Sri Lanka, Proceeding of Plantation Crop Research Symposium, Sri Lanka.

Partridge, M. D., Rickman, D. S., Ali, K. & Rose, M. O. (2008). The Geographic Diversity of U.S. Nonmetropolitan Growth Dynamics: A Geographically Weighted Regression Approach, *Journal of Land Economics,* 84(2), 241-266.

Rubber Development Department (2011). *Census of Rubber lands 2010/2011*, Department of Rubber Development, Ministry of Plantation Industries, Colombo, Sri Lanka.

Sri Lanka Customs Department (2016). *Rubber Exports Earnings*. Customs Department, Colombo, Sri Lanka.

Wijesuriya, B. W., Dissanayake, D. M. A. P, Gunaratne, P. K. K. S., Samarappuli, L., Herath, H. M. L. K. & Abeywaradena, O. V. (2012). Will smallholder rubber farming be viable in Moneragala district? Evidence from status of existing rubber lands, *Proceeding of Plantation Crop Research Symposium*, Sri Lanka.