GWR-PM - Spatial variation relationship analysis with Geographically Weighted Regression (GWR) – An application at Peninsular Malaysia

Jamhuri J1, Azhar B M S2, Puan C L3, and Norizah K3

1 Master student, Faculty of Forestry, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.
2 Senior Lecturer, Department of Forest Management, Faculty of Forestry, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.
3 Senior Lecturer, Department of Production, Faculty of Forestry, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.
norizah_k@upm.edu.my

Abstract. GWR-PM has been developed exclusively for decision makers in Peninsular Malaysia and the purpose is to provide them with additional flexibility in analysing spatial variation. While GWR extension analysis in ArcMap application has a universal coordinate system, GWR-PM is specifically designed with Peninsular Malaysia’s coordinate system of Kertau RSO Malaya Meter. This paper presents the development of GWR-PM model by using a model builder, the application of which is to examine the forest fire risk at North Selangor Peat Swamp Forest. This model can be extended and improved by using ArcGIS language of phyton.

1. Introduction
A spatial pattern is a perceptual structure, placement, or arrangement of objects on earth. It includes the space in between those objects. Patterns may be recognised because of their arrangement, maybe in a line or by a clustering of points. Spatial patterns have been used by scientists, analysts and other professionals for planning, managing and estimation. Many researchers have used the traditionally regression method but there is a lack of information such as location and attribute data, besides weightage values not taken into account. Recently, with a new evolution of technology in spatial analysis, Geographically Weighted Regression (GWR) has been introduced to solve such problems related to spatial non-stationary. Geographically Weighted Regression (GWR) is a regression method capable of handling various relationships of variables in local spatial patterns for modelling, examining, monitoring and decision making [1]. The final outputs of spatial analysis using GWR provide users with spatial non-stationary, or in other words, variation of variables in local spatial patterns and a map of the spatial variation in relationships [1][2].

There are several studies using GWR in spatial regression which have been applied in forestry. For example, [3] used GWR to analyse the factors that drive the afforestation in Northern Vietnam. In this study, they used locations of timber demand as dependent variable and heterogeneity such as population, distance, industry and forest as independent variables. In the studies of [4], GWR was applied to investigate the impact of scales on prediction of rainfall uncertainty. This was carried out by modelling
the relationship between vegetation and climate; they used vegetation as dependent variable and independent variables were precipitation and temperature. Besides that, [5] employed GWR analysis to estimate the net primary production (NPP) of the Chinese forest ecosystems. They used the Chinese forest NPP, derived from plot inventory data covering the entire country as dependent variable while independent variables were temperature, precipitation, altitude, and yearly Time-Integrated Normalised Difference Vegetation Index (TINDVI). GWR analysis is not limited to applications in wildlife studies. For example, [6] used GWR to investigate the effects of local spatial heterogeneity on multivariate relationships of white-tailed deer distribution, using land cover patch metrics and climate factors. The distribution data of deer were collected by using the numbers of pellets consumed at different places.

On the other hand, [7] used GWR to examine spatially non-stationary and scale-dependent relationships between urban landscape fragmentation and related factors. This study was intended to investigate the applicability of GWR in modelling the relationships between effective mesh size and related factors, and then examine their spatial non-stationarity and the scale-dependence. They used effective mesh size as dependent variable; distance to main roads, distance to district centres and slope were explanatory variables. [7] used an extension in ArcToolbox version ArcGIS 9.3 to run GWR analysis. In addition, a study by [8] was intended to improve the ecosystem service in Fuzhou City, China; this was done by providing geographic variations using GWR analysis to prioritise areas for conservation or construction, and design an ecological corridor. They used population density in Fuzhou City, China as dependent variable; proportion of urban and rural construction land, proportion of transportation land and proportion of forest land were explanatory variables used for GWR analysis in Arc GIS 10.0. ArcGIS provides such extension application under spatial statistics and helps to visualise and interpret spatial non-stationarity of dataset.

Dynamic changes in the Peninsular Malaysia forest ecosystem have brought new challenges to the process of decision making in forest management system. Multiple biotic and/or abiotic factors and additional side constraints such as uncertainty of their variability in space require a detailed explanation and regression analysis for spatially varying relationships. To date, there are several studies on spatial patterns in Malaysia such as land use planning, parcel management, census data, topographic mapping, environmental monitoring and natural resources management. An example of such studies was done by [9] in relation to a community’s forest dependency and its effects on the forest resources and wildlife abundances in Sarawak, Malaysia. They used heterogeneity as variables to study the effects of the community’s forest dependency on the forest resources and wildlife abundances in Sarawak. Other than that, [10] studied the pattern of the changes of the conservation policy of tigers in Malaysia. They used camera traps to collect visual data and the analysis was done by using spatial autocorrelation. However, both studies are limited to non-spatial display of data which cannot explain the relationships between some sets of variables and spatial locational variables; they may help to identify the nature of relationships that exist between variables by taking spatial location into consideration with GWR analyses, and the spatial variability between relationships across the study area can be detected [11]. Incorporating the explanatory variables into the GWR command is however a problem to an amateur or beginner user of GWR analysis.

Hence, this paper presents an approach to develop a toolbox’s extension of GWR for spatial non-stationary analyses with its related applications. A new extension of the GWR analysis development described in this paper provides users with an additional flexibility to choose input features of selected regions and pre-processing of all the spatial features before varying relationships analysis. Input feature tools with sequence help users to determine the right variables in relationships used in modelling and mapping with GWR analysis; this procedure has a special application to forest management in Peninsular Malaysia.
2. Model Development
The processes of capturing and managing the spatial data to be queried are an initial phase for spatial analysis in GIS. Decision makers are responsible for those requirements; identifying the issues and mitigating measures until the final analysis can be modelled and mapped. General requirements for the GWR model development in Peninsular Malaysia forests can be described as below:

2.1. Define the problem - as spatial non-stationarity analysis.
2.2. Identify the issue - as locational or spatial variations
2.3. Mitigation measures - as explanatory variables which believed or expected to give effect on the variations to the issue highlighted.

Thus, the relationship between the significant problems, issues and mitigation measures can be illustrated in the Toolbox of ArcMap application as shown (Figure 1) below;

![Diagram](image)

**Figure 1.** A simplified model of GWR analysis for Peninsular Malaysia.

2.1. Feature dataset of vector and raster data.

Feature datasets are used to present spatial variations to be analysed in GWR. The datasets consist of dependent and explanatory variables. Feature datasets are used because the location of the feature data will be used for GWR inquiry. Besides that, feature datasets have an attribute table that can be used to input dependent and explanatory variables. The raster data are used to overlay with the feature dataset for visual references.
2.2. Define projection
For the purpose of uniformity, feature datasets are referenced with the projected coordinate systems of rectified skew orthomorphic (RSO) Malaya Meters format. Malaysia uses RSO because it has a base geographic in 2 Dimensions (2D) under UTM zone 47N as projected Coordinate References System (CRS). The Kertau (RSO) / RSO Malaya (m) was revised on 14th August 2006. It was defined by information from Defence Geographic Centre (DFC) and used only for metrication of RSO grid.

2.3. Vector and/or Raster data analysis
Geoprocessing is a GIS operation used to manipulate spatial data. A typical geoprocessing operation takes an input dataset, performs an operation on that dataset, and returns the result of the operation as an output dataset. There are two types of geoprocessing in ArcGIS, which are geoprocessing and raster geoprocessing (Table 1). Geoprocessing process includes buffering, clipping, intersecting, union, merging and dissolving. On the other hand, raster geoprocessing includes raster clip (raster to vector clip), composite band, computing pan sharp weights, creating ortho-corrected raster dataset, creating pan-sharpened raster dataset and extracting sub-dataset. In this process, a raster dataset needs to be clipped with feature dataset to fit the area of study, as reducing the size speeds up ArcGIS processing. The spatial pattern will maximise production at the end point in every angle.

Table 1. Geoprocessing process available in ArcToolbox of ArcMap version 10.3 [12].

| Geoprocessing        |                                                                 |
|----------------------|-----------------------------------------------------------------|
| **Buffer**           | Creates buffer polygons around input features to a specified distance. |
| **Clip**             | Clip is a process to cut out a piece of one feature class using one or more of the features in another feature class as a cookie cutter. |
| **Intersect**        | Computes a geometric intersection of the input features. Features or portions of features which overlap in all layers and/or feature classes will be written to the output feature class. |
| **Union**            | Computes a geometric union of the input features. All features and their attributes will be written to the output feature class. |
| **Merge**            | Combines multiple input datasets of the same data type into a single, new output dataset. This tool can combine point, line, or polygon feature classes or tables. |
| **Dissolve**         | Aggregates features based on specified attributes. |

**Raster Geoprocessing**

| Geoprocessing               |                                                                 |
|-----------------------------|-----------------------------------------------------------------|
| **Clip**                    | Cuts out a portion of a raster dataset, mosaic dataset, or image service layer. |
| **Composite Band**          | Creates a single raster dataset from multiple bands              |
| **Compute Pan sharp Weights**| Calculates an optimal set of pan-sharpened weights for new or custom sensor data. |
| **Create Ortho Corrected Raster Dataset** | Incorporates elevation data and image metadata to accurately line up imagery. |
| **Create Pan-sharpened Raster Dataset** | Fuses a high-resolution panchromatic raster dataset with a lower-resolution multiband raster dataset to create a red-green-blue (RGB) raster with the resolution of the panchromatic raster. |
| **Extract Sub-dataset**     | Extracts raster datasets stored within a sub-dataset raster file. |
| **Raster To DTED**          | Splits a raster dataset into separate files based on the DTED tiling structure. |
| **Resample**                | Changes the spatial resolution of your raster dataset and sets rules for aggregating or interpolating values across the new pixel sizes. |
| **Split Raster**            | Divides a raster dataset into smaller pieces, by tiles or features from a polygon. |
3. GWR
Feature datasets are added as location data and are arranged as dependent variables and factors of influence to the dependent variables are fixed as explanatory variables. The result will be attributed with spatial pattern map, coefficient value and R2 value. The coefficient value describes to better understand of location sample either as a statistically significant hot spot or a statistically significant cold spot, while R2 value explains the significantly fraction between 0.0 and 1.0. Figure 2 shows the interface of GWR analysis developed by using the model builder in Arctoolbox of ArcMap version 10.3.

![Figure 2: Interface of GWR modelling analysis.](image)

4. Model applications
As a demonstration, we applied our model to Raja Musa Forest Reserve (RMFR), a part of North Selangor Peat Swamp Forest (NSPSF), to analyse the hotspot areas of high risk forest fire (Figure 3). The locations of the hotspot areas with frequent occurrences of fire were set as dependent variables; heterogeneity factors were taken into account as explanatory variables such as distance to road, distance to river, distance to oil palm, distance to paddy field, and distance to continuous forest. These data were collected by using the landscape metric measurement method in Google Earth Pro free version. Each variable was measured manually in each point of forest fire hotspot. Also, Landsat 8 image with 30 meter resolution of RMFR was used for Kernal mapping.
5. Results and discussion

The GWR analysis provides $R^2$ and coefficient raster to show the significantly of the model tested and the hotspot areas of the study sites. The results obtained from model application on hotspot areas of high risk forest fire in NSPSF show that 76.47% of $R^2$ is significantly related to heterogeneity variables which are distance to road (DTRD), distance to river (DTRV), distance to oil palm (DTOP), distance to paddy (DTPD), and distance to continuous forest (DTCF). Based on these results, all the variables are positively higher which is coefficient values (positive or negative) of hotspot is 4.25404. But, the highest coefficient value for single variable is distance to river DTRV which is the coefficient value of hotspot -9.94182 and followed by DTCF -9.25813, distance to paddy field DTPD 8.24629, distance to oil palm DTOP -7.71606 and lastly distance to road DTRD -0.000303741 (Figure 4). The result from the DTPD shows that, the closer paddy with the fire hotspot areas, the higher potential the area to get burn. This show the significant because paddy are easily to get burn especially in dry season (summer). Conversely, the result from DTRV show that the closer river with the fire hotspot areas, the lower potential of the area to get burn. This show the significant between fire hotspot areas and river because fire are less burning in wet area. In additionally, the same result to DTCF, DTOP and DTRD which are lower potential to burn when close and high potential of burning when far from fire hotspot areas. Forest and oil palm are dense area and ground level are covered by canopy which are help in maintain soil moisture and could prevent from fire burning while area with no canopy covered are directly exposed to sunlight heat which could make it dry and easily to get burn. Finally, distance to road not directly affect the fire because most of the road is surrounded by wild grass that grows in limited areas. Figure 5 below shows
the hotspots of forest fire areas in RMFR in the year of 2015. In regard to this matter, the results show the ability of GWR coefficients to vary over spaces; GWR often produces better residual sum of squares, which is an indication of better model fit [13] [14].

![Figure 4](image)

**Figure 4.** Result of the GWR analysis: (A) Interception of Variables, (B) distance to continuous forest (DTCF), (C) distance to road (DTRD), (D) distance to paddy field (DTPD), (E) distance to oil palm (DTOP), and (F) distance to river (DTRV).
6. Conclusion
As a conclusion, GWR provides better understanding in the form of kernel mapping such as in Figure 4 that displays the patterns of spatial map compared with other regressions which do not take location into account. Ignoring spatial variations under these conditions could produce results of biased estimation, misleading significant tests and sub-optimal predictions. Thus, GWR provides an effective alternative option by making use of spatial location as proxy variables. Therefore, GWR can achieve significantly improved results over those of other regression models. Furthermore, this model development fits well with GWR analysis, and this actually eases the analysis process. According to [15] more than 290 million hectares of land burned annually (between 79% and 91% of total global burned area). So, this modelling could lead to better management practices, especially in Forest Department. Currently, most of the practices in the department rely on elusive analysis; the modelling of this study actually improves the analysis by providing illustrations of spatial pattern maps that will be easy to understand.

7. References
[1] Fotheringham, A.S., Brunsdon, C., Charlton, M., 2002. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. Wiley, West Sussex.
[2] Luo, J., Wei, Y.H.D., 2009. Modeling spatial variations of urban growth patterns in Chinese cities: the case of Nanjing. Landscape and Urban Planning 91, 51–64.
[3] Clement, F., Orange, D., Williams, M., Mulley, C., & Epprecht, M. (2009). Drivers of afforestation in Northern Vietnam: Assessing local variations using geographically weighted regression. Applied Geography, 29(4), 561–576. http://doi.org/10.1016/j.apgeog.2009.01.003
[4] Propastin, P., Kappas, M., & Erasmi, S. (2008). *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, 3, 73–94. http://doi.org/10.2902/1725-0463.2008.03.art6

[5] Wang, Q., Ni, J., & Tenhunen, J. (2005). *Application of a geographically-weighted regression analysis to estimate net primary production of Chinese forest ecosystems*. Global Ecology and Biogeography, 14(4), 379–393. http://doi.org/10.1111/j.1466-822X.2005.00153.x

[6] Shi, H., Laurent, E.J., LeBouton, J., Racevskis, L., Hall, K.R., Donovan, M., Doepker, R.V., Walters, M.B., b, Lupi, F., Liu, J., 2006. *Local spatial modeling of white-tailed deer distribution*. Ecological Modeling, 190, 171–189.

[7] Gao, J., & Li, S. (2011). *Detecting spatially non-stationary and scale-dependent relationships between urban landscape fragmentation and related factors using Geographically Weighted Regression*. Applied Geography, 31(1), 292–302. http://doi.org/10.1016/j.appgeog.2010.06.003

[8] Hu, X., Hong, W., Qiu, R., Hong, T., Chen, C., & Wu, C. (2015). *Geographic variations of ecosystem service intensity in Fuzhou City, China*. Science of The Total Environment, 512-513, 215–226. http://doi.org/10.1016/j.scitotenv.2015.01.035

[9] Nelson, J., Muhammed, N., & Rashid, R. A. (2015). *Community’s forest dependency and its effects towards the forest resources and wildlife abundances in Sarawak, Malaysia, 4509(November)*. http://doi.org/10.1080/13504509.2015.1072117

[10] Rayan, D. M., & Linkie, M. (2015). *Conserving tigers in Malaysia: A science-driven approach for eliciting conservation policy change*. Biological Conservation, 184, 18–26. http://doi.org/10.1016/j.biocon.2014.12.024

[11] Shrestha, P. M. (2006). *Comparison of Ordinary Least Square Regression, Spatial Autoregression, and Geographically Weighted Regression for Modeling Forest Structural Attributes Using a Geographical Information System (GIS)/Remote Sensing (RS) Approach*. Master Thesis. Retrieved from http://people.ucalgary.ca/~mcermid/Docs/Theses/Shrestha_2006.pdf

[12] ArcGIS (2006). *ArcGIS Desktop Help 9.3 - An overview of commonly used tools*. Retrieved June 27, 2006, from http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=An_overview_of_commonly_used_tools

[13] Brunsdon, C., Fotheringham, a S., & Charlton, M. E. (1996). *Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity*. Geographical Analysis, 28(4), 281–298. http://doi.org/10.1111/j.1538-4632.1996.tb00936.x

[14] Zhang, L., Shi, H., 2004. *Local modeling of tree growth by geographically weighted regression*. Forest Science, 50, 225–244.

[15] Van Lierop, P., Lindquist, E., Sathyapala, S., & Franceschini, G. (2015). *Global forest area disturbance from fire, insect pests, diseases and severe weather events*. Forest Ecology and Management, 352, 78–88. http://doi.org/10.1016/j.foreco.2015.06.010