# Investigation of Rainfall-NDVI Spatial Variation - Geographical Weighted Regression Approach

Rohayu H. NARASHID, Ruslan RAINIS, and Zullyadini A. RAHAMAN

***Abstract*****—**Vegetation is one of the factors that affect the rainfall spatial patterns. To show the variation due to vegetation effect, the relationship between rainfall and NDVI should be spatially analyzed at local scale. In this study, the local regression technique of GWR was used to investigate the spatial relationship between rainfall-vegetation derived from(2000-2011) rainfall amount of 175 stations and seasonal NDVI of Landsat 7 ETM+ during southwest monsoon (SWM). Results of Moran's I spatial autocorrelation and Local Moran's I prove the existence of variation in rainfall spatial patterns. This indicator supports the use of GWR technique to explore the local variation of rainfall-NDVI relationship. The GWR results ($R^2$ = 0.49 and 0.71) have improved the global approach by Ordinary Least Square (OLS) - ($R^2$ = 0.12 and 0.28). Thus, the spatial relationships of rainfall-NDVI are possible to be analyzed at an individual location with the local approach by GWR.

***Keywords*****—**Rainfall spatial patterns, spatial non-stationarity, Normalized Difference Vegetation Index (NDVI), spatial autocorrelation and Geographically Weighted Regression (GWR)

## I. INTRODUCTION

Studies on rainfall spatial patterns always been related to the land-surface characteristics. Vegetation is one of the environmental factors on the land surface that also may influence rainfall spatial patterns of the area. In most cases, vegetation is always affected by precipitation. However, rainfall spatial patterns can also be influenced by vegetation cover due to land degradation (i.e. Deforestation and urbanization).As indicated by[1], the effect of latent heat due to deforestation has changed the patterns and distributionsof precipitation. Recent studies of precipitation in urban areas (e.g.[2] and [3]) reveal the impact of urbanization on the local precipitation for urban areas of developed countries since the spatial and temporal rainfall patterns are influenced by urbanization. Thus, vegetation degradation of the area may be considered as the influencial variables of rainfall local variability.

To investigate the vegetation degradation impact to rainfall variation, the analysis on the spatial relationship between rainfall and vegetation is necessary to be carried out. The measurement of vegetation amount may be obtained from vegetation indices or Normalized Difference Vegetation Index (NDVI).

Rohayu H. NARASHID, PhD student at Universiti Sains Malaysia (USM), Malaysia.

Ruslan RAINIS, Director of Center for Research Initiative - Liberal Arts & Social Sciences at USM, Malaysia.

Zullyadini A. RAHAMAN, Universiti Sains Malaysia (USM), Malaysia.

NDVI was first implemented by [4]to monitor vegetation system.Basically, NDVI derived from earth observation system are very useful to describe surface vegetation behaviour[5]. NDVI algorithm of satellite images can be derived from visible red (0.58-0.68 micrometers) and near-infrared (0.75-1.1 micrometers) band/channel with the subtraction of red bands (RED) from the near-infrared (NIR) and divides it by the sum of near-infrared and red bands (see (1)).

$$NDVI= (NIR-RED) / (NIR+RED) \quad (1)$$

Commonly, time-series NDVI have been used in relation to rainfall patterns of the area. Several researchers such as [6], [7]:[8]), [9]and [10]have utilize time series NDVI derived from the NOAA Advanced Very High Resolution Radiometer (AVHRR) as dependent variables in NDVI-rainfall relationship studies. Since the low spatial resolution images such as NOAA represents NDVI values for large coverage (i.e. 1km x 1km), the variation of rainfall due to vegetation factor is not really represented at local scale. In order to obtain the more specific NDVI values of locations, better spatial resolutions of remote sensing images are necessary to be used. Results of strong and more predictable relationship between rainfall and NDVI have been found by [11]when appropriate spatial scale is utilized. Thus, the use of multi-spectral satellite data with medium spatial resolution such as Landsat imagery seems to be more appropriate to serve a better NDVI values.

Most of the studies utilize global statistical regression, i.e. Ordinary Least Square (OLS) method to determine the spatial relationship between variables. Basically, the use of OLS is unsatisfied to prove the relationship of nonstationarity variables such as rainfall-NDVI. In many cases such as found in[6] and[12], the local variation is very poor to be detected based on the OLS approach since this method is suitable to quantify the relationship at global and regional scales. Thus, local statistical approach such as Geographically Weighted Regression (GWR) should be used to prove the local variation. GWRmay provide a more appropriate basis to investigate the relationship between variables ([13], [6] and [14]).Since GWR is useful to overcome the problem of non-stationarity, [8]expected the scale-dependent results for spatial non-stationarity relationship with a change in the spatial resolution.

In this paper, the spatial relationship between rainfall and vegetation at local level will be analyzed based on better resolution of NDVI, with the use of the local regression technique. The objectives are (1) to study rainfall spatial patterns of the area and (2) to explore the spatial relationship between rainfall and Landsat NDVI.

## II. MATERIALS AND METHODS

### *A. Study Areas and Datasets*

Generally, Malaysiareceived a large amount of rainfall with the uniform temperature and high humiditysince it is located in the tropical region. There are four seasons or monsoons in Peninsular Malaysia i.e. northeast monsoon (November to January), southwest monsoon (May to August) and two inter monsoon seasons at different region. [15]

The study area for this research covers the northern region of Peninsular Malaysia (see Fig. 1) which comprises of four (4) states, i.e. Perlis, Kedah, Pulau Pinang and small regions of Perak. It is situated within 6°38' 31" N latitudes and 100°13' 04" E longitudes to 2°41' 31" N latitudes and 101°42' 14" E longitudes from Northern Perlis to Sourthern Perak. Accordingly, there are 175 rainfall stations of the Malaysian Department of Irrigation and Drainage (DID) available within the study areas. To achieve the objectives of this study, the compulsory data of study areas to be used are (i) monthly rainfall depths and (ii) satellite images of Landsat 7 ETM+.

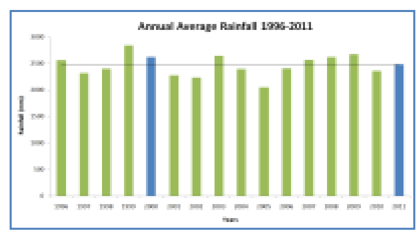

Fig. 1 Map of Study Areas

Rainfall spatial patternsand relationship of rainfall-NDVI were establishedduring southwest monsoon (SWM) of the years 2000-2011 due to the availability of Landsat scenes from this season of years 2000-2011.The rainfall descriptive statistics of the study area during SWMfor years1996-2011 are shown in Table 1 and Figure 2 respectively. Fig. 2 (a) show that the average annual rainfall of year 2000 and 2011 are calculated above annual average rainfall between the years 1996-2011. From the annual average rainfall during SWM in Fig. 2 (b), the increase amount has been found in year 2011 compared to 2000. Meanwhile the trend of monthly rainfall depths during the SWM were increased from June to September (refer to Fig. 2 (c)).

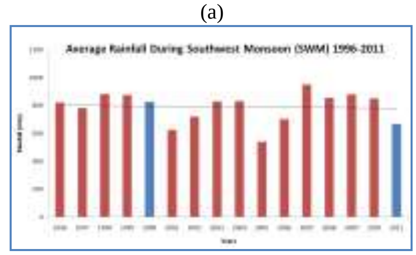
Satellite images of Landsat 7 ETM+ are compulsory for NDVI data preparation. Based on the selected study areas, five (5) scenes series of Landsat 7 ETM+ images acquired during SWM year 2000 and 2011 were used. Raw Landsat images of the study area were downloaded from United States Geological Survey (USGS) website. To produce initial images for NDVI, essential image pre-processing steps were implemented by using Erdas Imagine 2011.

TABLE I
SUMMARY OF RAINFALL DESCRIPTIVE STATISTICS DURING SWM2000-2011

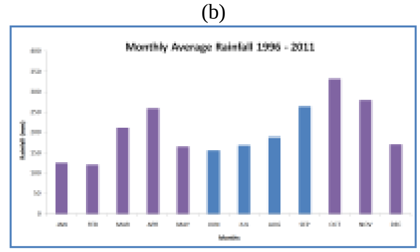
| **YEAR** | **Annual Average (mm)** | **Southwest Monsoon RF Statistics** | | | |
|---|---|---|---|---|---|
| | | ***Average (mm)*** | ***Minimum (mm)*** | ***Maximu m(mm)*** | ***Standard Deviation*** |
| 2000 | 2623 | 168 | 5.2 | 503 | 20.44 |
| 2001 | 2278 | 134 | 13.2 | 569 | 7.55 |
| 2002 | 2237 | 143 | 9.3 | 867 | 18.11 |
| 2003 | 2642 | 183 | 9 | 763 | 16.14 |
| 2004 | 2395 | 170 | 2 | 565 | 10.07 |
| 2005 | 2050 | 143 | 2 | 537 | 9.88 |
| 2006 | 2411 | 159 | 5.5 | 582 | 16.36 |
| 2007 | 2564 | 209 | 9 | 1018 | 13.81 |
| 2008 | 2617 | 179 | 7 | 671 | 11.57 |
| 2009 | 2671 | 367 | 2.5 | 589 | 46.50 |
| 2010 | 2358 | 212 | 10 | 603 | 7.37 |
| 2011 | 2484 | 170 | 5 | 645 | 27.62 |




(a)




(b)




(c)

Fig.2. Rainfall Averages

### *B. Spatial Autocorrelation Tests*

The spatial dependency identification of rainfall trend determines rainfall spatial pattern of the area. To investigate the spatial relationship between rainfall and NDVI, rainfall spatial pattern should be analyzed first. Thus, results of spatial pattern based on rainfall spatial autocorrelation can be used to analyze the relationship of rainfall amount in spatial dependence. In this study, Moran's statistics which is introduced by [16] have been used to examine the local level of rainfall spatial autocorrelation. The equation of Moran's spatial autocorrelation coefficient, I is shown in (2):

$$I=\frac{\sum_{i=1}^{J} n(R_i-\bar{R})(R_j-\bar{R})}{\sum_{i=1}^{n} J(R_i-\bar{R})^2} \quad (2)$$

where n is the total number of areas, J is the total number of joints,$R_i$ and $R_j$are the values of rainfall depths for two contiguous areas, and $\bar{R}$ is the overall mean of rainfall.

The significance of the local spatial patterns can be categorized as "High-High" or "Low-Low" for positive spatial autocorrelations which the high or low values surrounded by the high or low values of neighboring units. Meanwhile, spatial outliers of the location is determined based on "High-Low" or "Low-High" whereas high value in a low neighboring value and low value in a high neighboring value.[17]. Spatial autocorrelation test of Moran's I was carried out using ArcGIS spatial statistics tools.

### *C. Global and Local Regression Approach*

For spatial relationship, it is necessary to apply an appropriate regression technique based on spatial patterns of the variables. Basically, regression analysis can be used to describe relationships among variables.Thus, simple linear regression such as, OLS method can be used to predict global variation of rainfall. The relationship between rainfall (R) and its influencing factors such as vegetation (y) can be estimated based on Eq.3:

$$y = \beta_0 + \beta_1 R_1 + \cdots . + \beta_n R_n + \varepsilon \quad (3)$$

where$\beta_0 - \beta_n$ are the estimated parameter indicating the relationship between the y and R, and $R_1 - R_n$values of rainfall depths and ε is an error term.

Unfortunately, the conventional global regression model is insufficient to provide the information for the established spatial non-stationarity relationship at one location [18]. Thus, local approach should be applied. GWR, a local regression technique which is proposed by [19]:[20]have been widely used to explore the spatial non-stationarity of variables at local stage. Local variation in the relationship between rainfall (R) and its influencing factors such as vegetation (y) at individual location$(\mu, v)$is possible to be analyzed using GWR model (Eq.4):

$$y = \beta_0(\mu, v) + \beta_1(\mu, v)R_1 + \cdots . + \beta_n(\mu, v)R_n + \varepsilon \quad (4)$$

Based on NDVI values as the exploratory variable and rainfall depth during southwest monsoon as dependant variable, the relationship between rainfall and NDVI has been established at local level. GWR 4.0 software was used to generate GWR results for further analysis.

## III. RESULTS

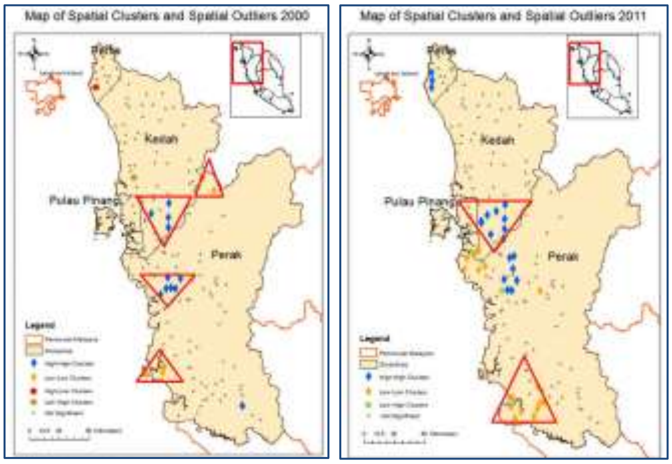
### *A. Rainfall spatial patterns of study area*

Rainfall spatial patterns can be described based on such spatial patterns of dispersed, random or clustering. The results of global Moran's I have shown the significant Moran index values (refer to Table 2). These measurements indicate the potential rainfall spatial variation patterns of the study area. The significant z-scores which are less than 1% likelihood reveal that the rainfall spatial patterns are varied randomly. This indicator supports the nature of non-stationarity of rainfall patterns. Thus, the selection of local regression technique, i.e. GWR is possible to explore the local variation of rainfall-NDVI relationship.

TABLE II
SUMMARY OF GLOBAL MORAN'S I FOR MEAN BETWEEN 2000 UNTIL 2011 DURING SWM

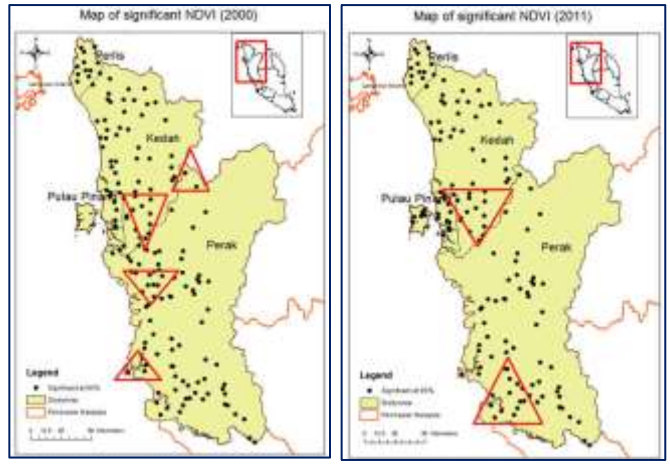
| YEAR | Moran's Index | Z Value |
|---|---|---|
| 2000 | 0.23645 | 5.10039 |
| 2001 | 0.57861 | 12.19330 |
| 2002 | 0.43487 | 9.29695 |
| 2003 | 0.44383 | 9.51500 |
| 2004 | 0.40170 | 8.50672 |
| 2005 | 0.18305 | 3.96715 |
| 2006 | 0.37987 | 8.06162 |
| 2007 | 0.27185 | 5.84887 |
| 2008 | 0.29072 | 6.21380 |
| 2009 | 0.65857 | 13.87223 |
| 2010 | 0.37418 | 7.94783 |
| 2011 | 0.45976 | 9.74575 |

To identify location of rainfall spatial variation, local Moran's I has been carried out. Fig.3 (a) and (b) illustrates the cluster and outliers of rainfall variations within the study area in year 2000 and 2011. During Southwest Monsoon in 2000, there are 11 stations found in high-high spatial cluster points which were located within southern Kedah (4 stations), northern Perak (6 stations) and southern Perak (1 station). There were 5 stationsdetected in Perak as low-low spatial cluster points which are found within the northeast (2 stations) and western Perak (3 stations) respectively. On the other hand, the spatial outliers which are determined by high-low and low-high spatial cluster points were located at Perlis and Kedah. In year 2011, the stations of high-high areas were increased from 11 to 19 stations within the westernPerlis (3 stations), southern Kedah (8 stations) and northern Perak (7 stations). However, the low-low spatial clusters were found within northwest (6 stations), southwest (9 stations) and center (1 station) region of Perak.There were two (2) spatial outlier points located in the eastern Perlis and northern Perak.




(a) (b)

Fig. 3 Maps of Rainfall Spatial Clusters and Outliers (a) Year 2000 (b) Year 2011




(a) (b)

Fig. 4 Maps of NDVI t-value (a) Year 2000 (b) Year 2011

The relationship between rainfall and NDVI derived from Landsat 7 ETM+ were established based on two (2) datasets of study area. It is found that, the results of GWR improved the global relationship based on Ordinary Least Square (OLS) technique. Both results show the significant improvement of $R^2$from OLS to GWR whereas$R^2$from OLS (0.12 and 0.28) to GWR (0.49 and 0.71) in year 2000 and 2011 respectively. The Akaike Information Criterion(AIC)has been used to assess the quality of the relationship established from OLS and GWR. The results of AIC were found significance when the AIC values from GWR are lower than AIC of OLS[13]. Table 3 showsthe comparison of $R^2$ and AIC between OLS and GWR.

TABLE III
SUMMARY OF OLS AND GWR RESULTS

| YEAR | OLS | | GWR | |
|---|---|---|---|---|
| | ***R*²** | ***AIC*** | ***R*²** | ***AIC*** |
| 2000 | 0.12 | 1908.98 | 0.49 | 1867.57 |
| 2011 | 0.28 | 1828.45 | 0.71 | 1712.12 |

For the meaningful interpretations and more predictable relationship between rainfall-NDVI at local level, the results of GWR should be spatially distributed on the map based on the significance levelindicated by local t-values [21]. Thus, GWR maps of NDVI t-valuewere produced based on the integrated used of GWR 4.0 and ArcGIS 10. With the 95% level of confidence, significant location of rainfall-NDVI were mapped (refer to Fig. 4). In year 2000, a total of 158 locations were found significance at 95% level of confidence. However the locations of significance 95% level in 2011 are decreased to 141 stations.

It is found that, several significance locations of 95% confidence levelwere coincide with the high-highand low-low spatial cluster points of rainfall stations, indicated bythe clusters of Moran's I. The coincidence locations were shown by trianglesin Fig. 3 and Fig. 4 respectively. Most of the coincidence stations for both years were located within the centre part of the study area i.e. Southern Kedah and Northern Perak.

## IV. CONCLUSION

Local investigations of spatial non-stationaritybetween the phenomena such as spatial rainfall and vegetation significantly improve based on the local approach.This finding supports the evidence of spatial heterogeneity in the relationship between rainfall and vegetation and the advantages of local approach. The relationship of NDVI-rainfall represented by OLS regression may not be applied to predict the location information ona single location. The spatial relationships between rainfall and NDVI of better resolution are possible to be analyzed at individual location with the local approach by GWR. The coincidence stations which are representedby Local Moran's I and NDVI t-value indicate that vegetation may influence the rainfall spatial variation of the locations. Thus, spatial autocorrelation test of the variable is necessary to be carried out in order to prove the existence of variation in the phenomenon spatial patterns and to support the results of the GWR. However, thisfindingonly shows the significant locations of rainfall spatial variations due to vegetation during SWM. The locations of the variation might be changed due to other seasons such as during the northeast monsoon (NEM). Future work will concentrate on the local modelling of rainfall spatial variation with other monsoon seasons and other influencing factors.

## ACKNOWLEDGMENT

Cooperation of Malaysian Drainage and Irrigation Department for providing rainfall data used in this study is gratefully acknowledged. The authors thank to UniversitiTeknologi MARA (UiTM) Perlis and Ministry of Higher Education (MOHE), Malaysia who funded this study.

## REFERENCES

[1] D. Avissar, R and Werth, "Global Hydroclimatological Teleconnections Resulting from Tropical Deforestation," *J. Hydrometeorol.*, vol. 6, no. 1993, pp. 134–145, 2005.

[2] Shepherd J.M and Mote T.L, "Urban effects on rainfall variability: Potential implications for Georgia's water supply," in *Proc. of the 2009 Georgia Water Resources*, 2009, no. 2007, p. 6.

[3] C. M. Kishtawal, D. Niyogi, M. Tewari, R. a. Pielke, and J. M. Shepherd, “Urbanization signature in the observed heavy rainfall climatology over India,” *Int. J. Climatol.*, vol. 30, no. 13, pp. 1908–1916, Nov. 2010.

[4] D. W. Rouse, J. W., Jr.; Haas, R. H.; Schell, J. A.; Deering, “Monitoring Vegetation Systems in the Great Plains with Erts,” in *Third Earth Resources Technology Satellite-1 Symposium, NASA SP-351I*, 1973, pp. 309–317.

[5] C. J. Tucker, “Red and photographic infrared linear combinations for monitoring vegetation,” *Remote Sens. Environ.*, vol. 8, no. 2, pp. 127–150, May 1979.

[6] G. M. Foody, “Geographical weighting as a further refinement to regression modelling: An example focused on the NDVI-rainfall relationship,” *Remote Sens. Environ.*, vol. 88, pp. 283–293, 2003.

[7] P. Propastin, N. Muratova, and M. Kappas, “Reducing uncertainty in analysis of relationship between vegetation patterns and precipitation,” *Proc. 7th Int. Symp. Spat. Accuracy Assess. Nat. Resour. Environ. Sci.*, pp. 459–468, 2006.

[8] P. Propastin, M. Kappas, and S. Erasmi, “Application of geographically weighted regression to investigate the impact of scale on prediction uncertainty by modelling relationship between vegetation and climate,” *Int. J. Spat. Data Infrastructures …*, vol. 3, pp. 73–94, 2008.

[9] S. A. Hashemi, “Investigation of Relationship Between Rainfall and Vegetation Index by Using NOAA / AVHRR Satellite Images,” *World Appl. Sci. J.*, vol. 14, no. 11, pp. 1678–1682, 2011.

[10] U. Usman, S. a Yelwa, S. U. Gulumbe, A. Danbaba, and R. Nir, “Modelling Relationship between NDVI and Climatic Variables Using Geographically Weighted Regression,” *J. Math. Sci. Appl.*, vol. 1, no. 2, pp. 24–28, 2013.

[11] J. Wang, P. M. Rich, and K. P. Price, “Temporal responses of NDVI to precipitation and temperature in the central Great Plains, USA,” *Int. J. Remote Sens.*, vol. 24, no. 1, pp. 2345–2364, 2003.

[12] P. A. Propastin and M. Kappas, “Reducing Uncertainty in Modeling the NDVI-Precipitation Relationship: A Comparative Study Using Global and Local Regression Techniques,” *GIScience Remote Sens.*, vol. 45, no. 1, pp. 47–67, 2008.

[13] A. S. Fotheringham, C. Brunsdon, and M. . Charlton, *Geographically Weighted Regression: the Analysis of Spatially Varying Relationships*. Chichester: Wiley, 2002.

[14] Q. Wang, J. Ni, and J. Tenhunen, “Application of a geographically-weighted regression analysis to estimate net primary production of Chinese forest ecosystems,” *Glob. Ecol. Biogeogr.*, vol. 14, pp. 379–393, 2005.

[15] Malaysian Meteorological Department, “2009 Climate Change Scenario for Malaysia 2001 - 2099,” 2009.

[16] P. a P. Moran, “Notes on continuous stochastic phenomena.,” *Biometrika*, vol. 37, no. 1, pp. 17–23, 1950.

[17] Anselin, Luc and S. J. Rey, *Perspectives on Spatial Data Analysis*. London: Springer, 2010.

[18] M. C. A. Stewart Fotheringham, Chris Brunsdon, *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. 2002.

[19] C. Brunsdon, S. Fotheringham, M. Charlton, C. Brunsdont, and S. Fotheringham, “Geographically Weighted Regression,” *Handb. Appl. Spat. Anal.*, vol. 28, no. 4, pp. 281–298, 1996.

[20] C. Brunsdon, S. Fotheringham, M. Charlton, C. Brunsdont, and S. Fotheringham, “weighted regression-modelling Geographically spatial non-stationarity,” *J. R. Stat. Soc. Ser. D (The Stat.*, vol. 47, no. 3, pp. 431–443, 1998.

[21] S. a. Matthews and T.-C. Yang, “Mapping the results of local statistics,” *Demogr. Res.*, vol. 26, pp. 151–166, Mar. 2012.


About Author (s):

**Rohayu H. NARASHID** MSc in Built Environment (UniversitiTeknologi MARA (UiTM), Shah Alam, 2005), BSc in Surveying Science and Geomatic (H), UniversitiTeknologi MARA (UiTM), Shah Alam, 2002).

Rohayu has more than 9 years experiences as a lecturer in Surveying Science and Geomatic in UniversitiTeknologi MARA (UiTMPs), Perlis since 2006. Currently, she is a PhD student at Universiti Sains Malaysia (USM). Her research interest is Geographical Information System and environmental Remote Sensing.

**Ruslan RAINIS, PhD.**Ph.D (City & Regional Planning) - Ohio State University, Master (City and Regional Planning) - Ohio State University, B.A. (Urban and Regional Planning) - Miami University, Ohio.

Professor DrRuslan is aGeographical Information Scientist with 25 years of experiences doing research, teaching, and consultancy in various topics related to GIS and remote sensing applications in business, urban planning, spatial demography, humanities and social sciences. He has served as a lecturer at Section of Geography,UniversitiSains Malaysia (USM) since 1991. Since 2013, he is a Director of Center for Research Initiative - Liberal Arts & Social Sciences at USM.

**Zullyadini A. RAHAMAN, PhD.**Ph.D. - UniversitiSains Malaysia (USM M.A. - UniversitiKebangsaan Malaysia (UKM), B.A. Ed. - UniversitiSains Malaysia (USM).

Dr Zullyadini has experiences over 12 years doing research, teaching, and consultancy in Catchment Hydrology, Geomorphology, Soil Erosion, Physical Geography and Hydrological Modelling.He has served as a lecturer at Section of Geography,UniversitiSains Malaysia (USM) since 2003.