

# Spatial Mapping of Dengue Incidence: A Case Study in Hulu Langat District, Selangor, Malaysia

Er, A.C., Rosli, M.H., Asmahani A., Mohamad Naim M.R., Harsuzilawati M.

**Abstract**—Dengue is a mosquito-borne infection that has peaked to an alarming rate in recent decades. It can be found in tropical and sub-tropical climate. In Malaysia, dengue has been declared as one of the national health threat to the public. This study aimed to map the spatial distributions of dengue cases in the district of Hulu Langat, Selangor via a combination of Geographic Information System (GIS) and spatial statistic tools. Data related to dengue was gathered from the various government health agencies. The location of dengue cases was geocoded using a handheld GPS Juno SB Trimble. A total of 197 dengue cases occurring in 2003 were used in this study. Those data then was aggregated into sub-district level and then converted into GIS format. The study also used population or demographic data as well as the boundary of Hulu Langat. To assess the spatial distribution of dengue cases three spatial statistics method (*Moran's I*, average nearest neighborhood (ANN) and kernel density estimation) were applied together with spatial analysis in the GIS environment. Those three indices were used to analyze the spatial distribution and average distance of dengue incidence and to locate the hot spot of dengue cases. The results indicated that the dengue cases was clustered ( $p < 0.01$ ) when analyze using *Moran's I* with z scores 5.03. The results from ANN analysis showed that the average nearest neighbor ratio is less than 1 which is 0.518755 ( $p < 0.0001$ ). From this result, we can expect the dengue cases pattern in Hulu Langat district is exhibiting a cluster pattern. The z-score for dengue incidence within the district is -13.0525 ( $p < 0.0001$ ). It was also found that the significant spatial autocorrelation of dengue incidences occurs at an average distance of 380.81 meters ( $p < 0.0001$ ). Several locations especially residential area also had been identified as the hot spots of dengue cases in the district.

**Keywords**—Dengue, Geographic Information System (GIS), Spatial Analysis, Spatial Statistics

Er, A.C. is a Lecturer in the School of Social, Development and Environmental Studies, Faculty of Social Sciences and Humanities, National University of Malaysia (corresponding author to provide phone:603-8921-3945; fax:603-89251168; e-mail: everr@ukm.my).

Rosli, M.H. is a Research Officer in the Sports Academy, University Putra Malaysia. (e-mail: mhafiz@putra.upm.edu.my).

Asmahani A. and Harsuzilawati M, are post graduate students in the School of Social, Development and Environmental Studies, Faculty of Social Sciences and Humanities, National University of Malaysia. (e-mail: anie267@gmail.com, md\_bambee@yahoo.com.my).

Mohamad Naim M.R. is a post graduate student in the Environmental Health Programme, Faculty of Allied Sciences, National University of Malaysia (e-mail: naimracd@gmail.com).

## I. INTRODUCTION

DENGUE is a mosquito-borne infection that has become a major international public health concern and significance in recent decades. Dengue is found in tropical and sub-tropical climates regions around the world, mostly and predominantly in urban and semi-urban areas. Currently, dengue haemorrhagic fever (DHF), sometimes a potentially lethal complication, affects most Asian countries and has become a leading cause of serious illness, hospitalization and death among children in the region. There is no specific treatment for dengue, however, with frequent and appropriate medical care; the lives of patients with the more serious dengue haemorrhagic fever (DHF) could be saved. The only way to prevent the transmission of dengue virus is to combat the carrier [1].

The incidence of dengue has grown dramatically around the world in recent decades. Dengue, places 2.5 billion people in more than 100 countries at risk which two fifths of the world's population, causing an estimated 50 to 100 million cases of dengue infections worldwide annually and thousands of deaths [1], [2]. Dengue virus is transmitted primarily by the yellow fever mosquito *Aedes aegypti* species which can be found predominantly in urban centers and in certain rural areas of Malaysia [3]. Dengue can be generally categorized into two types; the Dengue Fever (DF) or classic dengue and the Dengue Haemorrhagic Fever (DHF), which may evolve into Dengue Shock Syndrome (DSS) [4], [5].

The mapping of disease can be used to pinpoint the areas where outbreaks originate and effectively target high-risk areas for early prevention control. Geographic information systems (GIS) and related technologies have emerged as a new generation of information systems with the capability to manage spatial dimensions together with time, people and other dimensions of interest, which is not possible with the former information systems. GIS are currently recognized as a set of strategic and analytic tools for public health, so the design and implementation of an information system for dengue control with GIS capacity should be considered. A GIS based approach is proposed in this paper to determine the relationship between dengue cases and their spatial pattern.

Epidemiologists normally used the ratio of case numbers at a particular time to past case occurrences using the mean or median [6]. However, since dengue cases varies from one place to another, the spatial and time components must also be taken into consideration. Spatial – temporal interactions

among health events is an important component in epidemiological and public health study [6]. When dealing with problems of space, the step beyond simple cartography and mapping is spatial analysis, which in geographical research is the tool used to compare the spatial distribution of a set of features to a hypothetically-based random spatial distribution [7]. These spatial distributions, or patterns, are of interest to many areas of geographic research because they can help identify and quantify patterns of features in space so that the underlying cause of the distribution can be determined [8]. In health researches, spatial analysis is used to detect and quantify the patterns of disease distribution that may offer an insight into a disease's epidemiology [9].

Although spatial analytical techniques rarely give reasons why spatial patterns occur, they do identify where or the location of the *spatial* patterns occur. Within the realms of health research, this provides a useful means to hypothesize about the health outcomes or to identify spatial issues that need to be further investigated [10]. Thus, in this study a spatial analysis was applied in order to map the spatial distribution of dengue cases in Hulu Langat through spatial statistics method.

## II. METHODOLOGY

### A. Study Area

Hulu Langat is one of the nine districts in Selangor that consists six other *Mukim* or sub-district which are *Ampang*, *Hulu Langat*, *Cheras*, *Kajang*, *Semenyih*, *Hulu Semenyih* and *Beranang*. Hulu Langat is located at the upper part of Langat River Basin and it is the fifth biggest district (484.32 km<sup>2</sup>) in Selangor with total population of 915, 667 people in 2005. This study used '*Mukim*' or sub-district, the smallest local governing unit in Malaysia, as the spatial zoning boundary. All of the sub-district were under jurisdiction of *Majlis Perbandaran Kajang* (MPKj) except for *Ampang* which under jurisdiction of *Majlis Perbandaran Ampang Jaya* (MPAJ). Thus, in this study we exclude the *Ampang* area.



Fig. 1: Map of Hulu Langat District

### B. Data Collection

In order to achieve the objective of this study, several data are required to be incorporated in GIS environment such as dengue case data and demographic data. These data were gathered from various sources.

### C. Sources of Dengue Case Data

Dengue data was obtained from Hulu Langat District Health Office, Selangor for year 2003. All of the data was sorted according to each sub-district and years. This is a secondary data review of all dengue cases notified to Hulu Langat District Health Office, Selangor for year 2003. All the cases were documented in VEKPRO (Vector program) database. From that, it was extracted to Microsoft Excel program. Figure 2 shows the number of dengue cases for each sub-district in Hulu Langat. Attributes such as gender, age, race, occupation, address, date of admitted and date of discharge, type of dengue, locality of incidence took place and number of week. During the data smoothing process, we excluded the uncompleted data to avoid any biasness. In situ data collection was also carried out using handheld GPS Juno SB Trimble to locate the spatial location (coordinate) of the dengue incidence. In order to synchronize all of the digital data coordinate system, we decided to use World Geodetic System (WGS 84) which serve the x, y of an object by longitude and latitude respectively. The coordinate or point of the dengue incidence was essential in this study as it will reflect the spatial correlation of dengue between each sub-district. After collecting the coordinate in the field, the data that were stored in the memory card from the GPS was downloaded into a computer using GPS Pathfinder software version 4.3. To optimize the data collection reliability, we decided to move the entire coordinate to the center of the house or dengue area rather than using the actual GPS position. After *smoothing* the coordinate, it was transferred into the GIS software.

### D. Implementation of GIS and Spatial Analysis

In this study, the spatial distribution of dengue incidence within Hulu Langat district was examined using spatial statistic method. First, a global spatial autocorrelation method which is *Moran's I* used to test whether the dengue cases within Hulu Langat District is spatially correlated or not. *Moran's I* measures spatial autocorrelation and it is usually applied to area units where numerical ratio or interval data are available [6]. It is determined by calculating a mean for observation and then comparing the value of each incident with the value at all other locations [11].

It can be defined as:

$$I = \frac{n \sum_{i,j} W_{ij} (x_i - \bar{x})(x_j - \bar{x})}{w \sum (x_i - \bar{x})^2} \quad (1)$$

Where  $N$  is the number of cases,  $\bar{x}$  is the mean of the variable,  $x_i$  is the variable value at a particular location  $i$ ,  $x_j$  is the variable value at another location  $j$ , and  $W_{ij}$  is a weight indexing location of  $i$  relative to  $j$ . The value of Moran's  $I$  range from -1 for strong negative spatial autocorrelation to +1 for strong positive spatial autocorrelation. A value near 0 would indicate a spatially random pattern.

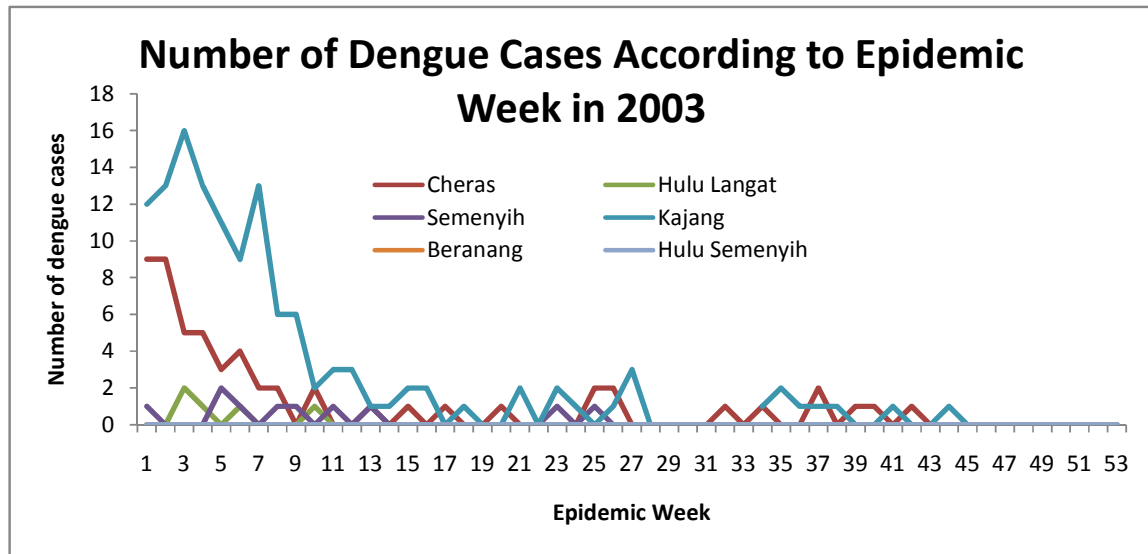


Fig. 2: Number of reported dengue cases for each sub-district in Hulu Langat district for year 2003

Distance analysis calculates statistics that are based not only in the location of points, but on the distance between points. Average nearest neighborhood (ANN) will be used to analyze whether the incidence is clustered or not. The ANN tool measures the distance between each feature centroid and its nearest neighbor's centroid location. It then averages all these nearest neighbor distances. If the average distance is less than the average for a hypothetical random distribution, the distribution of the features being analyzed is considered clustered. If the average distance is greater than a hypothetical random distribution, the features are considered dispersed. The average nearest neighbor ratio is calculated as the observed average distance divided by the expected average distance (with expected average distance being based on a hypothetical random distribution with the same number of features covering the same total area).

The average Nearest Neighbour is given as:

$$ANN = \frac{\bar{D}_O}{\bar{D}_E} \quad (2)$$

Where  $\bar{D}_O$  is the observed mean distance between each feature and their nearest neighbour:

$$\bar{D}_O = \frac{\sum_{i=1}^n d_i}{n} \quad (3)$$

And  $\bar{D}_E$  is the expected mean distance for the features given a random pattern:

$$\bar{D}_E = \frac{0.5}{\sqrt{n/A}} \quad (4)$$

In the previous equations,  $d_i$  equals the distance between feature  $i$  and its nearest feature,  $n$  corresponds to the total number of features and  $A$  is the total study area.

The  $z_{ANN}$ -score for the statistic is calculated as:

$$z_{ANN} = \frac{\bar{D}_O - \bar{D}_E}{SE} \quad (5)$$

Where:

$$SE = \frac{0.26136}{\sqrt{n^2/A}} \quad (6)$$

Lastly, a hot spot analysis using Kernel density estimation interpolation technique was done. Kernel estimation is a technique, used to generalize incident locations to the entire study area. According to [12], kernel density estimation is an effective tool to identify high-risk areas within point patterns of disease incidence by producing a smooth, continuous surface that defines the level of risk for that area. Kernel density estimation is an interpolation that is appropriate for individual point locations [11]. It calculates the density of point features around each output raster cell. According to [11] the kernel estimate is a better 'hotspot' identifier than the cluster analysis. Kernel density estimation is a useful method

as it helps to precisely identify the location, spatial extent and intensity of dengue disease hotspots.

### III. RESULTS AND DISCUSSION

#### A. Spatial Distribution

To assess the spatial distribution of dengue cases within the district of Hulu Langat, we used an index of spatial autocorrelation which is *Moran's I*. This index measures spatial autocorrelation based on both feature location and feature values simultaneously. *Moran's I* is an inferential statistic, which indicate that the result of the analyses must always interpreted within the context of the study null hypothesis. The null hypothesis for *Moran's I* stated that values of the analyzed attribute are randomly distributed among features in the dataset. In other words, the spatial process promoting the observed pattern of values is random chance. Result from *Moran's I* test on dengue incidence for Hulu Langat indicated that there was positive spatial autocorrelation exist among dengue incidence within the district. The *Moran's I* for dengue cases is 0.75 ( $p < 0.01$ ) while the z-score is 5.03 ( $p < 0.01$ ). From the result, we can conclude that the null hypothesis is rejected because the spatial distribution of dengue cases in the district is more spatially clustered than would be expected if underlying spatial process were random. Figure 3 shows the result of *Moran's I* analysis using GIS software.

#### B. Distance Analysis

In this study we applied (ANN) to analyze the dengue spatial distribution in Hulu Langat district. Using ArcGIS 9.3, this analysis gives three values: Nearest Neighbour Ratio, z-value and p-value as shown in Figure 4. It is found out that the dengue cases within the district are spatially clustered. Results from ANN analysis showed that the average nearest neighbor ratio is less than 1 which is 0.518755 ( $p < 0.0001$ ). From this result, we can expect the dengue cases pattern in Hulu Langat district is exhibiting a cluster pattern. The z-score for dengue incidence within the district is -13.0525 ( $p < 0.0001$ ). The z-score is a test of statistical significance that helps you decide whether or not to reject the null hypothesis. In this study, the null hypothesis is that there is no spatial pattern among dengue cases within Hulu Langat. With small z-score, there will be small probability which is less than 1% likelihood that this clustered pattern could be a result of random chance; so, we can reject the null hypothesis. By using ANN also, we can determine that the significant spatial autocorrelation of dengue incidences occurs at an average distance of 380.81 meters.

#### C. Hot Spot

Kernel density estimation was applied in this study to locate 'hot spot' for dengue cases in the district of Hulu Langat. The dark color area in Figure 5 is the hotspot identified with

maximum dengue incidence density. Hence, with the help of dengue density map we are able to target specific area within plot showing highest incidence.

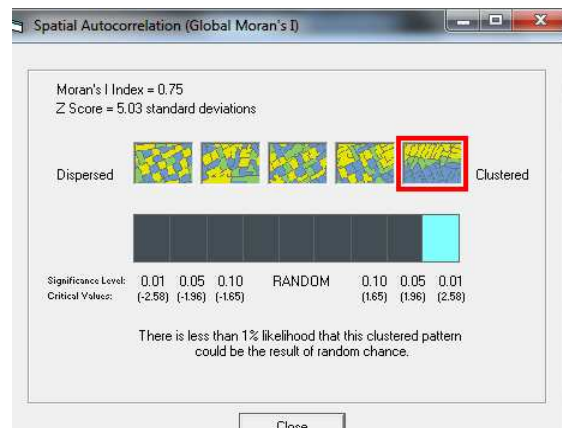


Fig. 3: Result of *Global Moran's I* for dengue cases within Hulu Langat district

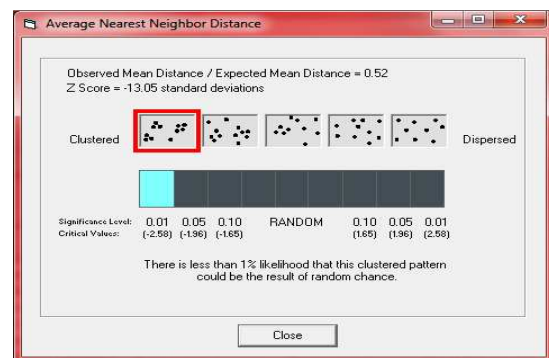


Fig. 4: Average Nearest Neighbour statistic result

Overall picture of dengue density variation within the plots can be known with the help of kernel density map. From the result, we can identify the area the most affected by dengue disease. Generally, most of the hot spot locations were located in *Kajang* sub-district. All of the detected hot spot in *Kajang* were situated at residential area such as *Sg. Jelok*, *Taman Impian Ehsan*, *Sg. Chua*, *Taman Kajang Utama* and *Taman Kajang Baru*. However, it was also found out that there was one location of high dengue density cases located in the sub-district of *Cheras* as shown in Fig. 5.

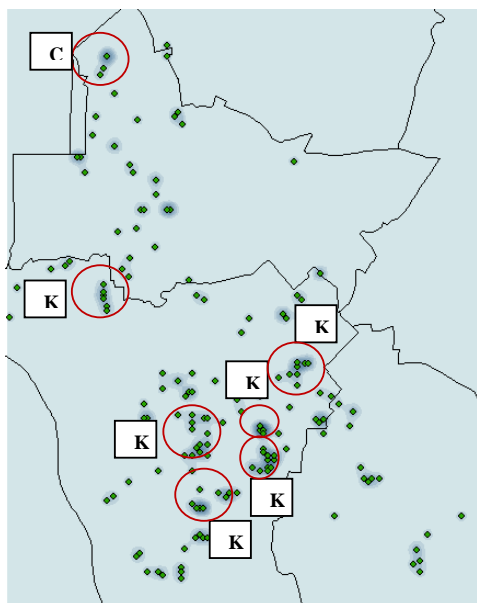


Fig. 5: Hot Spot locations of dengue cases in Hulu Langat. **C** indicate the area in *Cheras* sub-district while **K** indicate the area in *Kajang* sub-district

#### IV. CONCLUSION

This study analyzed the spatial distribution of dengue in Hulu Langat district in Selangor, Malaysia during the year 2003. The study was done by coupling the data manipulation and geographic presentation capability of GIS and spatial statistic to map the spatial distribution of dengue cases. It was found out that the dengue cases in the district were spatially autocorrelated through *Moran's I* indices. The dengue cases were clustered rather than distributed in random chances. This was supported by the ANN analysis that also stated the dengue cases within the district were highly clustered and it was found out that the significant spatial autocorrelation of dengue incidences occurs at an average distance of 380.81 meters. Using kernel density estimation in this study was a useful tool to locate the 'hot spot' or traditional use of incidence data cannot distinguish the spatial differences in risk areas from those of their neighboring areas, which crucial to the effective control of vector-borne infectious diseases [10]. From this study it was proven that using by GIS and spatial statistic tools, we can determine the spatial autocorrelation between dengue cases and population. Spatial mapping of dengue cases distribution also may assist health agencies, epidemiologist, public health officer, town planner and responsible authorities to combat dengue fever. Although GIS cannot cure the disease but it provided useful information as well as the ability to determined spatial relationship among dengue cases location through spatial analysis. However, this is only part of the approach to answering geographic questions in health research, as it takes spatial analysis in addition to a strong understanding of biological processes before underlying clues about health problems can be inferred [13].

#### REFERENCES

- [1] WHO – Regional Office for South-East Asia, SEARO (1999). Comprehensive Guidelines: Prevention and control of dengue and dengue haemorrhagic fever, WHO Regional Publication, SEARO No.29, 1999.
- [2] Gubler DJ, Meltzer M. (1999). Impact of dengue/dengue hemorrhagic fever on the developing world. *Adv Virus Res* 1999;53:35–70.
- [3] Mohd Din, M., M. Shaaban, et al. (2007). A Study of Dengue Disease Data by GIS Software in Urban Areas of Petaling Jaya Selatan. *GIS for Health and the Environment*: 206-213.
- [4] Derouich, M., A. Boutayeb, et al. (2003). "A model of dengue fever." *BioMedical Engineering OnLine* 2(1): 4.
- [5] Nakhapakorn, K. and N. Tripathi (2005). "An information value based analysis of physical and climatic factors affecting dengue fever and dengue haemorrhagic fever incidence." *International Journal of Health Geographics* 4(1): 13.
- [6] Nakhapakorn K, Jirakajohnkool S (2006). Temporal and spatial autocorrelation statistics of dengue fever. *Dengue Bull* 30: 177–183.
- [7] Mitchell, A. (2005). *The ESRI Guide to GIS Analysis: Volume 2: Spatial Measurements and Statistics*: ESRI Press.
- [8] Fotheringham, A. S., C. Brunsdon, et al. (2002). *Geographically Weighted Regression: the Analysis of Spatially Varying Relationships*, Wiley.
- [9] Srividya, A., Michael, E., Palaniyandi, M., Pani, S. P., & Das, P. K. (2002). A geostatistical analysis of the geographic distribution of lymphatic filariasis prevalence in southern India. *American Journal of Tropical Medicine and Hygiene*, 67(6).
- [10] Wen, T.-H., Lin, N. H., Chao, D.-Y., Hwang, K.-P., Kan, C.-C., Lin, K. C.-M., et al. (2010). Spatial-temporal patterns of dengue in areas at risk of dengue hemorrhagic fever in Kaohsiung, Taiwan, 2002. *International Journal of Infectious Diseases*, 14(4), e334-e343.
- [11] Levine, N. (2007). *Crimestat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (Version 3.0)*. Washington, D.C.: Ned Levine & Associates, Houston, TX and the National Institute of Justice.
- [12] Bithell, J. F. (1990). An application of density estimation to geographical epidemiology. *Statistics in Medicine*, 9, 691-701.
- [13] Malone, J. B. (2005). Biology-based mapping of vector-borne parasites by geographic information systems and remote sensing. *Parassitologia*, 47, 27-50.