

Geostatistical Modelling, Analysis and Mapping of Epidemiology of Dengue Fever in Johor State, Malaysia

Su Bee Seng¹, Albert K. Chong¹ & Antoni Moore²

¹School of Surveying
University of Otago, Dunedin,
New Zealand.

Email: subeeseng@gmail.com; chonga@albers.otago.ac.nz

²Antoni Moore
Department of Information Science
University of Otago, Dunedin
New Zealand.

Email: AMoore@infoscience.otago.ac.nz

Presented at SIRC 2005 – The 17th Annual Colloquium of the Spatial Information Research Centre
University of Otago, Dunedin, New Zealand
November 24th-25th 2005

ABSTRACT

Dengue fever, and especially the life-threatening form - DHF is an infectious mosquito-borne disease that places a heavy burden on public health systems in Malaysia as well as on most of the tropical countries around the world. Various environmental factors such as rainfall, temperature, living conditions, demography structure domestic waste management and population distribution are important in determining the mosquito survival and reproduction. A geostatistical modelling, analysis and mapping approach has been utilised in this research to understand the correlation between dengue fever prevalence, population distribution and meteorological factor, and the characteristics of space-time clusters in the Johor State. By supplementing GIS with geostatistical analysis and space-time permutation scan statistic tools, the spatial variation of dengue incidence can be mapped. Geographical weighted regression (GWR) analysis has revealed a strong ($R^2 = 0.87$) positive spatial association between dengue fever prevalence and population distribution in the Johor State. The dengue prevalence is expected to be higher in densely populated urban area, such as in Johor Bahru: however, there is a “rule” change in the Johor Bahru sub-district due to the positive impact from a dengue control and prevention programme. GWR analysis has also identified that ten to 14 days of accumulative rainfall is sufficient to support the mosquito breeding cycle and the dengue virus incubation period (vector + host) in the Johor Bahru district is 15 days. Space-time clusters showed that dengue transmission is a contagious type as the space-time extent is limited at 200m and 20 days and mainly involved household transmission. Results from this study reveal the ability of an augmented GIS surveillance system by incorporating the disease epidemiology and a geostatistical approach to provide reliable information for infectious disease management, control and surveillance. This research is the first study that has utilised GWR in infectious vector-borne disease, especially the attempt to “spatialise” the time in Hypothesis 2. In addition, it is also the first study which makes use of spatial-scan statistic permutation model to study the characteristics of dengue fever space-time clusters.

Keywords and phrases: Dengue fever, Geographical Weighed Regression (GWR), space-time, clustering, modelling, mapping, correlation, parameter estimates

1.0 INTRODUCTION

Spatial epidemiology covers the analysis of the prevalence and geographical distribution of a disease (Lawson and Williams, 2001). Issues such as data sampling, map interpretation and production, geostatistical analysis and modelling must be taken into consideration within this subject. Studies of spatial epidemiology often require the formulating of a mathematical model to describe the spread symbolically. Mathematical models are formed by summarizing the relationship between influencing factors and geographical distribution of the disease. The model can assist in predicting the impact of the outbreak and help implement an intervention plan.

Geographic Information Systems (GIS) has proven to be one of the most useful tools in public health research. It has been widely used in disease surveillance and monitoring, research hypotheses generation, identification of high-risk area and population at-risk, targeting resources and the monitoring of interventions (Gupta *et al.*, 2003). GIS provides an effective tool for visualization and spatial analysis of epidemiology data and environmental exposure. Recent studies have shown the increasing use of GIS as an important component in public health and epidemiology (Gupta *et al.*, 2003; Gatrell and Löytönen, 1998; Pearce, 1996).

Dengue fever (DF) is a mosquito-borne acute febrile viral disease characterized by sudden onset, fever, intense headache, myalgia, loss of appetite, rash and some non-specific signs and syndromes (WHO-SEARO, 1999). The female mosquito "*Aedes aegypti*" was found to be the most efficient vector. The *Aedes* female becomes infected when she takes the blood meal from an infected person within the viraemic phase of illness. The extrinsic incubation time of the *Aedes* female is about 8-12 days. After the extrinsic incubation period, the *Aedes* female is able to transmit the dengue virus to a human through her bite. The incubation within a human takes about three to 14 days (average 4-7 days).

DF remains an important public health issue in Malaysia due to its tropical climate, which is suitable for vector (mosquito) breeding sites. There are more than 20 cases reported weekly from every district of the Johor State according to the Johor weekly epidemiological bulletin. Within the state of Johor, the Johor Bahru district has the highest dengue cases in comparison to other districts. From 1st January 2004 until 31st December 2004, the cumulative figure of DF in Johor Bahru was 1356 cases out of the 3030 reported cases in the Johor State. In terms of the population, the Johor Bahru district covers 43% of the Johor State population.

The diffusion of dengue depends on the interaction between vector, parasite and human in the natural environment. The dengue virus cannot be transmitted directly from human to human. Effective vector control is the only solution for dengue control and prevention in situations where vaccines are unavailable. With its powerful analysis, modelling and mapping capabilities, GIS systems may serve as a decision-support tool for epidemic investigation, monitoring, simulation, prediction, prevention and resource allocation (Davenhall, 2002).

The overall aim of this research is to analyse the spatial pattern and diffusion of dengue fever cases by incorporating epidemiological and statistical techniques into a GIS system. This research will utilize the geostatistical analysis and spatial analytical tools in GIS to establish the relationship between dengue fever cases, population distribution, environmental impact and social-economic factors that contribute to dengue fever transmission.

This undertaking of building a GIS-based Dengue Fever Investigation and Surveillance System (DEFINSS) will aid in better health mapping and analysis compared to conventional methods. It will provide the epidemiologist with an effective tool to identify the environmental factors, mosquito breeding sites and population at risk related to the dengue outbreak. With detailed mapping and modelling of localized outbreaks, prevention and control activities can be implemented for instance. Furthermore, it can help to characterise the spread of the disease in time and space variation. The successful implementation of GIS technology in disease investigation and surveillance will help to increase the awareness of GIS technology in the public health sector.

2.0 METHODS

2.1 Data Collection, Integration and Management

Data from various organisations were collected, manipulated and integrated into the GIS system. Generally, spatial data consists of administrative boundaries (a total of 96 sub-districts) of Johor State, case event data (GPS points), satellite imagery and hardcopy topographic maps. Dengue fever patient records, meteorological (rainfall and evaporation), and population data were aspatial data utilised in this study.

All suspected and indeterminate cases of dengue fever reported in the Johor State for 2004 were used in this study. It should be noted that these cases were suspected cases only and could include other vector-borne diseases that are clinically similar to dengue fever infection. A reason for using suspect cases is because the antibody profiles differ at different stages of infection and the virus can only be isolated from serum drawn within the first few days of illness. Furthermore, dengue fever is endemic in Malaysia, and thus the possibility of suspected cases becoming confirmed cases is very high. The total number of cases was 3030 after excluded cases where the residential address was unspecified. 44% (1326 cases) of the dengue fever records are available as case event data collected from a Garmin eTrek handheld GPS. The total population for the study area is 3,029,300, resulting in a crude overall incidence rate of 100.02 per 100,000 person years.

ESRI ArcGIS V9.1 (ESRI, 1995-2005) was used to manage the spatial data. The geometry of spatial data was enforced after the topology was cleaned and built. This procedure resolved the “overshoot” or “undershoot” problem, for areal data in particular. Satellite imagery and aerial photos were georeferenced according to ground control points. Spatial and aspatial data were converted into a feature class group by feature dataset in a personal geodatabase. The Malaysian map projection was applied to all feature classes in ArcCatalog (ArcGIS V9.1) according to the data source. In Malaysia, cadastral data is normally available in *State Cassini* projection and topographic features are in *Rectified Skew Orthomorphic* (RSO) projection. ESRI’s ArcGIS is able to re-project data with a different projection and view them in the same project seamlessly.

2.2 Geographically Weighted Regression (GWR)

Knowledge regarding the extent of spatial association in the data being studied is very important in many spatial analyses (Getis and Ord, 1996). There are various ways (i.e. Moran’s *I*, Geary’s *C*) to identify the spatial association, however the appropriate technique should be acquired as based on the nature of the hypotheses. Recently, numbers of local statistics (i.e. Local Indicators of Spatial Associations (LISA) statistics (Anselin, 1995), Geographical Weighted Regression (GWR) (Fotheringham *et al.*, 2002), the *G* statistics (Getis and Ord, 1992), local Moran’s *I* and Geary’s *C*) have been developed to measure spatial dependency within its neighbours specified by distance or sample data instead of measuring the whole sample data within the study area. These statistics are able to portray the local variation in the study area where “hot spot” or “cold spot” can be easily identified.

GWR measures the spatial dependency (non-stationary) in a dataset and summarises the relationship by local regression parameters. Instead of studying the overall point pattern in the whole study region, GWR provides a more intuitive way to analyse the point pattern at a local scale (Fotheringham *et al.*, 2002). GWR technique is easily understood because of the traditional regression based framework. The results of GWR are checked with a statistical test such as Monte Carlo to verify the significance of potential spatial patterns. P-value for each variable is compared to the critical value at 95% CI (i.e. 0.05). The statistical inference helps to determine whether an observed pattern is due to random variation or a true spatial trend in the local model.

The combination of adaptive kernel with AIC minimisation bandwidth is adopted in the studies because the adaptive kernel is more appropriate to reflect the local variation resulting from the equal amount of sample data at each regression point as compared to the fixed kernel. Furthermore, it provides a minimum AIC value, which is considered as the strongest guide for model selection (Fotheringham 2005, Personal Communication). The output file from the GWR analysis can be linked to ArcGIS for mapping purposes.

2.2.1 Development of the Empirical Model

The Gaussian model is applied for Hypothesis 1 (dengue fever and population density) and Hypothesis 2 (dengue fever and meteorological factor). The Gaussian model in GWR is an ordinary least squares regression with a geographically weighted function. Data distribution was first examined through the *Normal QQ-Plot* function in ArcGIS 9.1, followed by data transformation to comply with the requirements of normality in the Gaussian model. Hypothesis 1 is about the relationship between population density and dengue fever cases in 96 sub-districts. This includes the analysis of population density in an alternatively form with population and area as separated factors. The possibility of non-linearity in the relationship between dengue fever prevalence and a predictor variable was explored by adding a polynomial term in Hypothesis 1 (i.e. LNPopDen²). Additionally, the effect of gender factor (in the population) is also considered in Hypothesis 1 by using the gender normal deviation (GenderNormDev) as an independent factor.

For Hypothesis 2, a variety of moving windows ranging from 10 to 14 days are used to obtain the accumulative rainfall, which is then used as the independent variable (X) to explore the effect of rainfall in two sub-districts (Senai/Kulai and Kluang) and its district level (Johor Bahru and Kluang). The daily precipitation evaporation factor is also taken into consideration in this study by subtracting the evaporation rate from the accumulative rainfall. Due to the average high temperature in Malaysia, the evaporation factor is crucial to determine the sufficient amount of water for mosquitoes to breed. It is postulated that 10 to 14 days accumulative rainfall is sufficient for the development from eggs to adult mosquitoes based on the mosquito's life cycle.

Hypothesis 2 attempts to find out the correlation between accumulative rainfall and the number of dengue fever incidence at a lag of 12 to 19 days (extrinsic + intrinsic incubation period). The lag of 12 to 19 days is believed to include the time interval for a human to show dengue fever infection symptoms after being bitten by an infectious mosquito. It is speculated that the dengue fever cases will be boosted 12 to 19 days after continuous rain spells. The centroid of each sub-district is used as the location variable in GWR analysis for Hypothesis 1. As an alternative of using geographical coordinates (X-Coordinate, Y-Coordinate) as a location variable, Hypothesis 2 introduces a new application of GWR in time series. Days in a year are used as the X and Y location variable.

In Hypothesis 1, the number of dengue fever cases is chosen to be the dependent variable and the population related factors as the independent variables. For Hypothesis 2, the number of dengue fever cases at district level or sub-district level was selected as the dependent variable and accumulative rainfall as the independent variable. GWR models were evaluated in terms of the improvement from global to local scale and spatial variability in local parameters. Only those local models that offer significant improvement (the difference between AICs is more than 3) over global models with significant local parameter estimates (Monte Carlo test) are valid for the most optimum model selection. Selection for the optimum GWR model was carried out by a combination of AIC value and goodness of fit criteria (R^2), however if there is a conflict, AIC has precedence.

GWR produces a set of local parameter (coefficient) estimates for each relationship at each location as a separated file. This parameter estimates file can be used as an input file to GIS software such as ArcGIS to visualize the spatial variability of parameter and standard error across the geographical space. Kriging interpolation (Geostatistical Analyst extension) is used to generate a surface of local parameter estimates from GWR. Prediction of dengue fever cases at any location within the Johor State is feasible by obtaining parameter values from the interpolated surface at that particular location. This application is very useful in disease control, management and surveillance.

2.3 Spatial Cluster Analysis (Hypothesis 3)

An analysis of the spatial distribution or dependencies of disease remains of the most important public health interest since the 1980s (Lawson, 2001). The spatial association between cases is subjected to the measure of geographical "closeness" or spatial proximity rather than a formal analysis (White 2005). Cluster analysis with statistical inference should be emphasized in case event studies. Since 2004, the Johor State Health Department has started to record the dengue fever notification as GPS points representing the geographical location of a case together with medical information. Spatial location of dengue fever cases is defined according to the patients' address. Only 44% (1326 cases) of the case event data in 2004 is available for cluster analysis when this research is conducted. Geocoding according to the patient's address is not feasible within the scope of this research due to the unavailability of a complete road network database. Therefore, the results from cluster analysis can only provide a general information

of the disease clusters' characteristics in terms of space and time, constrained by the availability of data, and should be interpreted in a conservative manner.

2.3.1 Spatial Clustering Methods

Statistical analysis of case event data is important to identify localised clusters of the disease that are excess of what would normally have been expected given the underlying population and demographic structure, and most importantly under the complete spatial randomness assumption (Besag *et al.*, 1991). The definition of clustering or what constitutes a cluster is often vague especially at an exploratory stage as the clusters under study can take a wide variety of forms which complicate the cluster analysis (Lawson, 2001).

In an exploratory analysis, the events distribution is regarded as randomly across the study area based on a uniform probability distribution. This means that the events are equally likely to occur at any location (Diggle *et al.*, 1995). A statistical significance test is determined by the level of significance to which the underlying distribution has deviated from a complete spatial randomness distribution. The hypothesis generating method in an exploratory analysis is often referred to as general clustering which is a global test across the whole study region whereas focused clustering is used as a confirmatory analysis involving focused tests in small areas (Bailey *et al.*, 1995).

Han *et al.* (2001) described cluster analysis as one of the data mining techniques and has categorised the clustering methods into four categories, namely, partitional, hierarchical, density-based and grid-based. Generally, the hierarchical clustering method requires the user to specify the threshold distance and a minimum number of points which constitutes a cluster. In the context of spatial epidemiology, a hot spot is determined by influential factors, not the number of incidents. A cluster can have any size within a certain time. Hierarchical clustering is an empirical derivative procedure with no theory or rationale behind the clusters, and is therefore not suitable for cluster analysis of dengue fever.

Similarly, the K-means partitioning clustering routine requires the user to specify the number of clusters to be obtained from the data and consequently divide the data into specific clusters according to the nearest distance to the centre of each cluster. The information of the number of clusters to be obtained is important to classify the spatial pattern. In the context of dengue fever disease, there is no *a priori* assumption of the number of clusters, and it is impossible to predict due to the complex relationship with environmental factor. Consequently, the partitioning clustering method was excluded from the consideration.

According to the Johor State Health Department, the definition of an outbreak is the occurrence of more than one case in the same locality, where the date of onset between the cases is less than 14 days. The outbreak is clear when no new case has been reported within 14 days from the date of onset of the last case in that particular locality. The space-time permutation model in SaTScan is able to capture this space-time association with dengue fever cases by utilising a density-based clustering method. The local intensity of the point pattern is measured with the flexibility in defining the spatial and temporal window. In conclusion, a space-time permutation scan statistic is implemented in SaTScan V5.0 (Kulldorff, 2004), and Crimestat V2.0 (Levine, 2002) was used to support the cluster analysis in this research. The former software was used to identify the significant localised clusters and the later was employed to understand the sequencing of sequential events in terms of time, distance and direction.

2.3.2 Space-Time Permutation Scan Statistic

The space-time pattern of dengue fever disease was retrospectively analysed with space-time scan statistic in SaTScan. There are various examples of the use of space-time scan statistic in spatial epidemiology applications, e.g. in the investigation of clusters of *giardiasis* (human intestinal infection) in Canada (Odoi *et al.*, 2004) and the analysis of the geographical distribution of *sporadic Creutzfeldt-Jakob* (rare fatal dementia) disease in France (Aignaux *et al.*, 2002). The space-time permutation model requires only case data with geographical coordinates and time for each case. Therefore, it is a most suitable model for the cluster analysis of dengue fever.

The space-time permutation scan statistic is defined by a cylindrical window consisting of a circular geographic base and height corresponding to time. The window is first centred in each location and gradually increases the radius of window from zero to an upper limit specified by the user. For each possible geographical location and size, the window also visits each possible time period which represents the number of days. By continuously changing the

cylinder window centre and radius, the window scans the study area for potential space-time clusters without any prior assumption about location, size and time frame. Since there is no population at-risk data involved in the space-time permutation probability model, the expected cases are calculated based on the cases. For each cylinder window, the alternative hypothesis implies that there is an elevated risk within the window as compared to the outside. Among all the cylinders evaluated, the cylinder with the maximum Poisson Generalized Likelihood Ratio (GLR) constitutes a most likely space-time cluster that is least likely to have occurred by chance.

Due to the large number of outbreaks in different locations, sizes, and time frames which were evaluated without population at-risk data, multiple testing was adjusted by generating a large number of random permutations of the spatial and temporal attributes for each case (Kulldorff *et al.*, 2005). The most likely cluster was calculated for each simulated data set, and the statistical significant (p-value) was obtained through Monte Carlo hypothesis testing (Dwass, 1957) by comparing the rank of GLR from the real and random data sets. This means that, the test of significance of the identified clusters is based on a likelihood ratio test whose p-value is obtained through Monte Carlo testing. Besides the most likely cluster, SaTScan also identifies and reports secondary clusters in dataset according to their GLR.

The main purpose of this space-time cluster analysis is to highlight the space-time characteristics of significant clusters which are crucial for dengue management, control and surveillance. The successful implementation of space-time permutation scan statistics in this research will encourage the use of an automated cluster detection technique in the Malaysia Health Department. Based on the Johor State outbreak report in 2004 (Vector Borne Diseases Control Unit, 2004), the longest period of an outbreak was around 80 days. Discussions with the epidemiologist (Personal Communication 2005) revealed that a cluster size of 400m is highly significant to the index case. In conclusion, a spatial window of 400m and a temporal window of 80 days have been used as the upper limit in clustering analysis. SaTScan will report the most likely and secondary clusters within this upper limit.

Correlated walk analysis (CWA) from Crimestat V2.0 was employed to generate the connection between individual events in a cluster. The diagram is very useful in the study of the temporal and spatial movement of dengue cases within a cluster. Hot-spots can be identified with ease from the flow diagram. The results from CWA are imported into ArcMap for further analysis and visualization. All GIS manipulation and mapping was performed in ArcGIS V9.1. The cluster information file from SaTScan was mapped on the Johor State sub-district boundary. Furthermore, the result from CWA was also incorporated into ArcMap to visualise the movement of case events within the cluster.

3.0 RESULTS

3.1 Correlation of Dengue Fever and Population Density (Hypothesis 1)

Figure 1 showed the variation of AIC value from the global (regression) to the local (GWR) model. As a result from the natural log transformation on Y-axis in model 4, the AIC value is much smaller (336.24) compared to the other model which used data in the original scale. Therefore model 4 is excluded from the figure to avoid misleading information. The difference in AICs value must be more than 3 in order to validate a significant improvement of the local model. Results in Hypothesis 1 point out that all the models show significant improvement from global to local AIC, except the model using population density (i.e. model 3) and the model with gender and area factor (i.e. model 8 and 9). Overall, model 1, 4, 6 and 7 qualified as candidates for optimum model selection.

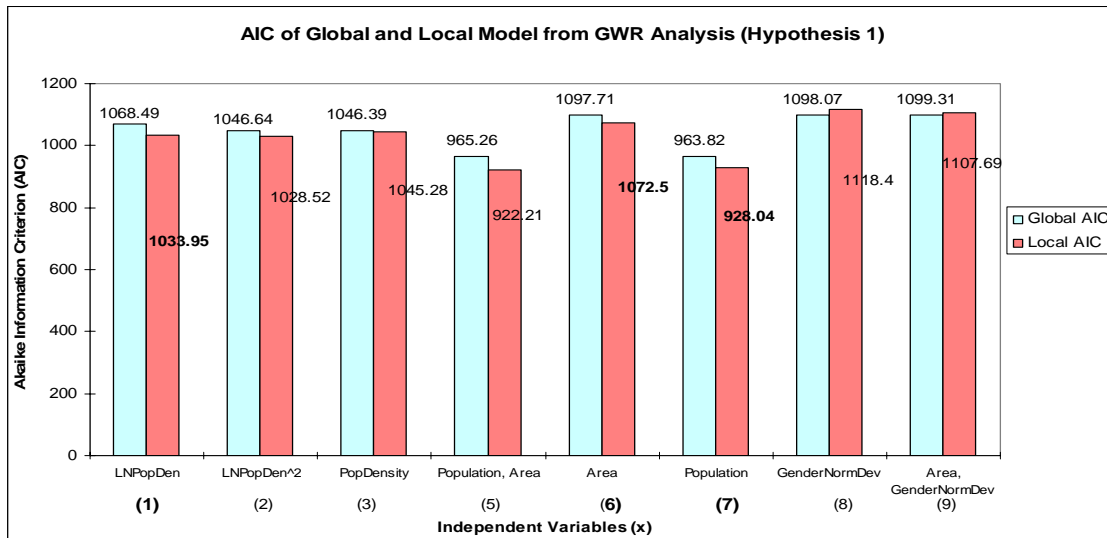


Figure 1: Global and local AIC score by independent variable(s) in Hypothesis 1.

GWR analysis found that the spatial variability of the *Population Density* parameter exhibited no significant relationship with dengue fever cases ($\rho = 0.80$) in local model 3. In this case, GWR had used 94 sample data to calibrate the local model as compared to 96 sample data in the global model. This suggested that the population density variable is globally fixed hence the local model does not show any significant improvement (AIC difference = 1.11). In addition, the spatial variability test also proves that the population density variable is stationary. Nevertheless, the global model of population density is significant at the 5% level (and 42% of the variation in dengue fever cases is explainable through the linear relationship).

Concluding from the GWR analysis above, model 1, 4, 6 and 7 are significant local models for the consideration in dengue fever modelling. The appropriate model should be applied according to the application and the availability of data. In this context, the population model (model 7) was chosen as the most optimum model. Figure 2 reveals that a similar number of cases are reported in different districts, having a huge difference in population count. In general, the Johor Bahru district has the biggest population (1,315,200), which is 3.5 times more compared to the Muar and Batu Pahat districts (374,900; 381,500) and 6.5 times more than the population in the Segamat district (201,900). The Mersing district has the lowest population in the Johor State (77,500).

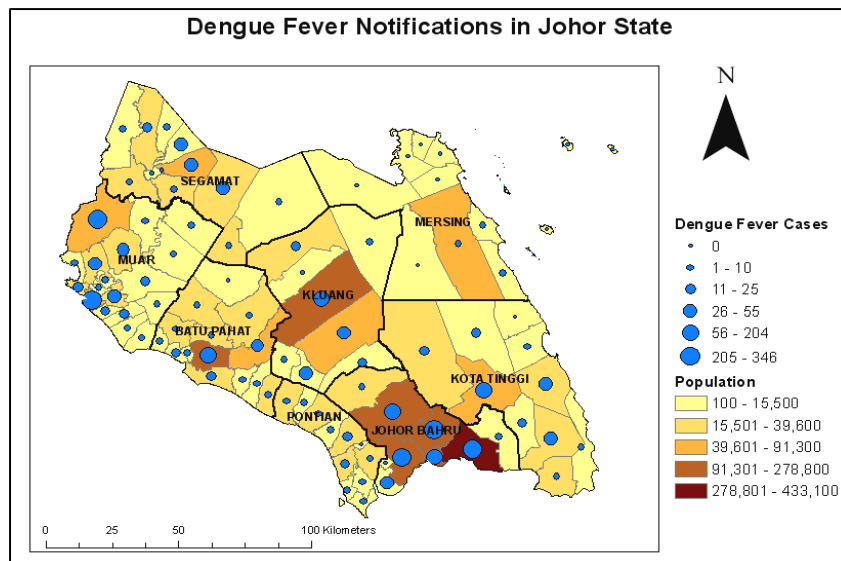


Figure 2: Dengue fever notifications in the Johor State.

In order to relate the impact of the population parameter to dengue fever cases, the local population parameter estimates (GWR result) were imported into ArcMap for mapping. The *Ordinary Kriging* technique is employed to generate a prediction surface in ArcMap from a population parameter estimate and superimposed with the dengue cases (Figure 3). The results indicate that there is a significantly positive correlation in the population parameter with a definite pattern over Johor State ranging from 0.0007 to 0.0035.

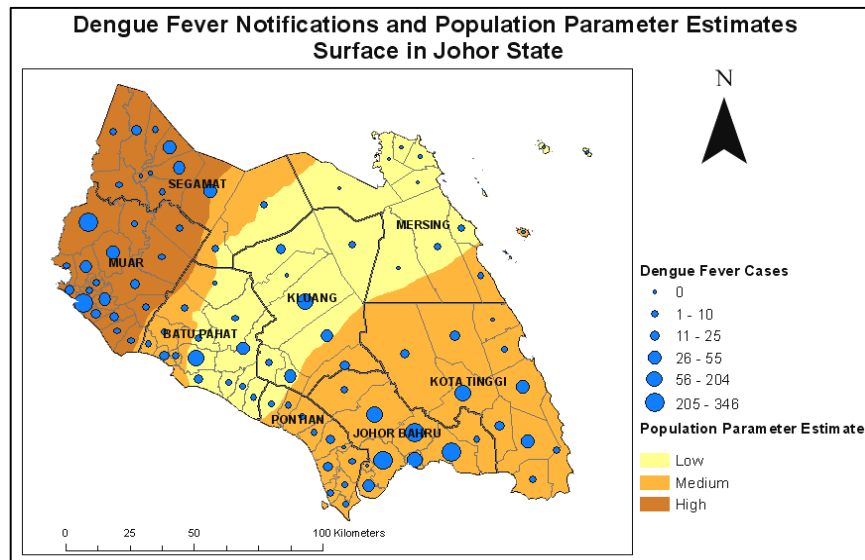


Figure 3: Dengue fever notifications and the population parameter estimates (*Ordinary Kriging*) in Johor State.

In general, the population parameter estimate can be divided in three zones, namely, low, medium and high. Dengue fever cases in the North-West district (Muar and Segamat) of the Johor State have a high correlation with the population parameter whereas the Centre district (Kluang, Batu Pahat, Mersing) have a low correlation. The correlation of the population parameter slowly increases from the Centre district towards the Southern district (Johor Bahru, Pontian, Kota Tinggi) of the Johor State. It is fascinating to discover that the population parameter estimate surface displayed a clear barrier between different districts. The parameter estimates surface show that the population in the Muar district have a higher risk to get infected by the dengue virus compared to the Batu Pahat district which has a similar population count.

The Segamat district can be compared to the Muar district with regard to topography, road network, and agricultural aspects. This may explain the similar trend of dengue fever transmission. It is speculated that accessibility to a road network, and thus a close proximity to towns, encourages people to travel around as part of their daily activities and facilitating the passive vector transmission by different types of transportation. Rural conditions, such as substandard housing and improper waste management as well as agricultural activities have fostered the dengue fever transmission, thus generating a high association with the population parameter estimates. The forest reserves and the less developed area towards the edge of the Muar and Segamat districts have become a natural barrier to stop the disease from spreading through the Batu Pahat, Kluang and Mersing districts.

Kluang has a lot of plantation and a few small recreation forests. Mersing on the other side, has a limited road connection with only one principal road from Kluang and a few more along the coastline (town area). Mersing is a hilly area with the lowest population in Johor State. As a result, only 10 cases were reported in Mersing due to its topography and population constraint. The Batu Pahat district has the second largest population and a well-structured road network in the Johor State. It is interesting that Batu Pahat has a low incidence rate although it has big a population. Perhaps the local district council has put in a lot of effort to keep the city clean. Overall, the low population density as well as an easy accessibility to the road network in the Centre district (Batu Pahat, Kluang and Mersing) have slowed down the transmission causing a low incidence rate in dengue cases, which also means that the population factor has a minimum impact on the transmission.

The southern district of the Johor State (Pontian, Johor Bahru and Kota Tinggi) is classified as a medium zone, where the population parameter estimates exhibit medium effect on dengue fever transmission. Johor Bahru is the second largest developing city after Kuala Lumpur (capital of Malaysia). 92% of the households are located in its urban areas. The Johor Bahru district has the largest population and it has reported 45% (1356 cases) of the total cases in the Johor State. It has been interesting to discover that the Johor Bahru district has the largest population with a medium risk of dengue fever infection, due to the positive impact from a dengue prevention program (i.e. COMBI).

By understanding the quantitative association of the population factor, prevention activity should be focused on areas with a high population parameter estimate (high risk) and close proximity to each other because dengue fever is able to spread across a sub-district in a short span of time. Aerial photos of high risk areas should be obtained to further explore population related factors such as housing distribution (clustered or disperse), vegetation and living environment for dengue prevention and control measure.

3.2 Correlation of Dengue Fever and Rainfall (Hypothesis 2)

Mosquito-borne disease transmission is closely related to climate for a number of reasons: standing water after periods of rain encourages the breeding of mosquitoes; a warm ambient temperature is crucial to the mosquito's life and gonotrophic cycle and virus replication (Patz *et al.*, 1998). The tropical climate as well as the tropical rainforest coverage in Malaysia have favourable impact on the year-round transmission of dengue fever through the provision of breeding sites and habitat for the vector. However, other environmental and socio-economic factors have complicated the direct relationship between rainfall and dengue fever transmission. Continuous rain, for example, may flush out larvae pools which in turn leads to a temporary reduction of vector populations.

Due to the availability and daily rainfall and dengue incidence data, Hypothesis 2 was carried out on a daily scale to exhibit the quantitative association between the two variables. It is believed that the daily scale study has been able to capture the correlation more accurately by taking daily variations into consideration compared to the weekly and monthly aggregated studies. Because of limited published meteorological data available in the Johor State, this study was focused on finding out the relationship of dengue infection with rainfall and evaporation rate in the Senai/Kulai and Kluang sub-districts and the district level for 2004.

GWR analysis in Hypothesis 2 attempts to find out the correlation of dengue fever prevalence and meteorological factors. Quantitative associations between 10 to 14 days of accumulative rainfall (independent variable, X) and dengue fever cases at 12 to 19 days lag (dependent variable, Y) are explored. In general, GWR has produced 40 local models in the district and sub-district levels respectively. The findings from Hypothesis 2 will provide an indication of the sufficient amount of accumulative rainfall necessary to support mosquitoes breeding as well as the period of dengue virus incubation in the study area.

GWR analysis in Johor Bahru district level reveals that the GWR models are only able to capture 9% to 10% of the association between dengue fever prevalence and meteorological factors. The model with the lowest AIC is not consistent at specific time lag across the accumulative rainfall, therefore an incubation period of the dengue virus could not be suggested. Furthermore, it is believed that the meteorological impact could be better explained by establishing the correlation with dengue fever prevalence in the district level. This is because 92% of the households in the Johor Bahru district are located in urban areas, which is free from the mountain range and rainforest effect. Therefore, it is postulated that areas within the Johor Bahru district will have the same impact in response to the meteorological factor.

Dengue fever cases from each sub-district were aggregated into the district level and GWR analysis was carried out to establish the relationship between accumulative rainfall and dengue fever prevalence at 12 to 19 days time lag. Figure 4 reveals that the model at 15 days time lag is found to be the model with the lowest AIC value consistently across the 10 to 14 days accumulative rainfall. The result is fascinating and the 15 days time lag has become an indication of the dengue virus incubation period in the Johor Bahru district. The R^2 value has increased from 0.46 at 10 days accumulative rainfall model and is maintained at 0.47 for the rest of the model at 15 days lag.

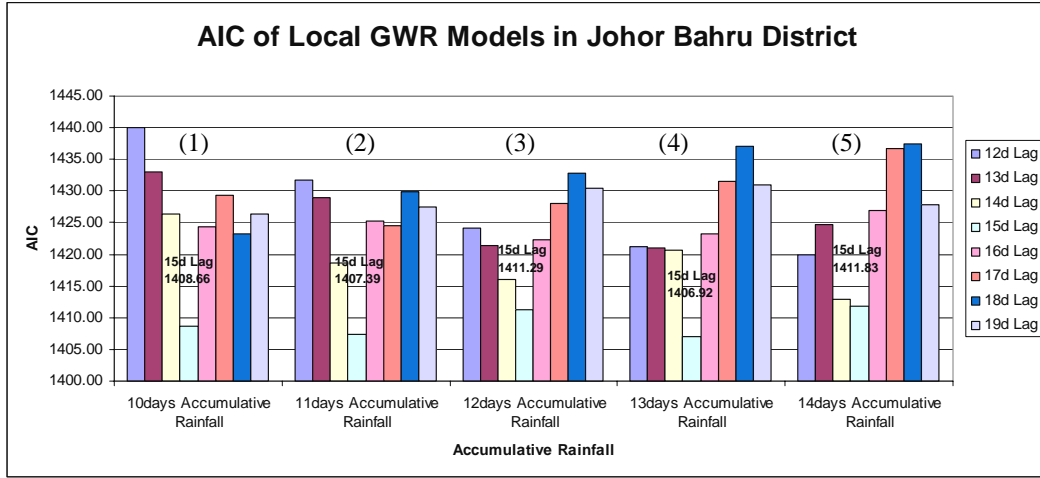


Figure 4: AIC of local models in the Johor Bahru district (Hypothesis 2).

To illustrate the findings of Hypothesis 2 from GWR analysis, model 3 (i.e. 13days accumulative rainfall & DF at 15days lag) is used as an example. The local equation of the relationship can be summarized as below, based on the Parameter 5-Numbers Summaries:-

$$DF_DHF_t = \beta_0(u_i, v_i) + \beta_1(u_i, v_i) (13daysAccumulativeRainfall)_{t-15days}$$

Where,

$$\beta_0 = -0.347 \text{ to } 4.73 \quad \beta_1 = -0.0237 \text{ to } 0.118$$

The equation explains dengue fever cases on a particular day in a year are mostly related to the 13 days accumulative rainfall on 15 days before, including that day. For instance, dengue fever cases at day 60 are related to 13days accumulative rainfall at day 46. This means that the accumulative rainfall and evaporation rate from day 34 to day 46 has contributed to the dengue fever cases at day 60. Long term epidemiological and meteorological data should be obtained as well as a retrospectively analysis to reaffirm the 15 days virus incubation period and the 10 to 14 days breeding cycle assumption. This would consequently produce a better correlation to explain the influence of meteorological factors to dengue fever prevalence. Seasonal or cyclic patterns of dengue fever transmission might be detected from the analysis of long term data.

Despite the moderate predictive power ($R^2 = 0.47$) between rainfall and dengue fever cases, the model suggests an application of potential public health interest, which may be useful for assessing and responding to the risk of an outbreak in a localized area based on accumulative rainfall (with evaporation effect) data of the previous period. It is undoubtedly better if other independent variables such as temperature and entomological surveillance data (i.e. *aedes* index and *pupal* index) are included.

A GWR analysis was carried out based on the same methodology and assumption employed in the Johor Bahru district. However, the time variability of 10 to 14 days accumulative rainfall was not significant in the Kluang district level. This might be due to the fact that the majority of the cases (53%) were reported within the Kluang sub-district. It will be more realistic to explore the correlation in the sub-district level. In the case of Kluang sub-district, GWR analysis reveals a very weak correlation between the meteorological impact and dengue fever prevalence. The models had a very regular and low R^2 at 0.12, which means only 12% of the variation of dengue fever prevalence in the Kluang sub-district can be explained through the model. Furthermore, the inconsistency correlation between dengue fever prevalence at the different time lag with the accumulative rainfall has complicated the interpretation of the GWR results. A similar trend in the correlation can hardly be observed. Hence, an association between the

temporal variation in dengue transmission and the variation in meteorological conditions, whilst plausible, could not be proven useful in the data that were analyzed.

To conclude, the findings from the GWR have confirmed the difficulties of correlating dengue fever with climatological factors. A more detailed investigation on the topography and socio-economic factors in the Kluang district is crucial to provide a possible explanation of the very weak correlation that was observed. A long-term study might reveal some interesting patterns of dengue transmission in the Kluang district compared to the Johor Bahru district.

3.3 Space-Time Cluster Analysis (Hypothesis 3)

Dengue fever is an infectious mosquito-borne disease with a limited extent of space and time. Uneven population distribution, socio-economic structure and mosquito' breeding sites create a spatial clustering in dengue fever. However, the fluctuation in mosquito populations, resulting from meteorological variables such as rainfall, temperature and humidity, has formed a temporal cluster in dengue fever. Understanding the space-time distribution of an infectious mosquito-borne disease like dengue fever is essential in the planning and implementing of effective control measures (Carter, 2002; Killeen *et al.*, 2003). The objective of this hypothesis is to establish cluster detection and mapping methodology in order to improve the disease investigation, surveillance and intervention.

Formal space-time cluster analysis identified a total of 31 clusters in the Johor State at a 5% significant level ($\rho = 0.05$). The results show that the Johor Bahru district has the highest number (11 clusters) of space-time clusters compared to the Muar and Pontian districts which only have two clusters. No space-time cluster exists in Mersing district with only 10 dengue cases per year. A scatter plot (Figure 5) was created to visualise the space-time distribution of significant clusters.

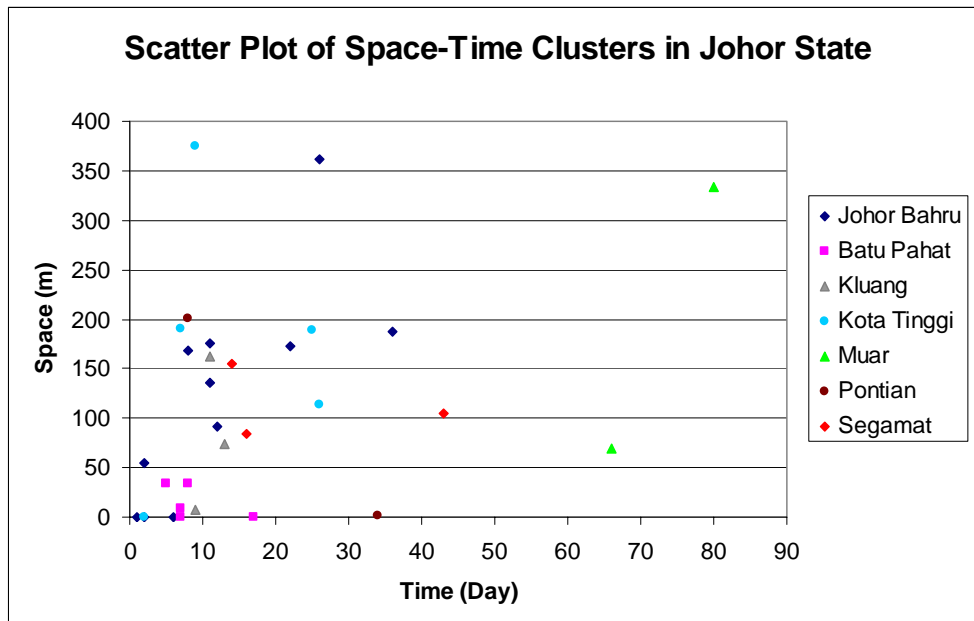


Figure 5: Scatterplot of significant space-time clusters in the Johor State (Hypothesis 3).

The space-time clustering analysis from Hypothesis 3 reveals a highly significance of the space-time association with dengue fever transmission. Figure 5 highlighted that most of the space-time clusters were limited at 200m and within 14 days. The space-time permutation scan statistic employed in this research is focused on detecting the space-time cluster and do not explore the correlation between the significant cluster. Even though the average nearest neighbor analysis has shown that the significant clusters in the study area (constraint by limited data) were not clustered in distance, however a detailed retrospective analysis of the inter-clusters relationship is worthwhile to

study further. It is believe that by doing so, the possible media of inter-community transmission could be suggested as well as the control and preventive measure.

Another important finding from Hypothesis 3 is that the household transmission is commonly involved in a dengue outbreak especially in students' and factory workers' hostels. Public awareness of the importance of dengue preventive measures within a household should be stressed to stop this transmission. The problem of the employment of foreign workers with low immunity in Malaysia is another serious issue in dengue control. Routine inspection of factories, construction sites and foreign workers' hostels should be carried out to minimize the risk of being infected.

3.4 Summary

A summary of the conclusion from each hypothesis was represented in Figure 6. In general, Hypothesis 1 has been rejected in a global model. However, the null hypothesis at the local model was accepted based on the GWR results as stated in Section 3.1. For Hypothesis 2, the results at the local model were evaluated, and the null hypothesis was rejected in the Kluang sub-district and the Johor Bahru district. A significant but weak correlation ($R^2=0.12$) was found in the Kluang sub-district, whereas dengue fever cases in the Johor Bahru district had a higher correlation ($R^2=0.47$) with accumulative rainfall. Furthermore, Hypothesis 3 was rejected based on the valuable results from space-time permutation scan statistic.

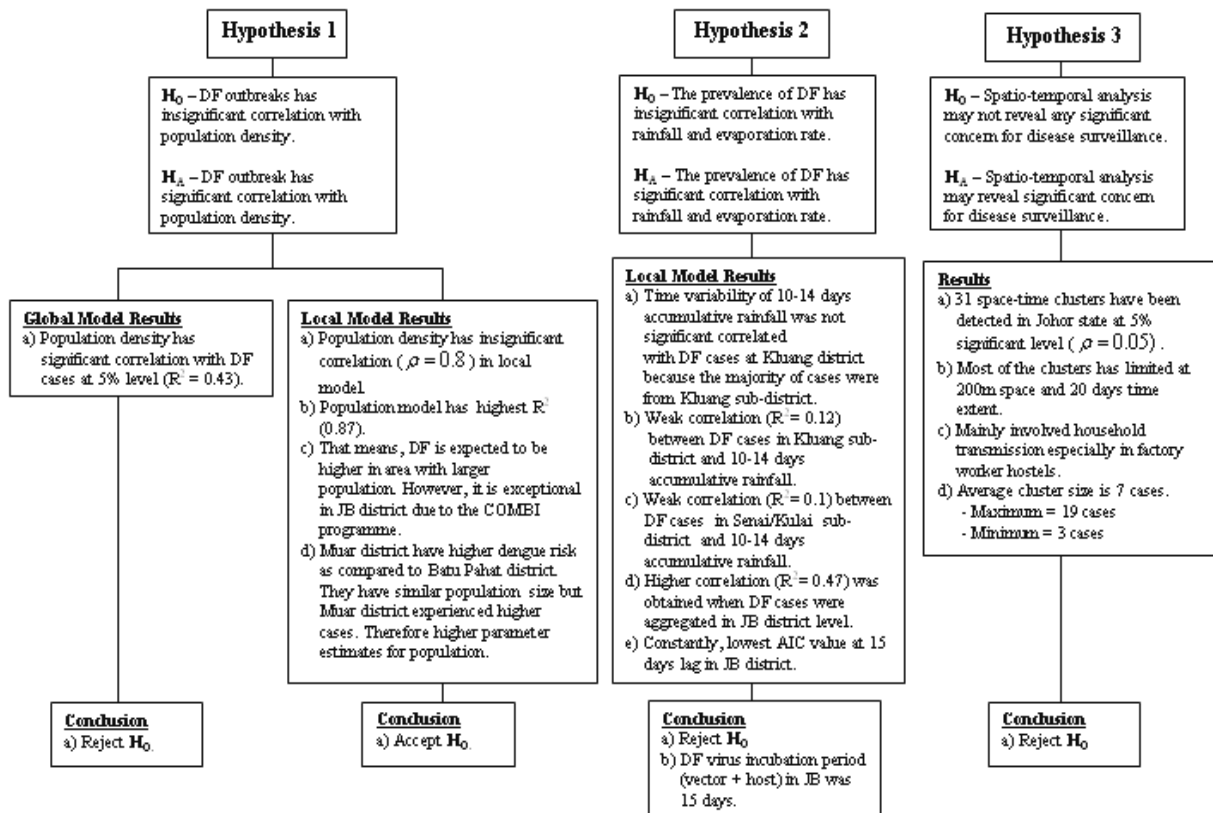


Figure 6: Summary of Hypothesis

4.0 DISCUSSION

The implementation of geographical weighted regression (GWR) analysis has improved the understanding of the linear relationship between population and dengue fever prevalence in the Johor State. In particular, the result from GWR has highlighted that the spatial distribution of dengue fever is closely related ($R^2 = 0.87$) to population distribution. This means that areas with a larger population are expected to have a higher dengue prevalence. However, an exceptional result was observed in Johor Bahru district where the dengue fever prevalence was considered medium compared to its large population due to the large scale dengue prevention programme.

Furthermore, the GWR analysis showed that the population in the Muar district has a higher risk compared to the Batu Pahat district which has a similar population. This finding is crucial from the aspect of dengue control and prevention. The study of the significant difference between the two districts in term of occupational risk, socio-economic, as well as environmental conditions might reveal important information regarding the risk of dengue infection in a heterogeneous environment.

Urbanization and population growth in areas such as the Johor Bahru district has fostered the dengue fever transmission especially in residential areas where inappropriate waste management at household level has created an abundance of artificial containers in surrounding premises which eventually become mosquitoes breeding sites. Furthermore, information from epidemiological data shows that 76% of the dengue fever cases in Johor Bahru district are reported within residential areas followed by 8% from squatter areas and 5% from traditional areas. This information has reaffirmed the association of dengue transmission with common breeding grounds associated with humans especially in developing and domestic environment such as the Johor Bahru district. Behaviour surveys to determine the key activities (i.e. gardening and waste management) which could encourage the dengue transmission might reveal fruitful results regarding the widespread occurrence of dengue fever in the domestic environment.

GWR analysis in (Hypothesis 2) illustrates an improved method to correlate the meteorological factor and dengue prevalence at a lag, based on the dengue virus incubation period. By incorporating the mosquito life cycle and virus incubation period in the localized model, the relationship between meteorological factors and dengue prevalence can be established at 5% significant level ($\rho = 0.05$). The result revealed a possibility of a 15 days virus incubation period and a 10-14 days mosquito breeding cycle in the Johor Bahru district. This finding might serve as an important measure to understand the dengue transmission in highly populated urban areas. However, it should be noted that the correlation of the meteorological factor and dengue prevalence is not always straightforward and might be complicated by other underlying factors. For instance, the result in the Kluang sub-district showed a very weak correlation ($R^2 = 0.12$). Besides, excessive amount of accumulative rainfall (>250mm) might have a negative impact on dengue prevalence such as found in the Johor Bahru district. Therefore, retrospective long term analysis is crucial to confirm the findings in this study.

Despite the number of previous studies cited in this research, convincing information supporting the association of dengue prevalence with rainfall is scarce, especially in Malaysia. Modelling of DF transmission and seasonal temperatures from 1988 to 1992 in Puerto Rico exposed a weak relationship between mean temperature and DF incidence (Hay *et al.*, 2002). Guha-Sapir and Schimmer (2005) distinguished the increasing problem of oversimplifications in the relationship between meteorological data and vector-borne diseases. Furthermore, mathematic models used to model the relationship have been criticised for not effectively account for rainfall, dynamic interaction of climatic variables, or relevant socio-economic factors (Reiter, 2001).

More detailed meteorological (i.e. temperature, humidity) and vector surveillance data (i.e. *Aedes* or container index) must be obtained in order to assess the forecasting ability of the model in Hypothesis 2. Additional research should extend from current study to a national scale to avoid oversimplifying the climatological impact on dengue transmission. Better understanding of the mosquito's gonotrophic cycle and virus latency period may be crucial in the planning of an efficient dengue control programme. Results from the space-time permutation scan statistic has revealed the limited space-time extent of dengue transmission at 200m and 20 days in Johor State. Based on the prior assumption of maximum spatial and temporal extent at 400m and 80 days, space-time permutation scan

statistics has identified 31 space-time clusters at 5% significant level ($\rho = 0.05$). Most of the clusters detected were located at medium or high population area, due to the fact that spatial clustering of disease is almost inevitable since human populations generally live in spatial clusters rather than random distribution of space.

However, space-time permutation scan statistic has accounted for the random distribution by the Monte Carlo hypothesis testing. This testing is also able to identify a significant space-time cluster which can provide important information on the epidemiology of dengue transmission. A common phenomenon - household transmission has been identified in most of the space-time clusters in the Johor State. Most of the space-time cluster involved at least one household transmission within very short distance and time. For instance, cluster C1259 involved 17 Nepal workers within 168.22m and 8 days.

Furthermore, most of the space-time clusters are observed in hostels and in residential areas. The problem of increased dengue transmission among foreign workers should be investigated in order to effectively address the main cause. The space-time cluster might be the result of an infective agent or an environmental episode on a local scale: local migration. Moreover, the space-time diagram such as the one generated from CWA serves as an important tool in tracking the space-time pattern of a cluster. By incorporating the epidemiology of dengue disease and geostatistical analysis into a GIS system, the spatial variation of the disease risk can be visualised with ease.

REFERENCES

- Aignaux, J.H., Cousens, S.N., Laupretre, N.D., Brandel, J.P., Salomon, D., Laplanche, J.L., Hauw, J.J. and Alperovitch, A. (2002) Analysis of the geographical distribution of *sporadic Creutzfeldt-Jakob* (rare fatal dementia) disease in France between 1992 and 1998. *International Epidemiology Association* 31, pp. 490-495.
- Anselin, L. (1995) Local indicators of spatial association-LISA. *Geographical Analysis* 27, pp. 93-115.
- Bailey, T.C. and Gatrell, A.C. (1995) *Interactive spatial data analysis*. England: Longman Group Limited.
- Besag, J. and Newall, J. (1991) The detection of cluster in rare diseases. *J R Stat Soc A* 154, pp. 143-155.
- Carter R. (2002) Spatial simulation of malaria transmission and its control by malaria transmission blocking vaccination. *Int J Parasitol* 32, pp. 1617-1624.
- Davenhall, B. (2002) Esri ArcUser Online. *ArcUser July-September 2002: Health geography*. <http://www.esri.com/news/arcuser/0702/overview.html> [accessed 20 November 2003].
- Diggle, P.J., Chetwynd, A.G., Haggkvist, R. and Morris, S. (1995) Second-order analysis of space-time clustering. *Statistical Methods in Medical Research* 4, pp. 124-136.
- Dwass, M. (1957) Modified randomization tests for nonparametric hypotheses. *Ann Math Stat* 28, pp. 181-187.
- ESRI (1999-2005) ArcGIS V9.1 *Environmental System Research Institute Inc. Redlands, California, USA*.
- Fotheringham, A.S., Brunson, C. and Charlton, M. (2002) *Geographically weighted regression: the analysis of spatially varying relationships*. England: John Wiley & Sons Ltd.
- Gatrell, A.C and Löytönen, M. (1998) *GIS and health*. London: Taylor & Francis.
- Getis, A. and Ord, J.K. (1992) The analysis of spatial association by use of distance statistics. *Geographical Analysis* 24, pp.189-206.
- Gupta, R., Jay, D. and Jain, R. (2003) Geographic Information System for the study and control of infectious disease. *GIS Development*. <http://www.gisdevelopment.net/application/health/overview/mi03113.htm> [accessed 11 May 2004].

Guha-Sapir, D. and Schimmer, B. (2005) Dengue fever: new paradigms for a changing epidemiology. *Emerging Themes in Epidemiology* 2005, Vol. 2, No 1. <http://www.ete-online.com/content/2/1/1> [accessed 6 April 2005].

Han, J., Kamber, M. and Tung, A.K.H. (2001) Spatial clustering methods in data mining. Chapter 8 In *Geographic data mining and knowledge discovery*. Edited by Miller, H.J. and Han, J. New York: Taylor and Francis Inc.

Hay, S.I., Cox, J., Rogers, D.J., Randolph, S.E., Stern, D.I., Shanks, G.D., Myers, M.F. and Snow, R.W. (2002) Climate change and the resurgence of malaria in the East African highlands. *Nature* 2002, Vol. 415, pp. 905-909.

Killeen, G.F., Knols, B.G. and Gu, W. (2003) Taking malaria transmission out of the bottle: Implications of mosquito dispersal for vector-control interventions. *Lancet Infect Dis* 3, pp. 297-303.

Kulldorff, M (2004) Software for the spatial and space-time scan statistics. SaTScan V5.0 National Cancer Institute and Farzad Mostashari at the New York City Department of Health and Mental Hygiene.

Kulldorff, M., Heffernan, R., Hartman, J., Assuncao, R. and Mostashari, F. (2005) A space-time permutation scan statistics for disease outbreak detection. *PLOS Medicine*, 2(3): e59.

Lawson, A.B (2001) Statistical methods in spatial epidemiology. England: John Wiley & Son, Ltd.

Lawson, A.B. and Williams, F.L.R. (2001) An introductory guide to disease mapping. England: John Wiley & Son, Ltd.

Levine, N. (2002) CrimeStat Version 2.0: A spatial statistic program for the analysis of crime incident locations. Annandale, VA: Ned Levine and Associates.

Odoi, A., Martin, S.W., Michel, P., Middleton, J.H. and Wilson, J. (2004) Investigation of clusters of giardiasis using GIS and spatial scan statistic. *International Journal of Health Geographics*. <http://www.ij-healthgeographics.com/content/3/1/11> [accessed 15 April 2004].

Patz, J. (1998) Global warming would foster spread of dengue fever into some temperate regions. *Science Daily Research News: March 1998* <http://www.sciencedaily.com/releases/1998/03/980310081157.htm> [accessed 20 October 2004].

Pearce, N. (1996) Traditional epidemiology, modern epidemiology and public health. *Am J Publ Health* 86, pp. 678-683.

Reiter, P. (2001) Climate change and mosquito-borne disease. *Environ Health Perspect* 2001, Vol. 109, pp. 141-161.

Vector Borne Diseases Control Unit, Johor Health Department (2004) "Dengue outbreak report in 2004".

White, P. (2005) "Spatial Data Cluster Analysis". Notes. Public Health Intelligence, Ministry of Health New Zealand.

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*.