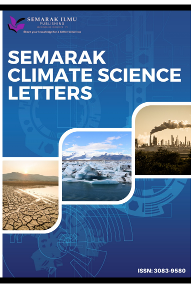
*Semarak Climate Science Letters* 1, Issue 1 (2024) 38-51

SEMARAK ILMU
PUBLISHING
202103268166(003316878-P)

Semarak Climate Science Letters

Journal homepage:
https://semarakilmu.my/index.php/scsl/index
ISSN: 3083-9580



# Geographically Varying Relationship of Dengue Incidence with Climate Factors in Malaysia

Jamaludin Suhaila[1,2,*], Ayuna Sulekan[2], Ruzaini Zulhusni Puslan[2], Foo Hui Mean[3]

[1] UTM Centre for Industrial and Applied Mathematics (UTM-CIAM), Ibnu Sina Institute for Scientific and Industrial Research, Universiti Teknologi Malaysia, 81310, Johor, Malaysia
[2] Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81310, Johor, Malaysia
[3] Institute of Statistical Science, Academia Sinica, Taipei,11529, Taiwan

**ARTICLE INFO**

***Article history:***
Received 19 September 2024
Received in revised form 31 October 2024
Accepted 15 November 2024
Available online 15 December 2024

***Keywords:***
Geographically Weighted Regression; Generalized Linear Model; dengue; stationarity; spatial autocorrelation

**ABSTRACT**

Dengue, a mosquito-borne arboviral disease transmitted by Aedes mosquitoes, has been a growing public health burden. Since dengue data is collected spatially, non-stationarity in the data series is possible. Therefore, we employ a Geographically Weighted Regression (GWR) model to ascertain the correlation between dengue incidence and location-specific predictor factors. Nevertheless, because GWR is based on conventional least squares, it has several drawbacks, including the inability to consider non-continuous variables, non-linear relationships, or non-Gaussian error distributions. Therefore, this study suggests a hybrid model called a Generalised Geographically Weighted Regression (GGWR), which combines Geographically Weighted Regression and a Generalised Linear Model (GLM) to address these flaws. This study aims to examine the influence of climate factors on the increase of dengue incidence in Malaysia and to establish their relationship using the proposed GGWR model. Stationarity and spatial autocorrelation assumptions will be evaluated using the Breusch-Pagan and Moran's tests, respectively, while normality will be assessed through the Jarque-Bera test. In addition, since the data series shows a problem of overdispersion, a negative binomial is proposed. Compared to GLM, the GGWR model offers a superior fit based on the minimum Akaike information criterion. This study also found that the GGWR negative binomial performed better than the GGWR Poisson regression based on the mean square errors in dealing with the overdispersion problem. The GGWR and GLM models demonstrated that climate factors substantially influence dengue incidence. The proposed GGWR model provides a more robust framework for understanding the spatial variability in dengue outbreaks, offering valuable insights for policymakers to devise more effective preventive strategies to mitigate the spread of the disease. The contributions of the study lie in developing and applying the GGWR model, which addresses key limitations of traditional GWR and GLM approaches, providing a more refined tool for analysing spatially varying relationships in public health data.

---

[*] *Corresponding author.*
*E-mail address: suhailasj@utm.my*

https://doi.org/10.37934/scsl.1.1.3851b

## 1. Introduction

Dengue has become a major issue for healthcare authorities, especially the World Health Organisation (WHO). The dengue virus causes dengue fever, an illness or disease that humans contract through mosquito bites. The symptoms of dengue fever often begin to appear approximately two weeks after infection and include a high fever, headache, vomiting, muscle and joint discomfort, and a rash. Generally, the fever takes five to seven days to recover, resulting in bleeding, a low number of blood platelets, leaking of plasma, or the syndrome of dengue shock. According to previous studies [1-3], dengue fever, a serious health problem in the tropics, has the potential to spread to other geographic areas, particularly Asia-Pacific countries such as Malaysia, Indonesia, Thailand, Vietnam, the Philippines, and South America.

Referring to the statistics reported by the World Health Organisation (WHO) [4], about half of the world's population is now at risk of dengue, with an estimated 100–400 million infections occurring each year. The dengue pandemic is currently a threat to approximately forty percent of the world's population, and countries with a climate that is considered subtropical are the most appropriate areas for dengue reproduction. In addition, several researchers discovered that dengue fever has increased in urbanised tropical and subtropical nations, such as Malaysia [5-7].

Numerous studies have shown a connection between climatic factors and infectious diseases, such as the studies by Githeko *et al.*,[8], Ten Bosch *et al.*,[9] and Tran *et al.*, [10]. Temperature and rainfall play critical roles in influencing the growth and survival of infectious agents, as well as in facilitating their transmission and spread. Githeko et al., [8] found that rainfall can influence the transport and spreading of the infectious agent, while temperature stimulates the growth and survival. Conversely, studies by Ten Bosch *et al.*, [9] and Tran *et al.*, [10] found that rainfall characteristics, which include the intensity and patterns of rainfalls, led to an increase in the number of dengue-affected areas. As mentioned by Wu *et al.*, [11], climate changes indeed impact the entire process of disease development for most recorded dengue incidences in statistics, including the reproduction and spreading of the disease pathogens and their hosts. Their findings have shown significant relationships between climate change factors and dengue incidence. In addition, a study by Lee *et al.*, [12] found that a higher temperature considerably increases the risk of domestic dengue epidemics in Korea. Similarly, Butterworth *et al.*, [13] studied the potential impact of climate change on dengue transmission in the Southeastern United States. They concluded that the cold winters on the US mainland inhibit dengue virus transmission.

When addressing global climate change, it is important to consider the factors that drive the development and spread of agent diseases. The findings and results of those studies by Mwanyika *et al.*,[14] and Caldwell *et al.*,[15] have shown that the occurrence and persistence of appropriate climate factors will enable dengue outbreaks. Liao *et al.*, [16] suggested that emerging infectious diseases continue to cause significant harm, and their frequency is expected to rise due to climate change. Hence, global strategies are needed to track, model, and project future emerging infectious disease behaviors. Direct effects due to climate change encompass the increasing frequency and intensity of extreme weather events, including variations in temperature and precipitation patterns [17]. For climate-sensitive infectious diseases, these alterations indicate a change in geographical and temporal distribution, seasonality, and transmission intensity [18].

Modelling the relationship between climate factors and the incidence of infectious disease has become the main challenge in climate-health studies. However, given that heterogeneity exists between the disease and climate variables, geographically weighted regression (GWR) is a common approach to solving heterogeneity by considering the variability of coefficients in diverse locations across the study area. GWR is a linear model that identifies relationships between dependent and independent variables that change across space, making it suitable for analysing non-stationary data. However, GWR fails to capture the non-linear behavior of the phenomena. The GWR model has been extensively employed for various diseases such as hand, foot, and mouth disease (HFMD) (Hong *et al*., [19], Sun *et al*.,[20]), dengue (Sulekan *et al*., [21] Sumanasinghe *et al*., [22]) and COVID-19 (Isazade *et al*., [23]; Jassim *et al*., [24]; Wu and Zhang [25]; Zhang *et al*.,[26].

Wang *et al*., [27] utilised the GWR model to address geographical dependency and heterogeneity within the data set. Their findings indicated that environmental factors, particularly human population density, the Normalised Difference Vegetation Index (NDVI), socioeconomic deprivation, and road density, exerted a more substantial influence than meteorological variables on the spatial variation of mosquito abundance at the urban level. On the other hand, Isazade *et al*., [23] identified the COVID-19 pandemic in Qom and Mazandaran provinces, Iran, using spatial analysis approaches. They used three models for assessing spatial autocorrelation: geographically weighted regression (GWR), ordinary least squares (OLS), and Moran's I. ArcGIS software executed all those models. Authorities and researchers use the spatial information from these modelling approaches to gain general insights for targeted investigations and policies in similar situations. Zhang *et al*., [26] examined the influence of sociodemographic variables on COVID-19 incidence across 342 cities in China from a geographical aspect. Their analysis used the local geographically weighted Poisson regression (GWPR) and Poisson regression models. Their findings indicated that cities with a higher gross domestic product (GDP), limited health resources, and shorter distances to Wuhan were at a higher risk for COVID-19.

In Malaysia, there are limited studies on GWR in spatial variation modelling [28-31]. Hazrin *et al*., [28] used GWR to analyze the health clinic requirement in a sub-district based on the population density and the number of high-level health clinics. They also map the spatial distribution of public and private health clinics. Eboy and Samat [29] used GWR as a space regression tool for modelling the valuation of property ratings over Kota Kinabalu, Sabah, while Jamhuri *et al*., [30] developed a toolbox to extend GWR to the forestry sector. On the other hand, Vaziri *et al*., [31] used GWR in the spatial distribution of poverty and in assessing all the factors that contribute most to the spatial configuration of Peninsula Malaysia.

However, GWR fails to handle non-continuous variables and non-Gaussian errors. The challenge in applying GWR arises from spatial data, which can involve various types such as binary or count data, making conventional GWR methods unsuitable. In many health problems, the response variable represents a count best modeled by the Poisson distribution, making classical GWR inappropriate for this data type. Therefore, to address the limitations of GWR, we employ a generalised geographically weighted regression (GGWR) to evaluate the impact of climate change on dengue incidence and concurrently model the spatially varying relationships between variables.

To the best of the authors' knowledge, the applications of geographically weighted regression in spatial modelling remain relatively novel in Malaysia, particularly in climate change and disease analysis. Therefore, this study seeks to propose a hybrid model that combines GWR with the Generalised Linear Model (GLM), resulting in the generalised geographically weighted regression. Furthermore, this study aims to assess the influence of climate factors on dengue incidence in Peninsular Malaysia and establish their spatial relationships using GWR modelling. The proposed GGWR model can then be a predictive tool for estimating dengue fever risks under varying climate change scenarios and identifying potential hotspot areas prone to outbreaks.

## 2. Data and Study Area

Weekly dengue data for each state in Peninsular Malaysia were obtained from the official open data portal (www.data.gov.my). However, the observed data is constrained due to limited open-access provisions from the Ministry of Health. Considering the availability of comprehensive weekly dengue records, the dengue data series from 2012 was selected for this analysis. The climatological factors, such as temperature ($X_1$), humidity ($X_2$), rainfall($X_3$), and wind speed($X_4$), were used as explanatory variables to examine the effect of climate factors on dengue incidences. These climate data were obtained from the Malaysian Meteorological Department from the following stations, as shown in Table 1.

**Table 1**
List of meteorological stations with their latitude and longitude

| Stations | Longitude | Latitude |
|---|---|---|
| Kluang | 2° 01' N | 103° 19' E |
| Alor Setar | 6° 12' N | 100° 24' E |
| Kota Bahru | 6° 10' N | 102° 18' E |
| Melaka | 2° 16' N | 102° 15' E |
| Ladang Air Hitam | 2° 56' N | 102° 25' E |
| Kuantan | 3° 46' N | 103° 13' E |
| Ipoh | 4° 34' N | 101° 06' E |
| Chuping | 6° 29' N | 100° 16' E |
| Bayan Lepas | 5° 18' N | 100° 16' E |
| Subang | 3° 08' N | 101° 33' E |
| Petaling | 3° 06' N | 101° 39' E |
| Kuala Terengganu | 5° 23' N | 103° 06' E |

## 3. Methodology

### *3.1 Model Diagnostic Checking*

In performing regression analysis, it is essential to evaluate the normality assumptions of the models. The Jarque-Bera test, commonly used to assess model bias, indicates whether the residuals follow a normal distribution. The method of Jarque-Bera is given as follows:

$$\text{Jarque} - \text{Bera} = \frac{n-12}{6}\left(S^2 + \frac{1}{4}(K-3)^2\right) \tag{1}$$

where $S$ refers to the skewness while $K$ represents the value of kurtosis. The higher the Jarque-Bera statistics, the higher the chances for the data to deviate from the norm.

Spatial autocorrelation analysis assesses whether observed variables at one location are independent of those at other locations. The presence of spatial autocorrelation indicates dependencies between variables across spatial locations. In addition, the Moran autocorrelation statistic was applied to detect whether there is spatial autocorrelation or clustering of the residuals, which violates the assumption of regression. Under the null hypothesis, it is assumed that no spatial autocorrelation exists within the data series, while the alternative hypothesis posits the presence of spatial autocorrelation in the series. Moran's I is given as

$$I = \frac{12}{S_0}\left(\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\omega_{i,j}z_i z_j}{\sum_{i=1}^{n} z_i^2}\right) \tag{2}$$

where $Z_i$ is the deviation of an attribute for location *i* from its mean $(x_i - \bar{X})$, $\omega_{i,j}$ is the spatial weight between location *i* and *j*, *n* is equal to the total number of observations, and $S_0$ is the sum of all the spatial weights, $S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} \omega_{i,j}$. The $Z_i$ score for the statistics can be computed as

$$Z_I = \frac{I - E[I]}{\sqrt{V[I]}} \tag{3}$$

where $E[I] = \frac{-1}{(n-1)}, V[I] = E[I^2] - (E[I])^2$ .The value of Moran's I index can vary between − 1 (perfect dispersion) to + 1 (a perfect positive autocorrelation). Note that a zero value indicates perfect spatial randomness.

A Koenker test employs a particular test known as the Breusch-Pagan test to determine whether the global model has non-stationarity. Because of the spatially autocorrelated environment, non-stationarity occurs when the connection between the study variables differs from region to region. The Breusch-Pagan test is a test for heteroscedasticity error in a linear model regression test. Heteroscedasticity occurs when the residual error variance is not uniform over the measured range values. If heteroscedasticity occurs, the outcome analysis may be inaccurate due to uneven variance in the variables utilized in the regression. The test follows the Chi-square distribution's features and evaluation, ${x^2}_v = n^* R^2$, where *n* is the total number of observations, $R^2$ is the coefficient of the determination of the regression of the squared residuals, and *v* is the degree of freedom.

### *3.2 Generalised Linear Model (GLM)*

A generalised linear model is an extension of classical linear models that represents a class of regression models that allows linear regression in various data types, including counts, binary, proportions, and continuous distribution. The structure of the generalized linear model can be easily seen by considering the classical linear model. $y_i = x_i^T \beta + \varepsilon_i$ where $y_i$ represents the $i^{\text{th}}$ observation of the dengue cases, $\beta$ is the vector of model parameters, $x_{ik}$ is referring to the $i^{\text{th}}$ observation of the $k^{\text{th}}$ climate factors variable and $\varepsilon_i$ be the independent normally distributed error terms with zero means and also assumed to be homoscedastic. The link function of GLM gives $g(\mu_i) = \eta_i = x_i^T \beta$.

Poisson GLM appears appropriate for modeling variables that describe counting data over a specified interval. The family of Poisson distribution with parameter $\mu$ is given as follows:

$$P(Y = y_i) = \frac{e^{-\mu_i} {\mu_i}^{y_i}}{y_i!}, y_i = 0,1,2,3, \ldots, \quad i = 1,2,3, \ldots, t \tag{4}$$

where $\mu_i$ is both the mean and variance of *Y*. The mean $\mu_i$ is expressed as a function of some explanatory variable through a log link function. The details can be found in Nelder and Wedderburn [32]. In the modelling process, by using the Poisson distribution, the problem of overdispersion might exist if the variance of the observation exceeds the mean. In these issues, the standard errors may be underestimated, leading to misleading conclusions about the regression parameters. Therefore, a negative binomial distribution will be used, as suggested by Breslow [33], since it is one of the ways to account for overdispersion problems in data and is considered a convenient and practical approach. A negative binomial distribution with the probability function is given as follows:

$$P(Y_i|y_i) = \int P(Y_i = y_i|\lambda_i)\, f(\lambda_i) d\lambda_i = \frac{\Gamma(y_i + v_i)}{\Gamma(y_i + 1)\Gamma(v_i)} \left(\frac{v_i}{v_i + \mu_i}\right)^{v_i} \left(\frac{\mu_i}{v_i + \mu_i}\right)^{y_i} \tag{5}$$

where the mean and variance are $E(Y_i) = \mu_i$ and $\mathrm{Var}(Y_i) = \mu_i + \mu_i^2 v_i^{-1}$, respectively.

### *3.3 Geographically Weighted Regression (GWR)*

Geographically weighted regression is a relatively simple technique that extends the traditional regression framework by allowing local rather than global parameters to be estimated so that the model is written as

$$Y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(\mu_{i,}v_i)x_{ik} + \varepsilon_i \tag{6}$$

where $(u_i, v_i)$ represents the coordinates of the *i*-th point in space and $\beta_k(u_i, v_i)$ represents a realization of the continuous function at point *i*. This spatially localised model assumes that the relationships between regression variables may vary over space; consequently, it generates a set of local line regression models rather than a global model, with estimates for every sample in space. The weights of each spatial unit vary as a function of the spatial relationship between them. A weighted calibration is applied so that more influence in the calibration is related to the point closer to *i*. Observations close to *i* are weighted more than observations farther away from *i;* that is

$$\hat{\beta}(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1}(X^T W(u_i, v_i) Y) \tag{7}$$

where $\hat{\beta}$ is an estimate of $\beta$ and $W(u_i, v_i)$ is the matrix of *n* by *n* whose off-diagonal elements are zero and whose diagonal elements denote the geographical weighting of observed data for point *i*. Local estimation of model parameters is derived using a subsample of data from nearby observations, which are weighted by a decreasing distance. In this way, the impacts of neighbouring samples are stronger than those farther away. A kernel bandwidth indicates the distance beyond which neighbours no longer influence local estimates.

Two choices of the weighting function, either computing using a Gaussian kernel function or a Bi-square kernel function, which is given as

$$W_{ij} = exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right) \tag{8}$$

$$W_{ij} = \left(1 - \left(\frac{d_{ij}}{b}\right)^2\right)^2 \text{ if } d_{ij} < b \quad \text{and } W_{ij} = 0, \text{ otherwise.} \tag{9}$$

where $d_{ij}$ is the distance between the location of observation *i* and *j,* with *b* is the bandwidth. Note that the bandwidth can be defined by a fixed number of the closest neighbours (adaptive bandwidth) or a fixed distance (known as fixed bandwidth). The bandwidth selection can be determined using the Akaike Information Criterion (AIC). Minimising the AIC provides a trade-off between goodness-of-fit and degrees of freedom. The AIC is defined for GWR as the following

$$AIC_c = 2nln(\hat{\sigma}) + nln(2\pi) + n\left\{\frac{n+tr(S)}{n-2-tr(S)}\right\} \tag{10}$$

where *n* is the sample size, $\hat{\sigma}$ is the estimated standard deviation of the error term, and $\mathrm{tr}(S)$ refers to the trace of the hat matrix, which is a function of the bandwidth. As the general rule, the lower the AIC, the closer the model approximation to reality.

### 3.4 Generalised Geographically Weighted Regression (GGWR)

Generalised geographically weighted regression is a version of the generalised linear model that incorporates spatial variability coefficients. In this case, it is used to describe the linear predictor with

$$\eta_i(u_i,v_i)=\beta_0(u_i,v_i)+\beta_1(u_i,v_i)X_{1,i}+\beta_2(u_i,v_i)X_{2,i}+\cdots+\beta_k(u_i,v_i)X_{k,i}=\sum_k\beta_k(u_i,v_i)X_{k,i} \quad (11)$$

Based on Nakaya *et al.*,[34], the geographically weighted Poisson regression (GWPR) can be defined as

$$\begin{aligned} &Y_i \sim Poisson(\mu_i)=Poisson\left(exp\big(\eta_i(u_i,v_i)\big)\right)\\ &log(\mu_i)=\eta_i(u_i,v_i)\\ &\mu_i=exp\big(\eta_i(u_i,v_i)\big)\end{aligned} \quad (12)$$

where Poisson ($\mu_i$) indicates the Poisson distribution with mean, $u_i$. A component $\rho_i$, which is known as the offset, is frequently included in the Poisson regression model as

$$\mu_i=\rho_i exp\big(\textstyle\sum_k\beta_k(u_i,v_i)\,X_{k,i}\big). \quad (13)$$

The GWPR model extends the traditional Poisson regression model, allowing for local rather than global parameter estimation. Poisson regression is commonly used for counting data, where the outcome variable is the count of events that follow a Poisson distribution. However, in the context of GWPR, the model's coefficients vary across geographic space. This flexibility is crucial for addressing spatial non-stationarity, where the influence of explanatory variables on the outcome may differ across locations.

Silva and Rodrigues [35] proposed GWR with a negative binomial distribution called GWNBR. The GWNBR model is an extension of the geographically weighted regression that caters to overdispersion in count data. Overdispersion occurs when the variance exceeds the mean, a common scenario in real-world datasets where count-based outcomes exhibit greater variability than the Poisson assumption allows. The GWNBR model, therefore, incorporates a negative binomial distribution, which has an extra parameter to account for overdispersion. The global model with family negative binomial is applied to a count model with overdispersion where

$$Y_i \sim NB\left[\rho_i exp\big(\textstyle\sum_k\beta_k(u_i,v_i)X_{k,i}\big),\alpha\right] \quad (14)$$

with $\rho_i$ denotes as an offset variable, $\alpha$ refers to the overdispersion parameter, $\beta_k$ is the parameter that is related to the explanatory variables, $X_k$ for *k*=1,2,3,…*K*, and $Y_i$ be the dependent variable when *i*=1,2,…,*n*.

## 4. Results and Discussion

### 4.1 Summary Statistics of Dengue Cases

Figure 1 displays a bar graph of the number of dengue cases with their respective coefficient of variations for each studied station in Peninsular Malaysia. Selangor recorded the highest number of dengue cases, while Perlis recorded the lowest. There is a high possibility that high urbanisation, industrialisation, and the rapid development of Selangor will influence the increasing number of dengue incidences in the area. The weekly variations range from 24% to 65%, with the lowest variability recorded for Johor and the highest in Perlis. The highest variability of the dengue cases in Perlis indicates that the weekly number of dengue cases is widely spread for the stations and varies in 2014.

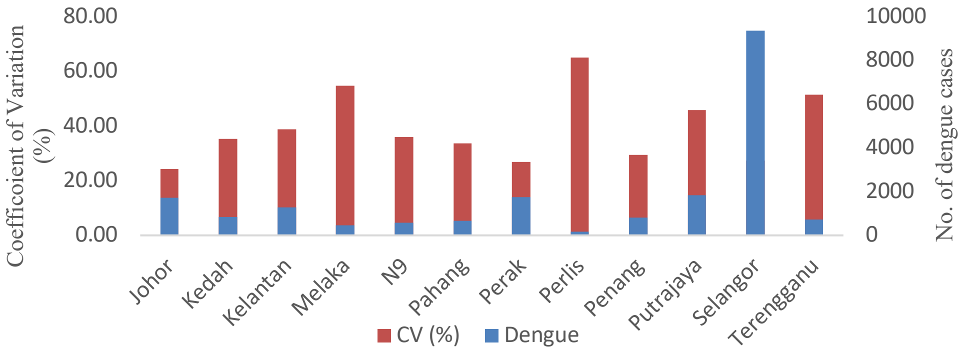


**Fig. 1.** The number of dengue cases of each station with their corresponding coefficient of variations

### 4.2 Collinearity Testing

Multicollinearity occurs when a regression model has a high intercorrelation between two or more independent variables. It will create repeated information that may lead to misleading results, especially when analysts attempt to determine which variables can be used most effectively. Moreover, multicollinearity yields a broader confidence interval and produces a less trustworthy probability. The Spearman's rho correlation coefficient is computed between the climate variables. Based on Table 2, temperature and humidity are significantly negatively correlated, indicating that these two variables may be collinear. Hence, two separate models are considered to overcome the existence of collinearity among regressors: the first model without temperature, $X_1$, and the second model without humidity, $X_2$.

**Table 2**
Spearman rank correlation coefficient

| Variables | $Y$ | $X_1$ | $X_2$ | $X_3$ | $X_4$ |
|---|---|---|---|---|---|
| $Y$ | 1.0000 | | | | |
| $X_1$ | 0.5173 | 1.0000 | | | |
| | (0.1212) | | | | |
| $X_2$ | -0.6454 | -0.7329 | 1.0000 | | |
| | **(0.0212)** | **(0.0067)** | | | |
| $X_3$ | 0.3658 | 0.3931 | -0.2227 | 1.0000 | |
| | (0.1236) | (0.2062) | (0.4866) | | |
| $X_4$ | -0.2605 | 0.2870 | -0.1149 | 0.5253 | 1.0000 |
| | (0.9462) | (0.3657) | (0.7223) | (0.0794) | |

### 4.3 Accessing the Significance of the Parameters of Generalised Linear Model

Next, the study will identify the key variables influencing dengue incidence. Given the count nature of the dengue data, the Generalised Linear Model (GLM) with a Poisson family was applied to develop two distinct models. Model 1 excludes temperature as a predictor, while Model 2 excludes humidity. Each model is independently examined throughout the study to provide comparative insights into the influence of these variables on dengue cases. The model equation of GLM Poisson and the corresponding AIC values are given as

$$\begin{aligned}&\text{Model 1: } Y_i = 22.058 - 0.188X_2 + 0.355X_3 - 1.472X_4 \quad \text{AIC}_1 = \mathbf{2891.1}\\&\text{Model 2: } Y_i = -11.633 + 0.696X_1 + 0.404X_3 - 1.939X_4 \quad \text{AIC}_2 = 2995.2\end{aligned} \tag{15}$$

All climate variables of each model are found to be significant for dengue incidence but with different magnitudes and signs. Since the data series showed an overdispersion pattern, a GLM negative binomial was implemented. The model equation of the GLM negative binomial can be written as

$$\begin{aligned}&\text{Model 1: } Y_i = 20.561 - 0.166X_2 + 0.289X_3 - 1.327X_4 \quad \text{AIC}_1 = \mathbf{195.8}\\&\text{Model 2: } Y_i = -9.164 + 0.601X_1 + 0.318X_3 - 1.439X_4 \quad \text{AIC}_2 = 199.2\end{aligned} \tag{16}$$

Based on a comparison between Model 1 and 2, the exclusion of temperature in Model 1 provides a better model fit (lower AIC), implying that temperature might have less predictive value than other variables when considered in conjunction with dengue cases. GLM with a negative binomial is a better approach than GLM with a Poisson, as the negative binomial distribution can deal with overdispersion in the data series. Besides, GLM with a negative binomial attained the lowest AIC values.

It appears that humidity ($X_2$) and wind speed ($X_4$) have a negative correlation with dengue cases, according to Eq. (15) and (16). A one-unit increase in humidity and wind speed leads to a decrease in the expected number of dengue cases. On the other hand, a one-unit increase in rainfall ($X_3$) leads to an increase in dengue cases. Similarly, a one-unit increase in temperature ($X_1$) will stimulate the growth and survival of Aedes. The coefficient for wind speed ($X_4$) is negative in both models but more substantial in Model 2, indicating a strong protective influence against dengue cases, especially when humidity is excluded. This comparative analysis shows how temperature and humidity influence dengue cases, with humidity's inclusion potentially resulting in a better model fit.

### *4.4 Residuals Diagnostic Checking*

Next, we performed a comprehensive residual analysis for each Generalised Linear Model (GLM) to assess the assumptions underlying the model's validity and identify potential issues with model fit. Specifically, we evaluated the residuals using three diagnostic tests: the Jarque-Bera test for normality, Moran's test for spatial autocorrelation, and the Breusch-Pagan test for homoscedasticity.

The Jarque-Bera test results indicated that the residuals of both models adhered to the assumption of normality, as they did not deviate significantly from a normal distribution. The results indicate a well-distributed distribution of the models' error terms, satisfying the normality assumption. However, Moran's test, which assesses spatial autocorrelation, revealed statistically significant levels of spatial dependency in the residuals. The spatial clustering of the residuals, which violates the assumption of independence, suggests the presence of unaccounted-for spatial effects in the data. Similarly, the Breusch-Pagan test, which assesses the homoscedasticity or constant variance of residuals, revealed that both models did not meet this assumption. The test results indicate heteroscedasticity, implying that the variance of residuals is not constant across observations, potentially leading to inefficiencies in parameter estimates.

Because of these results, the residuals' autocorrelation and heteroscedasticity show that a standard GLM is not quite right for this spatial dataset. As a result, a geographically weighted regression model is a more suitable approach for this series. GWR takes into account differences in local space by letting parameter estimates change from place to place. This fixes the problems with spatial autocorrelation and non-stationarity found in the residual diagnostics. This approach adequately models spatial heterogeneity, resulting in more robust and accurate insights from the data. Therefore, we propose a generalised geographically weighted regression for further analysis.

### 4.5 Spatial Weighting Functions of Generalised Geographically Weighted Regression

Several weighting functions can be used to calculate the model's parameter estimations of GWR. Table 3 shows the result of weighting function data with the bandwidth (*b*) values and the optimal AIC values. The best weighting function has the smallest bandwidth value (*b*) and the optimal AIC values. The results indicated that the best weighting function for the GWR model is an adaptive Gaussian function based on the minimum AIC.

**Table 3**
Results of bandwidth (b) values with optimal AIC values

| Model 1 | | |
|---|---|---|
| Weighting Function | Bandwidth (*b*) | AIC |
| Fixed Gaussian | 1.9659 | 50.6901 |
| Fixed Bisquare | 5.4086 | 42.6485 |
| **Adaptive Gaussian** | **0.6909** | **42.5823** |
| Adaptive Bisquare | 0.9451 | 66.5791 |
| Model 2 | | |
| Weighting Function | Bandwidth (*b*) | AIC |
| Fixed Gaussian | 1.6936 | 60.5967 |
| Fixed Bisquare | 4.9685 | 58.7625 |
| **Adaptive Gaussian** | **0.3292** | **57.8621** |
| Adaptive Bisquare | 0.9999 | 57.9247 |

The Gaussian function can be written as follows

$$\text{Model 1}: W_{ij} = exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{0.6909}\right)^2\right)$$

$$\text{(17)Model 2}: W_{ij} = exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{0.3292}\right)^2\right) \quad (18)$$

### 4.6 Generalised Geographically Weighted Regression Poisson and Negative Binomial

The study used the adaptive bandwidth parameters listed in Table 3 and the geographically weighted Poisson regression (GWPR) model to examine how climate factors affect the number of dengue cases in Peninsular Malaysia. This study also employed the geographically weighted negative binomial regression (GWNBR) model, as the data series showed signs of overdispersion. Both GGWR models were chosen due to their capacity to capture spatial heterogeneity, a critical factor when assessing health outcomes that may vary according to localised climate conditions and spatial distribution of cases.

Table 4 summarises the statistics of each variable of each model. For Model 1, the coefficient for humidity ($X_2$) is consistently negative across the quartiles, supporting the inverse relationship between humidity and dengue cases. The closed values indicate low variability within geographical locations, yet a strong and stable negative influence on dengue cases. The positive coefficient for rainfall ($X_3$) suggests a direct positive influence on dengue cases. The median at 0.3205 shows moderate variability, indicating that its effect is slightly less stable than that of humidity, but it remains positive across the range. The coefficient for wind speed ($X_4$) has a wider range, suggesting higher variability across the locations in its effect on dengue cases. However, it has a negative effect on dengue cases. The positive temperature coefficient ($X_1$) in Model 2 suggests that temperature positively affects dengue cases. However, the temperature varies across different locations in comparison to other variables. In Model 2, rainfall ($X_3$) also has a positive influence on dengue cases, but it varies more when humidity is excluded from the model. Model 2 also shows that the coefficient of wind speed ($X_4$) spans from negative to slightly positive. This suggests that wind speed is more variable in its effect on dengue cases when humidity is excluded, with some estimates even indicating a slight positive influence.

**Table 4**
Summary of the GGWR models

| Model | Variables | Min | 1st Quartile | Median | 3rd Quartile | Max |
|---|---|---|---|---|---|---|
| 1 | (Intercept) | 21.3349 | 21.7449 | 22.3721 | 23.5739 | 23.7851 |
| | $X_2$ | -0.2274 | -0.2223 | -0.1982 | -0.1844 | -0.1784 |
| | $X_3$ | 0.2943 | 0.3024 | 0.3205 | 0.3644 | 0.3706 |
| | $X_4$ | -1.2208 | -1.1780 | -1.0111 | -0.7327 | -0.6703 |
| 2 | (Intercept) | -18.3230 | -16.4894 | -15.2473 | -12.9920 | -11.9516 |
| | $X_1$ | 0.6509 | 0.6799 | 0.7967 | 0.8735 | 1.0046 |
| | $X_3$ | 0.2075 | 0.2330 | 0.2877 | 0.3960 | 0.4732 |
| | $X_4$ | -2.1620 | -1.8732 | -1.3137 | -0.5059 | 0.0341 |

Table 5 provides a comparative overview of the model results based on the mean square error (MSE) and root mean square error (RMSE) as key performance metrics. The comparative analysis highlights that the GWPR and GWNBR models significantly outperform traditional global models, such as the generalised linear model with Poisson and negative binomial specifications. Based on the findings, localised models (GWPR and GWNBR) offer improved accuracy and predictive performance for modelling relationships that consist of geographically or spatially varying data. Hence, the GGWR model is the best for determining the relationship between climate factors and dengue cases in Peninsular Malaysia.

**Table 5**
Model Evaluation of GGWR and GLM using MSE and RMSE

| | GWPR | | GLM Poisson | |
|---|---|---|---|---|
| Model | MSE | RMSE | MSE | RMSE |
| 1 | 0.2148 | 0.4635 | 231.52 | 15.22 |
| 2 | 0.1816 | 0.4261 | 394.48 | 19.86 |
| | GWNBR | | GLM Negative Binomial | |
| Model | MSE | RMSE | MSE | RMSE |
| 1 | 0.1684 | 0.4104 | 1.05 | 1.02 |
| 2 | 0.1358 | 0.3686 | 1.06 | 1.03 |

## 5. Conclusion

This study analyses the potential of geographically weighted regression as an alternative statistical model to examine the relationship between the variables that vary across locations. GWR can theoretically integrate geographical location, altitude, and other spatial estimates and reflect the non-stationary spatial relationship between these factors. By considering the spatial locations of the variables, the GWR provides the best way to describe the relationship between dengue incidence and climate variables. However, due to the categorisation of dengue series as count data, the hybrid model of GLM and GWR is more appropriate for modelling these types of series.

Consequently, the findings suggest that the geographically weighted generalised regression model is optimal for analysing climate disease relationships in this context. The adaptive bandwidth mechanism of the GGWR model allows it to accurately model the data by calibrating local parameter estimates that reflect the unique climatic and demographic factors affecting dengue transmission in each subregion. The geographically weighted regression with a negative binomial performed better than the geographically weighted Poisson regression in explaining the effects of climate variables on dengue incidence. These localised parameter estimates provide a refined understanding of spatial patterns, offering valuable insights for public health interventions to control and predict dengue outbreaks in regions with diverse climatic conditions, such as Peninsular Malaysia. Researchers have found that climate variables significantly impact the pattern of dengue cases. Temperature and rainfall positively affect dengue cases, while humidity and wind speed negatively influence them. In Model 1, humidity consistently shows a negative impact compared to temperature, potentially indicating a more consistent reduction in dengue cases when under control.

For future studies, it would be better to have more variations of variables that link to the nature of dengue incidence, such as duration of the infection and time for recovery, environmental conditions, medical intensive care unit cases, and hospitalised cases. Adding more variables allows for the acquisition and examination of more information. This work improves our understanding of the effects of climate conditions across Malaysia on the spatial variations of this epidemic. It helps local hygienic and environmental authorities make targeted joint interventions to prevent and control the epidemic.

## Acknowledgment

The authors would like to express their gratitude to the Ministry of Higher Education of Malaysia (MOHE) for the funding given under the Fundamental Research Grant Scheme (FRGS/1/2020/STG06/UTM/02/3) under vote 5F311. We are also grateful to Universiti Teknologi Malaysia for supporting this project.

## References

[1] Aye, Khin Saw, Komgrid Charngkaew, Ne Win, Kyaw Zin Wai, Kyaw Moe, Nuntaya Punyadee, Somchai Thiemmeca, Aroonroong Suttitheptumrong , Sanya Sukpanichnant, Malasit Prida, and Scott B Halstead. "Pathologic highlights of dengue hemorrhagic fever in 13 autopsy cases from Myanmar." *Human Pathology* 45, no.6 (2014): 1221–1233. https://doi.org/10.1016/j.humpath.2014.01.022

[2] Fagbami, Ademola H, and Anyebe B Onoja. "Dengue haemorrhagic fever: An emerging disease in Nigeria, West Africa." *Journal of Infection and Public Health* 11 no. 6 (2018): 57-762. https://doi.org/10.1016/j.jiph.2018.04.014

[3] Pasaribu, Ayodhia Pitaloka, Tsheten Tsheten, Muhammad Yamin, Yulia Maryani, Fahmi Fahmi, Archie C A Clements, Darren J Gray and Kinley Wangdi." Spatio-temporal patterns of dengue incidence in Medan City, North Sumatera, Indonesia." *Tropical Medicine and Infectious Disease* 6 no. 1 (2021): 30. https://doi.org/10.3390/tropicalmed6010030

[4] World Health Organization (WHO) 23 April 2024. https://www.who.int/news-room/fact-sheets/detail/dengue and severe-dengue

[5] Hii, Yien Ling, Rafdzah Ahmad Zaki, Nasrin Aghamohammadi, Joacim Rocklöv. "Research on Climate and Dengue in Malaysia: A Systematic Review." *Current Environmental Health Reports* 3 no. 1 (2016): 81–90. https://doi.org/10.1007/s40572-016-0078-z

[6] Jayaraj, Vivek Jason , Richard Avoi, Navindran Gopalakrishnan, Dhesi Baha Raja, and Yusri Umasa. "Developing a dengue prediction model based on climate in Tawau, Malaysia." *Acta Tropica* 197:105055. https://doi.org/10.1016/j.actatropica.2019.105055

[7] Stanaway, Jeffrey D, Donald S Shepard, Eduardo A Undurraga, Yara A Halasa, Luc E Coffeng, Oliver J Brady, Simon I Hay, Neeraj Bedi, Isabela M Bensenor, Carlos A Castañeda-Orjuela, Ting-Wu Chuang, Katherine B Gibney, Ziad A Memish, Anwar Rafay, Kingsley N Ukwaja, Naohiro Yonemoto, and Christopher JL Murray. "The global burden of dengue: an analysis from the Global Burden of Disease Study 2013." *The Lancet Infectious Diseases* 16 no. 6 (2016): 712–723. https://doi.org/10.1016/S1473-3099(16)00026-8

[8] Andrew K. Githeko, Steve W. Lindsay, Ulisses E. Confalonieri, and Jonathan A. Patz. "Climate change and vector-borne diseases: A regional analysis." *Bulletin of the World Health Organization* 78 no. 9 (2000): 1136–1147. PMID: 11019462; PMCID: PMC2560843.

[9] Ten Bosch, Quirine A, Brajendra K Singh, Muhammad R A Hassan, Dave D Chadee and Edwin Michael. "The Role of Serotype Interactions and Seasonality in Dengue Model Selection and Control: Insights from a Pattern Matching Approach." *PLOS Neglected Tropical Diseases* 10 no. 5 (2016): 1–25. https://doi.org/10.1371/journal.pntd.0004680

[10] Tran, Annelise, Grégory L'Ambert,Guillaume Lacour,Romain Benoît,Marie Demarchi,Myriam Cros,Priscilla Cailly, Mélaine Aubry-Kientz,Thomas Balenghien, and Pauline Ezanno. "A rainfall- and temperature- driven abundance model for Aedes albopictus populations." *International Journal of Environmental Research and Public Health* 10 no. 5 (2013): 1698–719. https://doi.org/10.3390/ijerph10051698

[11] Wu, Xiaoxu, Yongmei Lu, Sen Zhou, Lifan Chen, and Bing Xu. "Impact of climate change on human infectious diseases: Empirical evidence and human adaptation." *Environment International* 86 (2016): 14-23. https://doi.org/10.1016/j.envint.2015.09.007

[12] Lee, Hyojung, Jung Eun Kim ,Sunmi Lee, and Chang Hyeong Lee." Potential effects of climate change on dengue transmission dynamics in Korea." *Plos One* 13 no.6 (2018): e0199205. https://doi.org/10.1371/journal.pone.0199205

[13] Butterworth, Melinda K, Cory W Morin, and Andrew C Comrie. "An Analysis of the Potential Impact of Climate Change on Dengue Transmission in the Southeastern United States. *Environmental Health Perspective* 125 no.4 (2017): 579-585. https://doi.org/10.1289/EHP218

[14] Mwanyika, Gaspary, Leonard E G Mboera, Sima Rugarabamu, Baraka Ngingo, Calvin Sindato, Julius J Lutwama, Janusz T Paweska, and Gerald Misinzo. "Dengue virus infection and associated risk factors in Africa: A systematic review and meta-analysis." *Viruses* 13 no. 4 (2021):536. https://doi.org/10.3390/v13040536

[15] Caldwell, Jamie M, A. Desiree LaBeaud, Eric F. Lambin, Anna M. Stewart-Ibarra, Bryson A. Ndenga, Francis M. Mutuku, Amy R. Krystosik, Efraín Beltrán Ayala, Assaf Anyamba, Mercy J. Borbor-Cordova, Richard Damoah, Elysse N. Grossi-Soyster, Froilán Heras Heras, Harun N. Ngugi, Sadie J. Ryan, Melisa M. Shah, Rachel Sippy, and Erin A. Mordecai. "Climate predicts geographic and temporal variation in mosquito-borne disease dynamics on two continents." *Nature Communications* 12 no. 1 (2021):1233. https://doi.org/10.1038/s41467-021-21496-7

[16] Liao, Hongyan, Christopher J. Lyon, Binwu Ying and Tony Hu. "Climate change, its impact on emerging infectious diseases and new technologies to combat the challenge." *Emerging Microbes & Infections* 13 (2024): 2356143. https://doi.org/10.1080/22221751.2024.2356143

[17] Vijaya Kumar, Joyceni Cross, Erwan Hafizi Kasiman. "Flood Mitigation for Sungai Pinji, Ulu Kinta, Perak Using HEC-HMS and HEC-RAS Model." *Frontiers in Water and Environment* 2 no.1 (2024): 10-27.

[18] Grobusch, Lena C, and Martin P. Grobusch. "A hot topic at the environment-health nexus: investigating the impact of climate change on infectious diseases." *International Journal of Infectious Disease* 116 (2022): 7-9. https://doi.org/10.1016/j.ijid.2021.12.350

[19] Hong, Zhi-Min, Hu-Hu Wang,Yan-Juan Wang, and Wen-Rui Wang. "Spatiotemporal analysis of hand, foot and mouth disease data using time-lag geographically-weighted regression." *Geospatial Health* 15 no. 2(2020). https://doi.org/10.4081/gh.2020.849

[20] Sun, Jingtao, Sensen Wu, Zhen Yan, Yadong Li, Cheng Yan, Feng Zhang, Renyi Liu, and Zhenhong Du. "Using geographically weighted regression to study the seasonal influence of potential risk factors on the incidence of HFMD on the Chinese mainland." *ISPRS International Journal of Geo-Information* 10, no. 7 (2021): 448. https://doi.org/10.3390/ijgi10070448

[21] Sulekan, Ayuna, Jamaludin Suhaila, and Nurmarni Athirah Abdul Wahid. "Geographically weighted regression on dengue epidemic in peninsular Malaysia." In *Journal of Physics: Conference Series*, vol. 1988, no. 1, p. 012099. IOP Publishing, 2021. https://doi.org/10.1088/1742-6596/1988/1/012099

[22] Sumanasinghe, Nirosha, Armin Mikler, Chetan Tiwari and Jayantha Muthukudag. Geo-statistical Dengue Risk Model using GIS techniques to identify the risk prone areas by linking rainfall and population density factors in Sri Lanka. *Ceylon Journal of Science* 45 no. 3(2016): 39-46. https://doi.org/10.4038/cjs.v45i3.7399

[23] Isazade, Vahid ,Abdul Baser Qasimi, Pinliang Dong, Gordana Kaplan, and Esmail Isazade. "Integration of Moran's I, geographically weighted regression (GWR), and ordinary least square (OLS) models in spatiotemporal modeling of COVID-19 outbreak in Qom and Mazandaran Provinces, Iran." *Modeling Earth Systems and Environment* 9 (2023): 3923–3937. https://doi.org/10.1007/s40808-023-01729-y

[24] Jasim, Ihsan Abbas, Moheb Kamil Fileeh, Mustafa A. Ebrahhem, Laheab A. Al-Maliki, Sohaib K. Al-Mamoori and Nadhir Al-Ansari. "Geographically weighted regression model for physical, social, and economic factors affecting the COVID-19 pandemic spreading." *Environmental Science and Pollution Research* 29 (2022): 51507–51520. https://doi.org/10.1007/s11356-022-18564-w

[25] Wu, Xiu, and Jinting Zhang. "Exploration of spatial-temporal varying impacts on COVID-19 cumulative case in Texas using geographically weighted regression (GWR). "*Environmental Science and Pollution Research* 28 (2021): 43732–43746. https://doi.org/10.1007/s11356-021-13653-8

[26] Zhang, Huihui, Yini Liu, Fangyao Chen, Baibing Mi, Lingxia Zeng and Leilei Pei. "The effect of sociodemographic factors on COVID-19 incidence of 342 cities in China: a geographically weighted regression model analysis." *BMC Infectious Diseases* 21 (2021): 428. https://doi.org/10.1186/s12879-021-06128-1

[27] Wang, Fei, Yiyi Zhu, Hengduan Zhang, Junhua Fan, Peien Leng, Ji Zhou, Shenjun Yao, Dandan Yang, Yao Liu, Jingjing Wang, Juanyi Yao, Yibin Zhou, and Tongyan Zhao. "Spatial and temporal analyses of the influences of meteorological and environmental factors on Aedes albopictus (Diptera:Culicidae) population dynamics during the peak abundance period at a city scale." *Acta Tropica* 245 (2023): 106964. https://doi.org/10.1016/j.actatropica.2023.106964

[28] Hazrin, H., Y. Fadhli, A. Tahir, J. Safurah, M. N. Kamaliah, and M. Y. Noraini. "Spatial patterns of health clinic in Malaysia." *Health* 2013 (2013). https://doi.org/10.4236/health.2013.512287

[29] Eboy, Oliver Valentine, Narimah Samat. "Modeling property rating valuation using Geographical Weighted Regression (GWR) and Spatial Regression Model (SRM): The case of Kota Kinabalu, Sabah." *Geografia-Malaysian Journal of Society and Spac* 11, no. 11 (2015): 98-109. https://doi.org/10.1183/20734735.015614

[30] Jamhuri, J., B. M. S. Azhar, C. L. Puan, and K. Norizah. "GWR-PM-Spatial variation relationship analysis with Geographically Weighted Regression (GWR)-An application at Peninsular Malaysia." In *IOP Conference Series: Earth and Environmental Science*, vol. 37, no. 1, p. 012032. IOP Publishing, 2016. https://doi.org/10.1088/1755-1315/37/1/012032

[31] Vaziri, Mehrdad, Michael Acheampong, Joni Downs, and Rafee Majid Mohammad. "Poverty as a function of space: understanding the spatial configuration of poverty in Malaysia for Sustainable Development Goal number one." *GeoJournal* 84, no.5 (2019): 1317–1336. https://doi.org/10.1007/s10708-018-9926-8

[32] Nelder, A. J. A, and R. W. M Wedderburn. "Generalized Linear Models." *Journal of the Royal Statistical Society. Series A* 135 no. 3 (1972): 370-384. https://doi.org/10.2307/2344614

[33] Breslow, N. E. "Extra-Poisson Variation in Log-Linear Models." *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 33 no. 1 (1984): 38-44. https://doi.org/10.2307/2347661

[34] Nakaya, T, Fotheringham A. S, Brunsdon C, and Charlton M. "Geographically weighted Poisson regression for disease association mapping." *Statistics in Medicine* 24 no. 17 (2005): 2695–2717. https://doi.org/10.1002/sim.2129

[35] Silva, Alan Ricardo da, and Thais Carvalho Valadares Rodrigues. "Geographically Weighted Negative Binomial Regression incorporating overdispersion." *Statistics and Computing* 24 no. 5 (2014): 769–783. https://doi.org/10.1007/s11222-013-9401-9