Journal of Quality Measurement and Analysis *JQMA* 20(3) 2024, 163-180
e-ISSN: 2600-8602 *http://www.ukm.my/jqma*
https://doi.org/10.17576/jqma.2003.2024.12

A GENERALIZED MODELLING APPROACH TO ASSESS CLIMATE INFLUENCES ON HAND, FOOT, AND MOUTH DISEASE IN EAST COAST MALAYSIA

(Pendekatan Pemodelan Teritlak untuk Mengkaji Pengaruh Iklim Terhadap Penyakit Tangan, Kaki dan Mulut di Pantai Timur Malaysia)

NURMARNI ATHIRAH ABDUL WAHID*, JAMALUDIN SUHAILA
& HALIZA ABD RAHMAN

ABSTRACT

Hand, foot, and mouth disease (HFMD) outbreaks in Asia have increased since the late 1990s, causing severe and often fatal consequences. Several statistical approaches, such as Generalized Linear Models (GLM) and Generalized Additive Models (GAM), have been used in numerous studies to examine the association between climate factors and HFMD cases. However, the results vary by country. In Malaysia, these issues require further research, as there are only a few studies employing GLM and GAM approaches that focus on HFMD cases and climate factors, particularly in the East Coast region. Therefore, this study explores the association between HFMD and climate factors on Malaysia's East Coast using GLM and GAM with Negative Binomial to identify the best model for interpreting HFMD cases. The findings show that climate factors affect HFMD differently across states in East Coast Malaysia. The results show that the GAM Negative Binomial model best represents these issues. The temperatures between 26°C and 28°C will decrease the risk of HFMD cases in Pahang over the next two weeks. Besides, temperatures ranging from 25 to 27°C and 28.5 to 30°C significantly increased HFMD risk in Terengganu over the next two weeks. Nevertheless, Kelantan found no correlation between climate and HFMD. These findings can help local health authorities in developing a climate-based early warning system to minimize HFMD outbreaks in Malaysia's East Coast Region.

Keywords: generalized linear model; generalized additive model; climate change; disease

ABSTRAK

Penyakit tangan, kaki, dan mulut (HFMD) telah meningkat di negara Asia sejak akhir 1990-an, membawa kepada kesan negatif dan sering kali membawa maut. Beberapa pendekatan statistik, seperti Model Linear Teritlak (GLM) dan Model Tambahan Teritlak (GAM), telah digunakan dalam banyak kajian untuk mengkaji hubungan antara faktor iklim dan kes HFMD. Namun, hasil setiap kajian berbeza mengikut negara. Di Malaysia, isu ini memerlukan penyelidikan lanjut kerana hanya ada beberapa kajian yang menggunakan pendekatan GLM dan GAM untuk kes HFMD dan faktor iklim, terutama di Pantai Timur. Pendekatan GLM dan GAM Binomial Negatif digunakan untuk menentukan model terbaik bagi mewakili kes HFMD dan faktor iklim di Pantai Timur. Hasil kajian ini menunjukkan bahawa faktor iklim mempengaruhi kes penyakit HFMD secara berbeza mengikut negeri di Pantai Timur Malaysia. Kajian ini menunjukkan bahawa model GAM Binomial Negatif adalah yang terbaik untuk mentafsir isu ini. Suhu antara 26°C dan 28°C akan mengurangkan risiko kes HFMD di Pahang dalam tempoh dua minggu akan datang. Selain itu, suhu antara 25 hingga 27°C dan 28.5 hingga 30°C secara signifikan meningkatkan risiko kes HFMD di Terengganu dalam tempoh dua minggu akan datang. Manakala, tiada korelasi ditemui antara iklim dan kes penyakit HFMD di Kelantan. Hasil ini dapat membantu pihak berkuasa kesihatan tempatan membangunkan sistem amaran awal penyakit berasaskan perubahan iklim untuk mengurangkan wabak HFMD di Pantai Timur Malaysia.

Kata kunci: model linear teritlak; model tambahan teritlak; perubahan iklim; penyakit

1. Introduction

Hand, foot, and mouth disease (HFMD) is a common viral illness that rapidly spreads worldwide. In early 1957, Seddon made the first clinical diagnosis of HFMD in New Zealand (Flewett *et al.* 1963). The disease was named by Thomas Henry Flewett after a previous outbreak of HFMD in 1960 (Alsop *et al.* 1960). HFMD is caused by enteroviruses, the most common of which are Coxsackie virus A16 and Enterovirus 71. Between late June and July 1957, a total of 60 cases of HFMD caused by Coxsackievirus A16 were identified in Toronto, Canada (Robinson *et al.* 1958). The other HFMD virus, Enterovirus 71 (EV71), was found in 1969 in California (Schmidt *et al.* 1974). EV71 has been linked to sporadic outbreaks of HFMD as well as severe neurological conditions, including meningitis, encephalitis, and acute flaccid paralysis (AFP) (Chen *et al.* 2007).

This disease primarily affects children under the age of five, but it can also affect adolescents (Melnick 1984). Transmission of viruses occurs through saliva, blister fluid, and patient feces. The infection exhibits mild symptoms, initially showing as a rise in body temperature and the formation of blisters on the hands, feet, mouth, and tongue (Ministry of Health Malaysia 2012). The first sign of these symptoms will occur within a period of three to seven days following the infection. Initially, an individual may encounter a mild fever, poor appetite, swollen throat, and a general feeling of discomfort, commonly referred to as malaise (US Centers for Disease Control and Prevention 2021). HFMD caused by EV71 can result in severe neurological complications compared to other enterovirus serotypes, such as brainstem encephalitis and acute flaccid paralysis (Ooi *et al*. 2010; Wang & Liu, 2009). The majority of patients with neurological complications from EV71 infection are children, particularly those under the age of five (Chen *et al*. 2007; Wang *et al*. 1999).

The outbreak of HFMD in various countries, including China, Japan, Hong Kong, the Republic of Korea, Singapore, Thailand, Taiwan, Vietnam, the United States of America, Europe, Brazil, and Malaysia, has showed a significant challenge to global public health. A growing incidence of HFMD outbreaks in Asian countries has been observed in the past decade. These outbreaks have predominantly impacted children and have caused severe complications that have estimated thousands of fatalities (Xing *et al.* 2014). A significant HFMD outbreak in Malaysia triggered a series of outbreaks throughout the Asia-Pacific region. The HFMD outbreak, which originated in Sarawak in early April 1997, was mainly attributed to the EV71 infection. By June 1997, the disease had subsequently disseminated to Peninsular Malaysia (World Health Organization 2011). Despite having the most developed healthcare systems and modern technology, Malaysia has experienced the widespread spread of this disease.

Numerous studies have provided evidence indicating that climate change has significant effects on various health conditions including HFMD (Guo *et al.* 2023; Ibrahim *et al.* 2024; Wang *et al.* 2023; Yang *et al.* 2024). The findings indicate that there are variations in occurrences across different countries. In Rizhao, China (Wu *et al.* 2014) and South Korea (Kim *et al.* 2016), researchers have confirmed a significant non-linear relationship between humidity and HFMD cases. In contrast, studies in Taiwan found a positive linear correlation between the two variables (Chang *et al.* 2012). Furthermore, studies conducted in Hong Kong and Shandong Province, China indicate that increase in wind speeds enhances the risk of HFMD (Ma *et al.* 2010; Liao *et al.* 2015). Nevertheless, only a limited number of studies have provided evidence to support these claims. In addition, despite the fact that a significant correlation between rainfall and HFMD cases has been established in Singapore by Hii *et al.*

(2011), this finding contradicts a study in Japan, which found no evidence of correlation (Onozuka & Hashizume 2011). In summary, the results conducted in those countries were inconsistent. Therefore, it is important to also examine these potential influencing factors within the context of Malaysia.

In a study involving HFMD and climate factors, several statistical approaches were used. This includes various statistical techniques such as correlation analysis (Leong *et al.* 2011), time-series models (Du *et al.* 2017), classification and regression tree models (Du *et al.* 2016), and generalized modelling techniques (Onozuka & Hashizume 2011; Li *et al.* 2014; Kim *et al.* 2016; Chen *et al.* 2019). Prior studies have mostly employed generalized modelling techniques to explore the association between cases of HFMD and climate factors. These models include the Generalized Linear Model (GLM) and the Generalized Additive Model (GAM). The GLM is a widely employed modelling technique that can incorporate multiple statistical models, such as linear regression, logistic regression, and Poisson regression (Nelder & Wedderburn 1972). Furthermore, GLM allows for an extension of linear modelling concepts to involve a wider variety of response types, including count data and binary responses. Two studies used a generalized linear model (GLM) with a negative binomial distribution to examine the relationship between HFMD and climate change (Onozuka & Hashizume 2011; Li *et al.* 2014). Both studies employed negative binomial regression to address the issue of overdispersion in the datasets. However, the GLM modeling technique has certain limitations. One such limitation is its inability to adequately capture the complex non-linear and non-monotonic relationships that frequently occur in data structures.

Following that, several researchers have proposed using the GAM to model the non-linear effect with a non-Gaussian response. The GAM approach was used by several researchers to examine the non-linear connection that exists between HFMD incidence and climate factors (Chen *et al.* 2014; Kim *et al.* 2016; Chen *et al.* 2019). As stated by Hastie and Tibshirani (1995), the GAM approach relies on two key assumptions: the function is additive and the component is smooth. This modelling approach is a semi-parametric extension of GLM, which provides for flexible handling of non-linear covariate effects. For example, a study in China used Poisson auto-regression combined with GAM modeling to determine the non-linear relationship between climate variables and the incidence of HFMD (Chen *et al.* 2014). A penalized smoothing spline of time with six degrees of freedom per annum was incorporated into the model to account for the long-term trend and seasonal cycle. As a result, the study found that meteorological factors significantly contribute to HFMD transmission in Guangzhou, China (*Chen et al. 2014*).

Another study in China used a GAM with negative binomial family to evaluate the non-linear link between the weekly number of HFMD cases with the average temperature, relative humidity, and Baidu index (BDI) in two cities in China (Chen *et al.* 2019). The study shows that the GAM approach is useful in determining the exposure-response relationship for various types of data, particularly when exploring non-parametric relationships. In Malaysia, there are only a few studies focusing on HFMD cases and climate factors using the GLM and GAM approaches, particularly in the East Coast region. To the best of our knowledge, this is the only study that specifically examines the East Coast. Prior studies were limited to the Selangor central region of Malaysia (Wahid *et al.* 2021) and the country as a whole (Wahid *et al.* 2020). Therefore, the objective of this study is to explores the association between HFMD and climate factors along the East Coast of Malaysia through the use of the generalized approaches to modelling GLM and GAM. In the future, this research could assist health policymakers forecast future outbreaks, raise public awareness, and implement an effective

HFMD prevention strategy. Malaysians, particularly those residing in Pahang, Kelantan, and Terengganu, may therefore take additional measures to prevent the spread of the disease.

2. Research Method

2.1. *Study area*

The East Coast of Peninsular Malaysia includes three states which are Kelantan, Terengganu, and Pahang. The region covers approximately 132,490 km^2 and accounts for nearly 40% of Malaysia's total land area. Pahang is the largest state among the three East Coast states, and it also the largest state in Peninsular Malaysia. The East Coast region is highly seasonal, with strong monsoon winds and heavy rainfall along the coast annually from November to February.

2.2. *Data collection*

The daily HFMD cases for the three East Coast states of Malaysia were obtained from the Public Sector Open Data Portal Malaysia for the years 2010 to 2016. The data was made available by the Ministry of Health of Malaysia. The collected data was then processed into weekly resolutions to ensure consistency with the existing climate data. The climate data originated from the Malaysian Meteorological Department (MetMalaysia) and included weekly temperatures (in degrees Celsius), relative humidity (as a percentage), rainfall (in millimeters), and wind speed (in meters per second). The data used in this study covers the years 2010 to 2016 to align with the HFMD data.

2.3. *Statistical methods*

The analysis in this study is divided into several sections, which include multicollinearity testing, overdispersion testing, covariate selection for the model, model development, model evaluation, and model validation. The steps for modelling the association between HFMD cases and climate factors in Malaysia's East Coast region are described in greater detail in the subsequent section. This study also takes into account the delayed effects of climate factors and incubation periods. This is due to changes in climate conditions, which may have an impact on the HFMD characteristics. Several studies claim that using climate variables with a two-week lag period enhances the risk of HFMD (Ma *et al.* 2010; Hii *et al.* 2011; Kim *et al.* 2016). It is related to the incubation period for enteroviruses as well as the possibility of parental knowledge and response to children's signs and symptoms. All statistical analyses in this study were performed using the '*mgcv*' package in R programming software.

2.3.1. *Multicollinearity test*

Multicollinearity occurs when highly correlated factors in the regression model are examined. Following that, it is expected that there will be multicollinearity between the climate variables used in this study. Multicollinearity can lead to several potential issues, such as inflated standard errors. The standard errors of the coefficient estimates increase, making it difficult to determine the individual effect of each predictor variable. This may lead to wide confidence intervals and less reliable hypothesis tests (Neter *et al.* 1996). Additionally, unstable coefficients may arise, where the estimates become very sensitive to changes in the model, such as adding or removing variables. This can result in misleading interpretations, as small changes in the data can lead to large swings in the estimated coefficients (O'Brien 2007). Furthermore, multicollinearity complicates the identification of which variables are actually significant predictors of the dependent variable, as the shared variance between predictors can obscure their individual effects (Allison 1999). Thus, this study used the variance inflation factor (VIF), corrected variance inflation factor (CVIF), and tolerance (TOL) to detect multicollinearity. The climate variables are found to be highly correlated if the VIF value is greater than 5 (Kutner *et al.* 2005), the CVIF value is greater than or equal to 10, and the TOL value is close to 0 (Marquardt 1970). The next modelling phase will eliminate the highly correlated climate variables.

2.3.2. *Overdispersion test*

The test for overdispersion is an important component of regression analysis that must be addressed. Considering that the data used in this study comprises of count data, a Poisson distribution is the best approach. Nonetheless, when using Poisson regression to model data, overdispersion is frequently observed. The standard errors could possibly be underestimated due to the overdispersion issue, leading to incorrect inferences about the regression parameters. This problem arises when the variance of the response variable exceeds its expected value (Krzanowski 1998). Therefore, an overdispersion test developed by Cameron and Trivedi (1990) was used in this study to detect these issues. The following presents the hypothesis for the test.

H_0: There is no evidence of overdispersion in the datasets
H_1: There is evidence of overdispersion in the datasets

The rejection of the null hypothesis suggests that overdispersion is present in the datasets. To address this issue, a negative binomial regression model is recommended (Breslow 1984). The negative binomial distribution includes an extra parameter that provides increased flexibility in modeling variance, which can exceed the mean. In many situations involving count data, especially when the variance is significantly greater than the mean, the Poisson distribution can result in poor model fit and biased estimates (Cameron & Trivedi 2013). By incorporating this additional parameter, the negative binomial distribution effectively accommodates the extra variation, offering a more accurate representation of the underlying structure of the data (Hilbe 2011).

2.3.3. *Stepwise covariate selection method*

The stepwise method is a hybrid of forward and backward selection processes that allows for movements in both directions and the inclusion and exclusion of variables at multiple steps. The process can start with either a backward elimination and a forward selection process (Chowdhury & Turin 2020). This study employed the forward stepwise selection method, which is commonly used in various applications (Cheong *et al.* 2013). Additionally, in terms of model interpretability, forward stepwise selection begins with no predictors and adds them incrementally based on a chosen criterion. This approach provides clearer insights into how each predictor contributes to the model. In contrast, backward elimination starts with a full model, which can make it more challenging to interpret the incremental effect of each variable (Hastie *et al.* 2009). Furthermore, forward selection tends to be less computationally intensive, particularly when dealing with a large number of predictors. By starting with an empty model, it eliminates the need to evaluate all possible models, as is required in backward selection, making it especially advantageous for high-dimensional datasets (Miller 2002). Thus, a stepwise forward selection approach was used in selecting the climate variables for the selection of the optimal model. This selection of the best model is based on the Akaike Information Criterion (AIC) criterion which results in a sparse model (Akaike 1998). The best significant model is believed to be associated with a low value of AIC. AIC is calculated using the formula:

$$\text{AIC} = 2k - 2\ln(L) \tag{1}$$

where k is the number of estimated parameters in the model and L is the maximum value of the likelihood function for the models. The covariate selection process is as follows:

(1) Develop a GLM and GAM model for each of the climate factor. The lowest AIC value indicates the best single predictor.
(2) Develop a two-predictor model by adding the best climate factor identified in the first step as the first predictor, followed by the second predictor (other climate factors). The lowest AIC value indicates the best second predictor.
(3) Continue the process until adding predictors does not result in smaller than the earlier model.
(4) The selected predictors can be found in the model with the lowest AIC value.

2.3.4. *Generalized modelling techniques*

Two distinct generalized modelling approaches, namely the GLM and GAM, were employed in this study. The following section explains the specifics of each approach. The Poisson regression technique can be used to model response variables that describe either count or discrete data (Nelder & Wedderburn 1972). As the data on HFMD cases in Malaysia consists of non-negative integers and is not normally distributed, a Poisson model will be applied to this study.

(a) Generalized Linear Model

Generalized Linear Models (GLM) were formally introduced in 1972 by John Nelder and Robert Wedderburn. GLM is a more advanced version of classical linear models that allows general linear regression to be approached using a variety of response variables, including count, binary, proportions, and continuous distribution. This GLM is regarded as a valuable and widely used method due to its ability to handle a wide range of statistical problems (Nelder & Wedderburn 1972). The general form of a GLM can be formally written as:

$$g(\mu_i) = \beta_0 + \beta_1(x_{i1}) + \ldots + \beta_p(x_{ip}) \tag{2}$$

where $i = 1, 2, 3, \ldots, n$ and $g(\mu_i)$ is a link function that relates the mean of the response variable μ_i to the linear predictor. $x_1, \ldots, x_p$ represent the independent variables, or predictors, where p is the number of predictors in the model. The symbol of β_0 is the intercept, and the coefficients are denoted as $\beta_1, \ldots, \beta_p$.

(b) Generalized Additive Model

Generalized Additive Models (GAM) are non-parametric regression models developed in 1986 by Trevor Hastie and Robert Tibshirani. This method is an extension of the GLM approach by substituting the predictor, which consists of a sum of smooth functions (Hastie & Tibshirani 1990). The GAM modelling approach is frequently employed when there is no prior basis for selecting a particular response function, such as linear, quadratic, or other (Wood 2017). Besides, the GAM approach allows for considerable flexibility in describing response predictors for both linear and non-linear relationships. The GAM requires the resolution of three issues that are absent in linear modeling. These issues include the representation of smooth functions, the controllable degree of smoothness, and the selection of the most appropriate level of smoothness in a data-driven way (Wood 2017). The general structure of a GAM can be formally written as:

$$g(\mu_i) = \beta_0 + s_1(x_{i1}) + \ldots + s_p(x_{ip}) \tag{3}$$

where $i = 1, 2, 3, \ldots, n$. β_0 is the intercept and $g(\mu_i)$ is a link function that relates the mean of the response variable μ_i to the linear predictor. $x_1, \ldots, x_p$ represents the independent variables, or predictors, where p is the number of predictors in the model. $s_1(x_{i1}) + \ldots + s_p(x_{ip})$ are smooth function of the predictors $x_1, \ldots, x_p$. These functions can be splines or other smooth functions that allow for non-linear relationships between the predictors and the response variable.

The smooth function, s, is composed of the sum of the basis function, B, which has been chosen for its convenient properties, and its corresponding regression coefficients, β (Wood 2017). The following equation represents a smooth function.

$$s_p(x_{ip}) = \sum_{i=1}^{m_p} \beta_{ip} B_{ip}(x_p) \tag{4}$$

where m_p is the number of basis function for the p-th predictor, β_{ip} are the coefficients associated with the i-th basis function for the p-th predictor, and $B_{ip}(x_p)$ are the basis function, which could be splines, polynomial function, or other smooth functions.

2.3.5. *Model development of HFMD and climate factors*

In this study, the GLM and the GAM with Poisson family are two statistical modelling approaches used to study the impact of climate factors on HFMD in Malaysia's East Coast region. For GLM and GAM models, the general equations are denoted by Eqs. (5) and (6), respectively. The primary difference between the two models lies in the fact that GLM expands upon the classical linear model by incorporating response variables from any exponential family. In contrast, GAM offers the linear predictor substantial flexibility through the inclusion of local smooth functions. Thus, the GAM method is capable of accounting for the nonlinear relationship between HFMD and climate factors in this study. The complete model for GLM and GAM employed in the analysis of each state on the East Coast is represented by the equation below.

(a) GLM full model

$$\begin{aligned} \ln[E(HFMD_i)] = {} & \beta_0 + \beta_1(\text{Temperature}_i)_{t-2} + \beta_2(\text{Humidity}_i)_{t-2} + \\ & \beta_3(\text{Rainfall}_i)_{t-2} + \beta_4(\text{Wind speed}_i)_{t-2} + \beta_5(\text{Time}_i) \end{aligned} \tag{5}$$

(b) GAM full model

$$\begin{aligned} \ln[E(HFMD_i)] = {} & \beta_0 + \text{s}(\text{Temperature}_i)_{t-2} + \text{s}(\text{Humidity}_i)_{t-2} \\ & + \text{s}(\text{Rainfall}_i)_{t-2} + \text{s}(\text{Wind speed}_i)_{t-2} \\ & + \text{s}(\text{Time}_i, \text{df} = 4/\text{year}) \end{aligned} \tag{6}$$

In these two equations, i represents the week of HFMD cases, $i = 1, 2, 3, \dots, 365$, β_0 is the intercept, and $\mathrm{s(variable)}_{t-2}$ denotes the effect of all climate variables over a two-week period (lag time). The fundamental difference between these two models comes from the substitution of the parameter in Eq. (5) with a smoothing function $s(.)$, which can be seen in Eq. (6) above. In order to control the seasonal impact and long-term trend of HFMD in weekly cases, a time variable was incorporated into both models. Seasonality and the long-term trend pertain to time series data that exhibits a consistent upward or downward pattern, often involving a number of years (Siegel 2016). For GAM, a smoothing spline of time with four degrees of freedom (df) per year was applied. In the GAM model, a cyclic cubic regression spline is used as a smoothing spline, which is particularly suitable for cyclical or seasonal data (Underwood 2009; Sigauke *et al.* 2019). As mentioned previously, this study employed the cyclic cubic regression spline due to the seasonal patterns observed in the Malaysian HFMD and climate data.

In this study, both GLM and GAM models employed a log link function for the expected value of the response variable (HFMD cases). This choice ensures that the predicted values remain positive, which is essential when modeling counts or other strictly positive outcomes (McCullagh 1989). The log link function also enables a multiplicative relationship between the predictors and the response variable, allowing for easier interpretation of the effects of predictors as proportional changes rather than absolute changes (Hilbe 2011). Additionally, the log link function stabilizes variance in situations where the response variable exhibits positive skew, facilitating more robust modeling of complex relationships within the data (Hardin & Hilbe 2007).

2.3.6. *Criteria for the best model selection*

In evaluating the performance of statistical models, root mean square error (RMSE) and mean absolute error (MAE) have been widely used in meteorology and climate research studies (Chai & Draxler 2014). Following that, the RMSE and MAE values were used to measure the accuracy of each model in this study. The model with the best performance is the indicated by the estimated model that has the lowest values for both criteria. The RMSE and MAE can be calculated using the following formulas.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (7)$$

$$\mathrm{MAE} = \frac{1}{n}\left|\sum_{i=1}^{n} y_i - \hat{y}_i\right| \qquad (8)$$

2.3.7. *Model validation*

The best approach for time series cross-validation is the rolling basis cross-validation introduced by Hyndman. This technique of cross-validation is applicable when dealing with time-series data and dependent observations (Hyndman 2014). A rolling basis of cross-validation was employed in this analysis process. The process of cross-validation is described in greater detail below.

(1) Split the HFMD data into subsets based on years from 2010 to 2016.
(2) Start the rolling basis cross-validation using the observation of the HFMD data in 2010 for training purposes and test the model using the observation in 2011.

(3) Calculate the accuracy of the testing data for the year 2011 using the RMSE.
(4) Use the observation from 2010 to 2011 as training data, and then test the model with the observation from 2012 as testing data.
(5) Repeat the process using all previous years data until the last set of observations in 2016.
(6) Compute the accuracy of the model by averaging the RMSE over the six test sets.

As a result, the model's accuracy can be determined using the formula as shown below.

$$\mathrm{RMSE_{cv}} = \frac{\mathrm{RMSE_{2011}}+\mathrm{RMSE_{2012}}+\mathrm{RMSE_{2013}}+\mathrm{RMSE_{2014}}+\mathrm{RMSE_{2015}}+\mathrm{RMSE_{2016}}}{6} \tag{9}$$

3. Results

3.1. *Multicollinearity test for each state in East Coast region*

Multicollinearity occurs when two or more climate factors in the model show a significant correlation with one another. As a result, this section examines issues related to multicollinearity. The findings regarding multicollinearity for the three states representing the East Coast region are summarized in Table 1. The results indicate that the VIF, CVIF, and TOL values for each state's climate factors are all less than 5 and 10, and greater than 0, respectively. This demonstrates that multicollinearity does not exist among the variables in the datasets for each state. Consequently, each climate factor can be incorporated into the modeling phase.

Table 1: Multicollinearity test for climate factors in each East Coast Region

States	Multicollinearity Test				Overdispersion Test		
	Climate Factors	VIF	Tolerance	CVIF	Dispersion	*z*-values	*p*-values
Pahang	Temperature	1.9639	0.5092	1.9437	10.8328	4.9255	0.0000***
	Humidity	1.9169	0.5217	1.8972			
	Rainfall	1.3655	0.7324	1.3515			
	Wind speed	1.3448	0.7436	1.3310			
Kelantan	Temperature	1.6334	0.6122	1.6312	26.2840	4.1974	0.0000***
	Humidity	2.2907	0.4365	2.2876			
	Rainfall	1.7311	0.5777	1.7288			
	Wind speed	1.1815	0.8464	1.1799			
Terengganu	Temperature	1.6115	0.6206	1.6208	10.3708	4.1804	0.0000***
	Humidity	1.9771	0.5058	1.9886			
	Rainfall	1.7461	0.5727	1.7563			
	Wind speed	1.2226	0.8179	1.2297			

Significant codes: 0.05 '***'

3.2. *Overdispersion test for each state*

Overdispersion is a common problem in Poisson regression modelling. The problem could occur when the variance of the response variable exceeds its expected value. Thus, the overdispersion issues were addressed in this study by first checking their presence in the datasets before proceeding with the analysis. Based on the test mentioned before, the rejection of the null hypothesis implies that the datasets are overdispersed. The results of the overdispersion test for each state in the East Coast region of Malaysia are presented in Table 1. The dispersion parameter is high and the *p*-values for all tests in each state are less than 0.05, providing sufficient evidence to reject the null hypothesis. The results imply that overdispersion problems are present in the datasets for each state. Consequently, in order to address these difficulties, a Negative Binomial regression model has been suggested. Thus, in the following analysis, GLM and GAM of the Negative Binomial were performed.

3.3. *Covariate selection for GLM and GAM*

The modelling analysis begins with the covariate selection for GLM and GAM. In this study, the covariate selection was done using a stepwise forward method for identifying the best optimal GLM and GAM models for HFMD-Climate. Table 2 shows the covariate selection for the Pahang GLM and GAM models. A single-predictor GLM model was first developed for each climate factor; the model, which includes temperature, showed the smallest AIC value with in comparison to the other models. This indicates that the temperature is the most important factor in the model. Following that, a model with two predictors was developed, followed by models with three and four predictors. The lowest AIC values, on the other hand, result in a model that consists of only temperature. This implies that the temperature factor is only one element of the optimal GLM model for Pahang.

Table 2 summarized the covariates included in the GLM and GAM models. The results suggest that the GLM model for each state takes into account different climate factors. Climate factors such as temperature, humidity, and wind speed are required for the Terengganu model in the East Coast region. In contrast, Kelantan revealed a different result, as the model only requires humidity and rainfall. Additionally, the Pahang model consists only of a temperature parameter. The GAM model produced consistent results with only temperature as significant variable for all states in the East Coast region. Notably, the best model exclusively included only temperature factor. Overall, the covariate selection analysis demonstrated that the GLM model showed a different impact of climate factors on HFMD, while the GAM model detected similar variables. These findings could be influenced by the GLM and GAM models' unique characteristics. Therefore, model comparison is required to find the best model with climate factors that affect significantly HFMD cases in the East Coast region of Malaysia.

Table 2: Summary of the significant covariates based on GLM and GAM analysis

States	GLM Negative Binomial	GAM Negative Binomial
Pahang	Temperature	Temperature
Kelantan	Humidity, Rainfall	Temperature
Terengganu	Temperature, Humidity, Wind speed	Temperature

3.4. *Model evaluation for GLM and GAM*

In continuation of the preceding section, a comparison is made between the GLM and GAM optimal models in order to determine which model provides a more accurate description of the HFMD incidence and climate factors for each state. In this study, the seasonality effect and long-term trend of HFMD cases are taken into consideration by including the covariate of time in both models. Based on Table 3 to 5, the GAM model, consisting of a smooth function of the climate factors, is used to examine the non-linear association between the HFMD and the climate factors. In contrast, the GLM model only takes into account a linear relationship. In the GAM model, the abbreviation edf stands for effective degree of freedom. An edf value of one signifies a linear correlation between HFMD incidence and climate factors, whereas a value greater than one suggests a non-linear relationship. If the edf for a smoothing spline of the climate factors is less than one, it is eliminated since it suggests a linear relationship (Wood 2015).

Tables 3 to 5 show the results of the best GLM and GAM Negative Binomial models for each state. Table 3 presents the findings of the GLM Negative Binomial and GAM Negative Binomial models, showing that both models consider the same climate factors affecting HFMD cases in Pahang. A significant association between HFMD and time was found, as the p-value was less than 0.05. This indicates that the HFMD cases in Pahang follows a long-term trend and exhibits a seasonal effect. In GLM Negative Binomial, the temperature with a two-week lag period has a significant positive linear association with HFMD cases. The model indicates an increase in temperature is associated with 15.96% increase in the risk of developing HFMD cases over the subsequent two weeks [$(e^{0.1481} - 1) \times 100\% = 15.96\%$]. The GAM Negative Binomial showed a significant non-linear association between temperature and HFMD cases in Pahang, as shown by p-values less than 0.05. The value of edf in the GAM Negative Binomial is 2.238, indicating significant proof for a non-linear association between HFMD cases and temperature at a two-week lag.

Table 3: Model evaluation for Pahang

Model	GLM Negative Binomial				GAM Negative Binomial			
Linear terms	Estimate		Standard Error	p-value	Estimate		Standard Error	p-value
Constant	-9.1456		1.7640	0.0000***	1.4004		0.0429	0.0000***
Temperature	0.1481		0.0543	0.0064***			-	
Time	0.4305		0.0773	0.0000***			-	
Smooth terms	edf	Ref.df	Chi-square	p-value	edf	Ref.df	Chi-square	p-value
s(Temperature)			-		2.238	18	7.591	0.0080***
s(Time)			-		21.703	26	414.957	0.0000***
RMSE			7.5151				5.4991	
MAE			4.8158				3.2383	

Significant codes: 0.05 '***'; edf is the effective degree of freedom; Ref. df is reference degree of freedom

Table 4 shows the results for the states of Kelantan. The GAM Negative Binomial model indicates that temperature at a two-week lag has no significant association with HFMD, as the p-value is greater than 0.05. Contrary findings obtained from the GLM model show a significant association between HFMD cases and both humidity and rainfall. According to the GLM Negative Binomial model, a two-week delay in humidity increased the risk of HFMD incidence in Kelantan by 6.49% [$(e^{0.0629} - 1) \times 100\% = 6.49\%$], whereas a two-week delay in rainfall decreased the risk of HFMD incidence by 1.67% [$(e^{0.0166} - 1) \times 100\% = 1.67\%$]. Besides, HFMD cases and time demonstrate a significant relationship in both models, with a p-value less than 0.05. This suggests the presence of a long-term trend and seasonal effects in HFMD cases in Kelantan.

Table 5 presents the findings of modelling HFMD and climate factors in Terengganu by applying GLM Negative Binomial and GAM Negative Binomial. Both models indicate a significant relationship between HFMD and time as the p-value was less than 0.05, implying that HFMD cases in Terengganu exhibit a long-term trend and seasonal effect. The p-values for humidity and wind speed at a two-week lag period in the GLM Negative Binomial model are statistically significant, as the p-values are less than 0.05. This implies there is an association between humidity and wind speed in the following two weeks and HFMD cases in Terengganu. The risk of HFMD will decrease by 6.13% [$(e^{0.0595} - 1) \times 100\% = 6.13\%$] with an increase in humidity and by 47.40% 47.40% [$(e^{0.3880} - 1) \times 100\% = 47.40\%$] with an increase in wind speed, both with a two-week time lag. The GAM Negative Binomial model, however, identifies distinct and significant climate factors. The only variable that displays a non-linear correlation with HFMD cases is temperature, with *p*-values below 0.05 when considering a lag period of two weeks. The edf values for temperature exceed one, indicating that the relationship between HFMD and temperature exhibits non-linear patterns.

Table 4: Model evaluation for Kelantan

Model	GLM Negative Binomial				GAM Negative Binomial			
Linear terms	Estimate		Standard Error	*p*-value	Estimate		Standard Error	*p*-value
Constant	-9.1713		2.1753	0.0000***	1.6263		1.3214	0.218
Humidity	0.0629		0.0209	0.0026***			-	
Rainfall	-0.0166		0.0073	0.0237***			-	
Temperature			-		0.0039		0.0483	0.935
Time	0.41353		0.08613	0.0000***			-	
Smooth terms	edf	Ref.df	Chi-square	*p*-value	edf	Ref.df	Chi-square	*p*-value
s(Time)			-		22.64	26	1289	0.0000***
RMSE			16.8586				7.2853	
MAE			9.6908				3.7780	

Significant codes: 0.05 '***'; edf is the effective degree of freedom; Ref. df is reference degree of freedom

Table 5: Model evaluation for Terengganu

Model	GLM Negative Binomial				GAM Negative Binomial			
Linear terms	Estimate		Standard Error	*p*-value	Estimate		Standard Error	*p*-value
Constant	-0.0566		3.1439	0.9856	1.2828		0.0413	0.0000***
Temperature	0.12981		0.0721	0.0717			-	
Humidity	-0.0595		0.0160	0.0002***			-	
Wind speed	-0.3880		0.1212	0.0014***			-	
Time	0.6863		0.0741	0.0000***			-	
Smooth terms	edf	Ref.df	Chi-square	*p*-value	edf	Ref.df	Chi-square	*p*-value
s(Temperature)			-		3.148	18	16.55	0.0000***
s(Time)			-		21.441	26	608.31	0.0000***
RMSE			6.8046				4.6283	
MAE			4.2729				2.6461	

Significant codes: 0.05 '***'; edf is the effective degree of freedom; Ref. df is reference degree of freedom

In terms of the model comparison, the GAM Negative Binomial model performs better than the GLM Negative Binomial model in identifying the association between HFMD and climate factors in the East Coast region. This is evidenced by the smaller RMSE and MAE values. The results suggest that employing the GAM Negative Binomial model is a more suitable method for determining the association between HFMD incidence and climate factors in Pahang, Terengganu, and Kelantan. The GAM Negative Binomial model provides a comprehensive explanation of the non-linear relationship between HFMD incidence and climate factors.

3.5. *Model validation*

This study applies rolling basis cross-validation to identify the best and most accurate model for describing the association between HFMD incidence and climate factors in each state of the East Coast region of Malaysia. Table 6 shows the results of a rolling-basis cross-validation method for each model. The GLM and GAM Negative Binomial models are then compared by calculating the average values of the RMSE over the six test sets for each state. Comparing the two models in each state shows that the GAM Negative Binomial model has significantly lower RMSE values than the GLM Negative Binomial model. Thus, the GAM Negative Binomial approach is perhaps the most adequate model for describing the relationship between HFMD incidence and climate factors in the East Coast region of Malaysia. As noted by Hastie and Tibshirani (1986), the GAM model allows for nonlinear relationships between predictors and the response variable through smooth functions. This flexibility enables the model to capture complex patterns in the data that linear models might overlook, thereby enhancing predictive performance. Additionally, GAMs tend to be less sensitive to outliers compared to traditional linear models due to their flexible nature. In this study, the HFMD data may contain extreme values or noise that could disproportionately affect model performance. Therefore, the robustness of GAMs may help address these issues (Wood 2017).

Table 6: Rolling basis cross-validation of each model for East Coast region

RMSE	Pahang		Terengganu		Kelantan	
	GLM	GAM	GLM	GAM	GLM	GAM
1	3.4999	5.4736	2.2602	.5061	3.7705	4.1883
2	9.6821	8.3778	7.0395	6.0322	42.5385	41.7094
3	5.9739	3.7638	5.2979	2.8556	58.5671	6.1104
4	3.8034	3.9107	5.7894	7.6691	17.8728	12.3538
5	4.6113	4.6167	7.3107	2.0936	17.3243	2.5569
6	18.5447	17.6937	16.8619	18.6782	18.5569	21.8334
Average RMSE	7.6859	**7.3061**	7.4266	**6.8058**	26.4384	**14.7920**

3.6. *Final model*

The best model for each state on the East Coast has been identified through the process of model validation. Eqs. (10)-(12) were constructed from the best models presented in Tables 3 to 5. Eq. (10) provides the equations related to the GAM Negative Binomial models for Pahang. Figure 1a) displays the smoothed temperature and HFMD incidence curves from the GAM Negative Binomial model. A GAM Negative Binomial model suggests that temperatures between 26.5°C and 28°C will decrease the risk of HFMD in Pahang over the next two weeks. Meanwhile, the risk of HFMD will increase when temperatures exceed 28°C.

$$\ln[E(HFMD_{\text{PAHANG}})] = 1.4004 + \text{s(Temperature)}_{t-2} + \text{s(Time)} \qquad (10)$$

Moreover, the model equations for the GAM Negative Binomial models for Kelantan and Terengganu are given by Eqs. (11) and (12), respectively. Figure 1(b) depicts the smoothed association between of HFMD incidence and temperature in Terengganu. It demonstrates that temperatures between 27°C and 28.5°C substantially decreases the risk of HFMD over the next two weeks. In contrast, the risk of HFMD will increase when the temperature exceeds 28°C.

$$\ln[E(HFMD_{\text{KELANTAN}})] = 1.6263 + 0.0039\text{Temperature}_{\,t-2} + \text{s(Time)} \qquad (11)$$

$$\ln[E(HFMD_{\text{TERENGGANU}})] = 1.2828 + \text{s(Temperature)}_{t-2} + \text{s(Time)} \qquad (12)$$

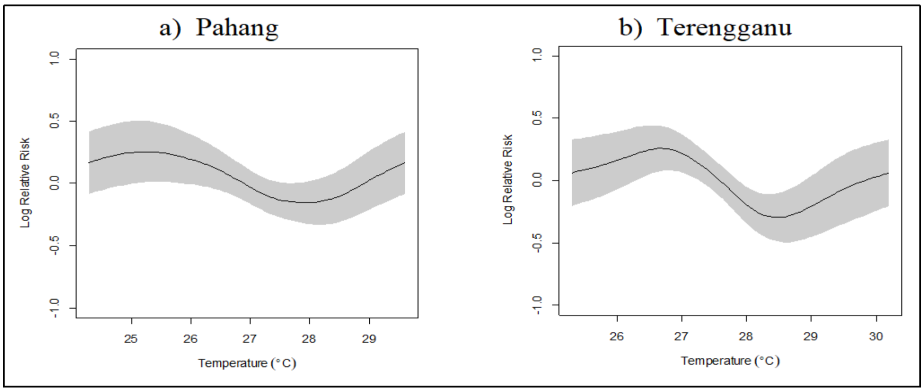

Figure 1: GAM smoothed association between HFMD and climate temperature for Pahang and Terengganu

The final model indicates that climate factors, specifically temperature with a two-week lag effect, are the primary predictors of HFMD cases in the East Coast region of Malaysia. The findings of this study align with previous research conducted in China (Chen *et al.* 2014; Wu *et al.* 2014; *Du et al.* 2017), Singapore (Hii *et al.* 2011), South Korea (Kim *et al.* 2016), Vietnam (Thanh *et al.* 2016), and Japan (Onozuka & Hashizume 2011). These studies indicate that climate variables, particularly temperature, play a significant role in the occurrence of HFMD cases. However, several studies conducted in various regions of China have shown that the incidence of HFMD is not significantly effect by temperature (Liao *et al.* 2015; Chen *et al.* 2019). These concerns are affected by a variety of factors, including the geographical location of each region or state, monsoon, and topographical effects.

The effect of the temperature on HFMD cases on the East Coast of Malaysia, however, is different between states. In Pahang and Terengganu, the association between HFMD and temperature shows almost the same wave-shaped curve. The curve shows that the risk of HFMD cases decreased temperatures ranging over the next two weeks with temperatire ranging from 26°C to 28°C. In contrast, the risk of HFMD increases when the temperature exceeds 28°C. These results may be explained by the fact that temperatures could affect the virus's ability to develop and remain alive (Chang *et al.* 2012). For instance, the virus exhibited a higher rate of reproduction and increased longevity at high temperatures compared to low temperatures. Consequently, the increase in HFMD cases was immediate and prolonged in areas with high temperatures. However, HFMD cases in Kelantan are not significantly affected by temperature. The variability of the findings in this study may be affected by location-specific factors. Further inquiry is required to determine the causes of the inconsistencies. According to some studies, temperature affects population behaviors and activities, as well as the survival and transmission of pathogenic microorganisms in the environment, which in turn affects the dynamics of infection transmission (Lin *et al.* 2013). These significant elements might be connected to behavior in people. People may decide to stop all activities, feel uncomfortable in crowded places when the temperature rises, and stay indoors (Kim *et al.* 2016). In these circumstances, the risk of HFMD cases will be minimized by the absence of contact with others.

4. Conclusion

In summary, the study found that climate factors have a significant impact on HFMD incidence; however, the effect varies by state in Malaysia's East Coast region. The findings show that temperature is the primary determinant of HFMD incidence in this region.

Furthermore, the GAM techniques provided the most substantial evidence of a nonlinear association between HFMD and climate factors in each state. This study also indicates that incorporating a smooth nonlinear function into the model gave GAM valuable additional flexibility for examining the non-parametric relationship between HFMD incidence and climate parameters. The study's findings have several implications for future HFMD prevention strategies in East Coast Malaysia. For climate-based monitoring, implementing an early warning system that tracks temperature changes can help predict HFMD outbreaks, allowing for timely interventions. Understanding that temperature affects HFMD differently across states enables the development of tailored strategies. For instance, specific temperature thresholds in Pahang and Terengganu can guide preventive measures. Raising awareness about the relationship between climate and HFMD can also encourage communities to adopt preventive measures during high-risk periods. Policymakers can integrate climate data into public health planning to enhance the overall effectiveness of HFMD prevention strategies. Local authorities could implement public awareness campaigns to educate communities about the connection between climate conditions and HFMD, empowering caregivers to recognize symptoms and adopt preventive measures. Enhanced hygiene and sanitation efforts, particularly during high-risk periods identified by climate data, are essential in public spaces, schools, and childcare facilities. These measures can significantly reduce HFMD incidence, particularly in the East Coast region.

Some limitations were acknowledged in this study. Although the GAM yielded superior results compared to the GLM, both models are constrained by the assumption of independence among response variables. The weekly HFMD data utilized in this study are characteristic of time-series data, which frequently exhibit issues of autocorrelation. When autocorrelation is present, it can result in biased parameter estimates and inaccurate conclusions. Specifically, autocorrelation may inflate the significance of predictors, leading to an overestimation of their effects, while also underestimating the model's standard errors, thereby producing overly optimistic confidence intervals. To address these limitations, future research should consider employing models that explicitly account for autocorrelation, such as Generalized Least Squares (GLS) or Mixed-Effect Models. These models are capable of incorporating random effects to account for temporal or spatial correlations within the data. Additionally, it is imperative that future studies conduct thorough exploratory data analyses to identify potential sources of autocorrelation prior to modeling. This proactive approach can inform the selection of appropriate modeling strategies and ensure that the underlying assumptions are adequately met. Furthermore, researchers might also investigate additional variables that could influence HFMD cases, including population density, air pollution, geographical location, monsoon patterns, topographical factors, and socioeconomic conditions.

Acknowledgments

The authors would like to express their greatest appreciation to the Ministry of Health, Malaysia, for providing the HFMD data and the Malaysian Meteorological Department for providing the climate data. This work was supported by UiTM Pahang Branch, under the Dana Dalam Negeri Research Grant Scheme, Turn Thesis Grant to Paper (GT2P) 600-TNCPI 5/3/DDN (06) (028/2023).

References

Akaike H. 1998. Information theory and an extension of the maximum likelihood principle. In Parzen E., Tanabe K. & Kitagawa G. (eds.). *Selected Papers of Hirotugu Akaike*: 199-213. New York, NY: Springer.

Allison P.D. 1999. *Multiple Regression: A Primer*. Thousand Oaks, CA: Pine Forge Press.
Alsop J., Flewett T.H. & Foster J.R. 1960. "Hand-foot-and-mouth disease" in Birmingham in 1959. *British Medical Journal* **2**(5214): 1708-1711.
Breslow N.E. 1984. Extra-Poisson variation in log-linear models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **33**(1): 38-44.
Cameron A.C. & Trivedi P.K. 1990. Regression-based tests for overdispersion in the Poisson model. *Journal of Econometrics* **46**(3): 347-364.
Cameron A.C. & Trivedi P.K. 2013. *Regression Analysis of Count Data*. 2nd Ed. Cambridge, UK: Cambridge University Press.
Chai T. & Draxler R.R. 2014. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development Discussions* **7**: 1247–1250.
Chang H.-L., Chio C.-P., Su H.-J., Liao C.-M., Lin C.-Y., Shau W.-Y., Chi Y.-C., Cheng Y.-T., Chou Y.-L., Li C.-Y., Chen K.-L & Chen K.-T. 2012. The association between enterovirus 71 infections and meteorological parameters in Taiwan. *PLoS ONE* **7**(10): e46845.
Chen C., Lin H., Li X., Lang L., Xiao X., Ding P., He P., Zhang Y., Wang M. & Liu Q. 2014. Short-term effects of meteorological factors on children hand, foot and mouth disease in Guangzhou, China. *International journal of Biometeorology* **58**(7): 1605-1614.
Chen K.-T., Chang H.-L., Wang S.-T., Cheng Y.-T. & Yang J.-Y. 2007. Epidemiologic features of hand-foot-mouth disease and herpangina caused by enterovirus 71 in Taiwan, 1998–2005. *Pediatrics* **120**(2): e244-e252.
Chen S., Liu X., Wu Y., Xu G., Zhang X., Mei S., Zhang Z., O'Meara M., O'Gara M.C., Tan X. & Li L. 2019. The application of meteorological data and search index data in improving the prediction of HFMD: a study of two cities in Guangdong Province, China. *Science of The Total Environment* **652**: 1013-1021.
Cheong Y.L., Burkart K., Leitão P.J. & Lakes T. 2013. Assessing weather effects on dengue disease in Malaysia. *International Journal of Environmental Research and Public Health* **10**(12): 6319-6334.
Chowdhury M.Z.I. & Turin T.C. 2020. Variable selection strategies and its importance in clinical prediction modelling. *Family Medicine and Community Health* **8**(1): e000262.
Du Z., Xu L., Zhang W., Zhang D., Yu S. & Hao Y. 2017. Predicting the hand, foot, and mouth disease incidence using search engine query data and climate variables: an ecological study in Guangdong, China. *BMJ Open* **7**(10): e016263.
Du Z., Zhang W., Zhang D., Yu S. & Hao Y. 2016. The threshold effects of meteorological factors on hand, foot, and mouth disease (HFMD) in China, 2011. *Scientific Reports* **6**(1): 36351.
Flewett T.H., Warin R.P. & Clarke S.K. 1963. Hand, foot, and mouth disease' associated with Coxsackie A5 virus. *Journal of Clinical Pathology* **16**(1): 53-55.
Guo Z., Wang Y., Li Y. & Zhou L. 2023. Impact of meteorological factors on the incidence of hand-foot-mouth disease in Yangzhou from 2017 to 2022: a time series study. *Frontiers in Public Health* **11**: 1278516.
Hardin J.W. & Hilbe J.M. 2007. *Generalized Linear Models and Extensions*. 2nd Ed. College Station, Texas: Stata Press.
Hastie T. & Tibshirani R. 1986. Generalized additive models. *Statistical Science* **1**(3): 297-318.
Hastie T. & Tibshirani R. 1990. Exploring the nature of covariate effects in the proportional hazards model. *Biometrics* **46**(4): 1005-1016.
Hastie T. & Tibshirani R. 1995. Generalized additive models for medical research. *Statistical Methods in Medical Research* **4**(3): 187-196.
Hastie T., Tibshirani R. & Friedman J. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd Ed. New York: Springer.
Hii Y.L., Rocklöv J. & Ng N. 2011. Short-term effects of weather on hand, foot, and mouth disease. *PLoS ONE* **6**(2): e16796.
Hilbe J.M. 2011. *Negative Binomial Regression*. 2nd Ed. Cambridge, UK: Cambridge University Press.
Hyndman R.J. 2014. Measuring forecast accuracy. In Gilliland M., Tashman L. & Sglavo U. (eds.). *Business Forecasting: Practical Problems and Solutions*: 177-183. Hoboken: John Wiley & Sons.
Ibrahim A.O., Chen W.Y., Robert J.Z., Yi C.C. & Majid M.A. 2024. An Investigation of the Effect of Weather on Hand-Foot Mouth Disease Incidence in Sarawak. In Thiruchelvam V., Alfred R., Ismail Z.I.B.A., Haviluddin H. & Baharum A. (eds.). *Proceedings of the 4th International Conference on Advances in Computational Science and Engineering. ICACSE 2023. Lecture Notes in Electrical Engineering*: 407-419. Singapore: Springer Nature Singapore.
Kim B.I., Ki H., Park S., Cho E. & Chun B.C. 2016. Effect of climatic factors on hand, foot, and mouth disease in South Korea, 2010-2013. *PloS ONE* **11**(6): e0157500.
Krzanowski W.J. 1998. *An Introduction to Statistical Modelling*. Hoboken: Wiley.
Kutner M.H., Nachtsheim C.J., Neter J. & Li W. 2005. *Applied Linear Statistical Models*. 5th Ed. New York, NY: McGraw-Hill.

Leong P.F., Labadin J., Abd Rahman S.B. & Juan S.F.S. 2011. Quantifying the relationship between the climate and hand-foot-mouth disease (HFMD) incidences. In *2011 Fourth International Conference on Modeling, Simulation and Applied Optimization*, pp. 1-5.

Li Y., Zhang J. & Zhang X. 2014. Modeling and preventive measures of hand, foot and mouth disease (HFMD) in China. *International Journal of Environmental Research and Public Health* **11**(3): 3108-3117.

Liao Y., Ouyang R., Wang J. & Xu B. 2015. A study of spatiotemporal delay in hand, foot and mouth disease in response to weather variations based on SVD: a case study in Shandong Province, China. *BMC Public Health* **15**: 71.

Lin H., Zou H., Wang Q., Liu C., Lang L., Hou X. & Li Z. 2013. Short-term effect of El Nino-Southern Oscillation on pediatric hand, foot and mouth disease in Shenzhen, China. *PloS ONE* **8**(7): e65585.

Ma E., Lam T., Wong C. & Chuang S.K. 2010. Is hand, foot and mouth disease associated with meteorological parameters? *Epidemiology & Infection* **138**(12): 1779-1788.

Marquardt D.W. 1970. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics* **12**(3): 591-612.

McCullagh P. 1989. *Generalized Linear Models*. 2nd Ed. New York: Routledge.

Melnick J.L. 1984. Enterovirus type 71 infections: a varied clinical pattern sometimes mimicking paralytic poliomyelitis. *Reviews of Infectious Diseases* **6**: 387-390.

Miller A. 2002. *Subset Selection in Regression*. 2nd Ed. Boca Raton, FL: Chapman and Hall.

Ministry of Health Malaysia. 2012. Hand, foot, and mouth disease (HFMD): Nasihat kesihatan (advisory) HFMD. http://www.moh.gov.my/english.php/pages/view/171 (23 February 2019).

Nelder J.A. & Wedderburn R.W.M. 1972. Generalized linear models. *Journal of the Royal Statistical Society Series A: Statistics in Society* **135**(3): 370-384.

Neter J., Kutner M.H., Nachtsheim C.J. & Wasserman W. 1996. *Applied Linear Statistical Models*. 4th Ed. New York, NY: McGraw-Hill.

O'brien R.M. 2007. A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity* **41**: 673-690.

Onozuka D. & Hashizume M. 2011. The influence of temperature and humidity on the incidence of hand, foot, and mouth disease in Japan. *Science of the Total Environment* **410-411**: 119-125.

Ooi M.H., Wong S.C., Lewthwaite P., Cardosa M.J. & Solomon T. 2010. Clinical features, diagnosis, and management of enterovirus 71. *The Lancet Neurology* **9**(11): 1097-1105.

Robinson C.R., Doane F.W. & Rhodes A.J. 1958. Report of an outbreak of febrile illness with pharyngeal lesions and exanthem: Toronto, summer 1957—isolation of group A coxsackie virus. *Canadian Medical Association Journal* **79**(8): 615-621.

Schmidt N.J., Lennette E.H. & Ho H.H. 1974. An apparently new enterovirus isolated from patients with disease of the central nervous system. *Journal of Infectious Diseases* **129**(3): 304-309.

Siegel A.F. 2016. *Practical Business Statistics*. 7th Ed. Cambridge, Massachusetts: Academic Press.

Sigauke C., Kumar S., Maswanganyi N. & Ranganai E. 2019. Reliable predictions of peak electricity demand and reliability of power system management. In Anand A. & Ram M. (eds.). *System Reliability Management: Solutions and Technologies*: 137-160. Boca Raton, FL: CRC Press.

Thanh T.C., Giao N.T.Q., Khoa D.D. & Hien T.T. 2016. Effects of climate variations on hand-foot-mouth disease in Ho Chi Minh City. *Journal of Science and Technology* **54**(2A): 120-127.

Underwood F.M. 2009. Describing long-term trends in precipitation using generalized additive models. *Journal of Hydrology* **364**(3-4): 285-297.

US Centers for Disease Control and Prevention. 2021. About hand, foot, and mouth disease. https://www.cdc.gov/hand-foot-mouth/index.html (20 May 2021).

Wahid N.A.A., Suhaila J. & Rahman H.A. 2021. Effect of climate factors on the incidence of hand, foot, and mouth disease in Malaysia: A generalized additive mixed model. *Infectious Disease Modelling* **6**: 997-1008.

Wahid N.A.A., Suhaila J. & Sulekan A. 2020. Temperature effect on HFMD transmission in selangor, Malaysia. *Sains Malaysiana* **49**(10): 2587-2597.

Wang S.M., Liu C.C., Tseng H.W., Wang J.R., Huang C.C., Chen Y.J., Yang Y.J., Lin S.J. & Yeh T.F. 1999. Clinical spectrum of enterovirus 71 infection in children in southern Taiwan, with an emphasis on neurological complications. *Clinical Infectious Diseases* **29**(1): 184-190.

Wang S.M. & Liu C.C. 2009. Enterovirus 71: epidemiology, pathogenesis and management. *Expert Review of Anti-infective Therapy* **7**(6): 735-742.

Wang L., Xu C., Wang J., Qiao J., Wu N. & Li L. 2023. Spatiotemporal associations between hand, foot and mouth disease and meteorological factors over multiple climate zones. *International Journal of Biometeorology* **67**(9): 1493-1504.

Wood S. 2015. Package 'mgcv'. R Package Version 1, 29.

Wood S.N. 2017. *Generalized Additive Models: An Introduction with R*. 2nd Ed. New York: Chapman and Hall.

World Health Organization. 2011. A Guide to clinical management and public health response for hand, foot and mouth disease (HFMD). WHO Regional Office for the Western Pacific.

Wu H., Wang H., Wang Q., Xin Q. & Lin H. 2014. The effect of meteorological factors on adolescent hand, foot, and mouth disease and associated effect modifiers. *Global Health Action* **7**(1): 24664.

Xing W., Liao Q., Viboud C., Zhang J., Sun J., Wu J.T., Chang Z., Liu F., Fang V.J., Zheng Y., Cowling B.J., Varma J.K., Farrar J.J., Leung G.M. & Yu H. 2014. Hand, foot, and mouth disease in China, 2008–12: an epidemiological study. *The Lancet Infectious Diseases* **14**(4): 308-318.

Yang X., Wang J., Zhang G. & Yu Z. 2024. Short-term effects of extreme meteorological factors on hand, foot, and mouth disease infection during 2010–2017 in Jiangsu, China: A distributed lag non-linear analysis. *GeoHealth* **8**(4): e2023GH000942.

College of Computing, Informatics, and Mathematics
Universiti Teknologi MARA Pahang Branch, Jengka Campus,
26400, Pahang, MALAYSIA
E-mail: nurmarni@uitm.edu.my[*]

Department of Mathematical Sciences
Faculty of Science
Universiti Teknologi Malaysia
81310, Johor Bahru, MALAYSIA
E-mail: suhailasj@utm.my, halizarahman@utm.my

Received: 9 September 2024
Accepted: 26 October 2024

[*]Corresponding author