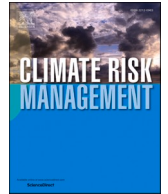




ELSEVIER

Contents lists available at ScienceDirect

Climate Risk Management

journal homepage: www.elsevier.com/locate/crm

Spatio-temporal detection for dengue outbreaks in the Central Region of Malaysia using climatic drivers at mesoscale and synoptic scale

Stan Yip^{a,*}, Norziha Che Him^b, Nur Izzah Jamil^{b,d}, Daihai He^a, Sujit K. Sahu^c

^a Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong

^b Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia, Pagoh, Johor, Malaysia

^c School of Mathematical Sciences, University of Southampton, Southampton, UK

^d Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Negeri Sembilan Branch, Rembau Campus

ARTICLE INFO

Keywords:

Bayesian spatio-temporal model

BYM2

dengue

early warning system

ENSO

El Niño Modoki

ABSTRACT

The disease dengue is associated with both mesoscale and synoptic scale meteorology. However, previous studies for south-east Asia have found a very limited association between synoptic variables and the reported number of dengue cases. Hence there is an urgent need to establish a more clear association with dengue incidence rates and the most relevant meteorological variables in order to institute an early warning system.

This article develops a rigorous Bayesian modelling framework to identify the most important covariates and their lagged effects for constructing an early warning system for the Central Region of Malaysia where the case rates have increased substantially in the recent past. Our modelling includes multiple synoptic scale Niño indices, which are related to the phenomenon of El Niño Southern Oscillation (ENSO), along with other relevant mesoscale environmental measurements and an unobserved variable derived from reanalysis data. An empirically well validated hierarchical Bayesian spatio-temporal is used to build a probabilistic early warning system for detecting an upcoming dengue epidemic.

Our study finds a 46.87% increase in dengue cases due to one degree increase in the central equatorial Pacific sea surface temperature with a lag time of six weeks. We discover the existence of a mild association with relative risk 0.9774 (CI: 0.9602, 0.9947) between the rate of cases and a distant lagged cooling effect in the region of coastal South America related to a phenomenon called El Niño Modoki. The Bayesian model also establishes that the synoptic meteorological drivers can enhance short-term early detection of dengue outbreaks and these can also potentially be used to provide longer-term forecasts.

1. Introduction

Dengue is a very harmful mosquito-borne viral infection worldwide. This viral infectious disease can lead to a wide spectrum of clinical manifestations such as acute onset high fever, muscle and joint pain, myalgia, cutaneous rash, hemorrhagic episodes and

* Corresponding author.

E-mail address: stan.yip@polyu.edu.hk (S. Yip).

<https://doi.org/10.1016/j.crm.2022.100429>

Received 18 November 2021; Received in revised form 15 February 2022; Accepted 30 March 2022

Available online 4 April 2022

2212-0963/© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

circulatory shock (Hasan et al., 2016). Thus the disease causes an enormous burden to public health systems. Bhatt (2013) estimate that there are 390 million total annual infections throughout the world. During the year 1990 to 2019, the number of dengue cases has increased by 85.5% according to Yang et al. (2021).

In Asia, dengue fever has been reported earliest by Skae (1902), followed by dengue hemorrhagic fever and dengue shock syndrome epidemics in the twentieth century (Henchal and Putnak, 1990). Gubler (1998) describes that the dengue virus (DENV) is transmitted by the bite of female *Aedes aegypti* mosquitoes. To a lesser extent, *Aedes albopictus* is also a vector of indoor transmission (Noor et al., 2018). Four serotypes of virus DENV-1, DENV-2, DENV-3 and DENV-4 following the human cycle are genetically similar (Mustafa et al., 2015).

Climate is a crucial determinant of dengue disease transmission by affecting its vector dynamics (Morin et al., 2013). Local and global climate not only influence the spatial distribution of infections (Johansson et al., 2009) but also the interannual variability (Cazelles et al., 2005). The climatic impact to a dengue outbreak is also known to be cumulative and delayed (Lowe et al., 2018). Hii et al. (2016) emphasise that dengue is a climate-sensitive infectious disease. The rapid change in climate drivers increases the risk of dengue outbreaks in the past decade. A climate-based early warning system (EWS) has the potential to enhance surveillance and control of the disease. Simple statistical models, which do not adequately account for complex spatio-dynamic dependencies, cannot be used to construct a reliable EWS. Hence there is an urgent need to develop complex Bayesian models which validate well in empirical studies.

In this article our study region is the Central Region in Malaysia which constitutes of the Selangor state together with the two adjacent federal territories, namely Kuala Lumpur and Putrajaya, see Fig. 1. This region contributes most to the national dengue hospitalisation numbers in Malaysia. The dengue infection rates in this region have increased significantly in the past decade as reported by AbdMajid et al. (2021) and Salim (2021).

A significant relationship has been found between dengue hospitalisation rates and covariates such as precipitation, temperature, number of monthly rain days and ENSO for 12 states of West Malaysia (Che Him et al., 2018). A similar study by Che Him et al. (2018) identifies two distinct spatial clusters via two generalised additive models (GAM) for nine districts of the state of Selangor. Ahmad et al.

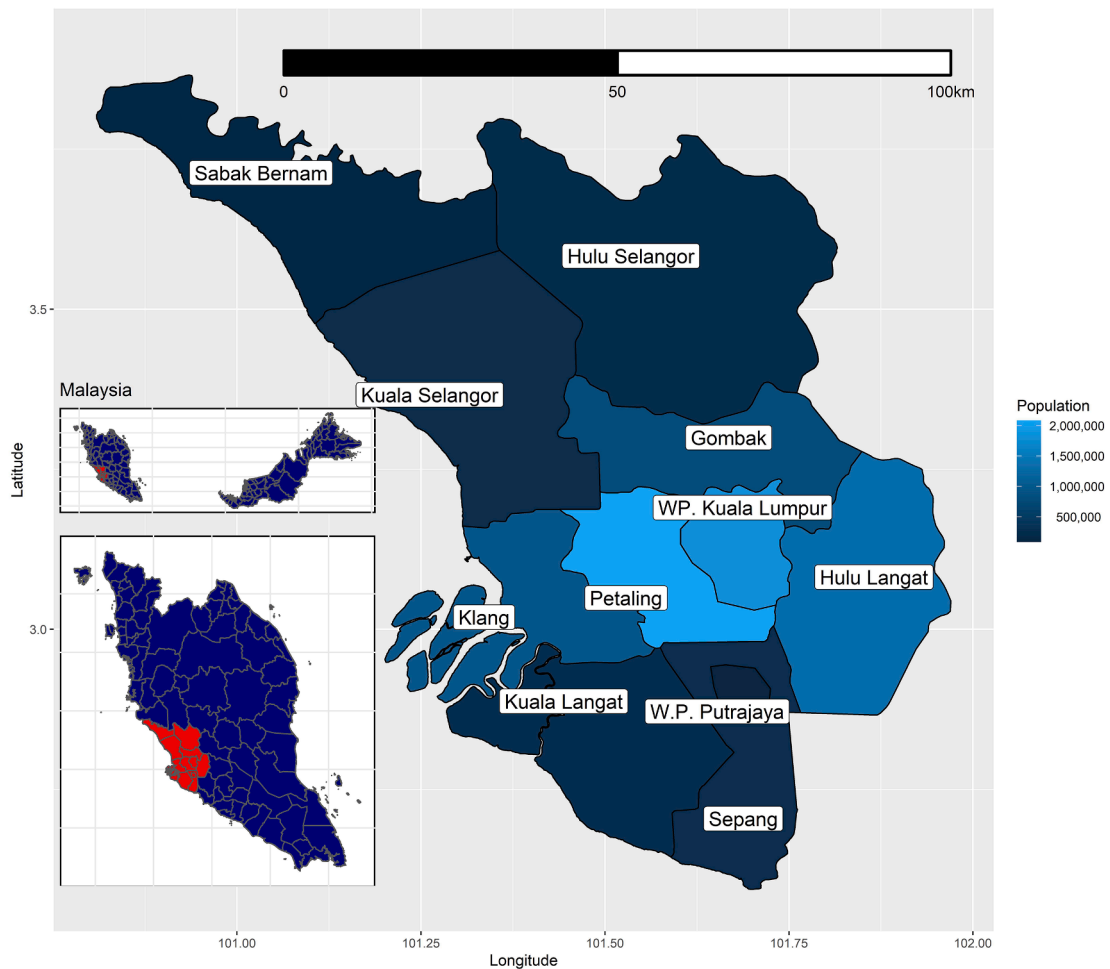


Fig. 1. Geographical distribution and population in the Central Region of Malaysia in 2019.

(2018) conducted a large scale study on dengue outbreaks in Malaysia by collecting mosquito samples using a device called ovitrap and rain gauge data. A linear regression model is used to identify the entomological, epidemiological and environmental drivers that contributed to the dengue outbreak of two locations in Selangor state. Salim (2021) develop a support vector machine model that incorporates environmental variables including temperature, wind speed, humidity, and rainfall to predict dengue outbreaks.

Hierarchical Bayesian spatio-temporal models for areal data are widely used in dengue disease mapping and prediction (Lowe et al., 2011; Lowe et al., 2013; Lowe et al., 2014; Stewart-Ibarra and Lowe, 2013). Using spatio-temporal model as a toolkit, Bayesian modelling can have a better capacity to handle explicit contribution from the covariates and latent spatio-temporal dependency. One popular choice of structured prior to capture spatially spill-over effect is a conditional intrinsic Gaussian autoregressive prior (CAR; Besag et al., 1991). Spatio-temporal autocorrelations are used to capture the facts that the disease rates in closer areal units and temporally close time periods tend to have more similar values (Lee et al., 2018). With the fact that spatial and temporal components are intrinsically interacted, a variety of CAR-based spatio-temporal model is developed to tackle many real-world applications as investigated by Bernardinelli et al. (1995), Knorr-Held (2000), Napier et al. (2016), Rushworth et al. (2014), Rushworth et al. (2017) and Sahu (2022).

ENSO is a global scale of climate variation where the cycles have lasted between two and seven years. Several previous studies have found significant associations between dengue outbreaks and ENSO in some specific study regions (Kovats et al., 2003). The oscillations of the sea surface temperature (SST) in different regions in the equatorial Pacific are used to define ENSO (Rasmusson and Carpenter, 1982). Ashok et al. (2007) define the anomalous warming events that occur in the central equatorial Pacific (Niño4 region) as an alternative type of El Niño called El Niño Modoki which is different from the conventional study region Niño3.4 (5°N–5°S, 170°W–120°W). (McGregor and Ebi, 2018) highlight that the contrasting rainfall fields for conventional El Niño and El Niño Modoki events hint at potential spatio-temporal inconsistencies of ENSO–health associations. Salimun et al. (2014) find that, although displayed much warmer SST anomalies in the Indian Ocean and regional seas in the Maritime Continent, the impact on the winter rainfall during conventional El Niño in boreal winter season over Peninsular Malaysia is minimal but significant higher during El Niño Modoki. Tangang et al. (2017) show that, during winter, a strong La Niña leads to a significant decrease in wet precipitation extremes over the Peninsular Malaysia due to the anomalous cyclonic circulation over strong La Niña. Nevertheless, Hanley et al. (2003) demonstrate that Niño4 index is more relevant to La Niña but poorly explain El Niño whilst the Niño1 + 2 index has the opposite characteristics. These two SST indices altogether cover different types of ENSO and their impact on dengue transmission.

Most high impact weather in synoptic scale occurs where the atmosphere is experiencing rising motion. The vertical velocity measured by the omega equation is associated with high impact weather and cyclones (Dostalek et al., 2017). In a study of the impact of meteorological factors to the air pollution in China, Hou et al. (2018) indicate that the vertical velocity has a short-term influence on PM_{2.5} level in the Pearl River Delta. It is expected that the unobserved meteorological variable would add value to our understanding of environmental association with the disease.

Wong et al., 2011 use a lagged 22 day mean air temperature to capture the second generation gonotrophic cycle of *Aedes* mosquitoes to predict ovitrap index. (Cheong et al., 2013) study the effects of temperature, rainfall and wind speed in Selangor with emphasis on their lag times. The lag times of 51 days minimum daily temperature and 28 days bi-weekly cumulated rainfall present a positive association with dengue hospitalisations. The effect from mesoscale local temperature and rainfall is related to some other major synoptic climate oscillations which influences the regional climate of Malaysia such as Indian Ocean Dipole (IOD; Tangang et al., 2012). IOD can happen in conjunction with ENSO or independently. Hong and Jin (2014) discover that the IOD-ENSO interaction is the cause of the generation of Super El Niños. Hameed et al. (2018) also show that the IOD lagged Niño3.4 by three to six months.

The paper by Hameed et al. (2018) conjectures that air pollution, especially ozone, has a profound effect on the mosquito vectors. Thiruchelvam et al. (2018) study the relationships between air quality and dengue hospitalisations. It is asserted that the air pollution index (API) levels do not have a significant effect on the reported cases. However, ozone is proven to have a repellent effect on both *Aedes aegypti* and *Aedes albopictus* (Wan-Norafikah et al., 2016). The API used by the Malaysian government follows the Pollutant Standard Index (PSI; Swamee and Tyagi, 1999) by the United States Environmental Protection Agency (USEPA). It considers five pollutants namely carbon dioxide, ozone, nitrogen dioxide, sulphur dioxide and particulate matter with a diameter of less than 10 microns. Individual air quality scores of each pollutant are assigned according to a gold standard. The API is the highest of those given individual scores. Its impact on humans has been thoroughly studied but its applicability to dengue transmission is still questionable. Without knowing which pollutant it refers to, the lagged value of API is meaningless for studying dengue incidence rates. For this reason, it is worthwhile to investigate each of the pollutants separately.

The remainder of this paper is organised as follows. Section 2 describes the data used in this study. Section 3 illustrates the components considered in the Bayesian spatio-temporal models. Model implementation, validation, evaluation of the EWS are discussed in Section 4. A detailed discussion of the results and their implications is provided in Section 5.

2. Data

2.1. Data source

Our data set contains the weekly counts of hospital admissions for dengue fever (Y_{kt}) in Selangor State and two federal territories in Malaysia (indexed by k) from 2013 to 2019 (indexed by t) obtained from the Ministry of Health (MOH) Malaysia. We also use relevant demographic information obtained from the Department of Statistics Malaysia (DOSM). Nine Selangor districts and two federal territories, namely Kuala Lumpur and Putrajaya, constitute our study region of the Central Region.

Environmental variables such as air pollution index, ozone concentration level (in parts per billion) and temperature are provided

by the Malaysian Department of Environment (DOE), Ministry of Environment and Water, whilst rainfall information (in mm) is provided by the Malaysian Meteorological Department (MetMalaysia). Ozone concentration data from the federal territories of Kuala Lumpur and Putrajaya are not available to us. That is why we have replaced these missing values by imputing the average values from the adjacent districts.

The Niño4 and Niño1 + 2 SST indices (Huang et al., 2021) capturing sea surface temperature anomalies in the central equatorial Pacific region (5°N-5°S, 160°E-150°W) and the region of coastal South America (0°-10°S, 90°W-80°W) are obtained from NOAA Climate Prediction Center.

Gridded (2.5° × 2.5°) reanalysis daily mean vertical velocity in pressure coordinates obtained from the NCAR/NCEP Reanalysis (Kalnay et al., 1996) is aggregated into weekly scale according to the epidemiological week (Epi week) defined by MOH.

Finally, the administrative district areal boundaries are extracted from The Humanitarian Data Exchange and all studied districts are within one (2.5° × 2.5°) grid cell in the reanalysis dataset.

2.2. Exploratory data analysis

2.2.1. Basic characteristics

A total of 414,284 dengue fever cases are reported in the nine districts of Selangor and two federal territories from January 2013 to December 2019 in the Central Region. The total numbers of cases vary from 26,422 in 2013 to 87,967 in 2019. Since the outbreaks after summer in 2013, there is no clear annual trend until a severe upsurge in 2019 which surpassed three fold of the total cases in 2013 (Fig. 2).

2.2.2. Temporal evolution and lagged effect dependency

The seasonality of dengue incidence rate (DIR) across Selangor is not as obvious as in other geographical regions, such as Thailand, in the existing literature see e.g. Lowe et al. (2016). A weak but consistent seasonality can be seen found in the study period. The weekly mean DIR peaks in the winter and finds another peak in the summer (Fig. 3).

Preceded by a weak El Niño event in 2014, historical observation shows that, the unusual 2015–2016 El Niño was one of the strongest El Niño in history (Lian et al., 2017). The SST anomalies in the central equatorial Pacific acts as a good predictor of El Niño but Trenberth et al. (2002) also suggest that El Niño should be identified along with the difference between the Niño4 and Niño12 indices. The difference measures the gradient in SST anomalies between the central and eastern equatorial Pacific termed El Modoki (Ashok et al., 2007). Fig. 4 shows that both Niño4 and Niño12 indices peaked in late 2015 when a strong El Niño event occurred. An El Niño Modoki event also began in late 2018 where the eastern equatorial SST became cooler. The DIR is closely related to this upward trend of both Niño1 + 2 and Niño4 indices during El Niño (Fig. 4). Taking out the effect of Niño4, the partial correlation between DIR and Niño1 + 2 index is 0.0578 only, although the Niño1 + 2 and Niño4 are highly correlated (Fig. 5). The Niño indices, representing the SST in the central/eastern equatorial Pacific, exert a delayed effect on Peninsular Malaysia local climate.

Following Cheong et al. (2013), a distributed lag non-linear model (DLNM; Gasparrini et al., 2010; Gasparrini, 2011) is used as an exploratory tool. The number of dengue hospitalisation and environmental variables are aggregated into a set of single region time-series. The lag time with quartic B-splines for the predictors and lag stratifications are then evaluated through a DLNM analysis. The relative risk (RR) at 90% quantile of temperature, ozone, rainfall, ozone, omega, Niño1 + 2 and Niño4 reach their maximum at a lag of 1, 10, 7, 15, 28, and 6 weeks respectively (Fig. 6). Capturing the effect of La Niña, for Niño4, the RR at 10% quantile reaches a local minimum at the 10 week lag. The formation of ozone is heavily influenced by sunlight and temperature (Ghazali et al., 2010). Since

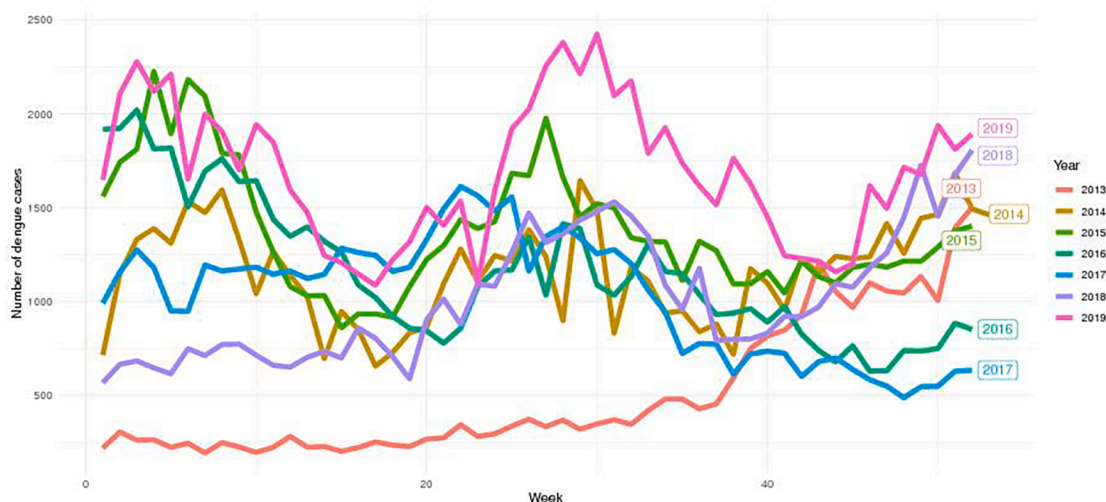


Fig. 2. Time series plots of dengue hospitalisations by year.

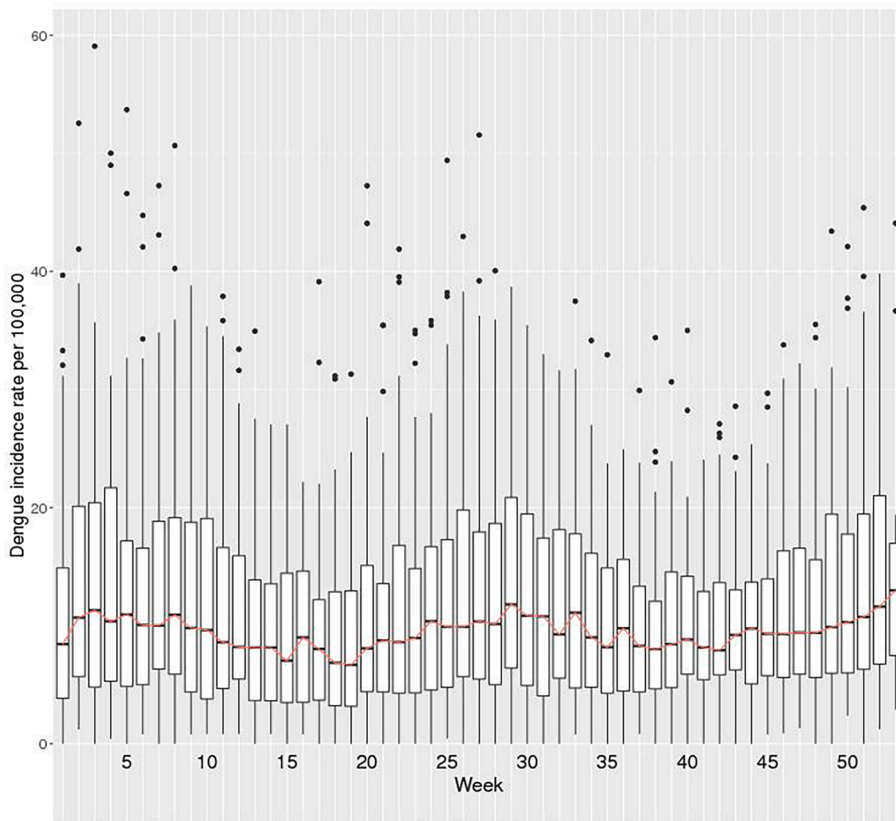


Fig. 3. Boxplots of weekly dengue hospitalisations in the Central Region, Malaysia, 2013–2019.

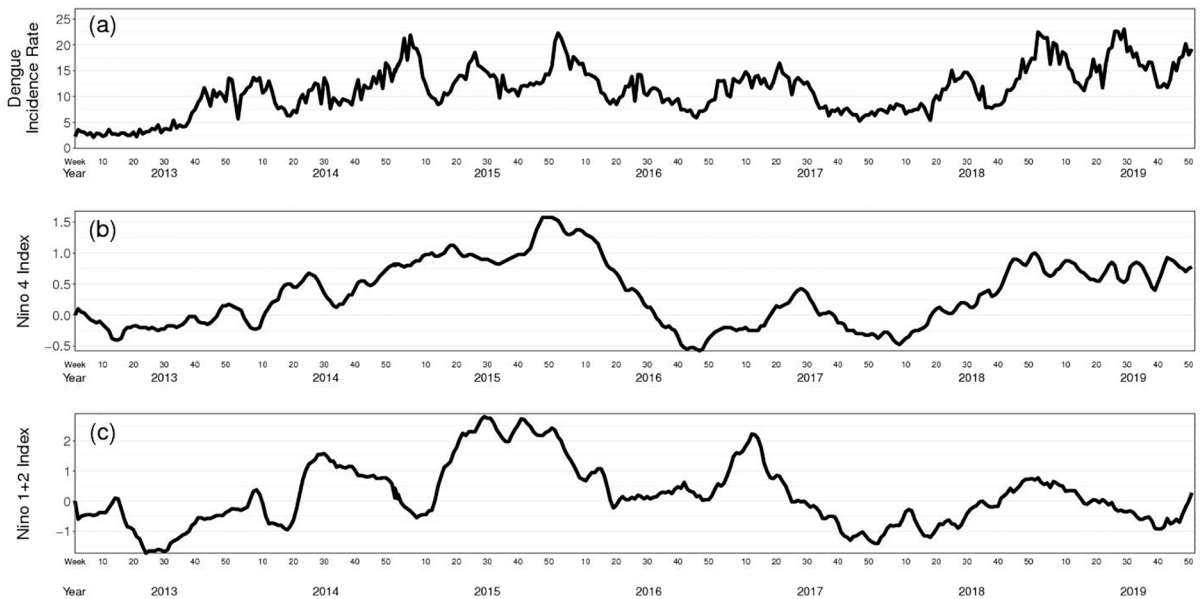


Fig. 4. Time series of (a) average DIR , (b) lagged four-week average of Niño4 index, (c) lagged four-week average of Niño 1 + 2 index in the Central Region, Malaysia for the period 2013–2019.

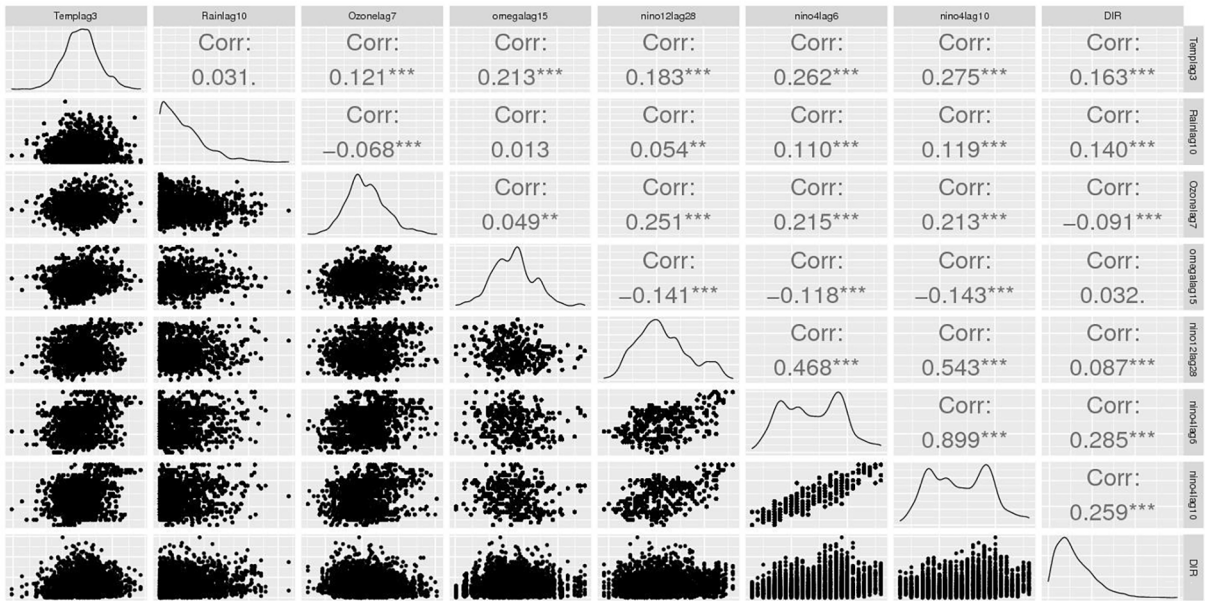


Fig. 5. Pairwise scatter plots of the DIR along with the covariates used in the models.

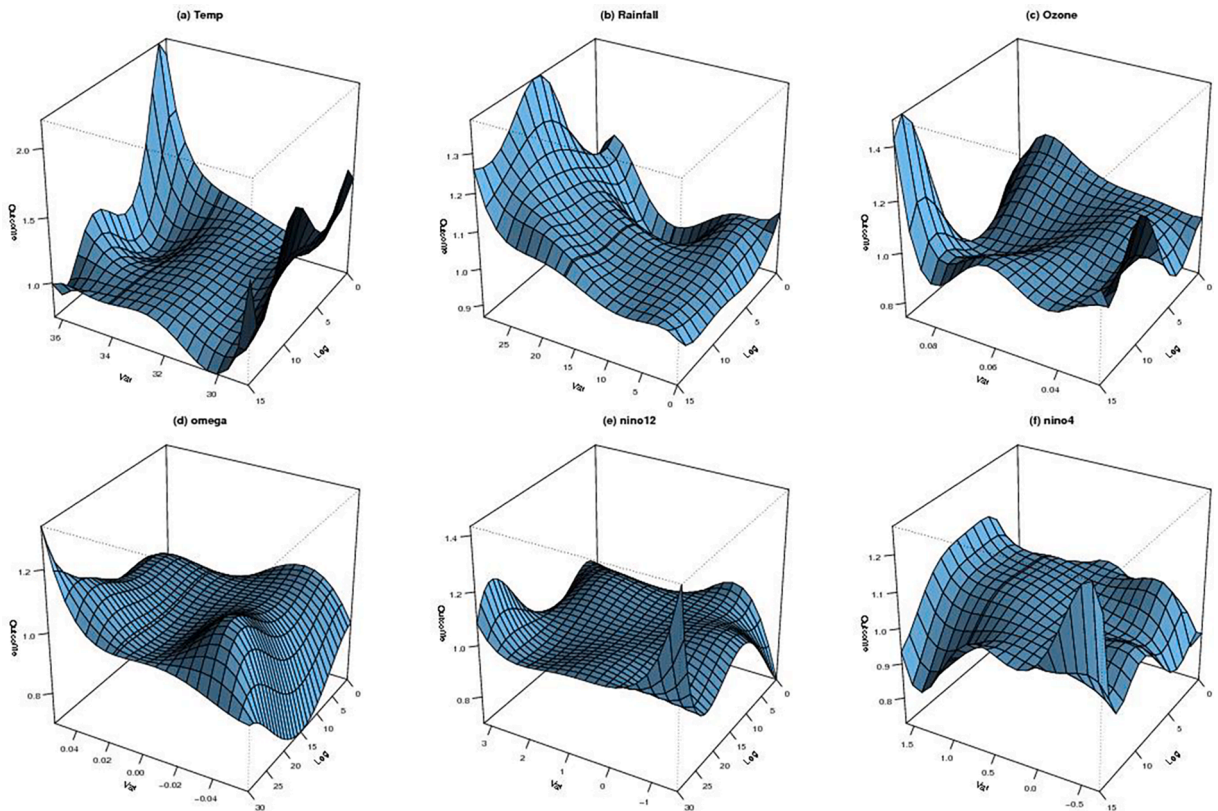


Fig. 6. RR surface of dengue hospitalisations by six variables. The variable *Temp* is the temperature, *Rain* is total precipitation, *Ozone* is the ground-level ozone concentration level, *omega* is vertical velocity of air motion derived from omega equation, *nino12* is the Niño1 + 2 index, *nino4* is the Niño4 index.

high temperature and presence of sunlight are the confounding factors, ozone has a strong immediate effect on dengue (Fig. 6c). The incremental cumulative RRs of rainfall has a monotonic increasing trend and have a long-range dependency throughout a long lag time. Fig. 6b shows a different pattern of short-term drought and wet scenario, with a very strong and immediate effect during drought (lag time 0–5) and a more delayed association with wet weather peak at a lag of 7 weeks. On the other hand, Fig. 6d shows a strong positive impact from the vertical velocity with a lag time of 9 weeks. It is understood that rainfall and vertical velocity are related to ground-level hydrology. A possible explanation is that drought makes people store water (Gagnon et al., 2001). Pontes et al. (2000) also suggest that household storage of water during the drought is correlated with the increase of *Aedes aegypti* vector abundance.

2.2.3. Regional variations

The Central Region area, especially for districts adjacent to Kuala Lumpur, became hyper-endemic of dengue transmission due to years of neglect (AhmadMeer et al., 2018). Fig. 7 plots a map of DIR, temperature, rainfall level, ground-level ozone concentration level from 2013 to 2019. The districts of Gombak, Petaling, Klang and Hulu Langat generally recorded higher DIR, mean temperature and ozone concentration level compared to other districts. However, the capital city Kuala Lumpur has shown significant reduction in cases although among the wettest in the region. This regional variation is regarded as a function of degree of urbanisation. An explicit formulation of this type of function is generally infeasible (Chandler, 2005). An entomological explanation to this variation is related to the abundance of the breeding areas of *Aedes aegypti* and *Aedes albopictus*. In an entomological surveillance study for two villages in Selangor, Noor et al. (2018) show that two species are indoor and outdoor breeders respectively. The transmission of the dengue vector is a combined effect of two species. Hence, this socio-economic difference between the two districts is a source of the step change in the cases count.

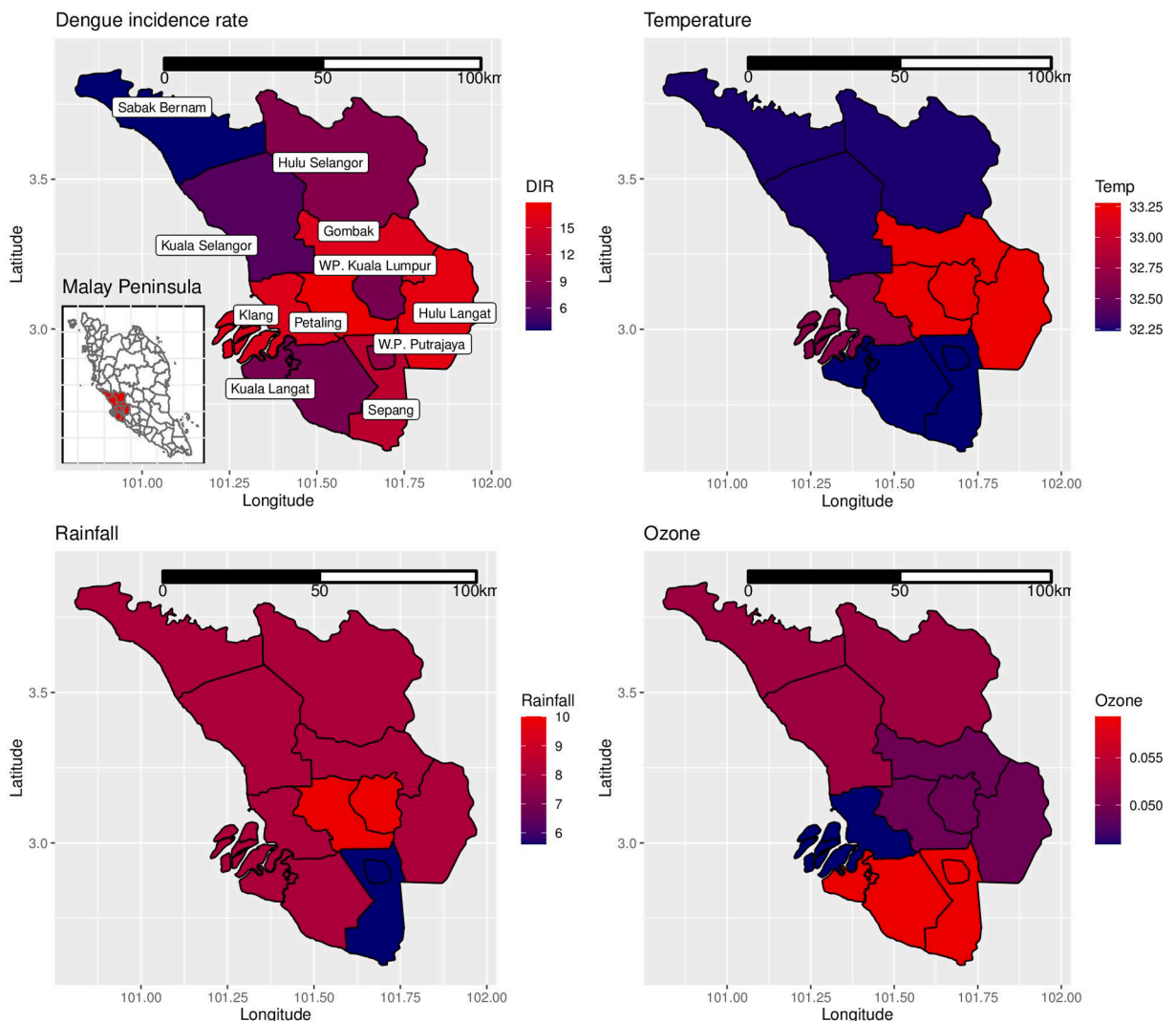


Fig. 7. Mean weekly DIR, temperature, rainfall, ground-level ozone concentration level from 2013 to 2019 by district.

Both the federal territories, Kuala Lumpur and Putrajaya, have lower rate of cases than the surrounding districts. This may be attributed to the increased activity of the enforcement agencies and anti-dengue campaigns conducted in the capital city (Hassan et al., 2012).

2.2.4. Spatial dependency

We first consider an independent Poisson generalised linear model for the disease count Y_{kt} defined in Section 2.1 and covariates found in the exploratory analysis through DLNM of the form:

$$\begin{aligned} Y_{kt} &\sim \text{Poisson}(\mu_{kt}), \\ \log(\mu_{kt}) &= \log(e_{kt}) + \log(\lambda_{kt}), \\ \log(\lambda_{kt}) &= \beta_0 + \beta_1 \text{Temp}_{k,t-3} + \beta_2 \text{Rain}_{k,t-10} + \beta_3 \text{Ozone}_{k,t-7} + \beta_4 \text{omega}_{t-15} + \\ &\quad \beta_5 \text{Nino12}_{t-28} + \beta_6 \text{Nino4}_{t-6} + \beta_7 \text{Nino4}_{t-10} + \beta_8 \text{Capital}_k, \end{aligned} \tag{1}$$

where μ_{kt} is the expected number of cases in the district k at time t , e_{kt} is the population size in the district k at time t , Niño12_t is the Niño1 + 2 index at time t , Niño4_t is the Niño4 index at time t , Ozone_{kt} is the ground-level ozone concentration level at time t in district k , Capital_k is a binary variable indicates whether the district is Kuala Lumpur, Temp_{kt} and Rain_{kt} is the temperature and total rainfall in the week t , omega_t is vertical velocity of air motion derived from weather model at time t . Note that here a lag of three weeks is used for temperature as one week is not a practical lag time for an EWS. We calculated the associated Moran's I statistics (Moran, 1950) for both spatial and spatio-temporal neighbourhood matrices. The statistics are 0.7075 and 0.72402 with p -values less than 0.001%. It indicates both the spatial and spatio-temporal variations have not adequately been captured through the generalised linear model.

2.2.5. Overdispersion

Overdispersion behaviour (Lawless, 1987) often exists in many disease count datasets. Lowe et al. (2011) suggested that a negative binomial model would adequately accommodate an extra-Poisson variation in the dengue case. We fit a negative binomial model using maximum likelihood estimation through a built-in R (R Core Team, 2021) function `glm.nb` with Eq. (1) and the estimated dispersion parameter is 2.61. The amount of overdispersion is quite high. It appears that the Poisson distribution is better suited to explain the "number of infected groups" rather than the total disease count. Represented as a compound Poisson distribution with a logarithmically distributed count per group (Quenouille, 1949), the negative binomial distribution turns out to be a reasonable model.

3. Model developments

The Bayesian hierarchical modelling approach is a flexible framework to describe the statistical properties in the previous Section. The components of the model formulation can be individually specified conditional to other parameters and data. In this Section, we will go through the key components of our candidate models.

3.1. Negative binomial regression

To overcome overdispersion, we use the negative binomial parametrisation which introduces r as a universal control parameter for overdispersion (Gelman et al., 1995). The probability mass function is given by:

$$p_{Y_{kt}}(y_{kt}) = \binom{y_{kt} + r - 1}{y_{kt}} \left(\frac{\mu_{kt}}{\mu_{kt} + r} \right)^{y_{kt}} \left(\frac{r}{\mu_{kt} + r} \right)^r, \tag{2}$$

where the mean and variance of the random variable are $E[Y_{kt}] = \mu_{kt}$ and $\text{Var}[Y_{kt}] = \mu_{kt} + \mu_{kt}^2/r$. As r goes to infinity, the distribution of Y_{kt} converges to the Poisson distribution.

3.2. Besag-York-Mollie model

The Besag-York-Mollie model (BYM; Besag et al., 1991; Besag and Kooperberg, 1995) specifies the additive relationship of the overall risk level as an intercept, the fixed effect by the covariates, the pure random effect θ_{kt} and the spatial variation component ϕ_k :

$$\log(\lambda_{kt}) = \beta_0 + \mathbf{x}'_{kt}\boldsymbol{\beta} + \theta_{kt} + \phi_k,$$

where θ_{kt} is a normally distributed unstructured error and ϕ_k is the structured error modelled by an intrinsic conditionally autoregressive model (ICAR). It has a conditional specification that is normally distributed with a mean equal to the average of its neighbours (ϕ_{k-j}) and its variance decreases as the number of neighbours d_k increases:

$$\phi_k | \phi_{k-j} \sim N \left(\frac{\sum_{k \sim j} \phi_k}{d_k}, \frac{\sigma_k^2}{d_k} \right).$$

An alternative form of BYM (BYM2) model proposed by Riebler et al. (2016), Simpson et al. (2017) and Morris et al., 2019 allows a clearer dependence structure with a spatial correlation parameter ranging from a full spatial neighbourhood dependent variation and

pure residual randomness in which the terms ϕ_k and θ_{kt} combined to one entity ϕ_{kt} :

$$\log(\lambda_{kt}) = \beta_0 + \mathbf{x}_{kt}'\boldsymbol{\beta} + \left(\phi_k^* \sqrt{\rho/s} + \theta_{kt}^* \sqrt{1 - \rho/s}\right)\sigma,$$

where $\text{logit}(\rho) \sim N(0, 1)$, ϕ_k^* is the ICAR model, $\theta_{kt}^* \sim N(0, 1)$, s is the scaling factor computed from the neighbourhood graph. Meanwhile, σ is the overall standard deviation of the two random effects.

3.3. Dynamic structure

Dynamic linear model (West and Harrison, 2006) is a special type of state-space model that enables a sequential model definition in the time series context and information propagates conditional to existing information. Taking a negative binomial model as a starting point, Eq. (2) is now a top-level observation equation, the spatio-temporal structure is defined as follows:

$$\begin{aligned} \text{Observation equation} \quad Y_{kt} &\sim \text{NB}(e_{kt} \lambda_{kt}, r), & r &\sim \Gamma(a, b) \\ \text{System equation} \quad \log(\lambda_{kt}) &= \alpha \log(\lambda_{k,t-1}) + \mathbf{x}_{kt}'\boldsymbol{\beta} + \phi_k + \omega_{kt}, & \boldsymbol{\phi} &\sim \text{BYM2}(s, \mathbf{W}), \\ & & \omega_{kt} &\sim N(0, \sigma_\omega^2), \\ \text{Initial information} \quad \log(\lambda_{k,0}) &\sim N[m_0, \sigma_0^2], & m_0 &\sim N(0, A), \\ & & \sigma_0^2 &\sim \text{IG}(a_0, b_0), \end{aligned}$$

where the overdispersion parameter follows a Gamma distribution with hyperparameters a and b , α is the autoregressive (AR) parameter to control temporal dependency between adjacent time points, ω_{kt} is the Gaussian distributed evolution error. Initial information is required for this temporal structure, s is the scaling parameter controls the proportion of a spatial and non-spatial variation, \mathbf{W} is the neighbourhood information formulated as a connected graph. The AR(1) model in the system equation could be understood as a moving average model of infinite order MA(∞) which aggregates all its lagged unexplained residuals as an additional piece of information. Sahu et al. (2009) impute the initial mean by the observed grand mean for a spatial point reference modelling problem. Alternatively, we choose to estimate the initial mean and set $\log(\lambda_{k0})$ to follow a normal distribution with a non-informative prior for both m_0 and σ_0^2 .

4. Modelling results

We consider five models with different levels of complexity (Table 1). The regression part of the model $\mathbf{x}_{kt}'\boldsymbol{\beta}$ is specified by the following setup:

$$\text{Temp}_{k,t-3} + \text{Rain}_{k,t-10} + \text{Ozone}_{k,t-7} + \text{omega}_{t-15} + \text{Nino12}_{t-28} + \text{Nino4}_{t-6} + \text{Nino4}_{t-10} + \text{Capital}_k.$$

No-U-turn sampler (NUTS) is used for Markov chain Monte Carlo (MCMC) sampling Hoffman et al. (2014). Nishio and Arakawa (2019) suggest that NUTS performance is better than Gibbs sampling due to the high effective sample sizes and low autocorrelations in some statistical applications.

4.1. Model assessment

Expected log pointwise predictive density (elpd; Vehtari et al. (2017)) is used to compare model performance. The criterion is estimated by leave-one-out cross-validation to mimic out-of-sample prediction data and the $\text{loaic} = -2\text{elpd}_{\text{loo}}$ will be used for reporting to provide typical Bayesian conventional scale of deviance information criterion (DIC; Spiegelhalter et al., 2002). The overall fit of each model is summarised in Table 1. The negative binomial family of models (Model B, C, D, E) outperforms the Poisson model (Model A). The negative binomial dynamic model (Model D) with the lowest loaic fits the data better than the other four models. The loaic of the negative binomial spatial dynamic model (Model E) and Model D differ by within one standard error.

Models B and C are well-specified because the effective number of parameters (pLOO; Vehtari et al., 2017) is smaller than the actual total number of parameters in the models whilst Model A is misspecified due to failure to capture overdispersion. Bürkner et al. (2020) show that elpd/loaic estimates are overly optimistic because the future observation has an influence to predictions of the past. Since the pLOO is the difference between elpd and the non-cross-validated log posterior predictive density, thus the pLOO is overestimated

Table 1

Model performance by the LOO information criterion (loaic), where pLOO is the estimated effective number of parameters of the model.

Model	System equation	loaic	pLOO
(A) Poisson	$\log(\lambda_{kt}) = \beta_0 + \mathbf{x}_{kt}'\boldsymbol{\beta}$	103074.7 ± 2601.8	251.3
(B) NB	$\log(\lambda_{kt}) = \beta_0 + \mathbf{x}_{kt}'\boldsymbol{\beta}$	35403.1 ± 165.4	9.0
(C) NB + spatial	$\log(\lambda_{kt}) = \beta_0 + \mathbf{x}_{kt}'\boldsymbol{\beta} + \phi_k$	33389.0 ± 180.9	20.6
(D) NB + dynamic	$\log(\lambda_{kt}) = \alpha \log(\lambda_{k,t-1}) + \mathbf{x}_{kt}'\boldsymbol{\beta}$	29180.6 ± 181.6	1053.7
(E) NB + spatial + dynamic	$\log(\lambda_{kt}) = \alpha \log(\lambda_{k,t-1}) + \mathbf{x}_{kt}'\boldsymbol{\beta} + \phi_k$	29180.9 ± 182.5	1056.6

under any dynamic setting. The evidence is inconclusive to determine whether Model D and E are well-specified or not. A further model validation procedure is required to check their validity.

4.2. Environmental and regional risk factors

Note that models A, B and C possess lower looic values when those are compared to the dynamic models. However, they preserve a considerable explanatory power. This is evident by taking a closer look at the coefficient estimates of Model C, which are presented in the form of relative risk (RR) in Table 2. The covariate lag(Nino4, 6) and lag(Nino4, 10) have a strong positive relationship with the disease, for each degree increase of the indices, the RRs increase by 46.87% and 8.44% respectively. Meanwhile, the Niño1 + 2 index of lag time 28 weeks decreases by 2.26% for each degree increase. The pollutant ozone has a negative effect on the disease. For every 10 ppb increase in concentration level, there is 3.13% decrease in dengue incidence. Kuala Lumpur has 40.40% expected cases lower than other regions. The local weather-related variables have a lesser impact on the RR with only 0.90% and -3.83% for a unit change in rainfall and temperature. The vertical velocity has a mild impact with only 3.61% RR increment for each 0.01 unit increase. The Niño4 index is the dominant factor and a negative temperature effect is seen as an adjustment to ENSO's impact. With a positive Niño4 and a negative Niño1 + 2 RR, although of different lag times, this provides a shred of indirect evidence that the central equatorial ENSO exerts a stronger impact on dengue disease than the convention ENSO.

4.3. Prediction for dengue epidemics and an early warning system

Four model validation criteria: root mean square error (RMSE), mean absolute error (MAE), continuous ranked probability score (CRPS; Hersbach, 2000), coverage at 95% nominal level (CVG; Sahu, 2022) are used for comparing out-of-sample model performance. The first two criteria evaluate the model performance in terms of mean response. The latter two are related to probabilistic forecasts. CRPS measures the discrepancy between the observations and the whole predictive distributions whilst the CVG detects underfitting and overfitting if the criterion drifts away to the nominal coverage probability of 95%. For the criteria RMSE, MAE and CRPS, better predictions correspond to their corresponding lower values. All these criteria values are calculated using the R package `bmstdr` developed by Sahu (2021).

Provided by European Centre for Medium-Range Weather Forecasts (ECMWF), a high resolution (HRES) integrated forecast system run every 12 h generates up to 10 days forecast (Owens and Hewson, 2018). In other words, replacing the daily mean temperature at lag time of 3 weeks by the forecasts provided by ECMWF, an EWS will have a capacity to produce outbreak detection signals at least four weeks in advance. One of the useful ways to disseminate outbreak detection is to use a visualisation called epidemic channel (Runge-Ranzinger, 2016). The probability of the dengue cases exceeding a certain threshold h in the region k at the time t is the complementary cumulative distribution of the random variable Y_{kt} . It can be computed by obtaining its empirical samples from the MCMC outputs with the equation:

$$P(Y_{kt} > h) \approx \frac{1}{M} \sum_{m=1}^M I_A(y_{kt}^{(m)}),$$

where $y_{kt}^{(m)}$ and M are the m th empirical samples of Y_{kt} and the number of empirical samples obtained from the MCMC algorithm respectively, $I(\cdot)$ is an indicator function that the set A is the condition when the empirical sample from the posterior predictive distribution exceeds the threshold h (i.e.: $A = \{x \in \mathbb{N}_0; x > h\}$). Once the predictive probability with a threshold level between 0.08–0.2 (Bowman, 2016) for future dengue cases exceeds a certain alarm value (e.g.: cases more than two standard deviation of the five-year average), an alarm signal forms when the weekly case numbers enter the “alarm zone”.

An out-of-sample probability forecast for the weekly reported cases in 11 districts and federal territories in the first four weeks in 2019 is generated from all model candidates. Table 3 summarises the values of the four model validation criteria from the fitted models using 2013–2018 data. Model D is the best model in terms of RMSE and MAE. Model E, although not being optimal in the first three criteria, appears to be the most adequate model if we consider its CVG. The sensitivity and specificity (Bowman et al., 2016; Lowe et al., 2016) represent the hit rate and true negative rate of an EWS. In order to achieve the goal of identifying potential outbreaks with high sensitivity, the probability threshold level is set to a relatively small value. This is also due to the concern that a single miss of a disease outbreak is costly from a disease surveillance point of view.

A receiver operating characteristic (ROC) analysis shows how sensitivity changes with the specificity. The optimal probability threshold level which balances the trade-off between sensitivity and specificity is defined as the probability threshold of issuing an alert that maximises the product of sensitivity and specificity (CZ value; Liu, 2012). The ROC curve achieves the maximum CZ value at the point the probability threshold $p = 0.15$ as the out-of-sample forecasts considered in the first four weeks in 2019 (Fig. 8). Setting the probability threshold level to 0.15, it means the posterior predictive distribution at 85 percentile exceeds the predefined alarm values of the reported cases greater than two standard deviations of the five-year average at each district driving an alarm signal. Using the same out-of-sample probability estimates for model validation statistics, Model E exhibits the highest sensitivity and a moderate specificity. A careful look at both Model D and E shows that the differences among the models with regard to the looic and model validation criteria are quite small. Model E appears to be preferable after evaluating the overall performance measures.

We find that the Niño4 index provides a high predictability for the number of dengue cases. It makes a longer-term prediction of dengue outbreaks feasible. Meng et al. (2020) show that a complexity-based approach allows us to forecast the magnitude of an ENSO

Table 2
Parameter estimates of RR for the negative binomial spatial model (Model C).

	RR	Credible Interval	
		2.5%	97.5%
lag(Temp, 3)	0.9617	0.9471	0.9761
lag(Rain, 10)	1.0090	1.0067	1.0114
lag(Ozone, 7) (10ppb)	0.9687	0.9561	0.9813
lag(omega, 15) (0.01Pas ⁻¹)	1.0361	1.0267	1.0456
lag(Nino12, 28)	0.9774	0.9602	0.9947
lag(Nino4, 6)	1.4687	1.3698	1.5733
lag(Nino4, 10)	1.0844	1.0081	1.1608
Capital	0.5960	0.1839	1.9244

Table 3
Model validation statistics and performance measures of early warning system derived from five models, A-E.

Model	RMSE	MAE	CRPS	CVG	Sensitivity	Specificity
(A) Poisson	222.63	203.62	7.97	2.27%	52.94%	44.44%
(B) NB	151.86	91.88	40.62	81.82%	52.94%	51.85%
(C) NB + spatial	135.81	84.50	32.69	72.73%	5.88%	88.89%
(D) NB + dynamic	44.72	28.92	29.08	100.00%	94.12%	62.96%
(E) NB + spatial + dynamic	51.84	33.44	38.03	95.45%	94.42%	70.37%

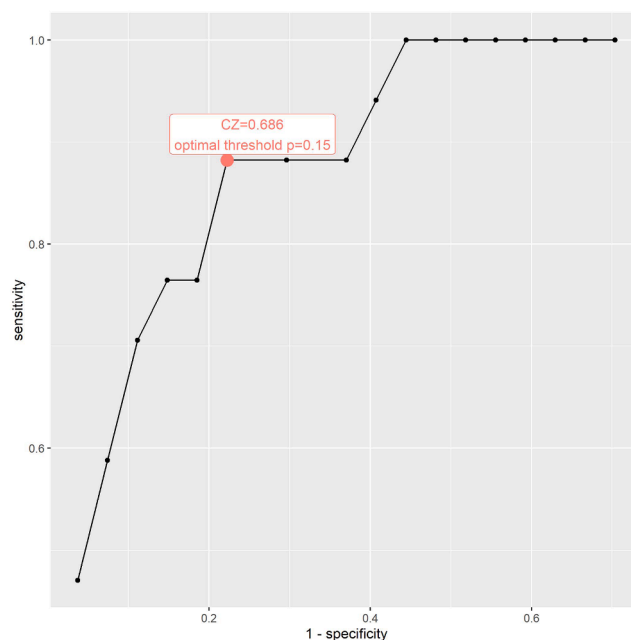


Fig. 8. The receiver operator characteristic curve for the number of dengue cases forecasts for the first four weeks in 2019 in the Central region of Malaysia.

event one year in advance. Ham et al. (2019) utilise a convolution neural network (CNN) to predict zonal SST (in their example, Niño3.4 region) by learning from historical simulations of a multi-model ensemble (Bellenger et al., 2014). Ballpark figures generated from a simpler EWS with ENSO information can be then assessed by the government agency. A longer-term climate uncertainty analysis (Yip et al., 2011; Northrop and Chandler, 2014) can be easily plugged into a disease mapping setting (e.g.: Baker et al., 2021).

5. Discussion

This paper presents a Bayesian spatio-temporal modelling framework leading to a full implementation of an EWS for dengue outbreaks from upstream data source to production. We propose to utilise some global and regional climate observation and variables derived from reanalysis data for a more accurate forecast. Exploratory data analysis methods show a long range of lag times is required

for some synoptic scale meteorological variables namely Niño12 and Niño4 indices. Our proposed holistic assessment goes beyond a single cross-validation metric. The whole assessment consists of calculation of out-of-sample predictive accuracy in multiple ways and alert signal evaluation. The methodology developed in this study can potentially be used to build a similar EWS in other countries or regions in Malaysia.

Dissimilar to horizontal wind, the vertical motion is often neglected due to its unobserved nature. In contrast to the environmental variables considered in previous studies, this study also considers a vertical velocity of air motion derived from reanalysis data and reveals to have a mild effect to the epidemics. Similar to many other studies, temperature and rainfall are used in the regression formula. Contrary to the findings of other studies (e.g. Lowe et al., 2016), the estimated coefficient of lagged temperature is negative. This appears to be a case of a local adjustment to a larger scale regional effect dominated by ENSO. The estimate of lagged ozone ties well with the biological argument based on *Aedes*'s gonotrophic cycle in Wong et al. (2011).

The RR estimates from the Model C exhibit that strong lagged anomalous warming in the Niño4 region has a strong positive effect on dengue hospitalisations. Consistent with our present findings, Gagnon et al. (2001) also report a significant positive correlation between El Niño and dengue epidemics in multiple countries. With a less-than-one RR for the lagged Niño1 + 2 index, cooling in the eastern tropical Pacific contributes to the increased dengue. This distinct relationship suggests that both El Niño and El Niño Modoki play a role in the epidemics. A previous study by Petrova et al. (2019) mentions that dengue epidemics can be associated with different teleconnections for different time lags. Dengue transmission is sensitive to the variability of rainfall due to its cumulative nature. A recent finding shows that a strong positive IOD which leads to drought (Amirudin et al., 2020) can be linked with the pre-existing El Niño Modoki with lead time up to one year (Doi et al., 2020). Our results align to these claims.

Due to data limitations, the impact from spatio-temporal variations of virus serotype are missing from the study. An anomalous upsurge happens twice in our study period, the first one occurred in the 2013 summer is verified by microbiology evidence (Ng et al., 2015). The second one observed in early 2019 is thought to be due to another serotype shift. A self-service EWS received a user feedback that change of predominant serotype alone attains a 50% of sensitivity of outbreak detection (Hussain-Alkhateeb et al., 2018). Currently a passive surveillance system is being applied in Malaysia. Hence, there is a very high possibility of underestimation of the public health burden because of the combined effect of delayed and under-reporting of cases. The actual number of cases can be four to five times of the number of reported cases (Shepard et al., 2012). Another data limitation due to the change of support problem (Gelfand et al., 2001) on the point-referenced rainfall and ozone concentration levels data concerns with the spatial misalignment of data used to infer district-level dengue counts. The point-to-district matching brings extra uncertainty into the prediction. However, as the data recorded are averaged out temporally by week, the effect can be seen as negligible.

Although the transmission dynamics is proven to be temperature-dependent (e.g.: Chen and Hsieh, 2012), the relationship between entomological parameters and the environment variables has not yet been clearly studied. A recent article by Sun et al. (2021) studies a residential-block-level dengue vector population in Singapore. It is shown that the *Aedes* abundance is heavily associated with the building age and managed vegetation cover. With modern geographic information systems (GIS) technology, these information can be incorporated in the future work.

Thanks to the flexibility of our modelling framework, joint modelling on multiple diseases is a possible methodological extension. Caminade et al. (2017) show the mosquito-borne transmission of Zika in South America is fuelled by the El Niño climate phenomenon. Funk et al. (2016) suggest, with their extensive sensitivity analysis, models for dengue transmission can be useful for handling the dynamics of Zika transmission. Held et al. (2005) demonstrate that the joint modelling approach on multiple diseases achieves a gain in precision of the RR estimates. Niriella et al. (2021) spot a sharp decrease in dengue cases for the second quarter of 2020 compared with pre-COVID-19 peaks in Sri Lanka. The drastic measures imposed by the Sri Lanka government regarding COVID-19 outbreaks help the reduction of hospitalisations. An identical pattern is also found during the first five phrases COVID-19 lockdown in Malaysia (Ong et al., 2021).

Contrast to the No-U-turn sampling scheme explained earlier, integrated nested Laplace approximation (INLA; Rue et al., 2009) algorithm is a fast alternative to our existing methodology (e.g.: Lowe et al., 2018; Lowe et al., 2021). The model A, B and C can be naturally fitted by INLA with a significant faster speed. However, The dynamic models (model D and E) are not currently implemented in the INLA R package (Lindgren and Rue, 2015) due to its complex state-space structure. A vector autoregression component (VAR; Spencer, 1993) can be also added to our current setup to incorporate dynamic lagged effect from other variables. Implemented in Stan language (Carpenter et al., 2017), conditional dependence such as spatial heterogeneity, temporal dynamics and covariate structure can be simply introduced and modified under the hierarchical Bayesian modelling paradigm, allowing for greater modelling flexibility.

6. Conclusion

The SST anomalies with a lag time of six weeks in the central equatorial Pacific region is the most crucial driver to the Central Region of Malaysia dengue hospitalisations. The EWS built on a Bayesian spatio-temporal hierarchical model yields reliable forecasts to help out dengue disease outbreak surveillance for at least four weeks in advance.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank Dr. Edith Cheng and Dr. Linda Lin for helpful comments. We gratefully acknowledge the Vector Borne Disease Sector, Disease Control Division, Ministry of Health of Malaysia, Department of Statistics of Malaysia, Department of Environment and Malaysian Meteorological Department for providing the information and dengue data used in this study.

References

- Abd Majid, N., Rainis, R., Sahani, M., Mohamed, A.F., Aziz, S.A.A.G., Nazi, N.M., 2021. Spatial pattern of dengue cases: An analysis in Bangi district, Selangor, Malaysia. *Geospatial Health*, 16 (1).
- Ahmad, R., Suzilah, I., Wan Najdah, W.M.A., Topek, O., Mustafakamal, I., Lee, H.L., 2018. Factors determining dengue outbreak in Malaysia. *PLoS One* 13 (2), e0193326.
- Ahmad Meer, A.M., Verasingam, K., Yen, Y.N., Chee, L.L., Chew, A.K., 2018. A review of dengue fever with special reference to Malaysia. *Asian J. Res. Infect. Diseases* 1–19.
- Amirudin, A.A., Salimun, E., Tangang, F., Juneng, L., Zuhairi, M., 2020. Differential influences of teleconnections from the Indian and Pacific Oceans on rainfall variability in Southeast Asia. *Atmosphere* 11 (9), 886.
- Ashok, K., Behera, S.K., Rao, S.A., Weng, H., Yamagata, T., 2007. El Niño modoki and its possible teleconnection. *J. Geophys. Res.: Oceans* 112 (C11).
- Baker, R.E., Yang, W., Vecchi, G.A., Metcalf, C.J.E., Grenfell, B.T., 2021. Assessing the influence of climate on wintertime SARS-CoV-2 outbreaks. *Nature Commun.* 12 (1), 1–7.
- Bellenger, H., Guilyardi, É., Leloup, J., Lengaigne, M., Vialard, J., 2014. ENSO representation in climate models: From cmip3 to cmip5. *Clim. Dyn.* 42 (7), 1999–2018.
- Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., Songini, M., 1995. Bayesian analysis of space–time variation in disease risk. *Stat. Med.* 14 (21–22), 2433–2443.
- Besag, J., Kooperberg, C., 1995. On conditional and intrinsic autoregressions. *Biometrika* 82 (4), 733–746.
- Besag, J., York, J., Mollié, A., 1991. Bayesian image restoration, with two applications in spatial statistics. *Ann. Inst. Stat. Math.* 43 (1), 1–20.
- Bhatt, S., et al., 2013. The global distribution and burden of dengue. *Nature* 496 (7446), 504–507.
- Bowman, L.R., et al., 2016. Alarm variables for Dengue outbreaks: A multi-centre study in Asia and Latin America. *PLoS One* 11 (6), e0157971.
- Bürkner, P.-C., Gabry, J., Vehtari, A., 2020. Approximate leave-future-out cross-validation for Bayesian time series models. *J. Stat. Comput. Simul.* 90 (14), 2499–2523.
- Caminade, C., Turner, J., Metelmann, S., Hesson, J.C., Blagrove, M.S., Solomon, T., Morse, A.P., Baylis, M., 2017. Global risk model for vector-borne transmission of Zika virus reveals the role of El Niño 2015. *Proc. Nat. Acad. Sci.* 114 (1), 119–124.
- Carpenter, B., et al., 2017. Stan: A probabilistic programming language. *J. Stat. Softw.* 76 (1), 1–32.
- Cazelles, B., Chavez, M., McMichael, A.J., Hales, S., 2005. Nonstationary influence of El Niño on the synchronous dengue epidemics in Thailand. *PLoS Med.* 2 (4), e106.
- Chandler, R.E., 2005. On the use of generalized linear models for interpreting climate variability. *Environmetrics* 16 (7), 699–715.
- Che Him, N., Kamardan, M.G., Rusiman, M.S., Sufahani, S., Mohamad, M., et al., 2018. Spatio-temporal modelling of dengue fever incidence in Malaysia. *J. Phys.: Conf. Ser.* 995, 012003. IOP Publishing.
- Che Him, N., Mohamad, N., Rusiman, M.S., Khalid, K., Shafi, M.A., 2018. Dengue incidence rate clustering by district in Selangor. *Int. J. Eng. Technol.* 7.4.30, 416–418.
- Chen, S.-C., Hsieh, M.-H., 2012. Modeling the transmission dynamics of dengue fever: implications of temperature effects. *Sci. Total Environ.* 431, 385–391.
- Cheong, Y.L., Burkart, K., Leitão, P.J., Lakes, T., 2013. Assessing weather effects on dengue disease in Malaysia. *Int. J. Environ. Res. Public Health* 10 (12), 6319–6334.
- Doi, T., Behera, S.K., Yamagata, T., 2020. Predictability of the super IOD event in 2019 and its link with El Niño Modoki. *Geophys. Res. Lett.* 47 (7) e2019GL086713.
- Dostalek, J.F., Schubert, W.H., DeMaria, M., 2017. Derivation and solution of the omega equation associated with a balance theory on the sphere. *J. Adv. Modeling Earth Syst.* 9 (8), 3045–3068.
- Funk, S., Kucharski, A.J., Camacho, A., Eggo, R.M., Yakob, L., Murray, L.M., Edmunds, W.J., 2016. Comparative analysis of dengue and Zika outbreaks reveals differences by setting and virus. *PLoS Neglected Tropical Diseases* 10 (12), e0005173.
- Gagnon, A.S., Bush, A.B., Smoyer-Tomic, K.E., 2001. Dengue epidemics and the El Niño southern oscillation. *Clim. Res.* 19 (1), 35–43.
- Gasparrini, A., 2011. Distributed lag linear and non-linear models in R: the package dlnm. *J. Stat. Softw.* 43 (8), 1.
- Gasparrini, A., Armstrong, B., Kenward, M.G., 2010. Distributed lag non-linear models. *Stat. Med.* 29 (21), 2224–2234.
- Gelfand, A.E., Zhu, L., Carlin, B.P., 2001. On the change of support problem for spatio-temporal data. *Biostatistics* 2 (1), 31–45.
- Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 1995. *Bayesian Data Analysis*. Chapman and Hall/CRC.
- Ghazali, N.A., Ramli, N.A., Yahaya, A.S., Yusof, N.F.F.M., Sansuddin, N., Al Madhoun, W.A., 2010. Transformation of nitrogen dioxide into ozone and prediction of ozone concentrations using multiple linear regression techniques. *Environ. Monit. Assess.* 165 (1), 475–489.
- Gubler, D.J., 1998. Dengue and dengue hemorrhagic fever. *Clin. Microbiol. Rev.* 11 (3), 480–496.
- Ham, Y.-G., Kim, J.-H., Luo, J.-J., 2019. Deep learning for multi-year ENSO forecasts. *Nature* 573 (7775), 568–572.
- Hameed, S.N., Jin, D., Thilakan, V., 2018. A model for super El Niños. *Nature Commun.* 9 (1), 1–15.
- Hanley, D.E., Bourassa, M.A., O'Brien, J.J., Smith, S.R., Spade, E.R., 2003. A quantitative evaluation of ENSO indices. *J. Clim.* 16 (8), 1249–1258.
- Hasan, S., Jamdar, S.F., Alalawi, M., Al Beajji, S.M.A.A., 2016. Dengue virus: A global human threat: Review of literature. *J. Int. Soc. Preventive Commun. Dentistry* 6 (1), 1.
- Hassan, H., Shohaimi, S., Hashim, N.R., 2012. Risk mapping of dengue in Selangor and Kuala Lumpur, Malaysia. *Geospatial Health* 7 (1), 21–25.
- Held, L., Natário, I., Fenton, S.E., Rue, H., Becker, N., 2005. Towards joint disease mapping. *Stat. Methods Med. Res.* 14 (1), 61–82.
- Henchal, E.A., Putnak, J.R., 1990. The dengue viruses. *Clin. Microbiol. Rev.* 3 (4), 376–396.
- Hersbach, H., 2000. Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather Forecasting* 15 (5), 559–570.
- Hii, Y.L., Zaki, R.A., Aghamohammadi, N., Rocklöv, J., 2016. Research on climate and dengue in Malaysia: a systematic review. *Curr. Environ. Health Rep.* 3 (1), 81–90.
- Hoffman, M.D., Gelman, A., et al., 2014. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* 15 (1), 1593–1623.
- Hong, L.-C., Jin, F.-F., 2014. A southern hemisphere booster of super El Niño. *Geophys. Res. Lett.* 41 (6), 2142–2149.
- Hou, X., Fei, D., Kang, H., Zhang, Y., Gao, J., 2018. Seasonal statistical analysis of the impact of meteorological factors on fine particle pollution in China in 2013–2017. *Nat. Hazards* 93 (2), 677–698.
- Huang, B., Liu, C., Banzon, V., Freeman, E., Graham, G., Hankins, B., Smith, T., Zhang, H.-M., 2021. Improvements of the daily optimum interpolation sea surface temperature (DOISST) version 2.1. *J. Clim.* 34 (8), 2923–2939.
- Hussain-Alkhateeb, L., et al., 2018. Early warning and response system (EWARS) for dengue outbreaks: Recent advancements towards widespread applications in critical settings. *PLoS one* 13 (5), e0196811.
- Johansson, M.A., Dominici, F., Glass, G.E., 2009. Local and global effects of climate on dengue transmission in Puerto Rico. *PLoS Neglected Tropical Diseases* 3 (2), e382.
- Kalnay, E., et al., 1996. The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* 77 (3), 437–472.
- Knorr-Held, L., 2000. Bayesian modelling of inseparable space-time variation in disease risk. *Stat. Med.* 19 (17–18), 2555–2567.

- Kovats, R.S., Bouma, M.J., Hajat, S., Worrall, E., Haines, A., 2003. El Niño and health. *Lancet* 362 (9394), 1481–1489.
- Lawless, J.F., 1987. Negative binomial and mixed poisson regression. *Can. J. Stat./La Revue Canadienne de Statistique* 209–225.
- Lee, D., Rushworth, A., Napier, G., 2018. Spatio-temporal areal unit modeling in R with conditional autoregressive priors using the CARBayesST package. *J. Stat. Softw.* 84 (1), 1–39.
- Lian, T., Chen, D., Tang, Y., 2017. Genesis of the 2014–2016 El Niño events. *Sci. China Earth Sci.* 60 (9), 1589–1600.
- Lindgren, F., Rue, H., 2015. Bayesian spatial modelling with r-inla. *J. Stat. Software* 63 (1), 1–25.
- Liu, X., 2012. Classification accuracy and cut point selection. *Stat. Med.* 31 (23), 2676–2686.
- Lowe, R., Bailey, T.C., Stephenson, D.B., Graham, R.J., Coelho, C.A., Carvalho, M.S., Barcellos, C., 2011. Spatio-temporal modelling of climate-sensitive disease risk: Towards an early warning system for dengue in Brazil. *Comput. Geosci.* 37 (3), 371–381.
- Lowe, R., Bailey, T.C., Stephenson, D.B., Jupp, T.E., Graham, R.J., Barcellos, C., Carvalho, M.S., 2013. The development of an early warning system for climate-sensitive disease risk with a focus on dengue epidemics in southeast Brazil. *Stat. Med.* 32 (5), 864–883.
- Lowe, R., Cazes, B., Paul, R., Rodó, X., 2016. Quantifying the added value of climate information in a spatio-temporal dengue model. *Stoch. Env. Res. Risk Assess.* 30 (8), 2067–2078.
- Lowe, R., et al., 2014. Dengue outlook for the World Cup in Brazil: an early warning model framework driven by real-time seasonal climate forecasts. *Lancet. Infect. Dis* 14 (7), 619–626.
- Lowe, R., et al., 2018. Nonlinear and delayed impacts of climate on dengue risk in Barbados: A modelling study. *PLoS Med.* 15 (7), e1002613.
- Lowe, R., et al., 2021. Combined effects of hydrometeorological hazards and urbanisation on dengue risk in Brazil: a spatiotemporal modelling study. *Lancet Planetary Health* 5 (4), e209–e219.
- McGregor, G.R., Ebi, K., 2018. El Niño Southern Oscillation (ENSO) and health: an overview for climate and health researchers. *Atmosphere* 9 (7), 282.
- Meng, J., Fan, J., Ludescher, J., Agarwal, A., Chen, X., Bunde, A., Kurths, J., Schellnhuber, H.J., 2020. Complexity-based approach for el Niño magnitude forecasting before the spring predictability barrier. *Proc. Nat. Acad. Sci.* 117 (1), 177–183.
- Moran, P.A., 1950. Notes on continuous stochastic phenomena. *Biometrika* 37 (1/2), 17–23.
- Morin, C.W., Comrie, A.C., Ernst, K., 2013. Climate and dengue transmission: evidence and implications. *Environ. Health Perspect.* 121 (11–12), 1264–1272.
- Morris, M., Wheeler-Martin, K., Simpson, S.J., Mooney, S.J., Gelman, A., DiMaggio, C., 2019. Bayesian hierarchical spatial models: Implementing the Besag York Mollie model in Stan. *Spatial Spatio-temporal Epidemiol.* 31 (100), 301.
- Mustafa, M., Rasotgi, V., Jain, S., Gupta, V., 2015. Discovery of fifth serotype of dengue virus (denv-5): A new public health dilemma in dengue control. *Med. J. Armed Forces India* 71 (1), 67–70.
- Napier, G., Lee, D., Robertson, C., Lawson, A., Pollock, K.G., 2016. A model to estimate the impact of changes in MMR vaccine uptake on inequalities in measles susceptibility in Scotland. *Stat. Methods Med. Res.* 25 (4), 1185–1200.
- Ng, L.-C., Koo, C., Mudin, R.N.B., Amin, F.M., Lee, K.-S., Kheong, C.C., 2015. 2013 dengue outbreaks in Singapore and Malaysia caused by different viral strains. *Am. J. Trop. Med. Hygiene* 92 (6), 1150.
- Niriella, M., Ediriweera, D., De Silva, A., Premarathna, B., Jayasinghe, S., de Silva, H., 2021. Dengue and leptospirosis infection during the coronavirus 2019 outbreak in Sri Lanka. *Trans. R. Soc. Trop. Med. Hyg.*
- Nishio, M., Arakawa, A., 2019. Performance of Hamiltonian Monte Carlo and No-U-Turn Sampler for estimating genetic parameters and breeding values. *Genetics Select. Evolut.* 51 (1), 1–12.
- Noor, A.A., Mohd Arif, A., Nazmi, W., Lee, H., 2018. Ovitrap surveillance of *Aedes aegypti* and *Aedes albopictus* in dengue endemic areas in Keramat and Shah Alam, Selangor in 2016. *IJUM Med. J. Malaysia* 17 (3).
- Northrop, P.J., Chandler, R.E., 2014. Quantifying sources of uncertainty in projections of future climate. *J. Clim.* 27 (23), 8793–8808.
- Ong, S.-Q., Ahmad, H., Mohd Ngesom, A.M., 2021. Implications of the COVID-19 lockdown on dengue transmission in Malaysia. *Infectious Disease Reports* 13 (1), 148–160.
- Owens, R., Hewson, T., 2018. *ECMWF Forecast User Guide*.
- Petrova, D., Lowe, R., Stewart-Ibarra, A., Ballester, J., Koopman, S.J., Rodó, X., 2019. Sensitivity of large dengue epidemics in Ecuador to long-lead predictions of El Niño. *Clim. Serv.* 15 (100), 096.
- Pontes, R., Freeman, J., Oliveira-Lima, J., Hodgson, J., Spielman, A., 2000. Vector densities that potentiate dengue outbreaks in a Brazilian city. *Am. J. Trop. Med. Hygiene* 62 (3), 378–383.
- Quenouille, M.H., 1949. A relation between the logarithmic, poisson, and negative binomial series. *Biometrics* 5 (2), 162–164.
- R Core Team, 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, R Foundation for Statistical Computing, url: <https://www.R-project.org/>.
- Rasmusson, E.M., Carpenter, T.H., 1982. Variations in tropical sea surface temperature and surface wind fields associated with the Southern Oscillation/El Niño. *Mon. Weather Rev.* 110 (5), 354–384.
- Riebler, A., Sørbye, S.H., Simpson, D., Rue, H., 2016. An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Stat. Methods Med. Res.* 25 (4), 1145–1165.
- Rue, H., Martino, S., Chopin, N., 2009. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *J. R. Stat. Soc.: Ser. B (Statistical Methodology)* 71 (2), 319–392.
- Runge-Ranzinger, S., 2016: *Technical handbook for dengue surveillance, dengue outbreak prediction/detection and outbreak response (model contingency plan)*. World Health Organization.
- Rushworth, A., Lee, D., Mitchell, R., 2014. A spatio-temporal model for estimating the long-term effects of air pollution on respiratory hospital admissions in Greater London. *Spatial Spatio-temporal Epidemiol.* 10, 29–38.
- Rushworth, A., Lee, D., Sarran, C., 2017. An adaptive spatiotemporal smoothing model for estimating trends and step changes in disease risk. *J. Roy. Stat. Soc.: Ser. C (Appl. Stat.)* 66 (1), 141–157.
- Sahu, S.K., 2021: *bmstdr: A r package for bayesian modeling of spatio-temporal data*. Technical Report, submitted.
- Sahu, S.K., 2022. Bayesian modeling of spatio-temporal data with R. Chapman and Hall/CRC Press, Boca Raton.
- Sahu, S.K., Yip, S., Holland, D.M., 2009. Improved space-time forecasting of next day ozone concentrations in the eastern US. *Atmos. Environ.* 43 (3), 494–501.
- Salim, N.A.M., et al., 2021. Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques. *Sci. Rep.* 11 (1), 1–9.
- Salimun, E., Tangang, F., Juneng, L., Behera, S.K., Yu, W., 2014. Differential impacts of conventional El Niño versus El Niño Modoki on Malaysian rainfall anomaly during winter monsoon. *Int. J. Climatol.* 34 (8), 2763–2774.
- Shepard, D.S., Undurraga, E.A., Lees, R.S., Halasa, Y., Lum, L.C.S., Ng, C.W., 2012. Use of multiple data sources to estimate the economic cost of dengue illness in Malaysia. *Am. J. Trop. Med. Hygiene* 87 (5), 796.
- Simpson, D., Rue, H., Riebler, A., Martins, T.G., Sørbye, S.H., 2017. Penalising model component complexity: A principled, practical approach to constructing priors. *Stat. Sci.* 32 (1), 1–28.
- Skae, F., 1902. Dengue fever in Penang. *Br. Med. J.* 2 (2185), 1581.
- Spencer, D.E., 1993. Developing a Bayesian vector autoregression forecasting model. *Int. J. Forecast.* 9 (3), 407–421.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. *J. R. Stat. Soc.: Ser. B (Statistical Methodology)* 64 (4), 583–639.
- Stewart-Ibarra, A.M., Lowe, R., 2013. Climate and non-climate drivers of dengue epidemics in southern coastal Ecuador. *Am. J. Trop. Med. Hygiene* 88 (5), 971.
- Sun, H., et al., 2021. Spatio-temporal analysis of the main dengue vector populations in Singapore. *Parasites Vectors* 14 (1), 1–11.
- Swamee, P.K., Tyagi, A., 1999. Formation of an air pollution index. *J. Air Waste Manage. Assoc.* 49 (1), 88–91.
- Tangang, F., Farzanmanesh, R., Mirzaei, A., Salimun, E., Jamaluddin, A.F., Juneng, L., 2017. Characteristics of precipitation extremes in Malaysia associated with El Niño and La Niña events. *Int. J. Climatol.* 37, 696–716.

- Tangang, F.T., Juneng, L., Salimun, E., Sei, K., Halimatun, M., et al., 2012. Climate change and variability over Malaysia: gaps in science and research information. *Sains Malaysiana* 41 (11), 1355–1366.
- Thiruchelvam, L., Dass, S.C., Zaki, R., Yahya, A., Asirvadam, V.S., 2018. Correlation analysis of air pollutant index levels and dengue cases across five different zones in Selangor, Malaysia. *Geospatial Health* 13 (1).
- Trenberth, K.E., Caron, J.M., Stepaniak, D.P., Worley, S., 2002. Evolution of el niño–southern oscillation and global atmospheric surface temperatures. *J. Geophys. Res.: Atmospheres* 107 (D8). AAC-5.
- Vehtari, A., Gelman, A., Gabry, J., 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* 27 (5), 1413–1432.
- Wan-Norafikah, O., Lee, H., Norazizah, A., Mohamad-Hafiz, A., 2016. Research note repellency effects of an ozone-producing air purifier against medically important insect vectors. *Trop. Biomed.* 33 (2), 396–402.
- West, M., Harrison, J., 2006. *Bayesian forecasting and dynamic models*. Springer Science & Business Media.
- Wong, M., Mok, H., Ma, H., Lee, M., Fok, M., 2011. A climate model for predicting the abundance of *Aedes* mosquitoes in Hong Kong. *Meteorol. Appl.* 18 (1), 105–110.
- Yang, X., Quam, M.B., Zhang, T., Sang, S., 2021. Global burden for dengue and the evolving pattern in the past 30 years. *J. Travel Med.* 28 (8), taab146.
- Yip, S., Ferro, C.A., Stephenson, D.B., Hawkins, E., 2011. A simple, coherent framework for partitioning uncertainty in climate predictions. *J. Clim.* 24 (17), 4634–4643.