Spatio-temporal detection of dengue outbreaks for the Malaysian Central Region using meteorological drivers at mesoscale and synoptic scale

Stan Yip*,
Department of Applied Mathematics,
The Hong Kong Polytechnic University, Hong Kong.

Norziha Che Him,
Department of Mathematics and Statistics,
Faculty of Applied Sciences and Technology,
Universiti Tun Hussein Onn Malaysia, Pagoh, Johor, Malaysia.

Nur Izzah Jamil,
Faculty of Computer and Mathematical Sciences,
Universiti Teknologi MARA, Negeri Sembilan branch, Rembau campus, Malaysia.

Daihai He,
Department of Applied Mathematics,
The Hong Kong Polytechnic University, Hong Kong.

Sujit K. Sahu,
School of Mathematical Sciences,
University of Southampton,
Southampton, UK.

September 23, 2021

Abstract

Background: The disease dengue is associated with both mesoscale and synoptic scale meteorology. Previous studies of this disease, especially for south-east Asia, have found very limited association between synoptic meteorological variables and the number of dengue hospitalisations. However, to tackle future severe outbreaks and to institute an early warning system, it will immensely beneficial to find and establish more clear association with rate of dengue hospitalisations and the most relevant meteorological variables.

Objectives: A rigorous Bayesian modelling framework is provided to identify the most important meteorological covariates and their lagged effects for developing an early warning system for dengue outbreaks in the Central Region of Malaysia.

Method: We obtain the dengue hospitalisations count data and the demographic information in the Central Region of Malaysia. Along with other mesoscale environmental measurements such as local temperature, precipitation and ozone concentration level, we also examine multiple synoptic scale Niño indices which are related to the phenomenon of El Niño Southern Oscillation and an unobserved meteorological variable derived from reanalysis data. A probabilistic early warning system is built based on a Bayesian spatio-temporal hierarchical model with a physically interpretable complex structure.

Results: Our study finds a 46.87% of increase in dengue hospitalisations due to one degree increase in the sea surface temperature anomalies in the central equatorial Pacific region with a lag time of six weeks. We also discover the existence of a mild association between the rate of cases and a distant lagged cooling effect of 28 weeks related to a phenomenon called El Niño Modoki. The proposed model also observes significant amount of association with some other meteorological parameters with disease rates. These associations are assessed by using an optimal Bayesian spatio-temporal model that outperforms other candidate models in terms of estimated out-of-sample predictive accuracy and performance in correctly issuing the warnings early.

*Corresponding Author. Address: Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, Email: stan.yip@polyu.edu.hk, Telephone number +852 3400 3909, Fax Number: +852 2362 9045

Discussion: With the novel spatial dynamic early warning system presented, our results show that the synoptic meteorological drivers can enhance short-term 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. 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). This viral infectious disease which 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 circulatory shock (Hasan et al., 2016). Its burden to the pubic health system is enormous. Bhatt et al. (2013) estimate there are 390 million total annual infections throughout the world. 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).

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 interanuual variability (Cazelles et al., 2005). The climatic impact to a dengue outbreak is also known to be cumulative and delayed (Lowe et al., 2018). Simple models cannot account for these complex spatial dynamic dependencies.

The Selangor state together with the two adjacent federal territories namely Kuala Lumpur and Putrajaya are collectively called the Central Region. It contributes most to the national dengue hospitalisation in Malaysia. In the region, the dengue infection rates have increased significantly in the past decade as reported by Abd Majid et al. (2021) and Salim et al. (2021). 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.

A significant relationship between dengue hospitalisations and covariates such as precipitation, temperature, number of monthly rain days and El Niño-Southern Oscillation (ENSO) phenomenon is found for 12 states of West Malaysia (Che Him et al., 2018b) . A similar study by Che Him et al. (2018a) identifies two distinct spatial clusters via two generalised additive models (GAM) for nine districts of the state of Selangor. Ahmad et al. (2018) conduct a

large scale study for 81 weeks including actively collected ovitrap and rain gauge data. A variant of 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 et al. (2021) develop a supporting vector machine model that incorporates environmental variables including temperature, wind speed, humidity, and rainfall to predict dengue outbreaks.

Bayesian spatio-temporal hierarchical modelling framework for areal data (Lowe et al., 2011, 2013, 2014; Stewart-Ibarra and Lowe, 2013) is widely used in dengue disease modelling and prediction. Using spatio-temporal model as a toolkit, it can have a better capacity to handle explicit contribution from covariates information 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). The spatio-temporal autocorrelation, as a source of information that 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, 2017) and Sahu (2021).

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. The model implementation, validation, evaluation of the EWS are discussed in Section 4. A detail discussion is delivered in Section 5.

2 Data

2.1 Environmental drivers related to dengue transmission

ENSO is a global scale of climate variation that the cycles have lasted between two and seven years. Several previous studies have found there is an association between epidemic dengue and ENSO in some world populations (Kovats et al., 2003). Different regions of sea surface temperature (SST) 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. Previously 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 on 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 positively associated with dengue hospitalisations.

This study conjectures that air pollution has a profound effect on the mosquito vectors especially ozone. Thiruchelvam et al. (2018) evaluate that the relationship between air quality and dengue hospitalisations. It is asserted that the air pollution index (API) levels do not have a significant effect on dengue hopitalizations. 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). The API is an index that the highest value of the sub-indices of five pollutants namely carbon dioxide, ozone, nitrogen dioxide, sulphur dioxide and particulate matter with a diameter of less than 10 microns taken. Its impact on humans has been thoroughly studied but its applicability to dengue transmission is questionable. Without knowing which pollutant it refers to, the lagged value of API is meaningless. For this reason, it is worthwhile to investigate individual pollutants separately.

2.2 Data source

The weekly counts of hospital admissions for dengue fever incidence (Y_{kt}) in Selangor State and two federal territories in Malaysia (indexed k) from 2013 to 2019 (indexed t) were obtained from the Ministry of Health (MOH) Malaysia. Relevant demographic information is obtained from the Department of Statistics Malaysia (DOSM). Nine Selangor districts and two federal territories namely Kuala Lumpur and Putrajaya have been considered.

Environmental variables such as air pollution index, ozone concentration level (in part per million) and temperature are provided by the Malaysian Department of Environment (DOE), Ministry of Environment and Water whilst precipitation information (in mm) is provided by the Malaysian Meteorological Department (MetMalaysia). All ozone concentration level information in the federal territories Kuala Lumpur and Putrajaya are missing. These are imputed by the average values of their 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 ($2.5° \times 2.5°$) 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 ($2.5° \times 2.5°$) grid cell in the reanalysis dataset.

2.3 Exploratory Data Analysis

2.3.1 Basic characteristics

A total of 414284 dengue fever cases was reported in 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 26422 in 2013 to 87967 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. 1).

2.3.2 Temporal evolution and lagged effect dependency

The seasonality of dengue incidence rate (DIR) across Selangor is not as obvious as in other geographical regions in existing literature such as Thailand in Lowe et al. (2016). The weekly mean DIR peaks in the winter and finds another peak in the summer (Fig. 2).

Preceded by a weak El Niño events in 2014, the unusual 2015-2016 El Niño was one of the strongest El Niño in history (Lian et al., 2017). The DIR is closely related to this upward trend of both Niño1+2 and Niño4 indices during El Niño (Fig. 3). 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. 4). These indices, refer to distant regions in the central/eastern equatorial Pacific, can be regarded as leading indicators of Peninsular Malaysia local climate.

Following Cheong et al. (2013), a distributed lag nonlinear model (DLNM; Gasparrini et al., 2009 and Gasparrini, 2011) is used as an exploratory tool. The dataset is aggregated into multiple single region time-series of dengue hospitalisations and environmental variables to evaluate the lag time with quartic B-splines for the predictors and lag stratifications. 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. 5). Capturing the effect of La Niña, for Niño4, the RR at 10% quantile reaches a local minimum at lag of 10 weeks. The formation of ozone is heavily influenced by sunlight and temperature (Ghazali et al., 2010). Since high temperature and presence of sunlight are the confounding factors, ozone has a strong immediate effect on dengue (Fig. 5c). The incremental cumulative RRs of rainfall has a monotonic increasing trend and have a long-range dependency throughout a long lag time. Fig. 5b 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. 5d 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.

2.3.3 Regional variations

The Central Region area, especially districts adjacent to Kuala Lumpur, becomes hyper-endemic for dengue transmission due to years of neglect (Ahmad Meer et al., 2018). Fig. 6 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 lower dengue hospitalisations 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 in-feasible (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 districts is a source of the step change in dengue hospitalisations.

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

2.3.4 Spatial dependency

Without explicit spatial and temporal dependent terms, consider a Poisson generalised linear model for the disease count Y_{kt} defined in Section 2.2 of the form:

$$\begin{aligned} Y_{kt} &\sim \text{Poisson}\left(\mu_{kt}\right), \\ \log \mu_{kt} &= \log \lambda_{kt} + \log e_{kt}, \\ \log \lambda_{kt} &= \beta_0 + \beta_1 \text{Temp}_{k,t-3} + \beta_2 \text{Rain}_{k,t-10} + \beta_3 \text{Ozone}_{k,t-7} + \beta_4 \text{omega}_{t-15} + \\ &\quad \beta_5 \text{Niño12}_{t-28} + \beta_6 \text{Niño4}_{t-6} + \beta_7 \text{Niño4}_{t-10} + \beta_8 \text{Capital}_k, \end{aligned} \tag{1}$$

where Y_{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 precipitation 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 calculate the associate Moran's I statistic (Moran, 1950) for the spatial neighbourhood matrices. The statistic is 0.7075 with a p-value of 0.001%. It indicates the spatial variation has not adequately been captured through the generalised linear model.

2.3.5 Overdispersion

Overdispersion behaviour (Lawless, 1987) often exists in many disease count datasets. It is suggested that a negative binomial model would nicely accommodate an extra-Poisson variation in the dengue case (Lowe et al., 2011). We fit a negative binomial model using maximum likelihood estimation through a built-in R (R Core Team, 2021) function glm.nb with Equation (1) and the estimated dispersion parameter is 2.61. The amount of overdispersion is quite high. Such a statistical property can be easily described by a person is more likely to be infected by disease through close contacts. 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 counts 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 in which introduces r as a universal control parameter for overdispersion (Gelman et al., 1995).

$$Y_{kt} \sim \mathrm{NB}(\mu_{kt}, r), \tag{2}$$

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

3.2 Besag-York-Mollié model

The Besag-York-Mollié model (BYM; Besag, York and Mollié, 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:

$$\log \lambda_{kt} = \beta_0 + x_{kt}\boldsymbol{\beta} + \theta_{kt} + \phi_k, \tag{3}$$

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 ($\phi_{k\sim j}$) and its variance decreased as the number of neighbours d_k increases:

$$p\left(\phi_k | \phi_{k\sim j}\right) = N\left(\frac{\sum_{k\sim j}\phi_k}{d_k}, \frac{\sigma_k^2}{d_k}\right). \tag{4}$$

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}:

$$\log \lambda_{kt} = \beta_0 + x_{kt}\boldsymbol{\beta} + \left(\phi_k^* \sqrt{\rho/s} + \theta_{kt}^* \sqrt{1-\rho/s}\right)\sigma, \tag{5}$$

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 two variations.

3.3 Dynamic structure

Dynamic linear model (West and Harrison, 2006) 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, the Equation (2) is now a top-level observation equation, the spatio-temporal structure is defined as follows:

$$\begin{array}{lll}
\text{Observation equation} & Y_{kt} \sim \text{NB}(\log \lambda_{kt} + \log e_{kt}, r), & r \sim \Gamma(a,b) \\
\text{System equation} & \log \lambda_{kt} = \alpha \log \lambda_{k,(t-1)} + x_{kt}\boldsymbol{\beta} + \phi_k + \omega_{kt}, & \boldsymbol{\phi} \sim BYM2\,(s, \mathbf{W})\,, \\
 & & \omega_{kt} \sim N(0, \sigma_\omega^2), \\
\text{Initial information} & \log \lambda_{k,0} \sim N[m_0, \sigma_0^2], & m_0 \sim N\,(0, A)\,, \\
 & & \sigma_0^2 \sim IG(a_0, b_0),
\end{array} \tag{6}$$

where the overdispersion parameter follows a Gamma distribution with hyperparameters a and b, α is the autoregressive 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, W is the neighbourhood information formulated as a connected graph. 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 $x_{kt}\boldsymbol{\beta}$ is specified by the following setup:

$$\text{Temp}_{k,t-3} + \text{Rain}_{k,t-10} + \text{Ozone}_{k,t-7} + \text{omega}_{t-15} + \text{Niño12}_{t-28} + \text{Niño4}_{t-6} + \text{Niño4}_{t-10} + \text{Capital}_k.$$

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 looic $= -2\widehat{\text{elpd}}_{\text{loo}}$ 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) outperform the Poisson model (Model A). The negative binomial dynamic model (Model D) with the lowest looic fits the data better than the other four models. The looic of the negative binomial spatial dynamic model (Model E) and Model D differ by within one standard error.

The Model B and Model 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/looic 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 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.

Model	System equation	looic	pLOO
(A) Poisson	$\log \lambda_{kt} = \beta_0 + x_{kt}\boldsymbol{\beta}$	103074.7 ± 2601.8	251.3
(B) NB	$\log \lambda_{kt} = \beta_0 + x_{kt}\boldsymbol{\beta}$	35403.1 ± 165.4	9.0
(C) NB + spatial	$\log \lambda_{kt} = \beta_0 + x_{kt}\boldsymbol{\beta} + \phi_k$	33389.0 ± 180.9	20.6
(D) NB + dynamic	$\log \lambda_{kt} = \alpha \log \lambda_{k,(t-1)} + x_{kt}\boldsymbol{\beta}$	29180.6 ± 181.6	1053.7
(E) NB + spatial + dynamic	$\log \lambda_{kt} = \alpha \log \lambda_{k,(t-1)} + x_{kt}\boldsymbol{\beta} + \phi_k$	29180.9 ± 182.5	1056.6

Table 1: Model performance by the LOO information criterion (looic), where pLOO is the estimated effective number of parameters of the model.

4.2 Environmental and regional risk factors

Although models A, B and C possess lower looic value, compared to dynamic models, they preserve a considerable explanatory power. Taking a closer look at the coefficient estimates

of Model C, the coefficient estimates in the form of relative risk (RR) is shown in Table 2. The covariate lag (Niño4, 6) and lag (Niño4, 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 parts per billion 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 is a shred of indirect evidence that the central equatorial ENSO exerts a stronger impact on dengue disease than the convention ENSO.

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

		Credible Interval	
	RR	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) (0.01Pa s^{-1})	1.0361	1.0267	1.0456
lag (Niño12, 28)	0.9774	0.9602	0.9947
lag (Niño4, 6)	1.4687	1.3698	1.5733
lag (Niño4, 10)	1.0844	1.0081	1.1608
Capital	0.5960	0.1839	1.9244

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, 2021) 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).

Atmospheric model high resolution (HRES) provided by European Centre for Medium-Range Weather Forecasts (ECMWF) 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). Once the predictive probability with a threshold level between $0.08 - 0.2$ (Bowman et al., 2016) for future dengue hospitalisations exceeds a certain alarm value (e.g.: dengue hospitalisations above 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 dengue hospitalisations 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, it 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. From a disease surveillance point a view, a single miss of a disease outbreak is costly, in order to achieve the goal of identifying potential outbreaks with high sensitivity, the probability threshold level is set to a relatively small value. Setting the probability threshold level to 0.15, it means the posterior predictive distribution at 85 percentile exceeds the predefined alarm values of dengue hospitalisations 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 a more preferable model after evaluating the overall performance measures.

We found that the most important RR comes from Niño4. It makes a longer-term prediction of dengue outbreaks feasible. Meng et al. (2020) shows that a complexity-based approach allows us to forecast the magnitude of an ENSO 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).

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%

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. Our results align to this

claim.

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).

Although the transmission dynamics is proven to be temperature-dependent (e.g.: Chen et al., 2012), the relationship between entomological parameters and the environment variables has not yet been clearly studied. A recent article by Sun et al. (2021) study 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 the modelling framework in this research article, 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 fueled 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 hospitalisations 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 dengue hospitalisations. An identical pattern is also found during the first five phrases COVID-19 lockdown in Malaysia (Ong et al., 2021). A vector autoregression component (VAR; Spencer, 1993) can be added to our current setup to incorporate lagged effect dynamically 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 Malaysian Central Region 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.

Acknowledgments

The authors thank Dr Edith Cheng 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., R. Rainis, M. Sahani, A. F. Mohamed, S. A. A. G. Aziz, and N. M. Nazi, 2021: Spatial pattern of dengue cases: An analysis in Bangi district, Selangor, Malaysia. *Geospatial Health*, **16 (1)**.

Ahmad, R., I. Suzilah, W. M. A. Wan Najdah, O. Topek, I. Mustafakamal, and H. L. Lee, 2018: Factors determining dengue outbreak in Malaysia. *PLoS One*, **13 (2)**, e0193 326.

Ahmad Meer, A. M., K. Verasingam, Y. N. Yen, L. L. Chee, and A. K. Chew, 2018: A review of dengue fever with special reference to Malaysia. *Asian Journal of Research in Infectious Diseases*, 1–19.

Ashok, K., S. K. Behera, S. A. Rao, H. Weng, and T. Yamagata, 2007: El niño modoki and its possible teleconnection. *Journal of Geophysical Research: Oceans*, **112 (C11)**.

Baker, R. E., W. Yang, G. A. Vecchi, C. J. E. Metcalf, and B. T. Grenfell, 2021: Assessing the influence of climate on wintertime SARS-CoV-2 outbreaks. *Nature communications*, **12 (1)**, 1–7.

Bellenger, H., É. Guilyardi, J. Leloup, M. Lengaigne, and J. Vialard, 2014: Enso representation in climate models: From cmip3 to cmip5. *Climate Dynamics*, **42 (7)**, 1999–2018.

Bernardinelli, L., D. Clayton, C. Pascutto, C. Montomoli, M. Ghislandi, and M. Songini, 1995: Bayesian analysis of space—time variation in disease risk. *Statistics in Medicine*, **14 (21-22)**, 2433–2443.

Besag, J. and C. Kooperberg, 1995: On conditional and intrinsic autoregressions. *Biometrika*, **82 (4)**, 733–746.

Besag, J., J. York, and A. Mollié, 1991: Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, **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)**, e0157 971.

Bürkner, P.-C., J. Gabry, and A. Vehtari, 2020: Approximate leave-future-out cross-validation for Bayesian time series models. *Journal of Statistical Computation and Simulation*, **90 (14)**, 2499–2523.

Caminade, C., J. Turner, S. Metelmann, J. C. Hesson, M. S. Blagrove, T. Solomon, A. P. Morse, and M. Baylis, 2017: Global risk model for vector-borne transmission of Zika virus reveals the role of El Niño 2015. *Proceedings of the National Academy of Sciences*, **114 (1)**, 119–124.

Carpenter, B., et al., 2017: Stan: A probabilistic programming language. *Journal of Statistical Software*, **76 (1)**, 1–32.

Cazelles, B., M. Chavez, A. J. McMichael, and S. Hales, 2005: Nonstationary influence of El Nino on the synchronous dengue epidemics in thailand. *PLoS Medicine*, **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., M. G. Kamardan, M. S. Rusiman, S. Sufahani, M. Mohamad, et al., 2018a: Spatio-temporal modelling of dengue fever incidence in Malaysia. *Journal of Physics: Conference Series*, IOP Publishing, Vol. 995, 012003.

Che Him, N., N. Mohamad, M. S. Rusiman, K. Khalid, and M. A. Shafi, 2018b: Dengue incidence rate clustering by district in Selangor. *International Journal of Engineering & Technology*, **7.4.30**, 416–418.

Chen, S.-C. and M.-H. Hsieh, 2012: Modeling the transmission dynamics of dengue fever: implications of temperature effects. *Science of the Total Environment*, **431**, 385–391.

Cheong, Y. L., K. Burkart, P. J. Leitão, and T. Lakes, 2013: Assessing weather effects on dengue disease in Malaysia. *International Journal of Environmental Research and Public Health*, **10 (12)**, 6319–6334.

Dostalek, J. F., W. H. Schubert, and M. DeMaria, 2017: Derivation and solution of the omega equation associated with a balance theory on the sphere. *Journal of Advances in Modeling Earth Systems*, **9 (8)**, 3045–3068.

Funk, S., A. J. Kucharski, A. Camacho, R. M. Eggo, L. Yakob, L. M. Murray, and W. J. Edmunds, 2016: Comparative analysis of dengue and Zika outbreaks reveals differences by setting and virus. *PLoS Neglected Tropical Diseases*, **10 (12)**, e0005 173.

Gagnon, A. S., A. B. Bush, and K. E. Smoyer-Tomic, 2001: Dengue epidemics and the El Niño southern oscillation. *Climate Research*, **19 (1)**, 35–43.

Gasparrini, A., 2011: Distributed lag linear and non-linear models in R: the package dlnm. *Journal of Statistical Software*, **43 (8)**, 1.

Gasparrini, A., B. Armstrong, and M. G. Kenward, 2010: Distributed lag non-linear models. *Statistics in Medicine*, **29 (21)**, 2224–2234.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin, 1995: *Bayesian Data Analysis*. Chapman and Hall/CRC.

Ghazali, N. A., N. A. Ramli, A. S. Yahaya, N. F. F. M. Yusof, N. Sansuddin, and W. A. Al Madhoun, 2010: Transformation of nitrogen dioxide into ozone and prediction of ozone concentrations using multiple linear regression techniques. *Environmental Monitoring and Assessment*, **165 (1)**, 475–489.

Gubler, D. J., 1998: Dengue and dengue hemorrhagic fever. *Clinical Microbiology Reviews*, **11 (3)**, 480–496.

Ham, Y.-G., J.-H. Kim, and J.-J. Luo, 2019: Deep learning for multi-year ENSO forecasts. *Nature*, **573 (7775)**, 568–572.

Hanley, D. E., M. A. Bourassa, J. J. O'Brien, S. R. Smith, and E. R. Spade, 2003: A quantitative evaluation of ENSO indices. *Journal of Climate*, **16 (8)**, 1249–1258.

Hasan, S., S. F. Jamdar, M. Alalowi, and S. M. A. A. Al Beaiji, 2016: Dengue virus: A global human threat: Review of literature. *Journal of International Society of Preventive & Community Dentistry*, **6 (1)**, 1.

Hassan, H., S. Shohaimi, and N. R. Hashim, 2012: Risk mapping of dengue in Selangor and Kuala Lumpur, Malaysia. *Geospatial Health*, **7 (1)**, 21–25.

Held, L., I. Natário, S. E. Fenton, H. Rue, and N. Becker, 2005: Towards joint disease mapping. *Statistical Methods in Medical Research*, **14 (1)**, 61–82.

Henchal, E. A. and J. R. Putnak, 1990: The dengue viruses. *Clinical microbiology reviews*, **3 (4)**, 376–396.

Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, **15 (5)**, 559–570.

Hii, Y. L., R. A. Zaki, N. Aghamohammadi, and J. Rocklöv, 2016: Research on climate and dengue in Malaysia: a systematic review. *Current Environmental Health Reports*, **3 (1)**, 81–90.

Hoffman, M. D., A. Gelman, et al., 2014: The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, **15 (1)**, 1593–1623.

Hou, X., D. Fei, H. Kang, Y. Zhang, and J. Gao, 2018: Seasonal statistical analysis of the impact of meteorological factors on fine particle pollution in China in 2013–2017. *Natural Hazards*, **93 (2)**, 677–698.

Huang, B., C. Liu, V. Banzon, E. Freeman, G. Graham, B. Hankins, T. Smith, and H.-M. Zhang, 2021: Improvements of the daily optimum interpolation sea surface temperature (DOISST) version 2.1. *Journal of Climate*, **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)**, e0196 811.

Johansson, M. A., F. Dominici, and G. E. Glass, 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. *Bulletin of the American Meteorological Society*, **77 (3)**, 437–472.

Knorr-Held, L., 2000: Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine*, **19 (17-18)**, 2555–2567.

Kovats, R. S., M. J. Bouma, S. Hajat, E. Worrall, and A. Haines, 2003: El Niño and health. *The Lancet*, **362 (9394)**, 1481–1489.

Lawless, J. F., 1987: Negative binomial and mixed poisson regression. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, 209–225.

Lee, D., A. Rushworth, and G. Napier, 2018: Spatio-temporal areal unit modeling in R with conditional autoregressive priors using the CARBayesST package. *Journal of Statistical Software*, **84 (1)**, 1–39.

Lian, T., D. Chen, and Y. Tang, 2017: Genesis of the 2014–2016 El Niño events. *Science China Earth Sciences*, **60 (9)**, 1589–1600.

Lowe, R., T. C. Bailey, D. B. Stephenson, R. J. Graham, C. A. Coelho, M. S. Carvalho, and C. Barcellos, 2011: Spatio-temporal modelling of climate-sensitive disease risk: Towards an early warning system for dengue in Brazil. *Computers & Geosciences*, **37 (3)**, 371–381.

Lowe, R., T. C. Bailey, D. B. Stephenson, T. E. Jupp, R. J. Graham, C. Barcellos, and M. S. Carvalho, 2013: The development of an early warning system for climate-sensitive disease risk with a focus on dengue epidemics in southeast Brazil. *Statistics in Medicine*, **32 (5)**, 864–883.

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. *The Lancet Infectious Diseases*, **14 (7)**, 619–626.

Lowe, R., et al., 2018: Nonlinear and delayed impacts of climate on dengue risk in Barbados: A modelling study. *PLoS medicine*, **15 (7)**, e1002 613.

McGregor, G. R. and K. Ebi, 2018: El Niño Southern Oscillation (ENSO) and health: an overview for climate and health researchers. *Atmosphere*, **9 (7)**, 282.

Meng, J., J. Fan, J. Ludescher, A. Agarwal, X. Chen, A. Bunde, J. Kurths, and H. J. Schellnhuber, 2020: Complexity-based approach for el niño magnitude forecasting before the spring predictability barrier. *Proceedings of the National Academy of Sciences*, **117 (1)**, 177–183.

Moran, P. A., 1950: Notes on continuous stochastic phenomena. *Biometrika*, **37 (1/2)**, 17–23.

Morin, C. W., A. C. Comrie, and K. Ernst, 2013: Climate and dengue transmission: evidence and implications. *Environmental Health Perspectives*, **121 (11-12)**, 1264–1272.

Morris, M., K. Wheeler-Martin, D. Simpson, S. J. Mooney, A. Gelman, and C. DiMaggio, 2019: Bayesian hierarchical spatial models: Implementing the Besag York Mollié model in stan. *Spatial and Spatio-temporal Epidemiology*, **31**, 100 301.

Mustafa, M., V. Rasotgi, S. Jain, and V. Gupta, 2015: Discovery of fifth serotype of dengue virus (denv-5): A new public health dilemma in dengue control. *Medical journal armed forces India*, **71 (1)**, 67–70.

Napier, G., D. Lee, C. Robertson, A. Lawson, and K. G. Pollock, 2016: A model to estimate the impact of changes in MMR vaccine uptake on inequalities in measles susceptibility in Scotland. *Statistical Methods in Medical Research*, **25 (4)**, 1185–1200.

Ng, L.-C., C. Koo, R. N. B. Mudin, F. M. Amin, K.-S. Lee, and C. C. Kheong, 2015: 2013 dengue outbreaks in Singapore and Malaysia caused by different viral strains. *The American Journal of Tropical Medicine and Hygiene*, **92 (6)**, 1150.

Niriella, M., D. Ediriweera, A. De Silva, B. Premarathna, S. Jayasinghe, and H. de Silva, 2021: Dengue and leptospirosis infection during the coronavirus 2019 outbreak in Sri Lanka. *Transactions of the Royal Society of Tropical Medicine and Hygiene*.

Nishio, M. and A. Arakawa, 2019: Performance of Hamiltonian Monte Carlo and No-U-Turn Sampler for estimating genetic parameters and breeding values. *Genetics Selection Evolution*, **51 (1)**, 1–12.

Noor, A. A., A. Mohd Arif, W. Nazni, and H. Lee, 2018: Ovitrap surveillance of Aedes aegypti and Aedes albopictus in dengue endemic areas in Keramat and Shah Alam, Selangor in 2016. *IIUM Medical Journal Malaysia*, **17 (3)**.

Northrop, P. J. and R. E. Chandler, 2014: Quantifying sources of uncertainty in projections of future climate. *Journal of Climate*, **27 (23)**, 8793–8808.

Ong, S.-Q., H. Ahmad, and A. M. Mohd Ngesom, 2021: Implications of the COVID-19 lockdown on dengue transmission in Malaysia. *Infectious Disease Reports*, **13 (1)**, 148–160.

Owens, R. and T. Hewson, 2018: ECMWF Forecast User Guide.

Petrova, D., R. Lowe, A. Stewart-Ibarra, J. Ballester, S. J. Koopman, and X. Rodó, 2019: Sensitivity of large dengue epidemics in Ecuador to long-lead predictions of El Niño. *Climate Services*, **15**, 100 096.

Pontes, R., J. Freeman, J. Oliveira-Lima, J. Hodgson, and A. Spielman, 2000: Vector densities that potentiate dengue outbreaks in a Brazilian city. *The American Journal of Tropical Medicine and 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. and T. H. Carpenter, 1982: Variations in tropical sea surface temperature and surface wind fields associated with the Southern Oscillation/El Niño. *Monthly Weather Review*, **110 (5)**, 354–384.

Riebler, A., S. H. Sørbye, D. Simpson, and H. Rue, 2016: An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research*, **25 (4)**, 1145–1165.

Runge-Ranzinger, S., 2016: Technical handbook for dengue surveillance, dengue outbreak prediction/detection and outbreak response (model contingency plan). World Health Organization.

Rushworth, A., D. Lee, and R. Mitchell, 2014: A spatio-temporal model for estimating the long-term effects of air pollution on respiratory hospital admissions in Greater London. *Spatial and Spatio-temporal Epidemiology*, **10**, 29–38.

Rushworth, A., D. Lee, and C. Sarran, 2017: An adaptive spatiotemporal smoothing model for estimating trends and step changes in disease risk. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **66 (1)**, 141–157.

Sahu, S. K., 2021: bmstdr: Bayesian modeling of spatio-temporal data with R. *Technical Report*, submitted.

Sahu, S. K., 2021 (in press): *Bayesian modeling of spatio-temporal data with R*. Boca Raton: Chapman and Hall/CRC Press.

Sahu, S. K., S. Yip, and D. M. Holland, 2009: Improved space–time forecasting of next day ozone concentrations in the eastern US. *Atmospheric Environment*, **43 (3)**, 494–501.

Salim, N. A. M., et al., 2021: Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques. *Scientific reports*, **11 (1)**, 1–9.

Salimun, E., F. Tangang, L. Juneng, S. K. Behera, and W. Yu, 2014: Differential impacts of conventional El Niño versus El Niño Modoki on Malaysian rainfall anomaly during winter monsoon. *International Journal of Climatology*, **34 (8)**, 2763–2774.

Simpson, D., H. Rue, A. Riebler, T. G. Martins, and S. H. Sørbye, 2017: Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, **32 (1)**, 1–28.

Skae, F., 1902: Dengue fever in Penang. *British Medical Journal*, **2 (2185)**, 1581.

Spencer, D. E., 1993: Developing a Bayesian vector autoregression forecasting model. *International Journal of Forecasting*, **9 (3)**, 407–421.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde, 2002: Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64 (4)**, 583–639.

Stewart-Ibarra, A. M. and R. Lowe, 2013: Climate and non-climate drivers of dengue epidemics in southern coastal Ecuador. *The American Journal of Tropical Medicine and 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. and A. Tyagi, 1999: Formation of an air pollution index. *Journal of the Air & Waste Management Association*, **49 (1)**, 88–91.

Tangang, F., R. Farzanmanesh, A. Mirzaei, E. Salimun, A. F. Jamaluddin, and L. Juneng, 2017: Characteristics of precipitation extremes in Malaysia associated with El Niño and La Niña events. *International Journal of Climatology*, **37**, 696–716.

Thiruchelvam, L., S. C. Dass, R. Zaki, A. Yahya, and V. S. Asirvadam, 2018: Correlation analysis of air pollutant index levels and dengue cases across five different zones in Selangor, Malaysia. *Geospatial Health*, **13 (1)**.

Vehtari, A., A. Gelman, and J. Gabry, 2017: Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, **27 (5)**, 1413–1432.

Wan-Norafikah, O., H. Lee, A. Norazizah, and A. Mohamad-Hafiz, 2016: Research note repellency effects of an ozone-producing air purifier against medically important insect vectors. *Tropical Biomedicine*, **33 (2)**, 396–402.

West, M. and J. Harrison, 2006: *Bayesian forecasting and dynamic models*. Springer Science & Business Media.

Wong, M., H. Mok, H. Ma, M. Lee, and M. Fok, 2011: A climate model for predicting the abundance of Aedes mosquitoes in Hong Kong. *Meteorological Applications*, **18 (1)**, 105–110.

Yip, S., C. A. Ferro, D. B. Stephenson, and E. Hawkins, 2011: A simple, coherent framework for partitioning uncertainty in climate predictions. *Journal of Climate*, **24 (17)**, 4634–4643.

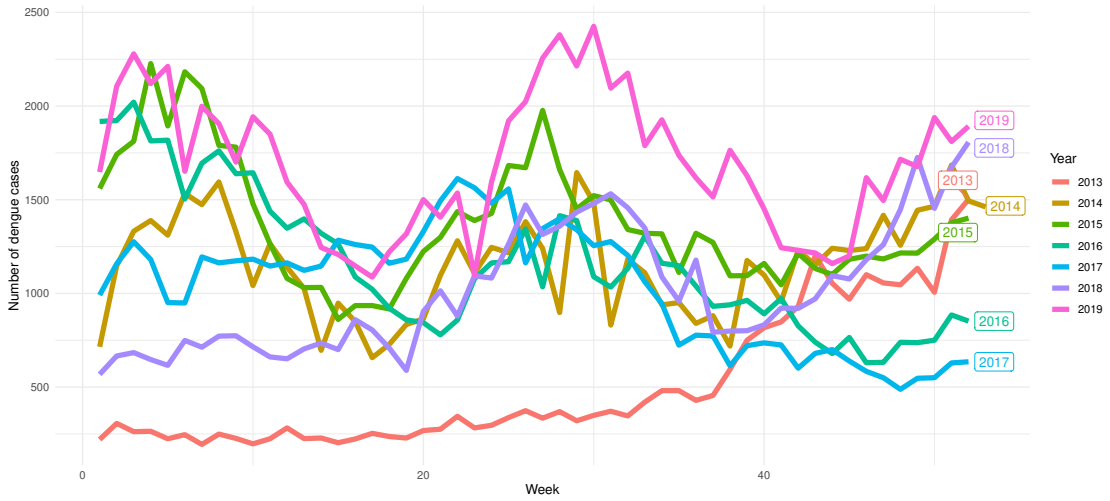


Figure 1: Time series plots of dengue hospitalisations by year.

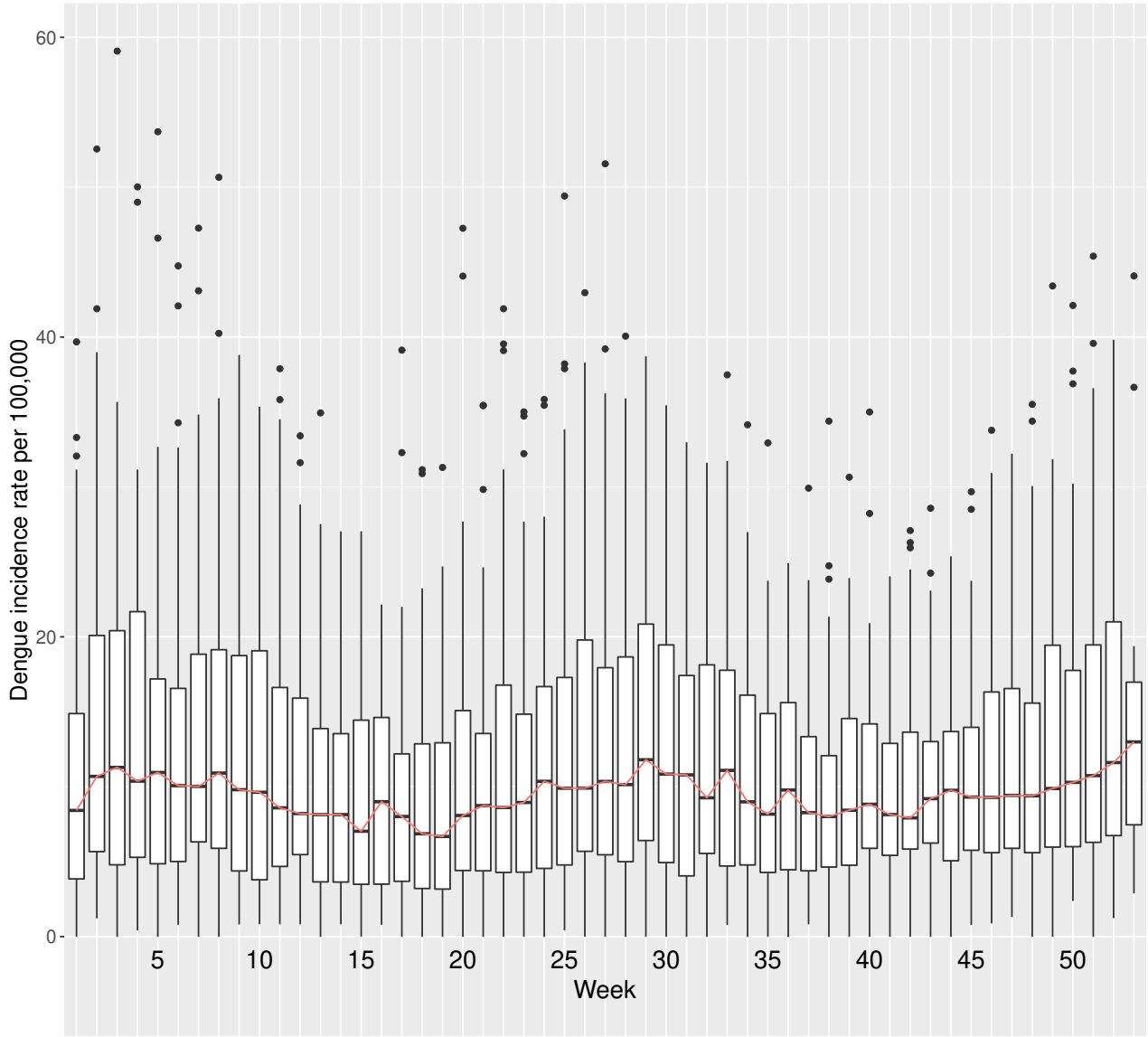


Figure 2: Boxplots of weekly dengue hospitalisations in the Central Region, Malaysia, 2013 - 2019.

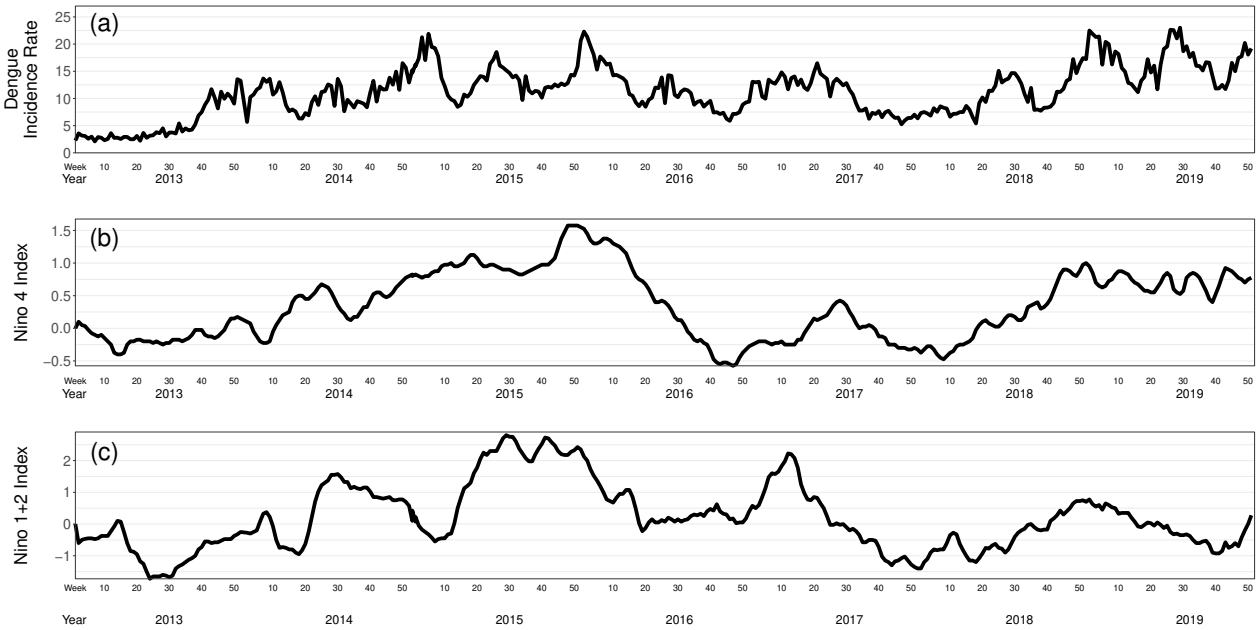


Figure 3: Time series of **(a)** average DIR , **(b)** lagged four-week average of Niño4 index, **(c)** lagged four-week average of Niño1+2 index in the Central Region, Malaysia for the period 2013-2019.

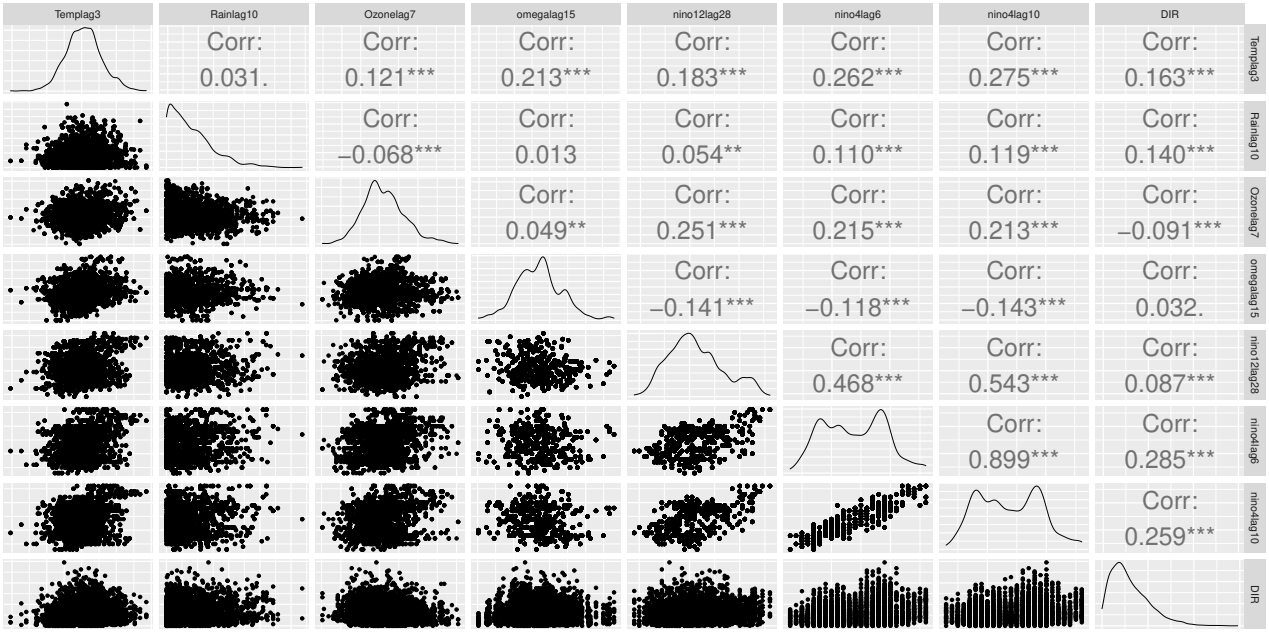


Figure 4: Pairwise scatter plots of the DIR along with the covariates used in the models.

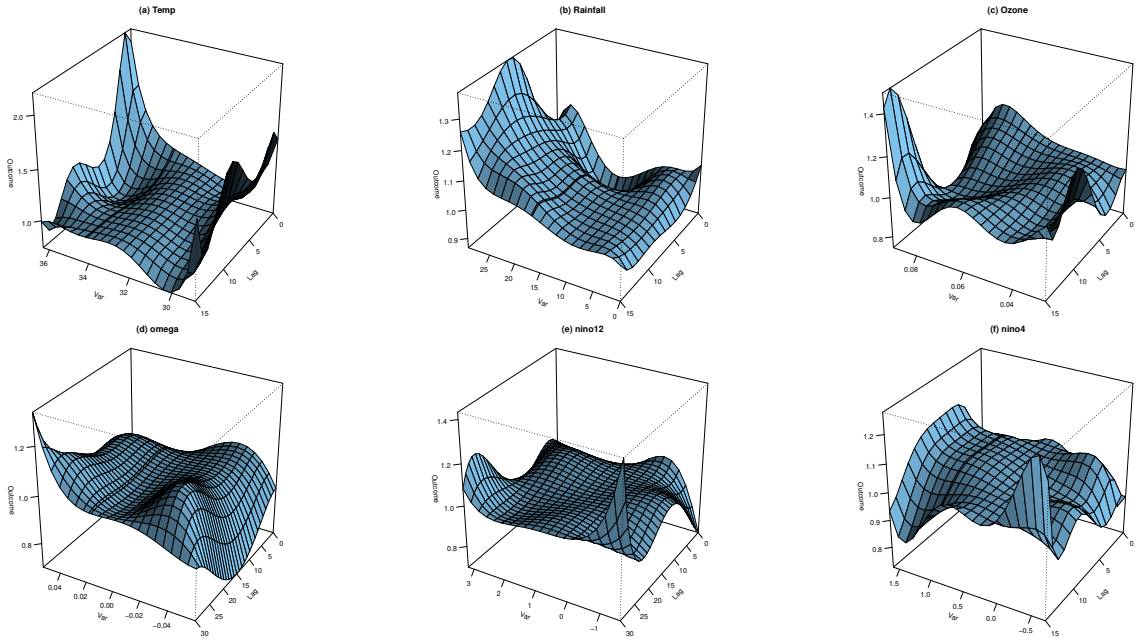


Figure 5: RR surface of dengue hospitalisations by six variables. The variable *Temp* is the temperature, *Rain* is total precipitation, Ozone is the ground-level ozone concentration level, omega is vertical velocity of air motion derived from omega equation, *nino12* is the Niño1+2 index, *nino4* is the Niño4 index.

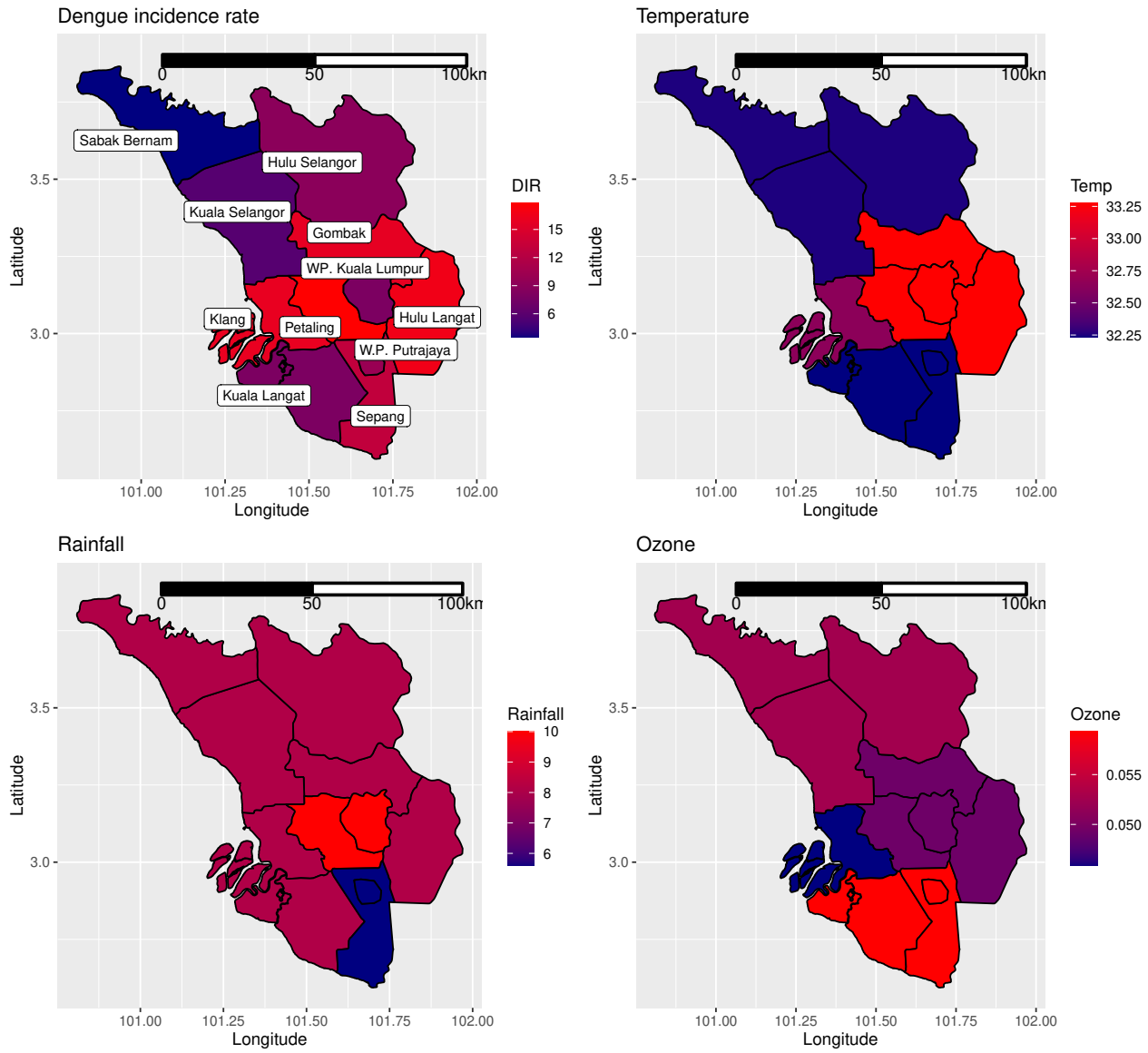
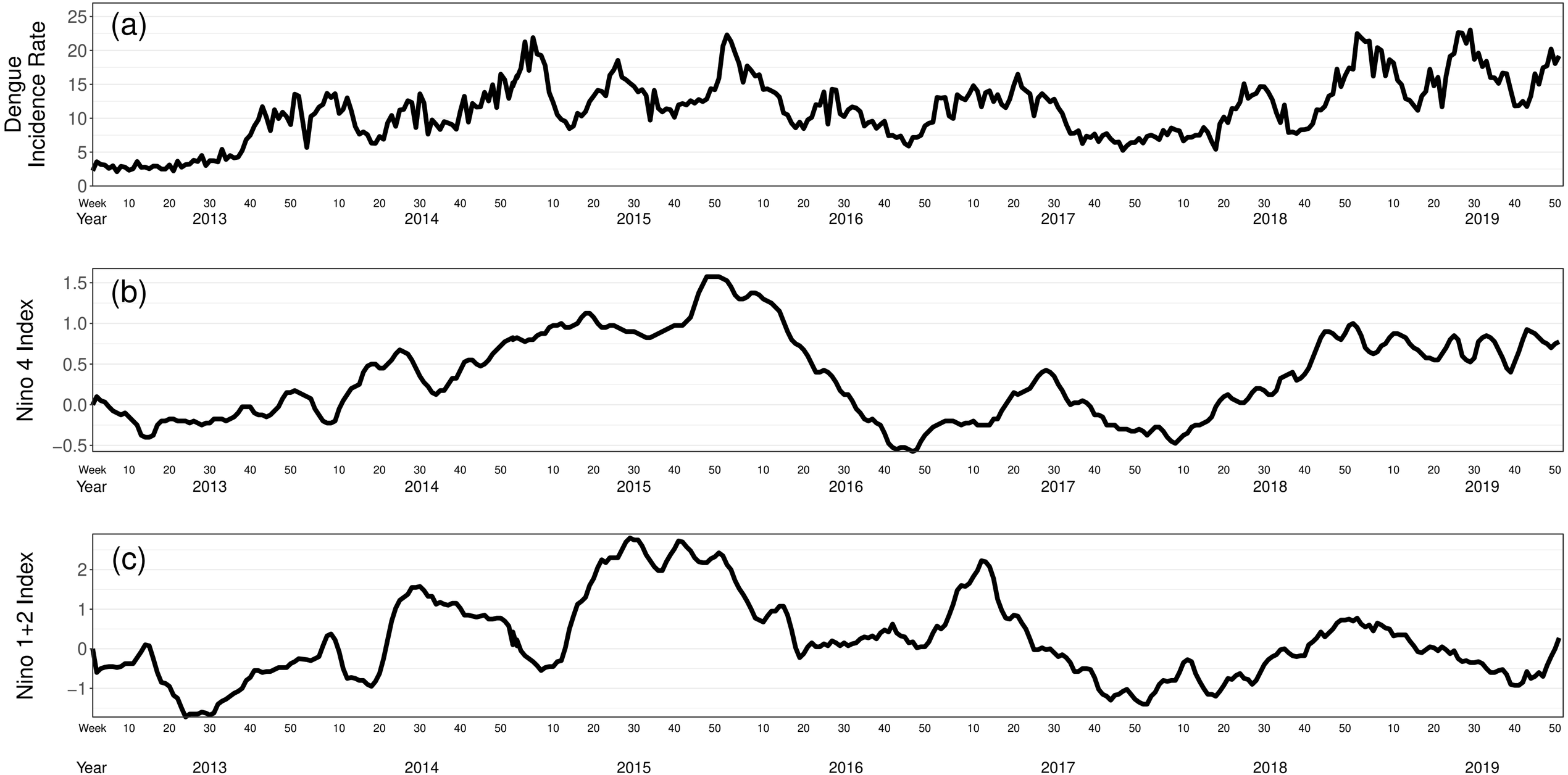
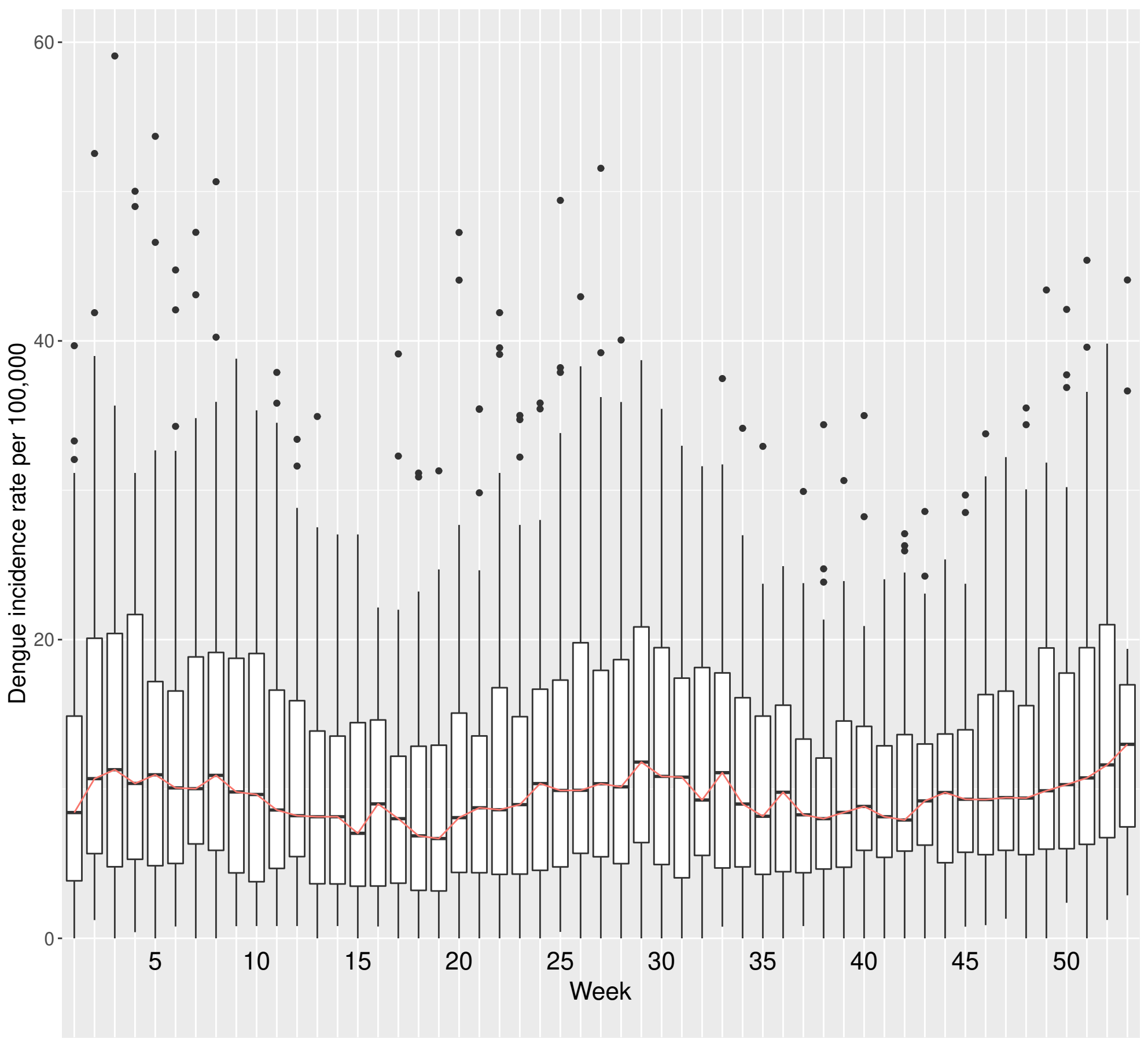
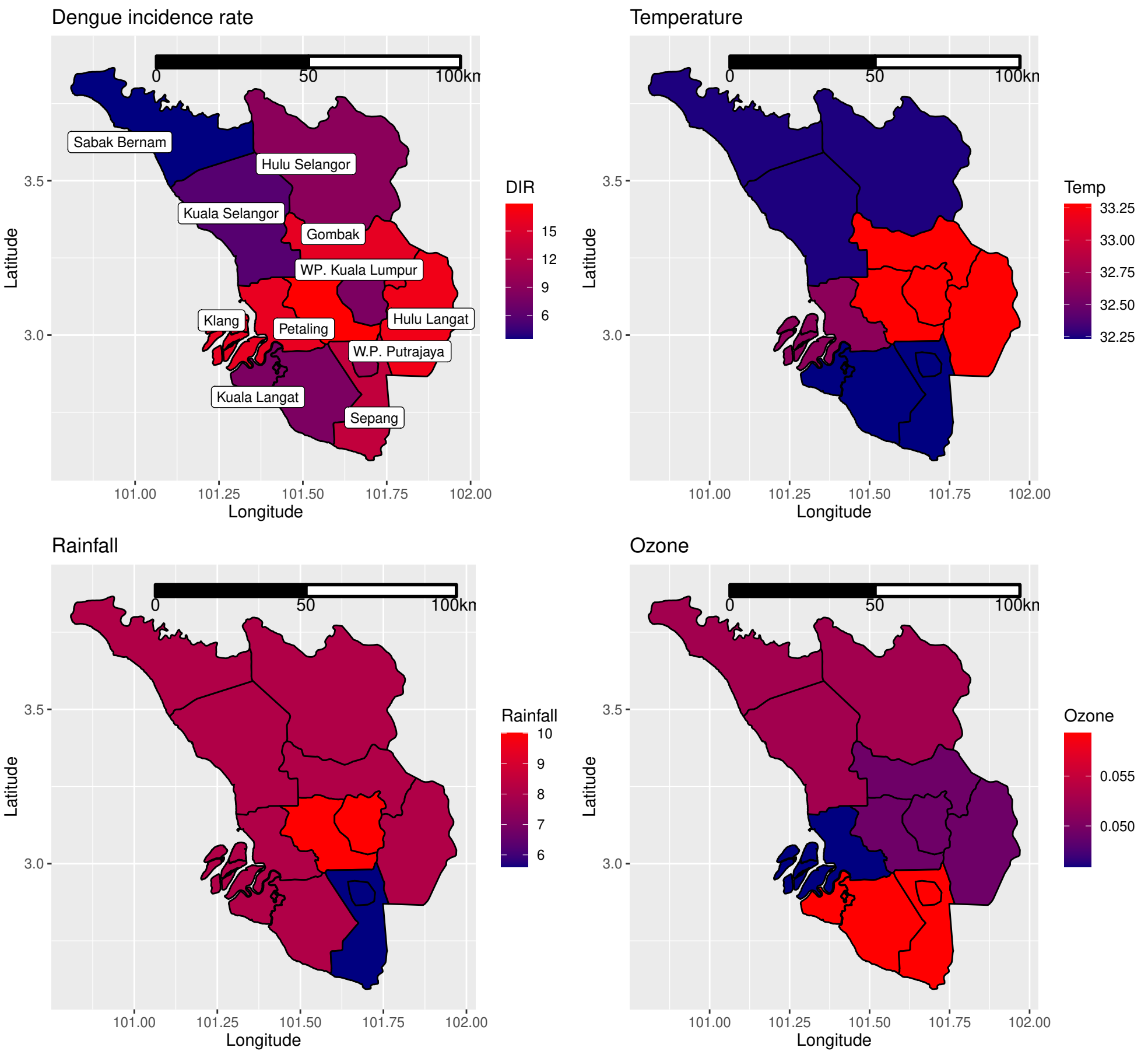
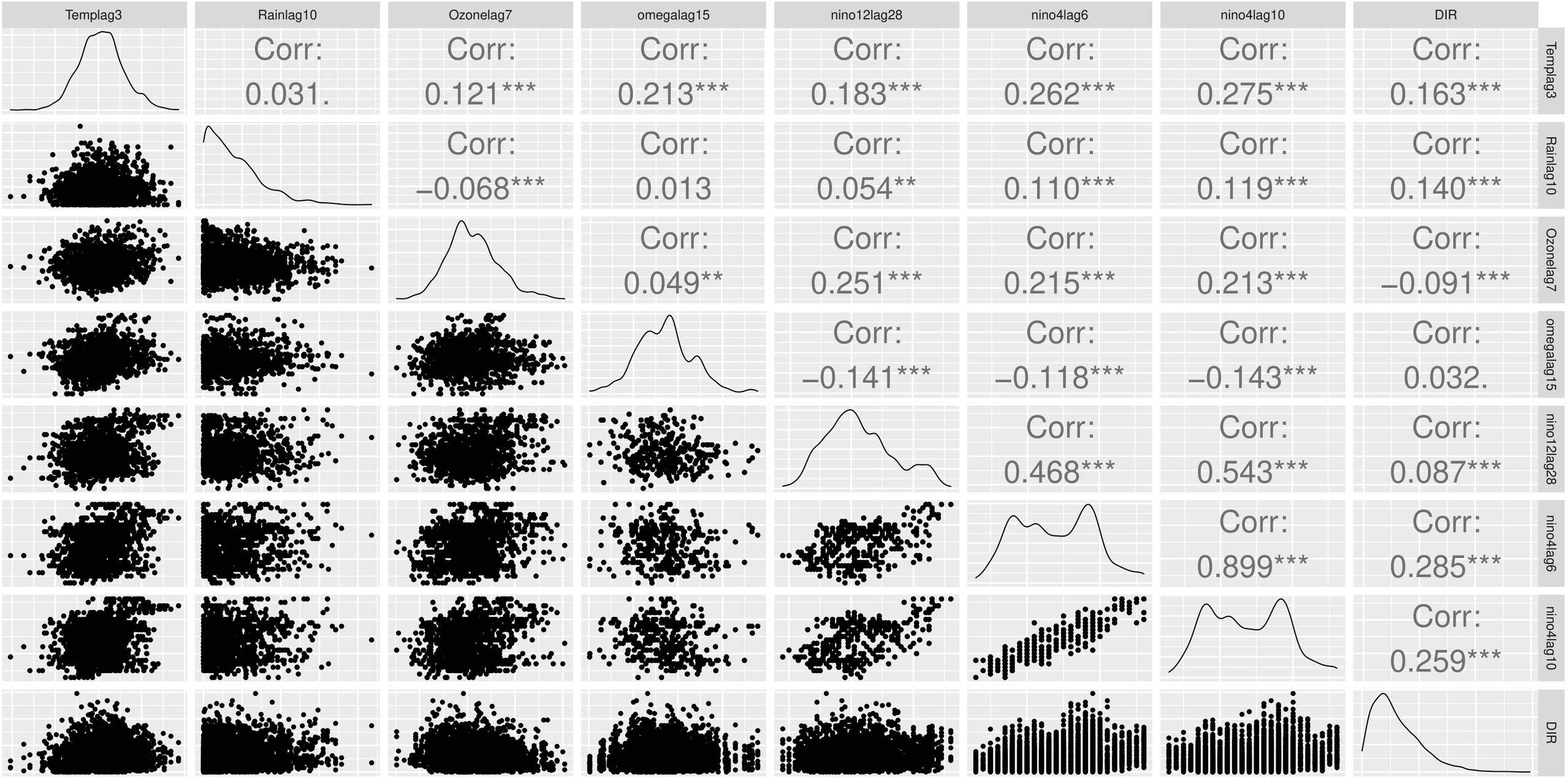


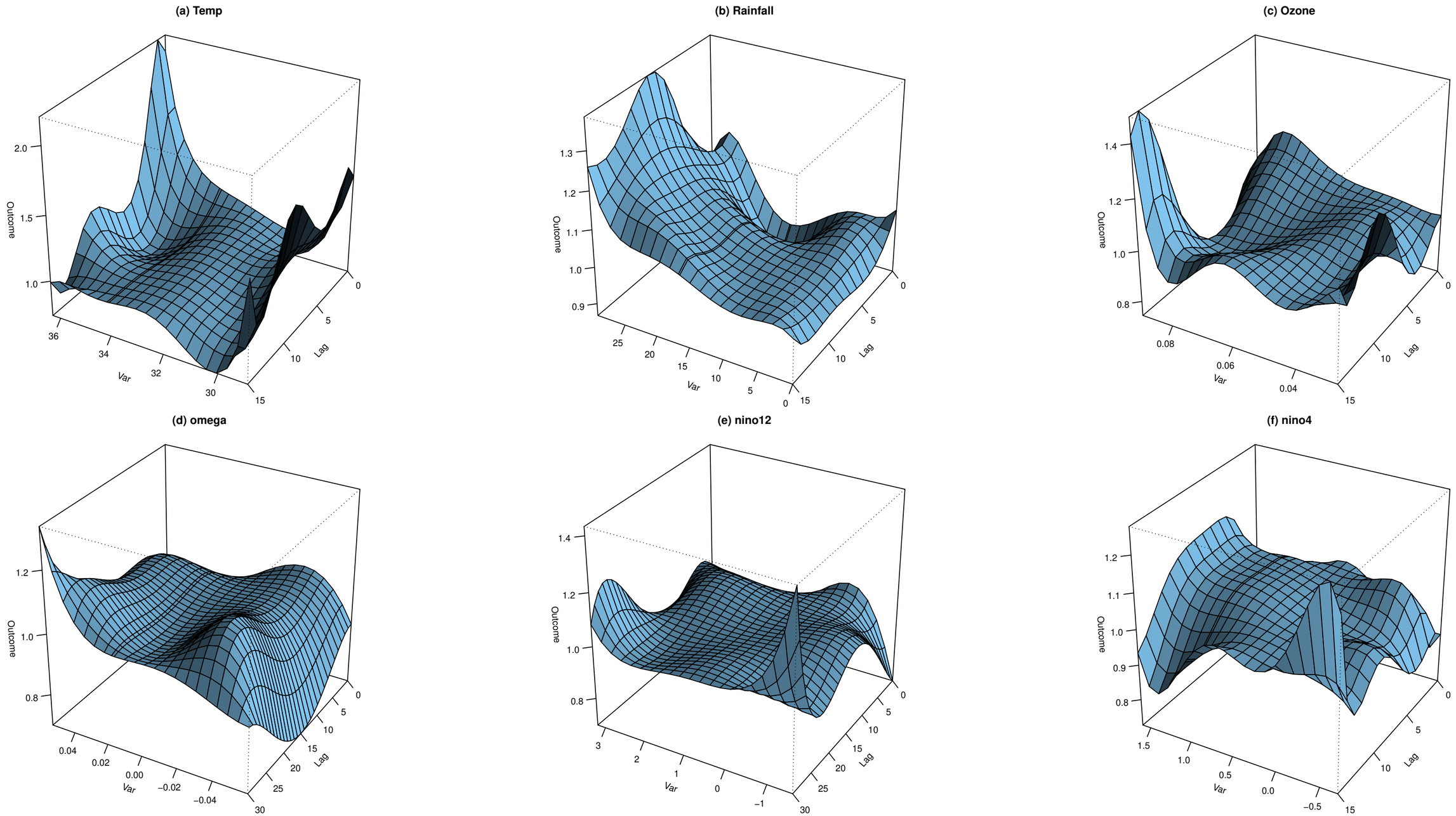
Figure 6: Mean weekly DIR, temperature, rainfall, ground-level ozone concentration level from 2013 to 2019 by district.

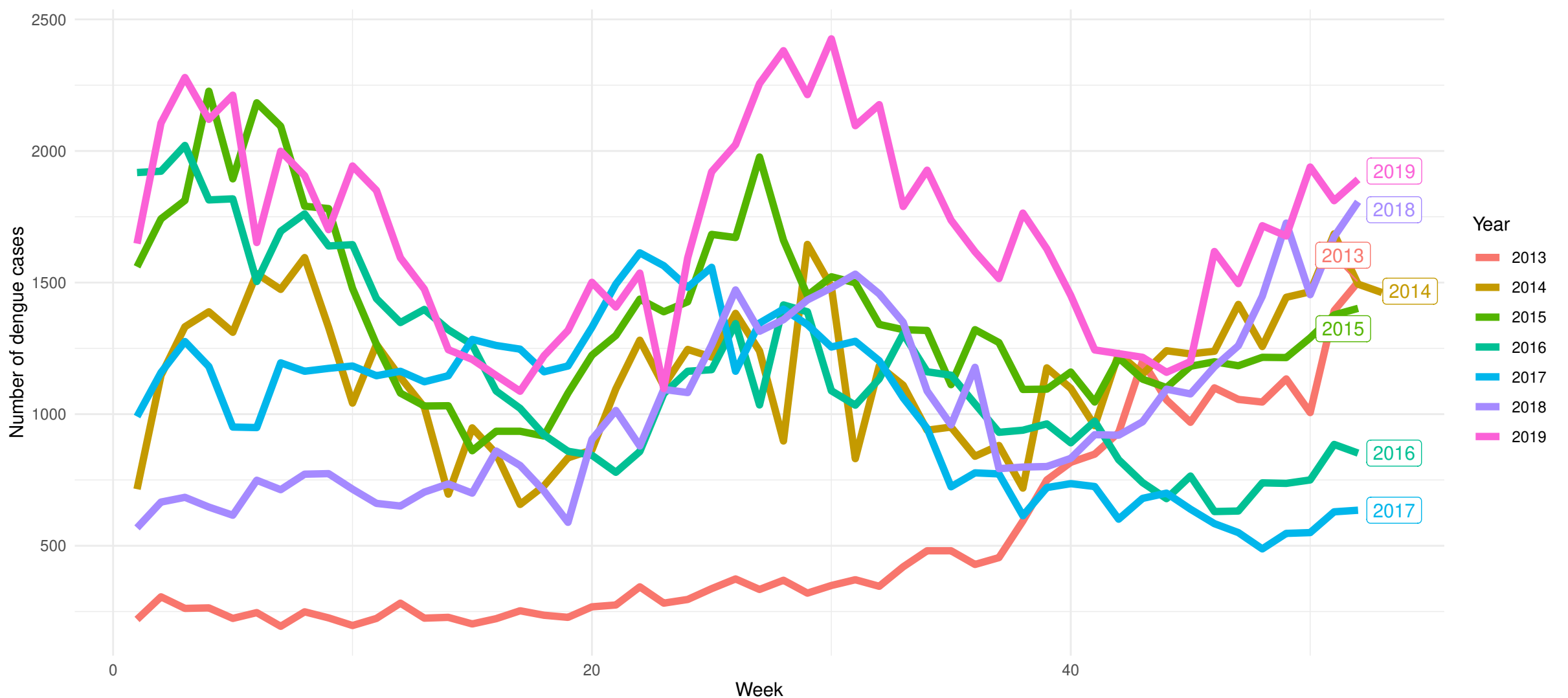

(a)
Dengue Incidence Rate
0
5
10
15
20
25
(b)
Nino 4 Index
−0.5
0.0
0.5
1.0
1.5
(c)
Nino 1+2 Index
−1
0
1
2
Week
10
20
30
40
50
Year
2013
2014
2015
2016
2017
2018
2019


60
40
20
0
Dengue incidence rate per 100,000
5
10
15
20
25
30
35
40
45
50
Week


Dengue incidence rate
0
50
100km
Sabak Bernam
Hulu Selangor
Kuala Selangor
Gombak
WP. Kuala Lumpur
Klang
Petaling
Hulu Langat
W.P. Putrajaya
Kuala Langat
Sepang
DIR
15
12
9
6
Latitude
3.5
3.0
101.00
101.25
101.50
101.75
102.00
Longitude
Temperature
Temp
33.25
33.00
32.75
32.50
32.25
Rainfall
Rainfall
10
9
8
7
6
Ozone
Ozone
0.055
0.050


Templag3
Rainlag10
Ozonelag7
omegalag15
nino12lag28
nino4lag6
nino4lag10
DIR
Templag3
Corr: 0.031.
Corr: 0.121***
Corr: 0.213***
Corr: 0.183***
Corr: 0.262***
Corr: 0.275***
Corr: 0.163***
Rainlag10
Corr: −0.068***
Corr: 0.013
Corr: 0.054**
Corr: 0.110***
Corr: 0.119***
Corr: 0.140***
Ozonelag7
Corr: 0.049**
Corr: 0.251***
Corr: 0.215***
Corr: 0.213***
Corr: −0.091***
omegalag15
Corr: −0.141***
Corr: −0.118***
Corr: −0.143***
Corr: 0.032.
nino12lag28
Corr: 0.468***
Corr: 0.543***
Corr: 0.087***
nino4lag6
Corr: 0.899***
Corr: 0.285***
nino4lag10
Corr: 0.259***
DIR


(a) Temp
Outcome
2.0
1.5
1.0
36
34
32
30
Var
0
5
10
15
Lag
(b) Rainfall
Outcome
1.3
1.2
1.1
1.0
0.9
25
20
15
10
5
0
Var
0
5
10
15
Lag
(c) Ozone
Outcome
1.4
1.2
1.0
0.8
0.08
0.06
0.04
Var
0
5
10
15
Lag
(d) omega
Outcome
1.2
1.0
0.8
0.04
0.02
0.00
−0.02
−0.04
Var
0
5
10
15
20
25
30
Lag
(e) nino12
Outcome
1.4
1.2
1.0
0.8
3
2
1
0
−1
Var
0
5
10
15
20
25
30
Lag
(f) nino4
Outcome
1.2
1.1
1.0
0.9
0.8
1.5
1.0
0.5
0.0
−0.5
Var
0
5
10
15
Lag


2500
2000
1500
1000
500
Number of dengue cases
0
20
40
Week
2019
2018
2013
2014
2015
2016
2017
Year
2013
2014
2015
2016
2017
2018
2019