Bayesian spatiotemporal mapping of relative dengue disease risk in Bandung, Indonesia

Jaya, Mindra; Folmer, Henk

Published in:
Journal of Geographical Systems

DOI:
10.1007/s10109-019-00311-4

Publication date:
2020

Journal of Geographical Systems (2020) 22:105–142
https://doi.org/10.1007/s10109-019-00311-4

ORIGINAL ARTICLE

Bayesian spatiotemporal mapping of relative dengue disease risk in Bandung, Indonesia

I. Gede Nyoman Mindra Jaya[1,2] · **Henk Folmer**[1,3]

Received: 6 September 2018 / Accepted: 17 July 2019 / Published online: 30 August 2019

Abstract

Dengue disease has serious health and socio-economic consequences. Mapping its occurrence at a fine spatiotemporal scale is a crucial element in the preparation of an early warning system for the prevention and control of dengue and other viral diseases. This paper presents a Bayesian spatiotemporal random effects (pure) model of relative dengue disease risk estimated by integrated nested Laplace approximation. Continuous isopleth mapping based on inverse distance weighting is applied to visualize the disease's geographical evolution. The model is applied to data for 30 districts in the city of Bandung, Indonesia, for the period January 2009 to December 2016. We compared the Poisson and the negative binomial distributions for the number of dengue cases, both combined with a model which included structured and unstructured spatial and temporal random effects and their interactions. Using several Bayesian and classical model performance criteria and stepwise backward selection, we chose the negative binomial distribution and the temporal model with spatiotemporal interaction for forecasting. The estimation results show that the relative risk decreased generally from 2014. However, it consistently increased in the north-western districts because of environmental and socio-economic conditions. We also found that every district has a different temporal pattern, indicating that district characteristics influence the temporal variation across space.

Keywords Dengue disease · Bayesian spatiotemporal random effects (pure) model · Integrated nested Laplace approximation (INLA) · Isopleth mapping · Bandung—Indonesia

The authors are indebted to Farah Kristiani, Mathematics Department, Parahyangan Catholic University, Bandung and Budi Nurani Ruchjana, Mathematics Department, Padjadjaran University, Bandung, for their helpful comments.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10109-019-00311-4) contains supplementary material, which is available to authorized users.

✉ I. Gede Nyoman Mindra Jaya
mindra@unpad.ac.id

Extended author information available on the last page of the article

JEL Classification C30 · I18

1 Introduction

Dengue disease is a mosquito-borne viral disease caused by the dengue virus (*DENV 1–4*). The disease is transmitted via the female *Aedes*-spp. mosquito from person to person (Acharya et al. 2016). Dengue is endemic in most tropical and subtropical countries, particularly in urban and semi-urban regions. The disease is spreading rapidly, and about half the world's population is now at risk (WHO 2018). The transmission of pathogens is promoted by factors such as population density, migration, commuting, trade, poor sanitation and limited access to clean water. Climate change and virus evolution have also played an important role in the spread of the disease (Murray et al. 2013; Zellweger et al. 2017). Epidemics tend to cause serious health problems, death and socio-economic losses such as impacts on medical care, hospitalization, a loss of productivity, school absenteeism and the unproductive consumption of time for unpaid caregivers (Jaya et al. 2017; Suaya et al. 2009; Widiyani 2013). The global cost of the dengue disease in 2013 was approximately USD 8.9 billion (Shepard et al. 2016).

Although vaccines exist, the usual strategy adopted by most dengue-prone countries is disease prevention, because vaccines or specific antiviral therapies are very expensive (WHO 2016). For example, in Indonesia—with a per capita income of USD 296 per month in 2017—the cost of a single injection is around USD 77 (1 USD = IDR 13,500). Three injections, administered every six months, are needed for vaccination to be effective (Detik 2016).

Prevention requires effective vector control to break the lifecycle of the mosquito (WHO 2012). Vector controls include environmental, biological or chemical operations (WHO 2009). A crucial step in prevention is the identification of regions where the a priori risk level is predicted to exceed a critical threshold for the ensuing periods. An early warning system is needed to ensure the success of vector control (CDC 2016; WHO 2018). Such a system requires statistical methods to generate forecasts for the risk by region and time (Ugarte et al. 2012).

Spatiotemporal maps which track the distribution of the disease are basic elements of an early warning system (Lawson 2006). They depict the spatiotemporal evolution of the incidence rate of the disease and thus provide clues for epidemiologists to expand their aetiologic hypotheses on disease outbreaks (Ugarte et al. 2012). Choropleth maps, based on patterns of shades or colours according to a prearranged key, are usually used to characterize regions. Each shading or colour type represents a range of values. However, choropleth maps show distributions that have discontinuities at the borderlines which render them less suitable for disease mapping, the purpose of which is to generate forecasts or to identify unknown risk factors that continuously vary spatially, such as rainfall, temperature or humidity (Emch et al. 2017).[1] To overcome these problems, isopleth maps, which typically show

[1] For further shortcomings of choropleth maps, we refer to Lawson (2006).

continuous change across space, may be used instead. They are applied in the present case study.

Modelling and forecasting of relative disease risk is a challenging task, especially in developing countries. The challenge comes from incomplete or inaccurate information on the main determinants. In particular, limited resources for surveillance tend to lead to incomplete data, and even if complete data on the risk factors are available, it may not directly relate to the incidence rate due to complex transmission pathways (Christenfeld et al. 2004; McMichael et al. 2013). As an alternative, we propose spatiotemporal random effects models (models without covariates), also known as *pure* or random intercept models, to overcome the information problems (Clayton and Kaldor 1987; Coly et al. 2015; Knorr-Held 2000; Waller et al. 1997; Waller and Carlin 2010).

A number of risk factors are available for the present case study, notably meteorological variables such as rainfall, temperature and humidity and socio-economic factors such as the larvae free index and the healthy house index. However, the objective of this paper is spatiotemporal *mapping*, i.e. to obtain out-of-bounds predictions (denoted as forecasts in the subsequent study) for each study area (see Schrödle and Held 2011; Ugarte et al. 2014; Wakefield 2007). For forecasting purposes, the inclusion of covariates is not needed; it can be based on univariate time series data. In addition, the inclusion of covariates may lead to complex models because of the stochastic variation in the relationship between the covariates and the interest variable, even if they are causally related. Another major advantage of mapping (based on a univariate time series) is that data on the variable of interest are often unique historical data and that corresponding data on covariates are not readily available (Linden et al. 2003; Peter and Silvia 2012). In the case of disease mapping, the information on the risk factors can be accounted for by means of structured and unstructured spatial and temporal random effects and their interactions (Wakefield 2007). The aim of 'spatial regression', on the other hand, is the estimation of the association between relative risk and potential risk factors.

The pure model is usually estimated by means of Bayesian methods because of their convenience for specifying the random components using a hierarchical structure on the parameters (Blangiardo et al. 2013). When no analytical solution is feasible, Bayesian spatiotemporal random effect models are usually estimated by means of Markov Chain Monte Carlo methods (MCMC) (Blangiardo et al. 2013). However, MCMC can be computationally challenging due to the complexity of the models, particularly the large number of parameters. Consequently, MCMC can lead to large Monte Carlo errors and consume a substantial amount of computation time (Arab 2015; Blangiardo et al. 2013; Ugarte et al. 2012). An alternative to MCMC is integrated nested Laplace approximation (INLA) (Blangiardo et al. 2013; Jaya et al. 2017). INLA reduces computation time and produces reliable parameter estimates which are equivalent to MCMC estimates (Bivand et al. 2015; Blangiardo and Cameletti 2015; Rue and Martino 2009; De Smedt et al. 2015).

We note that several recently published studies that forecast the relative risk of infectious diseases ignore temporal trends and seasonal patterns (Liu et al. 2017; Ugarte et al. 2012; Watson et al. 2017). This is a major omission since they are important characteristics from theoretical, methodological and policy perspectives

(Bauer et al. 2016; WHO 2009). As shown below, they can be accounted for in a straightforward manner using Bayesian methods and INLA.

The structure of the remainder of this paper is as follows. Section 2 presents the spatiotemporal dengue fever model and summarizes its estimation by INLA. Section 3 applies the method to Bandung, Indonesia. Section 4 presents the conclusions.

2 Methodology

2.1 The general Bayesian spatiotemporal disease model

The classical unbiased estimator of relative risk is the standardized incidence ratio (SIR), defined as the ratio of the number of cases in a region and the expected number. However, the SIR can be subject to a high level of unreliability due to a small number of cases in a region or a small population at risk (Lawson 2013; Ugarte et al. 2014). As a result, the estimate can be distorted. These drawbacks can be overcome by shrinkage, i.e. a reduction of the incidence of outliers, by smoothing the estimate using spatiotemporal methods including Bayesian spatiotemporal hierarchical models. To reduce the variability of the risk estimates, shrinkage borrows strength from neighbours, i.e. spatiotemporal dependency and heterogeneity. The notion of borrowing strength from adjacent regions is related to Tobler's (1970) law of geography—that nearby regions are more similar than regions further away from each other—and to the basic notion of spatial dependence in spatial econometrics (e.g. LeSage and Pace 2009). Borrowing strength from neighbours is also a key issue in small area estimation (SEA) (Handayani et al. 2018; Rao and Molina 2015).

During the last 20 years, a wide range of spatiotemporal models has been developed for disease modelling, most of them of the intrinsic conditional autoregressive (iCAR) type extending the well-known Besag, York and Mollie (BYM) model (Besag et al. 1991; Knorr-Held 2000; Waller et al. 1997). The BYM model (dengue disease) in this paper assumes that, conditional on the underlying relative risk, θ_{it}, the number of cases in each region i and at time t, y_{it}, follows a Poisson distribution with mean and variance equal to $\lambda_{it} = E_{it}\theta_{it}$

$$y_{it}|E_{it}\theta_{it} \sim \text{Poisson}\left(E_{it}\theta_{it}\right) i = 1, \ldots, n \quad \text{and} \quad t = 1, \ldots, T, \tag{1}$$

with E_{it} denoting the expected number of cases and θ_{it} the relative risk in region i and at time t. In the case of spatiotemporal data, the expected number of cases can be based on the following reference rates: (1) the average for each period (Abente et al. 2018; Jaya et al. 2017):

$$E_{it} = N_{it}\frac{\sum_{i=1}^{n} y_{it}}{\sum_{i=1}^{n} N_{it}} t = 1, \ldots, T, \tag{2}$$

and (2) the overall average (over across all periods):

$$E_{it} = N_{it}\frac{\sum_{i=1}^{n}\sum_{t=1}^{T} y_{it}}{\sum_{i=1}^{n}\sum_{t=1}^{T} N_{it}} i = 1, \ldots, n \quad \text{and} \quad t = 1, \ldots, T, \tag{3}$$

with N_{it} denoting the population in region i at time t, n the number of observed regions and T the number of observed periods. Following Abente et al. (2018), we apply the overall average in the case study.

Disease data frequently show an over-dispersion of zeros. This problem can be handled by the inclusion of a second parameter ε_{it} in the Poisson distribution governing the variance with ε_{it} following a Gamma distribution. Specifically (Mohebbi et al. 2014):

$$y_{it}\Big|E_{it}\theta_{it},\varepsilon_{it} \sim \text{Poisson}\left(E_{it}\theta_{it}\varepsilon_{it}\right) \quad \text{and} \quad \varepsilon_{it}\Big|\varrho \sim \text{Gamma}(\varrho,\varrho), \tag{4}$$

for $y_{it} = 0, 1, 2, \ldots$ and $\varrho > 0$. The combination of the Poisson–Gamma probability of y_{it} can be expressed as:

$$\begin{aligned} p\left(y_{it}|E_{it},\theta_{it},\varepsilon_{it}\right) &= \text{Gamma}(\varrho,\varrho)\text{Poisson}\left(E_{it}\theta_{it}\varepsilon_{it}\right) \\ &= \left(\frac{\varrho^{\varrho}\left(\varepsilon_{it}\right)^{\varrho-1}\exp\left(-\varrho\varepsilon_{it}\right)}{\Gamma\left(\varepsilon_{it}\right)}\right)\left(\frac{\exp\left(-E_{it}\theta_{it}\right)\left(E_{it}\theta_{it}\right)^{y_{it}}}{y_{it}!}\right). \end{aligned} \tag{5}$$

By integrating out the random effect ε_{it}, the marginal probability of y_{it} is obtained as the negative binomial (NB) distribution (Mohebbi et al. 2014):

$$p\left(y_{it}|E_{it}\theta_{it},\varrho\right) = \frac{\Gamma\left(y_{it}+\varrho\right)}{\Gamma\left(y_{it}+1\right)\Gamma(\varrho)}\left(\frac{E_{it}\theta_{it}}{E_{it}\theta_{it}+\varrho}\right)^{y_{it}}\left(\frac{\varrho}{E_{it}\theta_{it}+\varrho}\right)^{\varrho}. \tag{6}$$

The NB distribution has a mean $E\left(y_{it}\right) = E_{it}\theta_{it}$ and variance $\text{Var}\left(y_{it}\right) = E_{it}\theta_{it} + \left(E_{it}\theta_{it}\right)^2/\varrho$ with ϱ being the parameter of additional Poisson variation. The variance of the NB distribution is always greater than the mean. For $\varrho \to \infty$, the NB distribution converges on the Poisson distribution. The discussion below is in terms of the Poisson model. Generalization to the NB model is straightforward.

We now turn to the mean of the Poisson distribution, which we decompose by way of the natural logarithm link function:

$$\log\left(\lambda_{it}\right) = \log\left(E_{it}\right) + \log\left(\theta_{it}\right) i = 1, \ldots, n \quad \text{and} \quad t = 1, \ldots, T \tag{7}$$

The second component in Eq. (7), $\log\left(\theta_{it}\right)$, is the focus of further research. We model it as a *pure* model with the random effects consisting of structured and unstructured variance components. The structured variance components take account of the variation due to the correlation between spatial and temporal units, respectively, while the unstructured variance components represent the variation due to heteroscedasticity. Specifically:

$$\eta_{it} = \log\left(\theta_{it}\right) = \alpha + \omega_i + \upsilon_i + \phi_t + \gamma_t + \delta_{it} i = 1, \ldots, n \quad \text{and} \quad t = 1, \ldots, T, \tag{8}$$

where α is the intercept representing the overall relative risk; ω_i, υ_i, ϕ_t and γ_t are the spatially structured, spatially unstructured, temporally structured and temporally unstructured random effect components, respectively. δ_{it} represents spatiotemporal interaction.

The estimation of Eq. (8) can be done by means of frequentist or Bayesian methods. A frequentist method such as maximum likelihood (ML) is often unsatisfactory, especially for small regions, due to Poisson sampling variation (Ugarte et al. 2014). ML estimation of the region-specific risk and of the time-trend can be seriously affected by random variation, particularly when the data are scarce or incomplete (Bernardinelli et al. 1995). This problem can be overcome by Bayesian modelling, which allows the coherent incorporation of missing or uncertain data or information into the analysis (Lawson and Zhou 2005). Therefore, we use the hierarchical Bayesian approach in this paper, as summarized below.

Consider observed data in the ith region at time t, $\mathbf{y}_i = \left(y_{i1}, \ldots, y_{iT}\right)'$, generated from a probability distribution $p\left(\mathbf{y}_i|\boldsymbol{\Phi}, \boldsymbol{\tau}\right)$ with unknown parameters $\boldsymbol{\Phi} = \left(\alpha, \boldsymbol{\omega}', \boldsymbol{\nu}', \phi', \boldsymbol{\gamma}', \boldsymbol{\delta}'\right)'$. The unknown parameters $\boldsymbol{\Phi}$ are taken as random variables with priors $p(\boldsymbol{\Phi}|\boldsymbol{\tau})$, and with unknown hyper-parameters $\boldsymbol{\tau} = \left(\tau_\alpha, \tau_\omega, \tau_\nu, \tau_\phi, \tau_\gamma, \tau_\delta\right)'$ and hyper-priors $p(\boldsymbol{\tau})$. Under the assumption of independence, the likelihood function of the number of cases is identical to the joint density of $\mathbf{y}_i, i = 1, \ldots, n$. Accordingly, the joint posterior density of $\boldsymbol{\Phi}$ and $\boldsymbol{\tau}$, given $\mathbf{y}$, is provided by Jaya et al. (2017):

$$p(\boldsymbol{\Phi}, \boldsymbol{\tau}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\Phi}, \boldsymbol{\tau})p(\boldsymbol{\Phi}|\boldsymbol{\tau})p(\boldsymbol{\tau})}{p(\mathbf{y}|\boldsymbol{\tau})}, \tag{9}$$

where $p(\mathbf{y}|\boldsymbol{\Phi}, \boldsymbol{\tau})$ is the likelihood function of the number of cases y_{it}, and $p(\mathbf{y}|\boldsymbol{\tau})$ is the marginal likelihood of the data given hyper-parameters $\boldsymbol{\tau}$.$p(\mathbf{y}|\boldsymbol{\tau})$ is typically taken as a normalization constant as it does not depend on $\boldsymbol{\Phi}$. It can therefore be ignored in the estimations. Then, the posterior density can be specified as:

$$p(\boldsymbol{\Phi}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\Phi}, \boldsymbol{\tau})p(\boldsymbol{\Phi}|\boldsymbol{\tau})p(\boldsymbol{\tau}), \tag{10}$$

where the 'equal to' sign (=) is replaced by the 'proportional to' sign ($\propto$).

For the Poisson distribution, the likelihood function of the number of cases y_{it} can be expressed as:

$$p(\mathbf{y}|\boldsymbol{\Phi}, \boldsymbol{\tau}) = \prod_{i=1}^{n}\prod_{t=1}^{T} \frac{\exp\left(-E_{it}\theta_{it}\right)\left(E_{it}\theta_{it}\right)^{y_{it}}}{y_{it}!}. \tag{11}$$

The prior distributions are used to model the spatial and temporal effect components. The spatially structured random effect of region i $\left(\omega_i\right)$ is modelled using the intrinsic conditional autoregressive prior (Besag et al. 1991):

$$\omega_i|\boldsymbol{\omega}_{-i}, \tau_\omega, \mathbf{W} \sim \mathcal{N}\left(\frac{\sum_{j=1}^{n} w_{ij}\omega_j}{\sum_{i=1}^{n} w_{ij}}, \frac{1}{\tau_\omega \sum_{i=1}^{n} w_{ij}}\right)\forall t \quad \text{and} \quad i = 1, \ldots, n, \tag{12}$$

where $\boldsymbol{\omega}_{-i}$ indicates all the elements in $\boldsymbol{\omega}$ except the ith element, $\mathbf{W} = \left(w_{ij}\right)$ is the 'adjacency' matrix with $w_{ij} = 1$ if i and j are adjacent (i.e. are first-order contiguous) and $w_{ij} = 0$ otherwise, and τ_ω is the precision parameter of ω_i. The joint prior density function of $\boldsymbol{\omega} = \left(\omega_1, \ldots, \omega_n\right)'$ over time t is (Rue and Martino 2009):

$$\begin{aligned} p(\boldsymbol{\omega}|\tau_\omega) &\propto \tau_\omega^{\frac{n-1}{2}} \exp\left(-\frac{\tau_\omega}{2}\sum_{i\sim j}(\omega_i-\omega_j)^2\right) \\ &\propto \tau_\omega^{\frac{n-1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\omega}'\mathbf{Q}_\omega\boldsymbol{\omega}\right)\forall t, \end{aligned} \tag{13}$$

with the precision matrix $\mathbf{Q}_\omega = \tau_\omega \mathbf{R}_\omega$ and $\mathbf{R}_\omega$ the $n \times n$ spatial structure matrix defined as:

$$\mathbf{R}_\omega = \begin{cases} n_i & \text{if } i = j \\ -1 & \text{if } i \sim j \\ 0 & \text{otherwise} \end{cases}$$

where n_i is the number of neighbours of region i and $i \sim j$ denotes that regions i and j are neighbours.

The spatially unstructured random effect of region $i(v_i)$ follows an exchangeable normal distribution (i.e. a sequence of random variables that are independent and identically normally distributed (iid))

$$v_i|\tau_v \sim \mathcal{N}\left(0, \frac{1}{\tau_v}\right)\forall t \quad \text{and} \quad i = 1, \ldots, n, \tag{14}$$

where τ_v is the precision parameter of v_i. The joint density function of the vector $\boldsymbol{v} = (v_1, \ldots, v_n)'$ over time t is (Rue and Martino 2009):

$$\begin{aligned} p(\boldsymbol{v}|\tau_v) &\propto \tau_v^{\frac{n}{2}} \exp\left(-\frac{\tau_v}{2}\sum_{i=1}^{n} v_i^2\right) \\ &\propto \tau_v^{\frac{n}{2}} \exp\left(-\frac{1}{2}\boldsymbol{v}'\mathbf{Q}_v\boldsymbol{v}\right)\forall t, \end{aligned} \tag{15}$$

with $\mathbf{Q}_v = \tau_v \mathbf{I}_n$ being the precision matrix and $\mathbf{I}_n$ the $n \times n$ identity matrix.

Incidences of dengue disease normally increase when the breeding conditions for the *Aedes*-spp. mosquito are favourable, usually in the rainy season (Choi et al. 2016). Consequently, endemic outbreaks can have a regular, seasonal pattern (CDC 2014). In addition, there is likely to be a trend over long periods. The temporally structured component (ϕ_t) is thus the sum of a temporal trend (φ_t) and a seasonal component (κ_t):

$$\phi_t = \varphi_t + \kappa_t. \tag{16}$$

A common temporal trend (φ_t) is a random walk of order one (RW1):

$$\varphi_{t+1} - \varphi_t|\tau_\varphi \sim \mathcal{N}\left(0, \frac{1}{\tau_\varphi}\right)\forall i \quad \text{and} \quad t = 1, \ldots, T-1, \tag{17}$$

with τ_φ being the precision parameter. The joint density function of $\boldsymbol{\varphi} = \left(\varphi_1, \ldots, \varphi_T\right)'$ for region i is:

$$\begin{aligned} p\left(\boldsymbol{\varphi}|\tau_\varphi\right) &\propto \tau_\varphi^{\frac{(T-1)}{2}} \exp\left(-\frac{\tau_\varphi}{2}\sum_{t=1}^{T-1}\left(\varphi_{t+1}-\varphi_t\right)^2\right) \\ &\propto \tau_\varphi^{\frac{(T-1)}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\varphi}'\mathbf{Q}^{(\mathrm{RW1})}_{\varphi(T\times T)}\boldsymbol{\varphi}\right)\forall i, \end{aligned} \tag{18}$$

with the precision matrix $\mathbf{Q}^{(\mathrm{RW1})}_{\varphi(T\times T)} = \tau_\varphi \mathbf{R}^{(\mathrm{RW1})}_{\varphi(T\times T)}$ and $\mathbf{R}^{(\mathrm{RW1})}_{\varphi(T\times T)}$ being the $T \times T$ temporal trend structure matrix for the RW1 prior:

$$\mathbf{R}^{(RW1)}_{\phi(T\times T)} = \begin{pmatrix} 1 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & \ddots & \ddots & \ddots & & \\ & & -1 & 2 & -1 & \\ & & & -1 & 1 & \end{pmatrix}$$

where all empty cells of $\mathbf{R}^{(\mathrm{RW1})}_{\varphi(T\times T)}$ are zero.

A random walk of order two (RW2) could apply to φ_t instead of a RW1. This prior is more appropriate if the data has a pronounced linear trend. The temporal trend $\left(\varphi_t\right)$ of a RW2 is:

$$\varphi_t - 2\varphi_{t+1} + \varphi_{t+2}|\tau_\varphi \sim \mathcal{N}\left(0, \frac{1}{\tau_\varphi}\right)\forall i \quad \text{and} \quad t = 1, \ldots, T-2, \tag{19}$$

with joint density function for region i:

$$\begin{aligned} p\left(\boldsymbol{\varphi}|\tau_\varphi\right) &\propto \tau_\varphi^{\frac{(T-2)}{2}} \exp\left(-\frac{\tau_\varphi}{2}\sum_{t=1}^{T-2}\left(\varphi_t - 2\varphi_{t+1} + \varphi_{t+2}\right)^2\right) \\ &\propto \tau_\varphi^{\frac{(T-2)}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\varphi}'\mathbf{Q}^{(\mathrm{RW2})}_{\varphi(T\times T)}\boldsymbol{\varphi}\right)\forall i, \end{aligned} \tag{20}$$

with $\mathbf{Q}^{(\mathrm{RW2})}_{\varphi} = \tau_\varphi \mathbf{R}^{(\mathrm{RW2})}_{\varphi}$ being the precision matrix and $\mathbf{R}^{(\mathrm{RW2})}_{\varphi}$ the $T \times T$ temporal trend structure matrix of a RW2 prior:

$$\mathbf{R}^{(RW2)}_{\phi(T\times T)} = \begin{pmatrix} 1 & -2 & 1 & & & & \\ -2 & 5 & -4 & 1 & & & \\ 1 & -4 & 6 & -4 & 1 & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & 1 & -4 & 6 & -4 & 1 \\ & & & 1 & -4 & 5 & -2 \\ & & & & 1 & -2 & 1 \end{pmatrix}$$

The seasonal component (κ_t) with periodicity m (with m being smaller than the number of time periods T) is:

$$\kappa_t + \kappa_{t+1} + \cdots + \kappa_{t+m-1} | \tau_\kappa \sim \mathcal{N}\left(0, \frac{1}{\tau_\kappa}\right) \forall i \quad \text{and} \quad t = 1, \ldots, T-m+1, \quad (21)$$

where τ_κ is the precision parameter of κ_t. The joint density function of $\boldsymbol{\kappa} = \left(\kappa_1, \ldots, \kappa_T\right)'$ is:

$$\begin{aligned} \mathrm{p}\left(\boldsymbol{\kappa} | \tau_\kappa\right) &\propto \tau_\kappa^{\frac{(T-m+1)}{2}} \exp\left(-\frac{\tau_\kappa}{2} \sum_{t=1}^{T-m+1} \left(\kappa_t + \kappa_{t+1} + \ldots + \kappa_{t+m-1}\right)^2\right) \\ &\propto \tau_\kappa^{\frac{(T-m+1)}{2}} \exp\left(-\frac{1}{2} \boldsymbol{\kappa}' \mathbf{Q}_\kappa \boldsymbol{\kappa}\right) \forall i, \end{aligned} \quad (22)$$

with $\mathbf{Q}_\kappa = \tau_\kappa \mathbf{R}_\kappa$ being the precision matrix and $\mathbf{R}_\kappa$ the $T \times T$ temporal structure matrix for the seasonal prior:

$$\mathbf{R}_{\kappa(T\times T)} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 & & & & & & \\ 1 & 2 & 2 & \cdots & 2 & 1 & & & & & \\ 1 & 2 & \ddots & \ddots & \ddots & 2 & 1 & & & & \\ \vdots & \vdots & \ddots & m-1 & m-1 & \ddots & 2 & 1 & & & \\ 1 & 2 & \ddots & m-1 & m & m-1 & \ddots & \ddots & \ddots & & \\ & 1 & 2 & \ddots & \ddots & \ddots & \ddots & \ddots & 2 & 1 & \\ & & \ddots & \ddots & \ddots & m-1 & m & m-1 & \ddots & 2 & 1 \\ & & & 1 & 2 & \ddots & m-1 & m-1 & \ddots & \vdots & \vdots \\ & & & & 1 & 2 & \ddots & \ddots & \ddots & 2 & 1 \\ & & & & & 1 & 2 & \cdots & 2 & 2 & 1 \\ & & & & & & 1 & \cdots & 1 & 1 & 1 \end{pmatrix}$$

For example, for $m = 4$ we have:

$$\mathbf{R}_{\kappa(T\times T)} = \begin{pmatrix} 1 & 1 & 1 & 1 & & & & & \\ 1 & 2 & 2 & 2 & 1 & & & & \\ 1 & 2 & 3 & 3 & 2 & 1 & & & \\ 1 & 2 & 3 & 4 & 3 & 2 & 1 & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & 1 & 2 & 3 & 4 & 3 & 2 & 1 \\ & & & 1 & 2 & 3 & 3 & 2 & 1 \\ & & & & 1 & 2 & 2 & 2 & 1 \\ & & & & & 1 & 1 & 1 & 1 \end{pmatrix}$$

For the temporally unstructured component (γ_t) for region i, we assume an exchangeable normal distribution:

$$\gamma_t|\tau_\gamma \sim \mathcal{N}\left(0, \frac{1}{\tau_\gamma}\right)\forall i \quad \text{and} \quad t = 1, \ldots, T, \tag{23}$$

where τ_γ is the precision parameter of γ_t with joint density function of $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_T)'$ for region i given by:

$$\begin{aligned} p(\boldsymbol{\gamma}|\tau_\gamma) &\propto \tau_\gamma^{\frac{T}{2}} \exp\left(-\frac{\tau_\gamma}{2}\sum_{t=1}^{T}\gamma_t^2\right) \\ &\propto \tau_\gamma^{\frac{T}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\gamma}'\mathbf{Q}_\gamma\boldsymbol{\gamma}\right)\forall i, \end{aligned} \tag{24}$$

with $\mathbf{Q}_{\boldsymbol{\gamma}} = \tau_\gamma \mathbf{I}_T$ being the precision matrix and $\mathbf{I}_T$ the $T \times T$ identity matrix.

The random components $\boldsymbol{\omega}$, $\boldsymbol{\upsilon}$, $\boldsymbol{\varphi}$, $\boldsymbol{\kappa}$, and $\boldsymbol{\gamma}$ are described as the main effect, while $\boldsymbol{\delta}$ denotes the spatiotemporal interaction of the spatial and temporal main effects. Four types of interactions with corresponding priors for $\boldsymbol{\delta}$ have been proposed Knorr-Held (2000).

Type I Interaction Combines the spatially and temporally unstructured main effects (υ_i and γ_t). Then, $\boldsymbol{\delta}$ is independent and identically Gaussian distributed with a mean of zero and precision τ_δ with a structure matrix given by $\mathbf{R}_{\delta(\mathrm{I})} = \mathbf{R}_\upsilon \otimes \mathbf{R}_\gamma = \mathbf{I}_n \otimes \mathbf{I}_T = \mathbf{I}_{nT}$, where $\otimes$ denotes the Kronecker product. The density of the interaction component $\boldsymbol{\delta}$ is then:

$$\begin{aligned} p(\boldsymbol{\delta}|\tau_\delta) &\propto \tau_\delta^{\frac{nT}{2}} \exp\left(-\frac{\tau_\delta}{2}\sum_{i=1}^{n}\sum_{t=1}^{T}\delta_{it}^2\right) \\ &\propto \tau_\delta^{\frac{nT}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\delta}'\mathbf{Q}_{\delta(\mathrm{I})}\boldsymbol{\delta}\right) i = 1, \ldots, n \quad \text{and} \quad t = 1, \ldots, T \end{aligned} \tag{25}$$

where $\boldsymbol{\delta} = (\delta_{11}, \ldots., \delta_{nT})'$ and the precision matrix $\mathbf{Q}_{\delta(\mathrm{I})} = \tau_\delta \mathbf{R}_{\delta(\mathrm{I})}$. The interaction can be thought of as unobserved covariates for each observation *it* that do not have any structure in space and time.

Type II interaction Combines the spatially unstructured main effect (υ_i) and one of the temporally structured main effects $(\varphi_t$ or $\kappa_t)$. Due to the complexity of the computations, the temporal trend (φ_t) is usually used. Then, $\boldsymbol{\delta}_i = (\delta_{it}, \ldots, \delta_{iT})' \quad i = 1, 2, \ldots, n$ follows an independent RW1 or RW2 and $\boldsymbol{\delta}$ is independent and identically normally distributed with mean zero. The structure matrix of type II interaction is $\mathbf{R}_\delta = \mathbf{R}_\upsilon \otimes \mathbf{R}_\varphi$, where $\mathbf{R}_\upsilon = \mathbf{I}_n$ and $\mathbf{R}_\varphi$ is the temporal structure matrix of RW1 or RW2. This interaction assumes that the temporal trends are different from region to region but do not have any structure in a region. The structured matrix $\mathbf{R}_\delta$ has rank $n(T-1)$ for RW1 and $n(T-2)$ for RW1, and the prior for $\boldsymbol{\delta}$ is:

$$\begin{aligned} \text{RW1} : p(\boldsymbol{\delta}|\tau_\delta) &\propto \tau_\delta^{\frac{n(T-1)}{2}} \exp\left(-\frac{\tau_\delta}{2}\sum_{i=1}^{n}\sum_{t=1}^{T-1}\left(\delta_{i,t+1} - \delta_{i,t}\right)^2\right) \\ &\propto \tau_\delta^{\frac{n(T-1)}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\delta}'\mathbf{Q}_{\delta(\mathrm{II})}^{(\mathrm{RW1})}\boldsymbol{\delta}\right) i = 1, \ldots, n \quad \text{and} \quad t = 1, \ldots, T-1, \end{aligned} \tag{26}$$

$$\begin{aligned}\text{RW2}: p(\boldsymbol{\delta}|\tau_\delta) &\propto \tau_\delta^{\frac{n(T-2)}{2}} \exp\left(-\frac{\tau_\delta}{2}\sum_{i=1}^{n}\sum_{t=1}^{T-2}\left(\delta_{i,t}-2\delta_{i,t+1}+\delta_{i,t+2}\right)^2\right)\\ &\propto \tau_\delta^{\frac{n(T-2)}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\delta}'\mathbf{Q}_{\delta(\text{II})}^{(\text{RW2})}\boldsymbol{\delta}\right) i=1,\ldots,n \quad \text{and} \quad t=1,\ldots,T-2,\end{aligned} \tag{27}$$

where $\boldsymbol{\delta}=\left(\boldsymbol{\delta}_1',\ldots.,\boldsymbol{\delta}_n'\right)'$, and $\mathbf{Q}_{\delta(\text{II})}^{(\text{RW1})}=\tau_\delta\mathbf{R}_{\delta(\text{II})}^{(\text{RW1})}$ is the precision matrix for RW1 and $\mathbf{Q}_{\delta(\text{II})}^{(\text{RW2})}=\tau_\delta\mathbf{R}_{\delta(\text{II})}^{(\text{RW2})}$ for RW2, and where $\mathbf{R}_{\delta(\text{II})}^{(\text{RW1})}$ and $\mathbf{R}_{\delta(\text{II})}^{(\text{RW2})}$ are the temporal trend structure matrices for RW1 and RW2, respectively.

Type III interaction Combines the temporally unstructured main effect (γ_t) and the spatially structured main effect (ω_i). Then, $\boldsymbol{\delta}_t=\left(\delta_{1t},\ldots,\delta_{nt}\right)'$ $t=1,\ldots,T$ follows an independent iCAR prior with a mean of zero. The structure matrix of type III interaction is $\mathbf{R}_{\delta(\text{III})}=\mathbf{R}_\gamma\otimes\mathbf{R}_\omega$, where $\mathbf{R}_\gamma=\mathbf{I}_T$ and $\mathbf{R}_\omega$ is the spatial structure matrix defined through the iCAR prior specification. This interaction assumes that the spatially structured components are independent over time. The structured matrix $\mathbf{R}_{\delta(\text{III})}$ has rank $T(n-1)$, and the prior for $\boldsymbol{\delta}$ is provided by:

$$\begin{aligned}p(\boldsymbol{\delta}|\tau_\delta) &\propto \tau_\delta^{\frac{T(n-1)}{2}} \exp\left(-\frac{\tau_\delta}{2}\sum_{t=1}^{T}\sum_{i\sim j}^{n}\left(\delta_{it}-\delta_{jt}\right)^2\right)\\ &\propto \tau_\delta^{\frac{T(n-1)}{2}} \exp\left(-\frac{\tau_\delta}{2}\boldsymbol{\delta}'\mathbf{Q}_{\delta(III)}\boldsymbol{\delta}\right) i=1,\ldots,n \quad \text{and} \quad t=1,\ldots,T\end{aligned} \tag{28}$$

with $\boldsymbol{\delta}=\left(\boldsymbol{\delta}_1',\ldots,\boldsymbol{\delta}_T'\right)'$ and the precision matrix being $\mathbf{Q}_{\delta(\text{III})}=\tau_\delta\mathbf{R}_{\delta(\text{III})}$.

Type IV interaction Combines the spatially and temporally structured main effects (ω_i and φ_t), which imply that $\boldsymbol{\delta}=\left(\delta_{11},\ldots,\delta_{nt}\right)'i=1,\ldots,n$ and $t=1,\ldots,T$ is dependent over space and time. The temporal dependency structure for each region therefore depends on the temporal structure of the neighbouring regions. Then, $\boldsymbol{\delta}$ is independent and identically normally distributed with a mean of zero. The structure matrix of type IV interaction is $\mathbf{R}_{\delta(IV)}=\mathbf{R}_\omega\otimes\mathbf{R}_\varphi$. $\boldsymbol{R}_\omega$ denotes the spatial structure matrix defined by the iCAR prior with $\mathbf{R}_\varphi$ being the temporal trend structure matrix defined through RW1 or RW2. The structured matrix $\mathbf{R}_\delta$ has rank $(T-1)(n-1)$ for RW1 and $(T-2)(n-1)$ for RW2. The joint density prior for $\boldsymbol{\delta}$ is provided by:

$$\begin{aligned}\text{RW1}: p(\boldsymbol{\delta}|\tau_\delta) &\propto \tau_\delta^{\frac{n(T-1)}{2}} \exp\left(-\frac{\tau_\delta}{2}\sum_{t=1}^{T-1}\sum_{i\sim j}^{n}\left(\delta_{i,t+1}-\delta_{j,t+1}-\delta_{i,t}+\delta_{j,t}\right)^2\right)\\ &\propto \tau_\delta^{\frac{n(T-1)}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\delta}'\mathbf{Q}_{\delta(IV)}^{(\text{RW1})}\boldsymbol{\delta}\right) i=1,\ldots,n \quad \text{and} \quad t=1,\ldots,T-1\end{aligned} \tag{29}$$

$$\begin{aligned}\text{RW2}: p(\boldsymbol{\delta}|\tau_\delta) &\propto \tau_\delta^{\frac{n(T-2)}{2}} \exp\left(-\frac{\tau_\delta}{2}\sum_{t=1}^{T-2}\sum_{i\sim j}^{n}\left(\delta_{i,t}-\delta_{j,t}-2\delta_{i,t+1}+2\delta_{j,t+1}+\delta_{i,t+2}-\delta_{j,t+2}\right)^2\right)\\ &\propto \tau_\delta^{\frac{n(T-2)}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\delta}'\mathbf{Q}_{\delta(IV)}^{(\text{RW2})}\boldsymbol{\delta}\right) i=1,\ldots,n \quad \text{and} \quad t=1,\ldots,T-2\end{aligned} \tag{30}$$

with $\boldsymbol{\delta} = \left(\delta_{11}, \ldots, \delta_{nT}\right)'$, and where $\mathbf{Q}_{\delta(IV)}^{(\mathrm{RW1})} = \tau_\delta \mathbf{R}_{\delta(IV)}^{(\mathrm{RW1})}$ is the precision matrix for RW1 and $\mathbf{Q}_{\delta(IV)}^{(\mathrm{RW2})} = \tau_\delta \mathbf{R}_{\delta(IV)}^{(\mathrm{RW2})}$ for RW2, where $\mathbf{R}_{\delta(IV)}^{(\mathrm{RW1})}$ and $\mathbf{R}_{\delta(IV)}^{(\mathrm{RW2})}$ are the temporal trend structure matrices for the RW1 and RW2 priors, respectively.

Up to a proportionality constant, the product of the likelihood (for the Poisson distribution defined in Eq. 11) and the independent prior distributions for the unknown parameters $p(\alpha)$, $p(\boldsymbol{\omega})$, $p(\boldsymbol{\nu})$, $p(\boldsymbol{\varphi})$, $p(\boldsymbol{\kappa})$, $p(\boldsymbol{\gamma})$, $p(\boldsymbol{\delta})$, $p\left(\tau_\nu\right)$, $p\left(\tau_\omega\right)$, $p\left(\tau_\varphi\right)$, $p\left(\tau_\kappa\right)$, $p\left(\tau_\gamma\right)$, $p\left(\tau_\delta\right)$ yields the joint posterior distribution of the model parameters.

$$p\left(\alpha, \boldsymbol{\omega}, \boldsymbol{\nu}, \boldsymbol{\varphi}, \boldsymbol{\kappa}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \tau_\nu, \tau_\omega, \tau_\varphi, \tau_\kappa, \tau_\gamma, \tau_\delta | \mathbf{y}\right) \propto p\left(\mathbf{y} | \alpha, \boldsymbol{\omega}, \boldsymbol{\nu}, \boldsymbol{\varphi}, \boldsymbol{\kappa}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \tau_\nu, \tau_\omega, \tau_\varphi, \tau_\kappa, \tau_\gamma, \tau_\delta\right) p(\alpha) p(\boldsymbol{\omega}) p(\boldsymbol{\nu}) p(\boldsymbol{\varphi}) p(\boldsymbol{\kappa}) p(\boldsymbol{\gamma}) p(\boldsymbol{\delta}) p\left(\tau_\nu\right) p\left(\tau_\omega\right) p\left(\tau_\varphi\right) p\left(\tau_\kappa\right) p\left(\tau_\gamma\right) p\left(\tau_\delta\right). \tag{31}$$

The marginal posterior distributions are obtained from Eq. (31), and from which summary statistics (e.g. mean, median, mode or quantiles for credible intervals) can be derived.

2.2 Priors

We then specify the prior distributions of the intercept (α) and of the hyper-parameter $\left(\sigma_\omega^2 = 1/\tau_\omega, \sigma_\nu^2 = 1/\tau_\nu, \sigma_\varphi^2 = 1/\tau_\varphi, \sigma_{\boldsymbol{\kappa}}^2 = 1/\tau_{\boldsymbol{\kappa}}, \sigma_\gamma^2 = 1/\tau_\gamma, \sigma_{\boldsymbol{\delta}}^2 = 1/\tau_{\boldsymbol{\delta}}\right)$. Following Blangiardo et al. (2013), we specify a vague Gaussian prior distribution with zero mean and a large variance $\sigma_\alpha^2 = \tau_\alpha^{-1}$ for α, i.e. $\alpha \sim \mathcal{N}\left(0, 10^{-5}\right)$. There are several kinds of hyper-prior distributions for the hyper-parameters, notably the inverse Gamma (IG), penalized complexity (PC), half Cauchy (HC) and the uniform (U) distribution (Wang et al. 2018; Adin et al. 2018; Gelman 2006). For the inverse Gamma hyper-prior, we need to choose the shape and scale parameters a and b, respectively. Following Bernardinelli et al. (1995) and Schrödle and Held (2011), we set the shape parameter to equal one. The scale parameter can vary for each of the above hyper-parameters. For the temporal trend RW1 or RW2, Schrödle and Held (2011) recommended $\sigma_\varphi^2 \sim \mathrm{IG}(1, 0.00005)$ and $\mathrm{IG}(1, 0.01)$ for the precision parameters of the structured and unstructured spatial random effects, and for the seasonal and unstructured temporal random effects. Scaling is also required for the PC and HC hyper-priors. Wang et al. (2018) recommended the standard deviation of ln(SIR) as a scale for the PC hyper-prior, Gelman (2006) proposed 25 as scale parameter for the HC hyper-prior, and Adin et al. (2018) suggested $(0, \infty)$ for the uniform hyper-prior. The hyper-priors and hyper-parameters can affect the estimation results substantially and therefore need to be specified carefully so as to obtain reliable inferences (Ugarte et al. 2014). We apply sensitivity analysis to select the optimal hyper-priors and hyper-parameters in the case study.

2.3 Integrated nested Laplace approximation (INLA)

As mentioned in the Introduction, we apply INLA[2] estimation, which is based on numerical integration of the posterior (Rue and Martino 2009). It is a fast inference method for latent Gaussian models (LGMs), i.e. a subclass of structured (additive) regression models.[3]

A spatiotemporal disease model such as Eq. (8), defined as an LGM, is a three-stage hierarchical model. The first stage consists of the likelihood function of the observed variable $\mathbf{y}$, $p(\mathbf{y}|\boldsymbol{\Phi},\boldsymbol{\tau})$, where $\mathbf{y}$ is the observational n-vector, vector $\boldsymbol{\Phi}$ is the Gaussian field which contains all the latent (non-observable) model components $\boldsymbol{\Phi} = \left(\alpha, \boldsymbol{\omega}', \boldsymbol{\nu}', \boldsymbol{\varphi}', \boldsymbol{\kappa}', \boldsymbol{\gamma}', \boldsymbol{\delta}'\right)'$, and $\boldsymbol{\tau} = \left(\tau_\alpha, \tau_\omega, \tau_\nu, \tau_\varphi, \tau_\kappa, \tau_\gamma, \tau_\delta\right)'$ is the hyper-parameter vector of $\boldsymbol{\Phi}$. By assuming conditional independence of the Gaussian field (i.e. a Gaussian Markov random field (GMRF)[4]), the distribution of the nT observations is provided by the likelihood function:

$$p(\mathbf{y}|\boldsymbol{\Phi},\boldsymbol{\tau}) = \prod_{i=1}^{n} p\left(\mathbf{y}_i|\boldsymbol{\Phi}_i,\boldsymbol{\tau}\right) = \prod_{i=1}^{n}\prod_{t=1}^{T} p\left(y_{it}|\boldsymbol{\Phi}_{it},\boldsymbol{\tau}\right), \tag{32}$$

where each data point $\mathbf{y}_i = \left(y_{i1},\ldots,y_{iT}\right)'$ is connected to only one element $\boldsymbol{\Phi}_i = \left(\boldsymbol{\Phi}_{i1},\ldots,\boldsymbol{\Phi}_{iT}\right)'$ in the latent field $\boldsymbol{\Phi}$. This implies that the parameters are constant and $\mathbf{y}$ and $\boldsymbol{\Phi}$ have the same dimension.

The second stage consists of the conditional distribution of the latent Gaussian field $\boldsymbol{\Phi}$ given the hyper-parameter vector $\boldsymbol{\tau}$, $p(\boldsymbol{\Phi}|\boldsymbol{\tau})$, which follows a multivariate Gaussian distribution with a zero mean and a sparse precision matrix $\mathbf{Q}(\boldsymbol{\tau})$ (due to conditional independence assumption of $\boldsymbol{\Phi}$), i.e. $\boldsymbol{\Phi}|\boldsymbol{\tau} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}^{-1}(\boldsymbol{\tau})\right)$. The prior density of $\boldsymbol{\Phi}$ is:

$$p(\boldsymbol{\Phi}|\boldsymbol{\tau}) = (2\pi)^{nT}|\mathbf{Q}(\boldsymbol{\tau})|^{\frac{1}{2}}\exp\left(-\frac{1}{2}\boldsymbol{\Phi}'\mathbf{Q}(\boldsymbol{\tau})\boldsymbol{\Phi}\right), \tag{33}$$

where $|.|$ denotes the determinant.

The last stage involves $p(\boldsymbol{\tau})$, where the hyper-prior distribution of $\boldsymbol{\tau}$. $p(\boldsymbol{\tau})$ need not be Gaussian. According to Schrödle and Held (2011), an appropriate distribution for the precision parameters is the inverse Gamma.

[2] The INLA package can be downloaded for free at www.r-inla.org .

[3] LGMs are generalized linear models (GLMs) with the linear predictor being replaced by a possibly nonlinear structured (additive) predictor. LGMs are suitable for modelling temporal or spatial dependency (Hicketier 2015).

[4] For INLA to work properly and to achieve a substantial reduction in computation time, the LGM should have the following properties: (1) the latent field $\boldsymbol{\Phi}$ satisfies the conditional independence property, i.e. it should be a Gaussian Markov random field (GMRF). A GMRF $\boldsymbol{\Phi}$ is a Gaussian vector where $\boldsymbol{\Phi}_i$ and $\boldsymbol{\Phi}_j$ are conditionally independent, given the remaining elements $\boldsymbol{\Phi}_{-ij}$. Notation: $\boldsymbol{\Phi}_i \perp \boldsymbol{\Phi}_j | \boldsymbol{\Phi}_{-ij}$. The conditional independence property defines the zero pattern of the precision matrix $\mathbf{Q}(\boldsymbol{\tau})$ in that for a pair, i and j, with $j \neq i$, the corresponding element of the precision matrix is zero (Blangiardo and Cameletti 2015): $\boldsymbol{\Phi}_i \perp \boldsymbol{\Phi}_j \left| \boldsymbol{\Phi}_{-ij} \right. \Leftrightarrow Q_{ij}(\boldsymbol{\tau}) = 0$. Note that sparse matrices with large numbers of zeros are easier to invert. (2) The number of hyper-parameters is small.

For the hierarchical model described above, the joint posterior distribution of $\boldsymbol{\Phi}$ and $\boldsymbol{\tau}$ reads as:

$$\begin{aligned} p(\boldsymbol{\Phi}, \boldsymbol{\tau}|\mathbf{y}) &= \frac{p(\boldsymbol{\tau})p(\boldsymbol{\Phi}|\boldsymbol{\tau})p(\mathbf{y}|\boldsymbol{\Phi}, \boldsymbol{\tau})}{\boldsymbol{p}(\mathbf{y}|\boldsymbol{\tau})} \\ &\propto p(\boldsymbol{\tau})|\mathbf{Q}(\boldsymbol{\tau})|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\Phi}'\mathbf{Q}(\boldsymbol{\tau})\boldsymbol{\Phi} + \sum_{i=1}^{n} \log\left(\mathbf{y}_i|\boldsymbol{\Phi}_i, \boldsymbol{\tau}\right) \right). \end{aligned} \tag{34}$$

Instead of considering the full posterior distributions of $\boldsymbol{\Phi}$ and $\boldsymbol{\tau}$, INLA proceeds on the basis of approximations to the marginal posterior distributions $p\left(\boldsymbol{\Phi}_i|\mathbf{y}\right)$ and $p\left(\tau_k|\mathbf{y}\right)$, where $\tau_1 = \tau_\alpha, \tau_2 = \tau_\omega, \tau_3 = \tau_\nu, \tau_4 = \tau_\varphi, \tau_5 = \tau_\kappa, \tau_6 = \tau_\gamma$ and $\tau_7 = \tau_\delta$. The marginal posterior distribution of $\boldsymbol{\Phi}_i$ is:

$$p\left(\boldsymbol{\Phi}_i|\mathbf{y}\right) = \int p\left(\boldsymbol{\Phi}_i, \boldsymbol{\tau}|\mathbf{y}\right)\mathrm{d}\boldsymbol{\tau} = \int p\left(\boldsymbol{\Phi}_i|\boldsymbol{\tau}, \mathbf{y}\right)\mathrm{p}(\boldsymbol{\tau}|\mathbf{y})\mathrm{d}\boldsymbol{\tau} i = 1, \ldots, n, \tag{35}$$

where n is the number of spatial units. The marginal posterior distribution of τ_k is:

$$p\left(\tau_k|\mathbf{y}\right) = \int p(\boldsymbol{\tau}|\mathbf{y})\mathrm{d}\boldsymbol{\tau}_{-\mathrm{k}} \quad k = 1, \ldots, 7, \tag{36}$$

where $\boldsymbol{\tau}_{-k}$ indicates all the elements in $\boldsymbol{\tau}$ except the kth element.

To obtain Eqs. (35) and (36), the following tasks must be performed: (1) compute the marginal posterior distributions of the hyper-parameters $p\left(\boldsymbol{\tau}_k|\mathbf{y}\right)$, and (2) compute the conditional posterior distribution $p\left(\boldsymbol{\Phi}_i|\boldsymbol{\tau}, \mathbf{y}\right)$, which is needed to compute the marginal posterior of the parameters of $p\left(\boldsymbol{\Phi}_i|\mathbf{y}\right)$.

INLA numerically approximates the posteriors of interest based on Laplace transformation (Rue and Martino 2009). It approximates the integrand with a second-order Taylor series around the mode. For details, see Appendix '1'. Note that INLA uses approximations nested within each other (Rue et al. 2017). For the first task, $p(\boldsymbol{\tau}|\mathbf{y})$ is approximated using Laplace approximation (Rue and Martino 2009)[5]:

$$\begin{aligned} p(\boldsymbol{\tau}|\mathbf{y}) &= \frac{p(\boldsymbol{\Phi}, \boldsymbol{\tau}|\mathbf{y})}{p(\boldsymbol{\Phi}|\boldsymbol{\tau}, \mathbf{y})} = \frac{p(\mathbf{y}|\boldsymbol{\Phi}, \boldsymbol{\tau})p(\boldsymbol{\Phi}|\boldsymbol{\tau})p(\boldsymbol{\tau})}{p(\mathbf{y}|\boldsymbol{\tau})} \frac{1}{p(\boldsymbol{\Phi}|\boldsymbol{\tau}, \mathbf{y})} \\ &\propto \frac{p(\mathbf{y}|\boldsymbol{\Phi}, \boldsymbol{\tau})p(\boldsymbol{\Phi}|\boldsymbol{\tau})p(\boldsymbol{\tau})}{p(\boldsymbol{\Phi}|\boldsymbol{\tau}, \mathbf{y})} \approx \left.\frac{p(\mathbf{y}|\boldsymbol{\Phi}, \boldsymbol{\tau})p(\boldsymbol{\Phi}|\boldsymbol{\tau})p(\boldsymbol{\tau})}{p_{\mathrm{G}}(\boldsymbol{\Phi}|\boldsymbol{\tau}, \mathbf{y})}\right|_{\boldsymbol{\Phi}=\boldsymbol{\Phi}^*(\boldsymbol{\tau})} =: \tilde{p}(\boldsymbol{\tau}|\mathbf{y}), \end{aligned} \tag{37}$$

where $p(\boldsymbol{\Phi}|\boldsymbol{\tau}, \mathbf{y})$ is the conditional distribution of $\boldsymbol{\Phi}, p_{\mathrm{G}}(\boldsymbol{\Phi}|\boldsymbol{\tau}, \mathbf{y})$, which is its Gaussian approximation by Laplace transformation, $\boldsymbol{\Phi}^*(\boldsymbol{\tau})$ is the posterior mode (i.e. local maximum) of $p(\boldsymbol{\Phi}|\boldsymbol{\tau}, \mathbf{y})$ for $\boldsymbol{\tau}$, and $\tilde{p}(\boldsymbol{\tau}|\mathbf{y})$ is the Laplace approximation of $p(\boldsymbol{\tau}|\mathbf{y})$ (Tierney and Kadane 1986). For details see Appendix '2'.

[5] Using Bayes' theorem, the joint posterior distribution of the hyper-parameters can be written as $p(\boldsymbol{\tau}|\mathbf{y}) = p(\boldsymbol{\Phi}, \boldsymbol{\tau}|\mathbf{y})/p(\boldsymbol{\Phi}|\boldsymbol{\tau}, \mathbf{y})$ and the joint posterior distribution of $\boldsymbol{\Phi}$ and $\boldsymbol{\tau}$ as $p(\boldsymbol{\Phi}, \boldsymbol{\tau}|\mathbf{y}) = p(\mathbf{y}|\boldsymbol{\Phi}, \boldsymbol{\tau})$ $p(\boldsymbol{\Phi}|\boldsymbol{\tau})p(\boldsymbol{\tau})/p(\mathbf{y}, \boldsymbol{\tau}) \propto p(\mathbf{y}|\boldsymbol{\Phi}, \boldsymbol{\tau})p(\boldsymbol{\Phi}|\boldsymbol{\tau})p(\boldsymbol{\tau})$.

Note that the Gaussian approximation is accurate, since $p(\boldsymbol{\Phi}|\boldsymbol{\tau},\mathbf{y}) \propto \exp\Big(-\frac{1}{2}\boldsymbol{\Phi}'\mathbf{Q}(\boldsymbol{\tau})\boldsymbol{\Phi} + \sum_i \log p(\mathbf{y}_i|\boldsymbol{\Phi}_i,\boldsymbol{\tau})\Big)$ is almost Gaussian, as $\boldsymbol{\Phi}$ is a GMRF (assumed at stage 2). The resulting approximation will then be (Opitz 2017):

$$p_{\mathrm{G}}(\boldsymbol{\Phi}|\boldsymbol{\tau},\mathbf{y}) \propto \exp\left(-\frac{1}{2}\big(\boldsymbol{\Phi}-\boldsymbol{\Phi}^*(\boldsymbol{\tau})\big)'\big(\mathbf{Q}(\boldsymbol{\tau})+\mathrm{diag}(\mathbf{c})\big)\big(\boldsymbol{\Phi}-\boldsymbol{\Phi}^*(\boldsymbol{\tau})\big)\right), \tag{38}$$

where the vector $\mathbf{c}$ is the second-order term in the Taylor expansion of $\sum_i \log p\big(\mathbf{y}_i|\boldsymbol{\Phi}_i,\boldsymbol{\tau}\big)$, at modal value $\boldsymbol{\Phi}^*(\boldsymbol{\tau})$.

The second task is the computation of the marginal posterior conditional distribution $p\big(\boldsymbol{\Phi}_i|\boldsymbol{\tau},\mathbf{y}\big)$ which is needed to compute the marginal posterior $p\big(\boldsymbol{\Phi}_i|\mathbf{y}\big)$. It is more complex than (i), because $\boldsymbol{\Phi}$ consists of more elements than $\boldsymbol{\tau}$. There are three different strategies to approximate $p\big(\boldsymbol{\Phi}_i|\boldsymbol{\tau},\mathbf{y}\big)$: Gaussian, full Laplace and simplified Laplace approximation (Rue and Martino 2009). The Gaussian approximation $p_G(\boldsymbol{\Phi}|\boldsymbol{\tau},\mathbf{y})$ in Eq. (38) is used to obtain the marginal posterior $p_G\big(\boldsymbol{\Phi}_i|\boldsymbol{\tau},\mathbf{y}\big)$ directly. However, this approximation is generally not good if the true density of $p\big(\boldsymbol{\Phi}_i|\boldsymbol{\tau},\mathbf{y}\big)$ is not symmetric (Blangiardo and Cameletti 2015). The full Laplace approximation is a correction of Gaussian approximation. It rewrites the parameters vector as $\boldsymbol{\Phi} = \big(\boldsymbol{\Phi}_i,\boldsymbol{\Phi}_{-i}\big)$ and uses Gaussian approximation to obtain $\tilde{p}\big(\boldsymbol{\Phi}_i|\boldsymbol{\tau},\mathbf{y}\big)$. In particular, using the conditional probability rule, $p\big(\big(\boldsymbol{\Phi}_i,\boldsymbol{\Phi}_{-i}\big)|\boldsymbol{\tau},\mathbf{y}\big)$ can be expressed as:

$$p\big(\big(\boldsymbol{\Phi}_i,\boldsymbol{\Phi}_{-i}\big)|\boldsymbol{\tau},\mathbf{y}\big) = p\big(\boldsymbol{\Phi}_{-i}|\boldsymbol{\Phi}_i,\boldsymbol{\tau},\mathbf{y}\big)p\big(\boldsymbol{\Phi}_i|\boldsymbol{\tau},\mathbf{y}\big) \tag{39}$$

and with some simple manipulations, the following is obtained:

$$\begin{aligned} p\big(\boldsymbol{\Phi}_i|\boldsymbol{\tau},\mathbf{y}\big) &= \frac{p\big(\big(\boldsymbol{\Phi}_i,\boldsymbol{\Phi}_{-i}\big)|\boldsymbol{\tau},\mathbf{y}\big)}{p\big(\boldsymbol{\Phi}_{-i}|\boldsymbol{\Phi}_i,\boldsymbol{\tau},\mathbf{y}\big)} = p(\boldsymbol{\Phi}|\boldsymbol{\tau},\mathbf{y})\frac{1}{p\big(\boldsymbol{\Phi}_{-i}|\boldsymbol{\Phi}_i,\boldsymbol{\tau},\mathbf{y}\big)} \\ &= \frac{p(\boldsymbol{\Phi},\boldsymbol{\tau}|\mathbf{y})}{p(\boldsymbol{\tau}|\mathbf{y})\mathrm{p}(\mathbf{y})}\frac{1}{p\big(\boldsymbol{\Phi}_{-i}|\boldsymbol{\Phi}_i,\boldsymbol{\tau},\mathbf{y}\big)} \propto \frac{p(\boldsymbol{\Phi},\boldsymbol{\tau}|\mathbf{y})}{p(\boldsymbol{\tau}|\mathbf{y})}\frac{1}{p\big(\boldsymbol{\Phi}_{-i}|\boldsymbol{\Phi}_{\mathrm{i}},\boldsymbol{\tau},\mathbf{y}\big)} \\ &\propto \frac{p(\boldsymbol{\Phi},\boldsymbol{\tau}|\mathbf{y})}{p\big(\boldsymbol{\Phi}_{-i}|\boldsymbol{\Phi}_i,\boldsymbol{\tau},\mathbf{y}\big)} \approx \frac{p(\boldsymbol{\Phi},\boldsymbol{\tau}|\mathbf{y})}{p_{\mathrm{G}}\big(\boldsymbol{\Phi}_{-i}|\boldsymbol{\Phi}_i,\boldsymbol{\tau},\mathbf{y}\big)}\bigg|_{\boldsymbol{\Phi}_{-i}=\boldsymbol{\Phi}^*_{-i}(\boldsymbol{\Phi}_i,\boldsymbol{\tau})} =: \tilde{p}\big(\boldsymbol{\Phi}_i|\boldsymbol{\tau},\mathbf{y}\big), \end{aligned} \tag{40}$$

where $\boldsymbol{\Phi}_{-i}$ contains all the elements in $\boldsymbol{\Phi}$ except the ith element, $p_{\mathrm{G}}\big(\boldsymbol{\Phi}_{-i}|\boldsymbol{\Phi}_i,\boldsymbol{\tau},\mathbf{y}\big)$ is the Gaussian approximation of $p\big(\boldsymbol{\Phi}_{-i}|\boldsymbol{\Phi}_i,\boldsymbol{\tau},\mathbf{y}\big)$, with the whole expression being evaluated at $\boldsymbol{\Phi}^*_{-i}\big(\boldsymbol{\Phi}_i,\boldsymbol{\tau}\big)$, the mode of $p\big(\boldsymbol{\Phi}_{-i}|\boldsymbol{\Phi}_i,\boldsymbol{\tau},\mathbf{y}\big)$. The full Laplace approximation is very accurate because the random variable $\boldsymbol{\Phi}_{-i}|\boldsymbol{\Phi}_i,\boldsymbol{\tau},\mathbf{y}$ is generally reasonably Gaussian. However, the computational cost is very high because $p_{\mathrm{G}}\big(\boldsymbol{\Phi}_{-i}|\boldsymbol{\Phi}_i,\boldsymbol{\tau},\mathbf{y}\big)$ must be computed for each value of $\boldsymbol{\Phi}$ and $\boldsymbol{\tau}$. To avoid this problem, the third strategy is usually used, which is based on the Taylor series expansion of the full Laplace approximation $\tilde{p}\big(\boldsymbol{\Phi}_i|\boldsymbol{\tau},\mathbf{y}\big)$ in Eq. (40). This last strategy is sufficiently accurate for most applications (Blangiardo and Cameletti 2015). Having obtained $\tilde{p}\big(\boldsymbol{\Phi}_i|\boldsymbol{\tau},\mathbf{y}\big)$ and $\tilde{p}(\boldsymbol{\tau}|\mathbf{y})$, the marginal posterior distribution $p\big(\boldsymbol{\Phi}_i|\mathbf{y}\big)$ in Eq. (35) is approximated by:

$$\tilde{p}\big(\boldsymbol{\Phi}_i|\mathbf{y}\big) \approx \int \tilde{p}\big(\boldsymbol{\Phi}_i|\boldsymbol{\tau},\mathbf{y}\big)\tilde{p}(\boldsymbol{\tau}|\mathbf{y})\mathrm{d}\boldsymbol{\tau}, \tag{41}$$

where the integral can be solved numerically through a finite weighted sum:

$$\tilde{p}\left(\boldsymbol{\Phi}_i|\mathbf{y}\right) \approx \sum_j \tilde{p}\left(\boldsymbol{\Phi}_i|\boldsymbol{\tau}^{(j)},\mathbf{y}\right)\tilde{p}\left(\boldsymbol{\tau}^{(\mathrm{j})}|\mathbf{y}\right)\Delta_j j = 1,\ldots,J, \tag{42}$$

for a set of integration points $\left\{\boldsymbol{\tau}^{(j)}\right\}$ with a corresponding weight $\left\{\Delta_j\right\}$, with J denoting the number of evaluation points.[6]

2.4 Forecasting

The main objective of this paper to obtain forecasts of relative risk, $h = 1, 2, \ldots$ steps ahead from T, i.e. $\hat{y}_{i(T+h)}$, given the observed $\mathbf{y}$. The forecast values $\hat{y}_{i(T+h)}$ are obtained by fitting the model for $T + h$ observations with the forecasts defined as unobserved responses. The posterior predictive distribution needed for this purpose is obtained by integrating the posterior distribution $p\left(\boldsymbol{\Phi},\boldsymbol{\tau}|\hat{y}_{i(T+h)}\right) \propto p\left(\hat{y}_{i(T+h)}|\boldsymbol{\Phi},\boldsymbol{\tau}\right)p(\boldsymbol{\Phi}|\mathbf{y},\boldsymbol{\tau})$ over the parameters ($\boldsymbol{\Phi}$) in the latent field (Morrison et al. 2016):

$$p\left(\hat{y}_{i(T+h)}|\mathbf{y},\boldsymbol{\tau}\right) = \int p\left(\hat{y}_{i(T+h)}|\boldsymbol{\Phi},\boldsymbol{\tau}\right)\mathrm{p}(\boldsymbol{\Phi}|\mathbf{y},\boldsymbol{\tau})\mathrm{d}\boldsymbol{\Phi}, \tag{43}$$

where $\hat{y}_{i(T+h)}$ denotes the forecast value in the ith region at time t. To approximate the integral over the hyper-parameter $\boldsymbol{\Phi}$ in Eq. (43), a numerical method similar to the one used to approximate Eq. (42) can be used (Morrison et al. 2016).

Forecasting with INLA can be easily implemented by entering 'Not Available (NA)' for the$h$ observations we want to forecast, i.e. '$y_{i(T+h)}$ = NA' for the 'observation' at the time $T + h$ which needs to be forecasted (Wang et al. 2018). This is implemented in R-INLA by constructing vectors for the response variable and for the random effect components. Particularly, 'NA' is specified in the vector of observations for the h periods that the response variable is to be forecasted, $\left((1,\ldots,n)',\ldots,(1,\ldots,n)'\right)'$ in the corresponding vectors of the spatial random components, $\left((T+1,\ldots,T+1)',\ldots,(T+h,\ldots,T+h)'\right)'$ in the vectors of the temporal random components, and $((nT+1,\ldots,n(T+1))',\ldots,(n(T+h-1)+1,\ldots,n(T+h))')'$ in the vector of the random interaction term. For instance, $id\omega$ (index vector ω) shown in the second vector below denotes the vector of the spatially structured component (ω), and $id\phi$ denotes the vector of the temporally structured component. The data structured for the model yielded by Eq. (8) with interaction type I are expressed as follows (for further information, see Rue and Martino 2009):

[6] Evaluation points for $\tilde{p}(\boldsymbol{\tau}|\mathbf{y})$are needed for the numerical integration of Eq. (42). Two approaches have been proposed: grid search and central composite design (CCD). Grid strategy is the more accurate strategy; however, it is time-consuming. The CCD strategy is less time-consuming, but less accurate than the grid strategy (see for details, Rue and Martino 2009). We applied grid search. Finally, the marginal posterior mean or median in Eq. (42) is usually used to obtain estimates for the spatiotemporal model parameters.

$$\underset{\text{Observed y}}{\begin{pmatrix} y_{11} \\ \vdots \\ y_{n1} \\ y_{12} \\ \vdots \\ y_{n2} \\ \vdots \\ y_{1T} \\ \vdots \\ y_{nT} \\ y_{1T+1} = NA \\ \vdots \\ y_{nT+1} = NA \\ \vdots \\ y_{1T+h} = NA \\ \vdots \\ y_{nT+h} = NA \end{pmatrix}} = \underset{\text{id } \omega}{\begin{pmatrix} 1 \\ \vdots \\ n \\ 1 \\ \vdots \\ n \\ 1 \\ \vdots \\ n \\ 1 \\ \vdots \\ n \\ 1 \\ \vdots \\ n \end{pmatrix}} + \underset{\text{id } \upsilon}{\begin{pmatrix} 1 \\ \vdots \\ n \\ 1 \\ \vdots \\ n \\ 1 \\ \vdots \\ n \\ 1 \\ \vdots \\ n \\ 1 \\ \vdots \\ n \end{pmatrix}} + \underset{\text{id } \phi}{\begin{pmatrix} 1 \\ \vdots \\ 1 \\ 2 \\ \vdots \\ 2 \\ \vdots \\ T \\ \vdots \\ T \\ T+1 \\ \vdots \\ T+1 \\ \vdots \\ T+h \\ \vdots \\ T+h \end{pmatrix}} + \underset{\text{id } \gamma}{\begin{pmatrix} 1 \\ \vdots \\ 1 \\ 2 \\ \vdots \\ 2 \\ \vdots \\ T \\ \vdots \\ T \\ T+1 \\ \vdots \\ T+1 \\ \vdots \\ T+h \\ \vdots \\ T+h \end{pmatrix}} + \underset{\text{id } \delta}{\begin{pmatrix} 1 \\ \vdots \\ n \\ n+1 \\ \vdots \\ n+n \\ \vdots \\ n(T-1) \\ \vdots \\ nT \\ nT+1 \\ \vdots \\ n(T+1) \\ \vdots \\ n(T+h-1)+1 \\ \vdots \\ n(T+h) \end{pmatrix}}$$

2.5 Model selection criteria

General model Eq. (8) contains several sub-models. Several selection criteria are available to select the best model. The most common Bayesian criterion for predictive performance is the conditional predictive ordinate (CPO). This is defined as follows. Given a set of spatiotemporal observations $\mathbf{y} = \left(y_{11}, ..., y_{nT}\right)'$, the CPO_{it} for each observation is (Rodrigues and Assunção 2012):

$$\text{CPO}_{it} = p\left(\hat{y}_{it} = y_{it}|\mathbf{y}_{-it}\right) = \int p\left(\hat{y}_{it} = y_{it}|\mathbf{\Phi}\right)p\left(\mathbf{\Phi}|\mathbf{y}_{-it}\right)d\mathbf{\Phi}, \tag{44}$$

with $\hat{y}_{it}$ denoting the predicted number of cases in region i for time t, $\mathbf{\Phi}$ the all-parameter vector, $\mathbf{y}_{-it}$ the data vector without the itth observation, and $p\left(\mathbf{\Phi}|\mathbf{y}_{-it}\right)$ the posterior distribution of $\mathbf{\Phi}$. predicted without y_{it}. CPO_{it} is thus the cross-validated predictive probability mass at the observation y_{it}. A small CPO_{it} indicates that the itth observation is unlikely under the postulated model.

A related measure is CPO-failure$_{it}$. It is defined as (Blangiardo and Cameletti 2015):

$$\text{CPO} - \text{failure}_{it} = \sum_{j=1}^{J} \text{failure}_{it,j}\left(\boldsymbol{\tau}^{(j)}|\mathbf{y}\right)\Delta_j \quad j = 1, \ldots J, \tag{45}$$

where $\text{failure}_{it,j}p\left(\boldsymbol{\tau}^{(j)}|\mathbf{y}\right)$ indicates the misfit of $p\left(\boldsymbol{\tau}^{(j)}|\mathbf{y}\right)$ for observation y_{it} at the jth grid, and Δ_j is the corresponding weight. CPO-failure$_{it}$ is actually the expected failure of observation y_{it} over the posterior distribution for the hyper-parameter $\boldsymbol{\tau}$. It takes the value of one for a misfit and zero otherwise (Schrödle and Held 2011). The

total value of CPO-failure for all observations ranges from zero to nT. Based on the total value of the CPO-failure criterion, the model with the lowest value is selected.

Another CPO-based measure is the marginal predictive likelihood (MPL), defined as (Urtasun 2017):

$$\text{MPL} = \sum_{i=1}^{n} \sum_{t=1}^{T} \log\left(\text{CPO}_{it}\right). \tag{46}$$

The larger the MPL, the better the prediction.

The probability integral transform (PIT) is the value of the predicted cumulative distribution function at observation y_{it} (Urtasun 2017):

$$\text{PIT}_{it} = Pr\left(\hat{y}_{it} \leq y_{it}|\mathbf{y}_{-it}\right) = \int p\left(\hat{y}_{it} \leq y_{it}|\mathbf{\Phi}\right) p\left(\mathbf{\Phi}|\mathbf{y}_{-it}\right) \mathrm{d}\mathbf{\Phi}. \tag{47}$$

The PIT histogram indicates the model fit across all observations. The closer the PIT histogram is to the uniform distribution histogram, the better the fit (Hicketier 2015).

A measure which considers both fit and complexity is the deviance information criterion (DIC) (Spiegelhalter et al. 2002). It measures the performance of the model with parameters fixed to the posterior mean $\hat{\mathbf{\Phi}} = E\left[\mathbf{\Phi}|\mathbf{y}\right]$ and is defined as (Gelman et al. 2014):

$$\text{DIC} = D\left(\hat{\mathbf{\Phi}}\right) + 2\text{p}_{\text{DIC}}, \tag{48}$$

where $D\left(\hat{\mathbf{\Phi}}\right)$ is the model's deviance, i.e. $D\left(\hat{\mathbf{\Phi}}\right) = -2\log p\left(\mathbf{y}|\hat{\mathbf{\Phi}}\right)$, and p_{DIC} is the effective number of parameters indicating the model's complexity:

$$\text{p}_{\text{DIC}} = 2\left(\log p\left(\mathbf{y}|\hat{\mathbf{\Phi}}\right) - \text{E}_{\text{post}}\left[\log p(\mathbf{y}|\mathbf{\Phi})\right]\right). \tag{49}$$

The expectation $E_{\text{post}}\left[\log p(\mathbf{y}|\mathbf{\Phi})\right]$ is an average of $\mathbf{\Phi}$ over its posterior distribution. It can be calculated by simulation as $\frac{1}{S}\sum_{s=1}^{S} \log p(\mathbf{y}|\mathbf{\Phi}^s) s = 1, \ldots, S$ with $\mathbf{\Phi}^s$ denoting the sth draw from $p_{\text{post}}(\mathbf{\Phi}|\mathbf{y})$. Equation (49) can be negative. To overcome this problem, Gelman et al. (2014) proposed the following alternative definition:

$$\text{p}_{\text{DIC}} = 2\text{Var}_{\text{post}}\left[\log p(\mathbf{y}|\mathbf{\Phi})\right], \tag{50}$$

where $\text{Var}_{\text{post}}\left[\log p(\mathbf{y}|\mathbf{\Phi})\right] = \sum_{i=1}^{n}\sum_{t=1}^{T}\left\{E_{\text{post}}\left[\left(\log p\left(y_{it}|\mathbf{\Phi}\right)\right)^2\right] - E_{\text{post}}\left[\log p\left(y_{it}|\mathbf{\Phi}\right)\right]^2\right\}$. A lower DIC indicates a better fit.

An alternative to Eq. (48) is the Watanabe–Akaike information criterion (WAIC) (Watanabe 2010). The WAIC reads (Gelman et al. 2014; Utazi et al. 2018):

$$\text{WAIC} = -2\left(\text{D}_{\text{WAIC}} - \text{p}_{\text{WAIC}}\right), \tag{51}$$

where D_{WAIC} measures the fit of the model defined as $\text{D}_{\text{WAIC}} = \sum_{i=1}^{n}\sum_{t=1}^{T} \log E_{\mathbf{\Phi}|\mathbf{y}}\left[p\left(y_{it}|\mathbf{\Phi}\right)\right]$ and p_{WAIC} denotes the effective number of parameters defined as

$p_{\text{WAIC}} = \sum_{i=1}^{n}\sum_{t=1}^{T} \text{Var}_{\text{posterior}}\left[\log p\left(y_{it}|\boldsymbol{\Phi}\right)\right]$. The lower the WAIC, the better the fit. This is preferable to the DIC because it computes the effective number of parameters for the variance for each data point separately, and then takes the sum (Gelman et al. 2014).

Another model selection criterion is the Bayes factor (BF). It is defined as follows. Assume two models, M_1 and M_2. The distribution of the data $\mathbf{y}$ and the prior probability of model M_q can be written as $p(\mathbf{y}|M_q)$ and $p\left(M_q\right)$ $q = 1, 2$. The posterior probabilities of $M_q (q = 1, 2)$ are provided by:

$$p\left(M_q|\mathbf{y}\right) = \frac{p\left(\mathbf{y}|M_q\right)p\left(M_q\right)}{p(\mathbf{y})}. \tag{52}$$

The posterior odds in favour of model M_1 over the alternative M_2 are:

$$\frac{p\left(M_1|\mathbf{y}\right)}{p\left(M_2|\mathbf{y}\right)} = \left(\frac{p\left(\mathbf{y}|M_1\right)}{p\left(\mathbf{y}|M_2\right)}\right)\left(\frac{p\left(M_1\right)}{p\left(M_2\right)}\right). \tag{53}$$

The ratio of the marginal likelihoods $\left(p\left(\mathbf{y}|M_1\right)/p\left(\mathbf{y}|M_2\right)\right)$ in Eq. (53) is known as the Bayes factor (BF), where $p(\mathbf{y}|M_q)$ is the evidence for model M_q. When $\text{BF}_{12} > 1$, the data favour M_1 over M_2, and when $\text{BF}_{12} < 1$, the data favour M_2 (Chipman et al. 2001).

It can be complicated to implement the BF in the case of complex random effects models (i.e. models with more than one random effect). The solution is an approximation of the BF based on CPO values (Gelfand and Dey 1994), which is called the pseudo-Bayes factor (PBF). For models M_1 and M_2, the PBF reads (Gelfand 1996):

$$\begin{aligned}
\text{PBF} &= \prod_{i=1}^{n}\prod_{t=1}^{T}\left\{\frac{\text{CPO}_{it}\left(M_1\right)}{\text{CPO}_{it}\left(M_2\right)}\right\} = \prod_{i=1}^{n}\prod_{t=1}^{T}\left\{\frac{p\left(\hat{y}_{it}|\mathbf{y}_{-it}, M_1\right)}{p\left(\hat{y}_{it}|\mathbf{y}_{-it}, M_2\right)}\right\} \\
&= \exp\left(\log\left(\prod_{i=1}^{n}\prod_{t=1}^{T}\frac{p\left(\hat{y}_{it}|\mathbf{y}_{-it}, M_1\right)}{p\left(\hat{y}_{it}|\mathbf{y}_{-it}, M_2\right)}\right)\right) \\
&= \exp\left(\sum_{i=1}^{n}\sum_{t=1}^{T}\log p\left(\hat{y}_{it}|\mathbf{y}_{-it}, M_1\right) - \sum_{i=1}^{n}\sum_{t=1}^{T}\log p\left(\hat{y}_{it}|\mathbf{y}_{-it}, M_2\right)\right) \\
&= \exp\left(\text{MPL}\left(M_1\right) - \text{MPL}\left(M_2\right)\right)
\end{aligned} \tag{54}$$

A $\text{PBF} < 1$ indicates that the data favour M_2 over M_1.

Other measures of predictability are the mean absolute error (MAE), the root-mean-square error (RMSE) and the adjusted pseudo-coefficient of determination $\left(\widetilde{R}^2\right)$. They are defined as (Mohebbi et al. 2014; Urtasun 2017):

$$\text{MAE} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left|\hat{y}_{it} - y_{it}\right| \tag{55}$$

$$\text{RMSE} = \sqrt{\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\hat{y}_{it} - y_{it}\right)^2} \tag{56}$$

$$\widetilde{R}^2 = 1 - \left(\frac{nT-1}{nT-p_D}\right)\left(1-\text{R}^2\right) \text{ with } \quad R^2 = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\hat{y}_{it} - \bar{y}\right)^2}{\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it} - \bar{y}\right)^2} \tag{57}$$

where $\hat{y}_{it}$ denotes the predicted number of cases at region i at time t, and $\bar{y}$ is the mean of the observed values. Ceteris paribus, the lower the MAE and RMSE and the higher the $\widetilde{R}^2$, the better the fit.

Other tools for model evaluation check whether the observed residuals are white noise (residual analysis). One such tool is the autocorrelation function (ACF) plot, defined as:

$$\rho_i(l) = \frac{\frac{1}{T-l}\sum_{t=l+1}^{T}\left(r_{it} - \bar{r}_i\right)\left(r_{it-l} - \bar{r}_i\right)}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(r_{it} - \bar{r}_i\right)}\sqrt{\frac{1}{T-l}\left(r_{it-l} - \bar{r}_i\right)}} \quad i = 1, \ldots, n, \; t = 1, \ldots, T, \text{ and } \; l = 0, 1, 2, \ldots \tag{58}$$

where $\bar{r}_i$ is the sample mean of the residual for region i over T periods and $r_{it} = \left(y_{it} - \hat{y}_{it}\right)$. The ACF can be calculated under the (strong) null hypothesis that the time series is white noise (iid) (H_0: iid), or under the (weaker) hypothesis that it is generalized autoregressive conditional heteroscedasticity (GARCH) (H_0: GARCH). Different tests are needed for these two assumptions. For the hypothesis that the time series is iid, a portmanteau test such as the Ljung and Box test with test statistics (Francq and Zakoian 2010) can be applied:

$$Q_l^{\text{LB}} = T(T+2)\sum_{j=1}^{l}\hat{\rho}(j)/(T-j). \tag{59}$$

The strong white noise hypothesis is rejected if Q_l^{LB} is greater than the $(1-\alpha)$ quantile of χ_l^2. For the GARCH assumption, the corrected Ljung and Box portmanteau test is applied. The test statistic is (Francq and Zakoian 2010):

$$Q_l = T\hat{\boldsymbol{\rho}}(j)'\,\hat{\Sigma}_{\hat{\boldsymbol{\rho}}(j)}^{-1}\hat{\boldsymbol{\rho}}(j), \tag{60}$$

which has an asymptotic χ_l^2 distribution, where $\widehat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\rho}}(j)}$ is the asymptotic covariance matrix of $\hat{\boldsymbol{\rho}}(j)$, which can be obtained using nonparametric estimation. The hypothesis that the data are generated by a GARCH process is rejected if Q_l is greater than the $(1-\alpha)$ quantile of χ_l^2.

The spatiotemporal autocorrelation of the residuals can be evaluated using an extension of Moran's I, called Moran's spatiotemporal autocorrelation statistic (MoranST). It is defined as (Anderson and Ryan 2017):

$$\text{MoranST} = \frac{nT\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j=1}^{n}\sum_{s=1}^{T}\widetilde{w}_{(it,js)}\left(r_{it} - \bar{r}\right)\left(r_{js} - \bar{r}\right)}{\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j=1}^{n}\sum_{s=1}^{T}\widetilde{w}_{(it,js)}\sum_{i=1}^{n}\sum_{t}^{T}\left(r_{it} - \bar{r}\right)^2} \tag{61}$$

where $\bar{r}$ is the mean of the observed residuals r_{it} over T periods and n spatial units, and $\widetilde{w}_{(it,js)}$ is the weight accounting for the spatiotemporal autocorrelation between r_{it} and r_{js}, defined as:

$$\tilde{w}_{(it,js)} = \begin{cases} w_{ij} & \text{if} \quad t = s \\ 1 & \text{if} \quad i = j \text{ and } |t - s| = 1 \\ 0 & \text{otherwise} \end{cases}$$

where w_{ij} is one if regions i and j are neighbours, and zero otherwise. A MoranST which is close to one indicates a strong positive spatiotemporal autocorrelation of the spatiotemporal residuals. A value close to zero indicates white noise.

2.6 Mapping

As discussed in the introduction, we use continuous or isopleth maps for the visual representation of the spatial dengue distribution. For this purpose, successive interpolation is applied using predicted relative risk values, which are obtained from the underlying Bayesian spatiotemporal model. Note that the predicted values are denoted as observed in the mapping process (Tiwari 2013). An observed relative risk value is typically placed at the centroid of a spatial unit with coordinates (x, y), denoting grid point (x, y). Observed relative risk values at selected grid points are then used to generate new values at selected grid points via interpolation. The observed values along with the interpolated values are then converted into a smooth, continuous surface. There are two common interpolation methods, i.e. inverse distance weighting (IDW) and Kriging. We use IDW in this paper, which is relatively easy to apply and efficient, and provides good results in practice (Setianto and Triandini 2013; Tiwari 2013).

The basic assumption of IDW is that the observed value (known value) closest to the interpolation grid point has more influence on the interpolated value (unknown value) than those farther away. Following Revesz (2003), the general formulation of IDW interpolation for the value at grid point (x, y), at time t, is:

$$\widetilde{\theta}(x, y) = \sum_{l=1}^{L} \varpi_l \hat{\theta}_l; \varpi_l = \frac{d_l^{-p}}{\sum_{k=1}^{L} d_k^{-p}} \quad \text{for } l = 1, \ldots, L, \text{ and for all } t, \tag{62}$$

where $\widetilde{\theta}(x, y)$ is the interpolated value for grid point (x, y), ϖ_l is the weight assigned to the observed value $\widehat{\theta}_l$ at grid (x_l, y_l), L is the number of grid points used for interpolation of the value at (x, y), where L may be smaller than or equal to the total number of observed grid points (n), d_l is the distance between the observed grid (x_l, y_l) and the interpolation grid (x, y), and $p = 0, 1, 2, \ldots$ is the power of the weighting of $\widehat{\theta}_l$ with respect to $\widetilde{\theta}(x, y)$. To obtain smooth isopleth maps, they are usually divided into $10,000 \times 10,000$ grid cells. The optimal power p is determined by minimizing the root-mean-square prediction error (RMSPE):

$$\text{RMSPE} = \sqrt{\frac{1}{L}\sum_{l=1}^{L}\left(\widehat{\theta}_l - \widetilde{\theta}(x,y)\right)^p}. \tag{63}$$

The most common power is $p = 2$ which yields the inverse distance squared weighted interpolation (Lloyd 2010).

3 Application: dengue disease in the city of Bandung

The application is based on 30 districts in the city of Bandung, the provincial capital of West Java, Indonesia, over the period 2009–2016. The city is seriously affected by dengue disease every year. According to Bandung (2010–2017), 35,496 dengue cases were registered for the period 2009–2016.

3.1 Model selection

As a first step in the selection procedure, we compared the appropriateness of the Poisson and the negative binomial (NB) as distributions for the number of dengue cases in Bandung. We estimated the pure random effects model Eq. (8) based on the Poisson distribution and tested for over-dispersion. Due to the lack of strong prior knowledge, we used Gaussian distribution as a hyper-prior for the overall relative risk, i.e. $\alpha \sim \mathcal{N}\left(0, 10^5\right)$. Following Schrödle and Held (2011), we used IG(1, 0.01) as the hyper-prior for the precision parameters of the structured and unstructured spatial random effects, and for the seasonal and unstructured temporal random effects. We used IG(1, 0.00005) as the precision parameter of the RW1 and RW2 temporal trends. The hyper-prior for the precision parameter of the interaction effects followed from the above specifications, as described in Sect. 2. The over-dispersion parameter estimate was 0.0238 with a standard error estimate of 0.0038 and 95% credible interval (0.0177; 0.0327).[7] The interval did not contain zero. We therefore chose the NB model for further analysis.[8]

We then estimated model Eq. (8) based on the NB distribution for y_{it} with four different types of interactions, and used the model selection criteria in Sect. 2.5 to choose between RW1 and RW2. The outcomes for RW1 were generally better than for RW2. We continued the analysis to identify the best type of interaction. The results are shown in Table 1. They show that interaction type IV (the temporal dependency structure for each region depends on the temporal structure of

[7] The models were estimated using the R-INLA package (http://www.r-inla.org/).

[8] We also considered penalized complexity, half Cauchy and uniform distribution as hyper-priors for the various precision parameters in the decision over the Poisson and NB distributions and to choose the interaction type. Moreover, we performed a sensitivity analysis of the hyper-parameter values of the IG hyper-prior. The results consistently supported the choice of the IG as hyper-prior and the selected values of its hyper-parameters. The calculations are not presented here. They are available from the first author on request.

Table 1 Selection of interaction type

Hyper-prior	Type interaction	CPO Failure	DIC	WAIC	MPL	PBF	MAE	RMSE	$\tilde{R}^2$
Inverse gamma	I	0	17,490.21	17,500.95	−8753.33	0	4.37	6.27	0.57
	II	0	16,552.63	16,498.89	−8264.53	1.5×10^{-123}	3.15	4.44	0.69
	III	0	16,247.67	16,217.70	−8299.80	7×10^{-129}	2.19	2.93	0.66
	IV	0	15,866.68	15,910.80	−8004.75	–	2.58	3.53	0.77

The headers of the columns denote: *CPO* conditional predictive ordinate, *DIC* deviance information criterion, *MPL* marginal predictive likelihood, *PBF* pseudo-Bayes factor, *MA* mean absolute error, *RMSE* root-mean-square error, pseudo-determination coefficient ($\tilde{R}^2$)

Table 2 Summary statistics of the posterior mean of the standard deviations (SD) of the random effects and their critical ratios

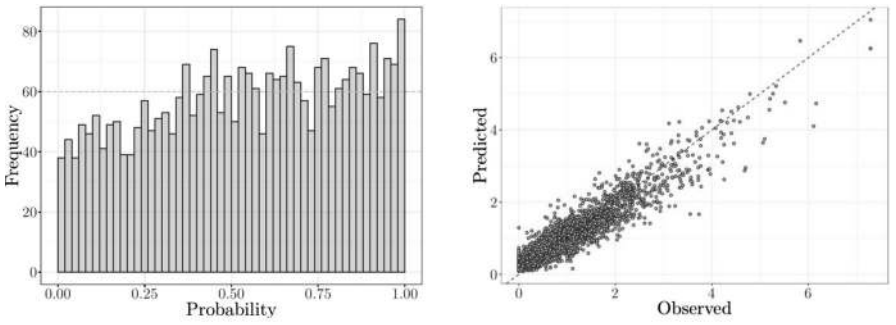
Component	Estimated mean	Estimated standard error	Q (0.025)	Q (0.975)	Critical ratios
SD of spatially unstructured (σ_{υ}^2)	0.147	0.094	0.054	0.401	1.563
SD of iCAR (σ_{ω}^2)	0.148	0.090	0.054	0.392	1.640
SD of temporally unstructured (σ_{γ}^2)	0.097	0.030	0.050	0.165	3.288
SD of RW1 (σ_{φ}^2)	0.560	0.086	0.419	0.754	6.540
SD of seasonal (σ_{κ}^2)	0.059	0.013	0.036	0.089	4.379
SD of interaction (σ_{δ}^2)	0.174	0.009	0.157	0.191	20.075

Critical ratios = estimated mean/estimated standard error, intrinsic conditional autoregressive (iCAR), random walk of order one (RW1)

neighbouring regions) has minimum DIC and WAIC values and the largest MPL. It also outperforms models II and III in terms of PBF. We therefore continued the analysis using model Eq. (8) and interaction type IV, and evaluated the structured and unstructured components using their critical ratios. The results are shown in Table 2. The spatially unstructured and the iCAR component have low critical ratios, indicating that their effects do not significantly differ between districts. Put differently, the relative risk is almost constant across space.

We re-estimated the model without the spatially structured and unstructured components but with type IV interaction (denoted as the temporal model in the subsequent study).[9] Support for the temporal model is presented in Fig. 1a which shows

[9] We re-estimated the reduced model in question, using the inverse Gamma, penalized complexity, half Cauchy and uniform distribution as hyper-priors for the various precision parameters. We found the half Cauchy to be the best hyper-prior for the reduced model, while the second best hyper-prior is inverse Gamma, given the model selection criterion. However, the precision parameter of the seasonal component was too high. Finally, we estimated the model in question by combining hyper-priors between the



(a) Probability integral transform **(b)** Observed θ_{it} versus predicted $\hat{\theta}_{it}$

Fig. 1 **a** The probability integral transform and **b** observed versus predicted relative risk

that the probability integral transform (PIT) histogram does not substantially differ from the uniform distribution histogram. Figure 1b shows that the observed $\boldsymbol{\theta}$ and predicted $\hat{\boldsymbol{\theta}}$ are strongly correlated, as indicated by the clustering of points along the diagonal.

We next estimated the autocorrelation functions (ACF) of the residuals for each district (using the *sarima* R-package). Figure 2a depicts the ACF for six selected districts in the city of Bandung.[10] Figure 2b displays the p values from the test that the residuals are white noise for each district. The ACF plots in Fig. 2a show that the model has adequately captured the patterns in the data, although in a few districts (e.g. Bandung Kulon) there is a slight autocorrelation in the residuals (as indicated by the significant spike in the ACF plot). Some improvements are therefore possible, even though they are unlikely to make a significant difference to the resulting forecasts. Figure 2b presents the p values for white noise tests using the corrected Ljung–Box test. This p value is larger than 0.01 for the majority of the districts. A similar result was obtained for the spatiotemporal Moran index. The value of the spatiotemporal Moran's index is relatively small, -0.111, indicating that there is virtually no spatiotemporal autocorrelation present in the residuals.

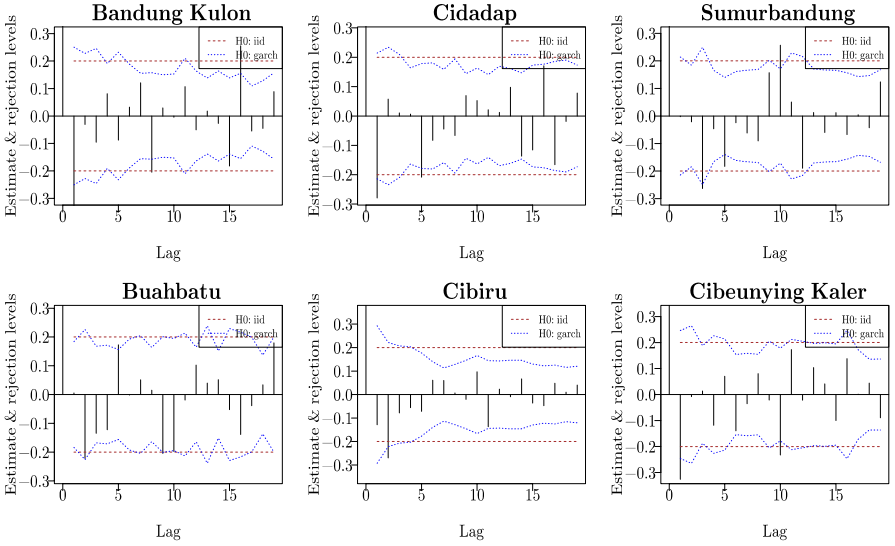
3.2 Final estimation results

Based on the temporal model identified in Sect. 3.1, we present estimated posterior summary statistics for 2009–2016 in Table 3. The table shows that the posterior mean fixed effect, i.e. the posterior mean log-relative risk, is -0.219, with posterior standard deviation equal to 0.029 and 95% credible interval (-0.276; -0.163). This means that the posterior mean relative risk of dengue disease is

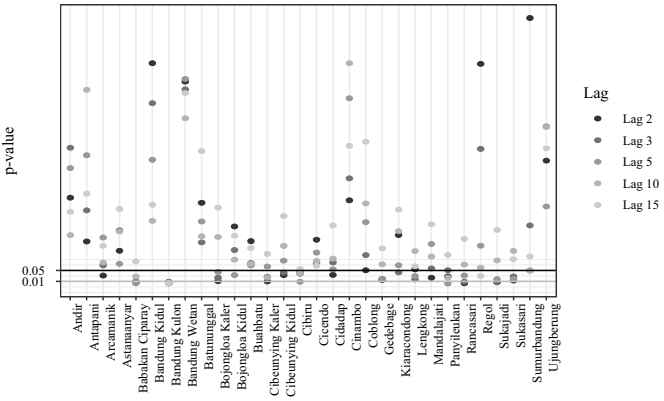
Footnote 9 (continued)

half Cauchy prior and inverse Gamma. Here, for the seasonal component we used inverse Gamma and for the other random components, the half Cauchy hyper-prior was applied. The details are not presented here. They are available from the first author on request.

[10] The results for the other districts are available from the first author on request.



(a) Autocorrelation function



(b) white noise test p-values

Fig. 2 **a** Autocorrelation function for six selected districts and **b** p values for the hypothesis of white noise for all 30 Bandung city districts

$\exp(-0.219) = 0.803$ for all districts and times. Put differently, if there were no spatial, temporal or spatiotemporal interaction effects, the relative risk in all 30 districts and periods would be equal to 0.803. However, the relative risk does have spatiotemporal characteristics, as shown in Table 4. The table shows the summary statistics for the posterior standard deviation for each random effect including the mean, its standard deviation, quantiles, critical ratio and the fraction of the overall variance (FV): $\text{FV}_h = \sigma_h^2 / \sum_{h=1}^{H} \sigma_h^2$. All the random effect components have a critical ratio

Table 3 Summary statistics of the posterior mean of the fixed effect (α)

Parameter	Mean	SD	0.025 quant	0.50 quant	0.975 quant
Intercept	−0.219	0.029	−0.276	−0.219	−0.163

Table 4 Summary statistics of the posterior mean of standard deviation (SD) of the random effects

Component	Mean	SD	0.025 quant	0.50 quant	0.975 quant	Critical ratio	Fraction of the variance (%)
SD of RW1 (σ_{φ})	0.6400	0.0947	0.4892	0.6270	0.8584	6.7594	85.7013
SD of seasonal (σ_{κ})	0.0574	0.0134	0.0373	0.0552	0.0894	4.2898	0.6813
SD of temporally unstructured (σ_{γ})	0.0894	0.0317	0.0359	0.0872	0.1566	2.8243	1.6332
SD of interaction (σ_{δ})	0.1709	0.0118	0.1459	0.1720	0.1910	14.5478	5.8730

greater than two, which indicates that all the components are relevant to explaining the variation in relative risk of dengue. The FV ranks the components and indicates that the temporal trend (RW1) (83%) is the most important component. The interaction effect is next. It indicates that the trend for each district depends on the temporal structure of neighbouring districts. The relatively low FVs for the other effects imply that they are less important. Specifically, the low FV of the seasonal effect indicates that only a small part of the variability of relative risk of dengue is explained by the seasonal component.

The above conclusions are illustrated in Fig. 3, which illustrates the estimated marginal posterior density of the standard deviation of each random effect. It shows that the estimated posterior distribution of the standard deviation of the interaction component (σ_{δ}) is highly concentrated. This is because of we have 3240 observations which provide information about this hyper-parameter, in contrast to only 108 observations for σ_{φ}, σ_{κ} and σ_{δ}. The large standard deviation for temporal trend indicates that there is substantial variation in the relative risk over the months.

Figure 4a–c presents the estimated posterior mean temporal effects and their 95% credible intervals. The estimated posterior mean of the temporally unstructured effect is virtually zero (see Fig. 4c). Moreover, its credible interval is narrow over the observation period. RW1 exhibits a strong nonlinear pattern which is different from zero (see Fig. 4a) This is in line with Table 4, indicating the temporal trend as being highly important in explaining the variability of the relative risk over time. The relative risk increased between 2010 and 2013 and then started to decline. The decline is probably related to increasing public awareness and increased preventive action (Tribunjabar 2016). Figure 4b shows that although on average the seasonal effect is close to zero, there is a strong seasonal pattern. This is further illustrated in Fig. 5 for 2016, which shows that the high-risk season is between January and May, after the peak of the rainy season, which is usually in February. The

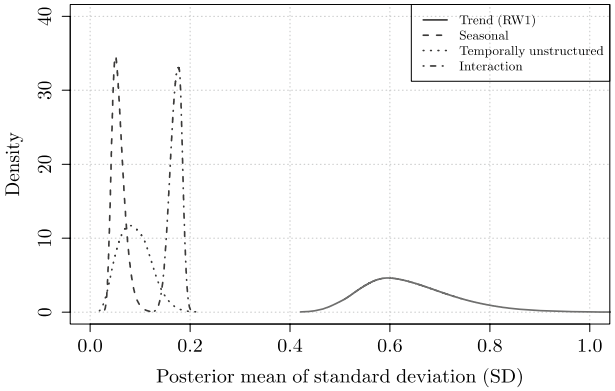


Fig. 3 Estimated marginal posterior density of the standard deviation of the temporal trend, seasonal effect **κ**, temporally unstructured effect **γ** and the interaction effect **δ**

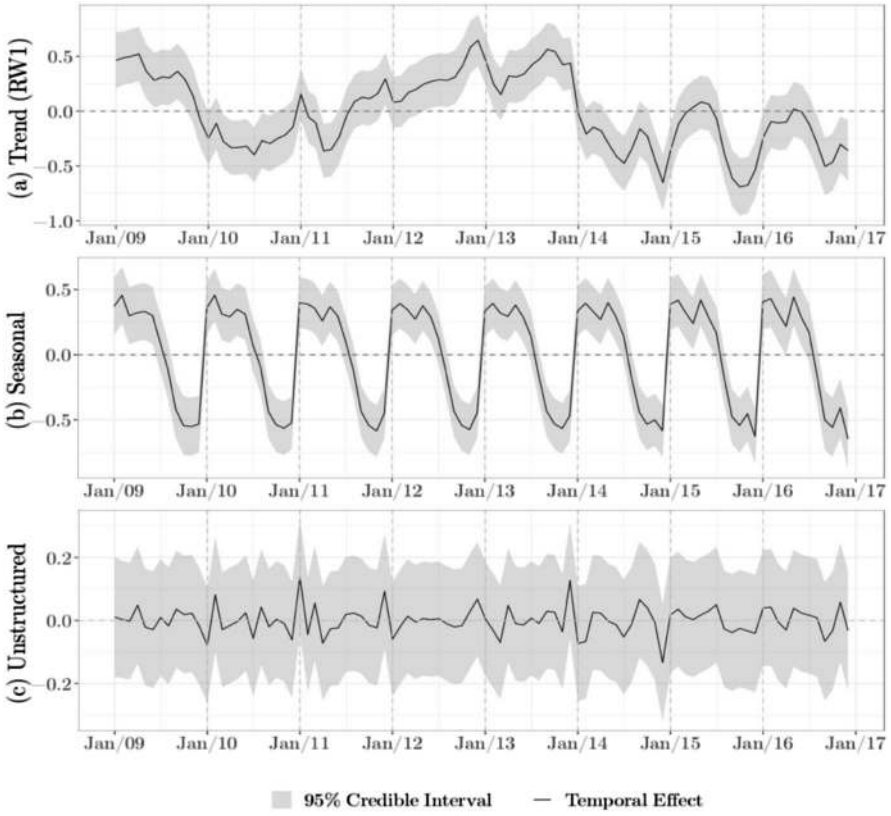


Fig. 4 Estimated posterior mean temporal effects: **a** trend RW1 **φ**, **b** seasonal effect **κ** and **c** unstructured effect **γ**

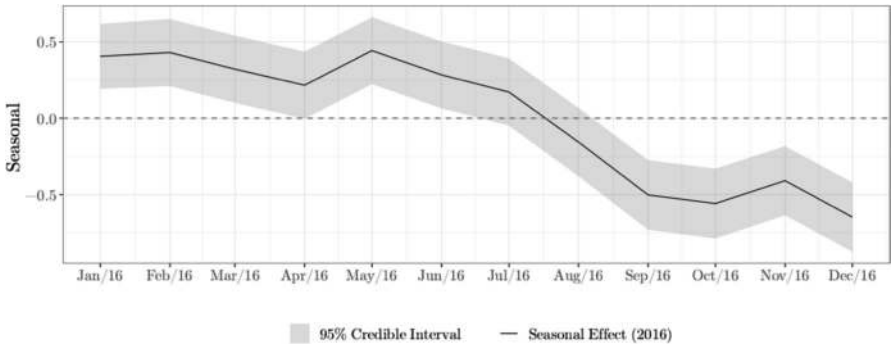


Fig. 5 Estimated posterior mean seasonal effect $\boldsymbol{\kappa}$ (2016)

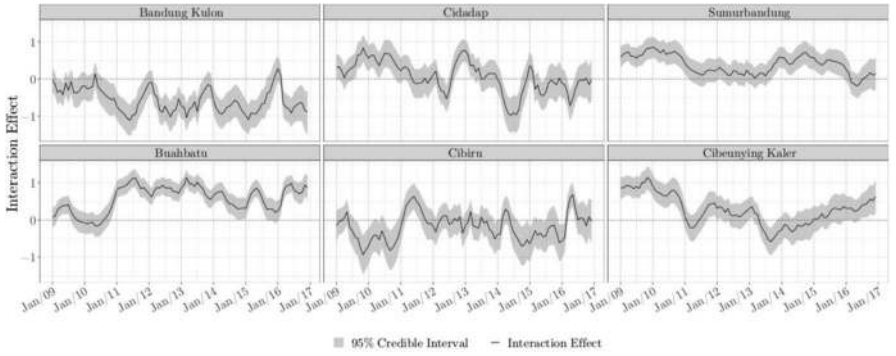


Fig. 6 Spatiotemporal interaction effects $\boldsymbol{\delta}$ for six selected districts

Aedes-spp. mosquitoes breed in the rainy season: their larvae become adults after three to four weeks and are then ready to transmit dengue disease (Kompas 2016). The low-risk season is between September and December. The seasonal pattern is very similar for the other years.[11]

Figure 6 depicts the spatiotemporal pattern for six selected districts in the city of Bandung[12]: Cidadap in the north, Bandung Kulon in the west, Buahbatu in the south, Cibiru in the east and Sumur Bandung and Cibeunying Kaler in central Bandung. The graphs show that there are substantial differences among the districts. Bandung Kulon displays a relatively constant trend over the years with no substantial relative changes from 2009 to 2016. Moreover, the temporal effect in Bandung Kulon is below zero most of the time, indicating that its risk is smaller than those in the other districts, notably in Sumur Bandung and Buah Batu. In the latter district, there is a strong increase in 2010. Moreover, the temporal effect stays above zero. In Cidadap and Cibiru, it strongly fluctuates around zero, which indicates that the relative risk

[11] The graphs can be obtained from the first author on request.

[12] The results for the other districts are available from the first author on request.

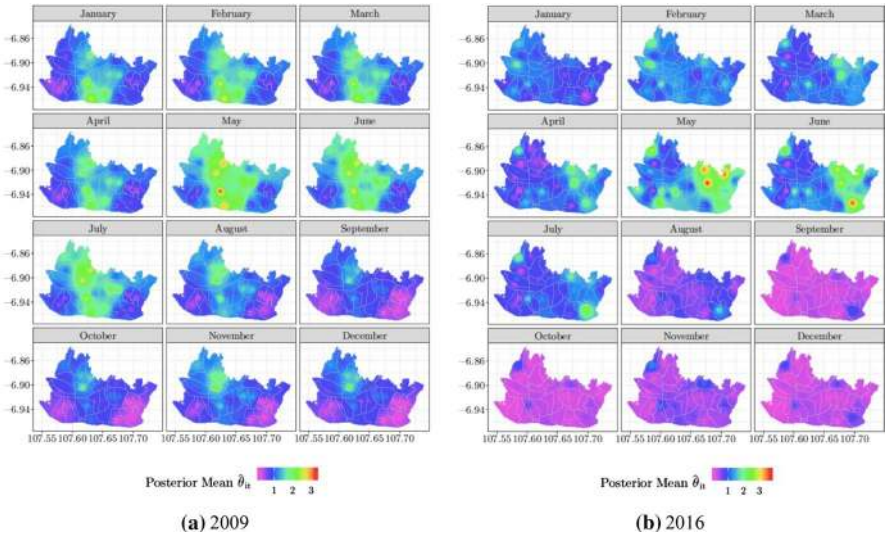


Fig. 7 Estimated relative risk for January to December 2009 and 2016

in both districts varies greatly over the years. The trend for Sumur Bandung is downward. However, it is generally greater than zero, indicating that the risk in this district is relatively high. The temporal trend for Cibeunying Kaler is strongly nonlinear. It strongly decreased from 2009 and strongly increased from around May 2013 onward. The above results are in line with Table 4, where the spatiotemporal interaction component was found to have the second largest variance and thus to have a substantial impact on spatiotemporal variability. Figure 7 summarizes the estimates of the relative risk for 2009 and 2016.[13] It shows that in 2009 the high-risk districts were mainly in the north, while in 2016 they had moved southwards.

3.3 Forecasts

Figure 8 displays plots of observed versus estimated relative risk for January 2009 to December 2016, and the forecast for 2017, for six selected districts. For the period January to December 2017, we found that Cidadap, Bandung Kulon, Sumur Bandung and Cibeunying Kaler displayed an upward trend, while Buahbatu and Cibiru a downward one.

Figure 9a presents the monthly forecasts for 2017. It shows that the predicted relative risk in the eastern and western districts of Bandung was relatively low all year round, and relatively high in the northern, central and southern districts. Furthermore, from January to August, the relative risk in the southern districts was forecasted to be twice as high as average ($\alpha = 0.803$). However, between September and December, a decrease was foreseen in these districts and an increase in the northern

[13] For the other years, the results are available from the first author upon request.

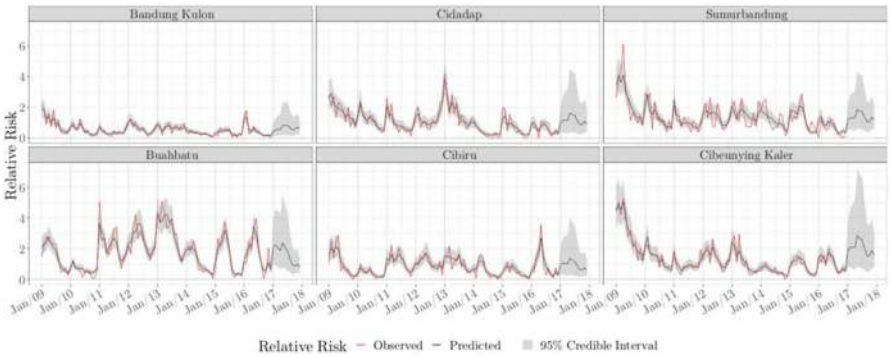


Fig. 8 Observed versus estimated relative risk, January 2009 to December 2016 and forecast for 2017, for six selected districts

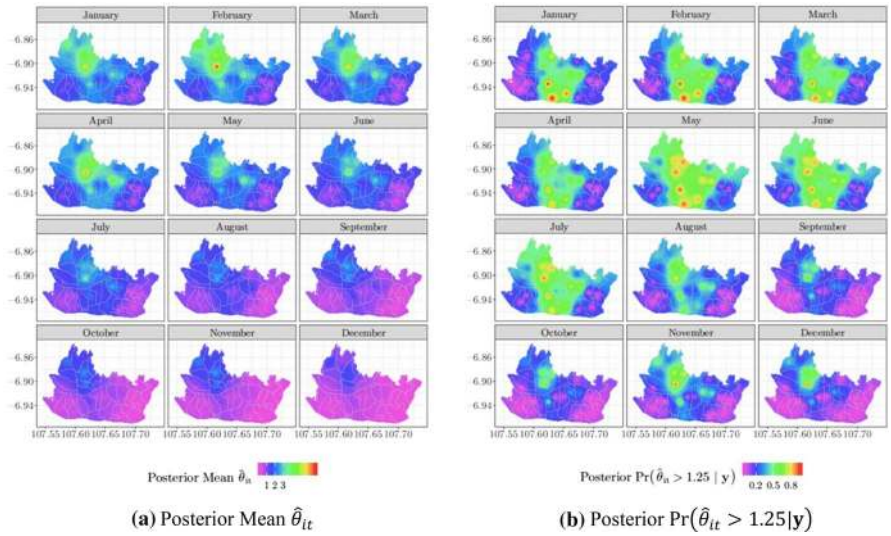


Fig. 9 Monthly forecasts of relative risk for January to December 2017. **a** Posterior mean $\hat{\theta}_{it}$ and **b** posterior $Pr(\hat{\theta}_{it} > 1.25|\mathbf{y})$

districts. For the north, a higher than average relative risk for each month was forecasted. Figure 9b illustrates the posterior probability forecast of relative risk greater than 1.25 for each district from January to December 2017. The spatial variation in monthly forecasts is due to geographical, weather and socio-economic conditions. High precipitation, moderate temperatures and high atmospheric humidity are favourable ecological conditions for breeding, birth, growth, mortality and activity of the vector (Rueda et al. 1990).

Several districts located in north Bandung displayed relatively high dengue risk almost every month. This is because northern Bandung is a mountainous area

approximately 800 m above sea level, which is densely covered in forests. It also has high precipitation, moderate temperatures and high atmospheric humidity all year round (Metro-Trend, 2019). The central districts also have high precipitation but lack dense vegetation, meaning that during the dry season—September to January—the weather conditions are less favourable for dengue because of the relatively low humidity. Socio-economic conditions also contribute to the high relative risk of central Bandung. It is the centre of business and government, has a high population density and high mobility, all of which facilitate the transmission of dengue by increasing the contact potential between infected mosquitoes and human hosts (Araujo et al. 2015). Another reason for the relatively high risk in central Bandung is related to the number of incidents reported. The population of central Bandung has a high level of education, high employment rate and high economic status. These conditions raise people's awareness of health risks and help them to find their way to healthcare. The registration rate in the area is higher than in districts with a less well-educated population (Wijayanti et al. 2016).

In districts adjacent to central Bandung, environmental factors contribute to the high relative risk. Taman Sari, for instance, is a slum area with poor access to safe water, high incidence of dengue and high population density (Sari 2018). Southern Bandung was forecasted to have a relatively high risk from January to September because of its high-density housing and poor water quality. These areas also border slum areas. The eastern and western districts are forecasted to have relative low risks. These districts have low population density and do not have much business or tourism activities. In addition, the eastern districts have relatively high temperatures and low precipitation.

4 Conclusions

Monitoring, modelling, prediction and isopleth mapping of the incidence rate of an epidemic provide the information for an early warning system needed to allocate resources effectively to the appropriate region and at the appropriate time. Spatial and temporal characteristics can be used to develop spatiotemporal models and maps of the disease distribution, which are instrumental to analysing and forecasting its diffusion.

Precise targeting requires information at fine spatial units and for short time intervals. However, there is usually a lack of information on covariates at this scale. This problem can be handled by applying pure models with spatial and temporal random effects and their interactions to account for the variation in the unobserved risk factors across space and time. Because of the large number of parameters that pure models support, we propose Bayesian Integrated Nested Laplace Analysis (INLA) to obtain the marginal predictive posterior distributions for estimation and forecasting. We furthermore propose isopleth mapping, which typically shows continuous change across space and thus smooths the disease's incidence rate and provides clues for epidemiologists to expand their aetiologic hypotheses on disease outbreaks.

The method was applied to monthly data for 30 districts at census tract level in the city of Bandung, Indonesia, for the period 2009–2016. The model was used to generate

monthly forecasts for 2017. The estimation and forecasting results and the isopleth maps showed a clear spatiotemporal picture. Although the dengue risk has generally decreased since 2014, some districts, especially in north-western Bandung, were estimated and forecasted to have relatively high risk. Geographical and weather conditions contributed to this. The north-western districts of Bandung are characterized by an elevation level of 800 m above sea level, dense vegetation, high precipitation and humidity and are an ideal breeding ground and habitat for the *Aedes*-spp. *mosquito* (Ebi and Nealon 2016). The relatively high risk level in central Bandung from January to August is related to high precipitation and socio-economic conditions, notably population density and high mobility. Moreover, the relatively high registration rate in the area is due to a relatively high level of education. The relatively low risk during the dry season is related to low humidity, which is due to the low vegetation density. Poverty, a lack of safe water and population density are related to the relatively high risk in the slum areas, whereas the opposite is true for the relatively low risk in the more prosperous districts in the east and west of the city. Enabling further research on the association between relative risk and potential risk factors is the main theme of spatial regression.

Specific actions can be considered for districts with forecasted high dengue risk levels, particularly control of the dengue vector (i.e. *Aedes*-spp. *mosquito*) and protecting people from mosquito bites. Such actions include the cleaning-up of abandoned properties and daily garbage collection to control the dengue vector. In addition, fumigating inside homes and using mosquito bed nets are needed to protect people from mosquito bites in the dengue-prone districts during the dengue season, between January and August.

Acknowledgements We are very grateful for the comments made by the editor and the reviewers as they have contributed to improve an earlier version of this article. We also thank the health office, city of Bandung, for supplying the aggregated dengue and population data per district. The authors thank Rector Padjadjaran University (Unpad). This paper is funded by ALG Unpad Contract: 2296/UN6.D/KS/2018 and RFU Unpad Contract: 1732 d/UN6.RKT/LT/2018.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Appendix 1: Laplace approximation (Rue and Martino 2009)

The Laplace approximation to the integral of a scalar function $p(\mathbf{\Phi})$, with $\mathbf{\Phi}$a n -dimensional vector, has the following form:

$$\int_{-\infty}^{\infty} p(\mathbf{\Phi})d\mathbf{\Phi} = \int_{-\infty}^{\infty} \exp\left(\log p(\mathbf{\Phi})\right)d\mathbf{\Phi} = \int_{-\infty}^{\infty} \exp\left(g(\mathbf{\Phi})\right)d\mathbf{\Phi}. \tag{64}$$

where $g(\mathbf{\Phi}) = \log p(\mathbf{\Phi})$. Since for unimodal functions the integral value is mainly determined by the behaviour around the mode of $g(\mathbf{\Phi})$, a second-order Taylor

approximation of $g(\boldsymbol{\Phi})$ can be substituted for $g(\boldsymbol{\Phi})$ to calculate an approximate value of the integral.

Let $\boldsymbol{\Phi}^*$ be the global maximum of $(\boldsymbol{\Phi})$ which is defined as:

$$\boldsymbol{\Phi}^* = \text{argmax}_{\boldsymbol{\Phi}} g(\boldsymbol{\Phi}), \text{then } \left.\frac{\partial g(\boldsymbol{\Phi})}{\partial \boldsymbol{\Phi}}\right|_{\boldsymbol{\Phi}=\boldsymbol{\Phi}^*} = 0 \tag{65}$$

and then $g(\boldsymbol{\Phi})$ is approximated as:

$$g(\boldsymbol{\Phi}) \approx g\left(\boldsymbol{\Phi}^*\right) + 0.5\left(\boldsymbol{\Phi} - \boldsymbol{\Phi}^*\right)'\mathbf{H}(g)\left(\boldsymbol{\Phi}^*\right)\left(\boldsymbol{\Phi} - \boldsymbol{\Phi}^*\right), \tag{66}$$

where $\mathbf{H}(g)\left(\boldsymbol{\Phi}^*\right)$ is the Hessian of $g\left(\boldsymbol{\Phi}^*\right)$, and Eq. (64) can be written:

$$\begin{aligned}
\int_{-\infty}^{\infty} p(\boldsymbol{\Phi})\mathrm{d}\boldsymbol{\Phi} &= \int_{-\infty}^{\infty} \exp\left(g\left(\boldsymbol{\Phi}^*\right) + 0.5\left(\boldsymbol{\Phi} - \boldsymbol{\Phi}^*\right)'\mathbf{H}(g)\left(\boldsymbol{\Phi}^*\right)\left(\boldsymbol{\Phi} - \boldsymbol{\Phi}^*\right)\right)\mathrm{d}\boldsymbol{\Phi} \\
&= \exp\left(g\left(\boldsymbol{\Phi}^*\right)\right)\int_{-\infty}^{\infty} \exp\left(0.5\left(\boldsymbol{\Phi} - \boldsymbol{\Phi}^*\right)'\mathbf{H}(g)\left(\boldsymbol{\Phi}^*\right)\left(\boldsymbol{\Phi} - \boldsymbol{\Phi}^*\right)\right)\mathrm{d}\boldsymbol{\Phi} \\
&= \exp\left(g\left(\boldsymbol{\Phi}^*\right)\right)\int_{-\infty}^{\infty} \exp\left(-0.5\left(\boldsymbol{\Phi} - \boldsymbol{\Phi}^*\right)'\left[-\mathbf{H}(g)\left(\boldsymbol{\Phi}^*\right)\right]\left(\boldsymbol{\Phi} - \boldsymbol{\Phi}^*\right)\right)\mathrm{d}\boldsymbol{\Phi} \\
&= \exp\left(g\left(\boldsymbol{\Phi}^*\right)\right)(2\pi)^{\frac{n}{2}}\left|\mathbf{H}(g)\left(\boldsymbol{\Phi}^*\right)\right|^{-\frac{1}{2}} \\
&\quad \int_{-\infty}^{\infty} (2\pi)^{-\frac{n}{2}}\left|\mathbf{H}(g)\left(\boldsymbol{\Phi}^*\right)\right|^{\frac{1}{2}} \exp\left(-0.5\left(\boldsymbol{\Phi} - \boldsymbol{\Phi}^*\right)'\left[-\mathbf{H}(g)\left(\boldsymbol{\Phi}^*\right)\right]\left(\boldsymbol{\Phi} - \boldsymbol{\Phi}^*\right)\right)\mathrm{d}\boldsymbol{\Phi},
\end{aligned} \tag{67}$$

where the integral can be associated with the density of a multivariate Gaussian distribution. In fact, by setting $-\mathbf{H}(g)\left(\boldsymbol{\Phi}^*\right) = \mathbf{Q}\left(\boldsymbol{\Phi}^*\right)$ where $\mathbf{Q}\left(\boldsymbol{\Phi}^*\right)$ denotes the precision matrix for the random vector $\boldsymbol{\Phi}^*$, we obtain:

$$\begin{aligned}
&\int_{-\infty}^{\infty} p(\boldsymbol{\Phi})d\boldsymbol{\Phi} \approx \exp\left(g\left(\boldsymbol{\Phi}^*\right)\right)(2\pi)^{\frac{n}{2}}\left|\mathbf{Q}\left(\boldsymbol{\Phi}^*\right)\right|^{-\frac{1}{2}} \\
&\int_{-\infty}^{\infty} (2\pi)^{-\frac{n}{2}}\left|\boldsymbol{Q}\left(\boldsymbol{\Phi}^*\right)\right|^{\frac{1}{2}} \exp\left(-0.5\left(\boldsymbol{\Phi} - \boldsymbol{\Phi}^*\right)'\mathbf{Q}\left(\boldsymbol{\Phi}^*\right)\left(\boldsymbol{\Phi} - \boldsymbol{\Phi}^*\right)\right)\mathrm{d}\boldsymbol{\Phi} \\
&\quad \approx (2\pi)^{\frac{n}{2}}\left|\mathbf{Q}\left(\boldsymbol{\Phi}^*\right)\right|^{-\frac{1}{2}} \exp\left(g\left(\boldsymbol{\Phi}^*\right)\right).
\end{aligned} \tag{68}$$

Appendix 2: Gaussian approximation

The conditional posterior distribution of $p(\boldsymbol{\Phi}|\boldsymbol{\tau}, \mathbf{y})$ is defined from the joint posterior distribution in Eq. (34):

$$p(\boldsymbol{\Phi}, \boldsymbol{\tau}|\boldsymbol{y}) \propto p(\boldsymbol{\tau})|\mathbf{Q}(\boldsymbol{\tau})|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\Phi}'\mathbf{Q}(\boldsymbol{\tau})\boldsymbol{\Phi} + \sum_{\mathrm{i}=1}^{\mathrm{n}} \log p\left(\mathbf{y}_i|\boldsymbol{\Phi}_i, \boldsymbol{\tau}\right) \right).$$

By ignoring elements without $\boldsymbol{\Phi}$, it can be written as:

$$p(\boldsymbol{\Phi}|\tau, y) \propto \exp\left(-\frac{1}{2}\boldsymbol{\Phi}' Q(\tau)\boldsymbol{\Phi} + \sum_{i=1}^{n} g_i\left(\boldsymbol{\Phi}_i\right) \right), \tag{69}$$

where the term $g_i\left(\boldsymbol{\Phi}_{\boldsymbol{i}}\right) = \log p\left(\mathbf{y}_{\boldsymbol{i}}|\boldsymbol{\Phi}_{\boldsymbol{i}}, \boldsymbol{\tau}\right)$.

The Gaussian approximation of Eq. (69), $p_G(\boldsymbol{\Phi}|\boldsymbol{\tau}, \boldsymbol{y})$, is reached by matching the mode and the curvature at the mode of $p(\boldsymbol{\Phi}|\boldsymbol{\tau}, \mathbf{y})$. The mode is computed iteratively by using a Newton–Raphson method. Let $\mu^{(0)}$ be the initial value of the mode and expand $g_i\left(\boldsymbol{\Phi}_{\boldsymbol{i}}\right)$ around $\boldsymbol{\mu}_i^{(0)} = \left(\mu_{i1}^{(0)}, \ldots, \mu_{iT}^{(0)}\right)'$ to the second-order Taylor expansion,

$$g_i\left(\boldsymbol{\Phi}_{\boldsymbol{i}}\right) \approx g_i\left(\boldsymbol{\mu}_i^{(0)}\right) + \mathbf{b}_{\boldsymbol{i}}'\boldsymbol{\Phi}_{\boldsymbol{i}} - \frac{1}{2}\mathbf{c}_{\boldsymbol{i}}'\boldsymbol{\Phi}_{\boldsymbol{i}}'\boldsymbol{\Phi}_{\boldsymbol{i}} \tag{70}$$

where $\mathbf{b}_i$ and $\mathbf{c}_i$ depend on $\boldsymbol{\mu}^{(0)}$. Inserting Eqs. (70) into (69) gives:

$$\begin{aligned} p_G(\boldsymbol{\Phi}|\tau, y) &\approx g_i\left(\mu_i^{(0)}\right) \exp\left(-\frac{1}{2}\boldsymbol{\Phi}'(Q + c)\boldsymbol{\Phi} + b'\boldsymbol{\Phi} \right) \\ &\propto \exp\left(-\frac{1}{2}\boldsymbol{\Phi}'(Q + c)\boldsymbol{\Phi} + b'\boldsymbol{\Phi} \right). \end{aligned} \tag{71}$$

A Gaussian approximation of $p_G(\boldsymbol{\Phi}|\boldsymbol{\tau}, \boldsymbol{y})$ is obtained, with the precision matrix $(\mathbf{Q} + \mathrm{diag}(\mathbf{c}))$ and mode $\boldsymbol{\mu}^{(1)}$, which is the solution of$(\mathbf{Q} + \mathrm{diag}(\mathbf{c}))\boldsymbol{\mu}^{(1)} = \mathbf{b}$. The process can then be iterated, with $\boldsymbol{\mu}^{(1)}$ as the new starting value, until it converges to a Gaussian distribution with, say, mean $\boldsymbol{\mu}^{(j)} \rightarrow \boldsymbol{\mu}^{(*)} = \boldsymbol{\Phi}^*$ and precision matrix $\mathbf{Q}^{(j)} \rightarrow \mathbf{Q}^{(*)} = \mathbf{Q} + \mathrm{diag}(\mathbf{c}^*), j = 1, 2, \ldots$, where an appropriate convergence criterion must be used (e.g. $\boldsymbol{\mu}^{(j+1)} - \boldsymbol{\mu}^{(j+1)} = 10^{-6}$).

The Gaussian approximation of the conditional posterior distribution $p(\boldsymbol{\Phi}|\boldsymbol{\tau}, \mathbf{y})$ is computationally very fast, but the approximation is generally not very good if the conditional posterior distribution is not to close to Gaussian (Hicketier 2015).

References

Abente LG, Aragonés N, García-Pérez J, Fernández NP (2018) Disease mapping and spatio-temporal analysis: importance of expected-case computation criteria. Geospat Health 9(1):27–33

Acharya KB, Cao C, Lakes T, Chen W, Naeem S (2016) Spatiotemporal analysis of dengue fever in Nepal from 2010 to 2014. BMC Public Health 16(1):849–858

Adin A, Lee D, Goicoa T, Ugarte MD (2018) A two-stage approach to estimate spatial and spatio-temporal disease risks in the presence of local discontinuities and clusters. Stat Methods Med. https://doi.org/10.1177/0962280218767975

Anderson C, Ryan LM (2017) A comparison of spatio-temporal disease mapping approaches including an application to ischaemic heart disease in New South Wales. Int J Environ Res Public Health 14(2):1–16

Arab A (2015) Spatial and spatio-temporal models for modeling epidemiological data with excess zeros. Int J Environ Res Public Health 12(9):10536–10548

Araujo RV, Albertini MR, Costa-da-Silva AL, Suesdek L, Franceschi NC, Bastos NM, Katz G, Cardoso VA, Castro BC, Capurro ML, Allegro VL (2015) São Paulo urban heat islands have a higher incidence of dengue than other urban areas. Braz J Infect Dis 19(2):146–155

Bandung HD (2010) Health profile of Bandung municipality in 2009. Bandung Government, Bandung

Bandung HD (2011) Health profile of Bandung municipality in 2010. Bandung Government, Bandung

Bandung HD (2012) Health profile of Bandung municipality in 2011. Bandung Government, Bandung

Bandung HD (2013) Health profile of Bandung municipality in 2012. Bandung Government, Bandung

Bandung HD (2014) Health profile of Bandung municipality in 2013. Bandung Government, Bandung

Bandung HD (2015) Health profile of Bandung municipality in 2014. Bandung Government, Bandung

Bandung HD (2016) Health profile of Bandung municipality in 2015. Bandung Government, Bandung

Bandung HD (2017) Health profile of Bandung municipality in 2016. Bandung Government, Bandung

Bauer C, Wakefield J, Rue H, Self S, Feng Z, Wang Y (2016) Bayesian penalized spline models for the analysis of spatio-temporal count data. Stat Med 35(11):1848–1865

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, York J, Mollié A (1991) Bayesian image restoration, with two applications in spatial statistics. Ann Inst Stat Math 43(1):1–20

Bivand RS, Gomez-Rubio V, Rue H (2015) Spatial data analysis with R-INLA with some extensions. J Stat Softw 63(20):1–31

Blangiardo M, Cameletti M (2015) Spatial and spatio-temporal Bayesian models with R-INLA. Wiley, Chichester

Blangiardo M, Cameletti M, Baio G, Rue H (2013) Spatial and spatio-temporal model with INLA. Spat Spatiotemporal Epidemiol 4(1):33–49

CDC (2014) Epidemiology. Centers for disease control and prevention. https://www.cdc.gov/dengue/epidemiology/index.html. Accessed 10 April 2018

CDC (2016) Surveillance and control of aedes aegypti and aedes albopictus in the United States. CDC, Atlantas

Chipman H, George EI, McCulloch RE (2001) The practical implementation of Bayesian model selection. In: Lahiri P (ed) Model selection-monograph series, vol 38. Institute of Mathematical Statistics Lecture Notes, Beachwood, pp 65–134

Choi Y, Tang SC, McIverL HM, Chan V, Abeyasinghe RR, Iddings S, Huy R (2016) Effects of weather factors on dengue fever incidence and implications for interventions in Cambodia. BMC Public Health 16(1):241–247

Christenfeld NJ, Sloan RP, Carroll D, Greenland S (2004) Risk factors, confounding, and the illusion of statistical control. Psychosom Med 66(6):868–875

Clayton D, Kaldor J (1987) Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics 43(3):671–681

Coly S, Garrido MC, Abrial D, Lafourcade AF (2015) Spatiotemporal disease mapping applied to infectious diseases. Proc Environ Sci 26(1):32–37

De Smedt T, Simons K, Van Nieuwenhuyse A, Molenberghs G (2015) Comparing MCMC and INLA for disease mapping with Bayesian hierarchical models. Arch Public Health 73(1):1

Detik (2016) Vaksin DBD masih mahal, belum semua RS menyediakan. Detik. https://health.detik.com/berita-detikhealth/d-3329213/vaksin-dbd-masih-mahal-belum-semua-rs-menyediakan. Accessed 1 March 2018

Ebi K, Nealon J (2016) Dengue in a changing climate. Environ Res 151(1):115–123

Emch M, Root ED, Carrel M (2017) Health and medical geography, 4th edn. The Guilford Press, New York

Francq C, Zakoian J (2010) GARCH models: structure, statistical inference and financial applications. Wiley, Chichester

Gelfand AE (1996) Model determination using sampling-based methods. In: Gilks WR, Richardson S, Spiegelhalter DJ (eds) Markov chain Monte Carlo in practice. Chapman & Hall, London, pp 145–162
Gelfand AE, Dey DK (1994) Bayesian model choice: asymptotics and exact calculations. J R Stat Soc B 56(3):501–514
Gelman A (2006) Prior distribution for variance parameters in hierarchical models. Bayesian Anal 1(3):515–533
Gelman A, Hwang J, Vehtari A (2014) Understanding predictive information criteria for Bayesian models. Stat Comput 24(6):997–1016
Handayani D, Folmer H, Kurnia A, Notodiputro AK (2018) The spatial empirical Bayes predictor of the small area mean for a lognormal variable of interest and spatially correlated random effects. Empir Econ 55(1):147–167
Hicketier A (2015) Spatio-temporal modeling of Hantavirus in Germany. Thesis, University of Stockholm, Sweden
Jaya IGNM, Folmer H, Ruchjana BN, Kristiani F, Andriyana Y (2017) Modeling of infectious diseases: a core research topic for the next hundred years. In: Jackson R, Schaeffer P (eds) Regional research frontiers, methodological advances, regional Systems modeling and open sciences, vol 2. Springer, West Virginia, pp 239–255
Knorr-Held L (2000) Bayesian modeling of inseparable space–time variation in disease risk. Stat Med 19(17–18):2555–2567
Kompas (2016) Health. Kompas. https://health.kompas.com/read/2016/02/04/155700723/Begini.Siklus.Hidup.Nyamuk.Aedes.Aegypti.Penyebar.DBD. Accessed 13 August 2018
Lawson A (2006) Statistical methods in spatial epidemiology. Wiley, London
Lawson A (2013) Bayesian disease mapping: hierarchical modelling in spatial epidemiology, 2nd edn. Chapman & Hall, London
Lawson A, Zhou H (2005) Spatial statistical modeling of disease outbreaks with particular reference to the UK foot and mouth disease (FMD) epidemic of 2001. Prev Vet Med 71(3–6):141–156
LeSage J, Pace K (2009) Introduction to spatial econometrics. Chapman & Hall, Boca Raton
Linden A, Adams J, Roberts N (2003) Evaluating disease management program effectiveness: an introduction to time-series analysis. Dis Manag 6(4):243–255
Liu Y, Watson SC, Gettings JR, Lund RB, Nordone SK, Yabsley MJ, McMahan CS (2017) A Bayesian spatio-temporal model for forecasting Anaplasma species seroprevalence in domestic dogs within the contiguous United States. PLoS ONE 12(7):1–18
Lloyd C (2010) Spatial data analysis. Oxford University Press, New York
McMichael A, Lendrum CD, Corvalán C, Ebi K, Githeko A, Scheraga J, Woodward A (2013) Climate change and human health risks and responses. WHO, Geneva
Metro-Trend (2019) Perkiraan cuaca di Bandung Kota. Metrotrend. https://id.meteotrend.com/forecast/id/bandung/. Accessed 10 Feb 2019
Mohebbi M, Wolfe R, Forbes A (2014) Disease mapping and regression with count data in the presence of overdispersion and spatial autocorrelation: a Bayesian model averaging. Int J Environ Res Public Health 11(1):883–902
Morrison KT, Shaddock G, Henderson SB, Buckeridge DL (2016) A latent process model for forecasting multiple time series in environmental public health surveillance. Stat Med 35(18):3085–3100
Murray NE, Quam MB, Wilder-Smith A (2013) Epidemiology of dengue: past, present and future prospects. Clin Epidemiol 5(1):299–309
Opitz T (2017) Latent Gaussian modeling and INLA: a review with focus on space-time applications. J Fr Stat Soc 158(3):62–85
Peter D, Silvia P (2012) ARIMA vs. ARIMAX-which approach is better to analyze and forecast macroeconomic time series? In: International conference mathematical methods in economics. Silesian University, Czech Republic, pp 136–140
Rao J, Molina I (2015) Small area estimation, 2nd edn. Wiley, Hoboken
Revesz P (2003) Constraint-based visualization of spatiotemporal database. In: Sarfraz M (ed) Advances in geometrics modeling. Wiley, New York, pp 263–275
Rodrigues E, Assunção R (2012) Bayesian spatial models with a mixture neighborhood structure. J Multivar Anal 109(1):88–102
Rue H, Martino S (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J R Stat Soc Ser B 71(2):319–392

Rue H, Riebler A, Sørbye SH, Illian JB, Simpson DP, Lindgren FK (2017) Bayesian computing with INLA: a review. Annu Rev Stat Appl 4(1):395–421
Rueda L, Patel K, Axtell R, Stinner R (1990) Temperature-dependent development and survival rates of culex quinquefasciatus and aedes aegypti (Diptera: Culicidae). J Med Entomol 27(5):892–898
Sari WC (2018) Selain DBD, waspada diare dan ISPA di musim ini. Pikiran Rakyat. http://www.pikiran-rakyat.com/bandung-raya/2018/04/26/selain-dbd-waspada-diare-dan-ispa-di-musim-ini-423428. Accessed 10 Feb 2019
Schrödle B, Held L (2011) Spatio-temporal disease mapping using INLA. Environmetrics 22(6):725–734
Setianto A, Triandini T (2013) Comparison of kriging and inverse distance weighted (IDW) interpolation methods in lineament extraction and analysis. J SE Asian Appl Geol 5(1):21–29
Shepard DS, Undurraga EA, Halasa YA, Jr S (2016) The global economic burden of dengue: a systematic analysis. Lancet Infect Dis 16(8):931–941
Spiegelhalter D, Best N, Carlin B, Linde AV (2002) Bayesian measures of model complexity and fit. J R Stat Soc B 6(4):583–639
Suaya JA, Shepard DS, Siqueira JB, Martelli CT, Lum LC, Tan LH, Kongsin S, Jiamton S, Garrido F, Montoya R, Armien B, Huy R, Castillo L, Caram M, Sah BK, Sughayyar R, Tyo KR, Halstead SB (2009) Cost of dengue cases in eight countries in the Americas and Asia: a prospective study. Am J Trop Med Hyg 80(5):846–855
Tierney L, Kadane J (1986) Accurate approximations for posterior moments and marginal densities. J Am Stat Assoc 81(393):82–86
Tiwari C (2013) Methods for creating smoothed maps of disease burdens. In: Boscoe F (ed) Geographic health data fundamental techniques for analysis. CPI Group (UK), Croydon, pp 125–141
Tobler W (1970) A computer movie simulating urban growth in the Detroit region. Econ Geogr 46(1):234–240
Tribunjabar (2016) Pantau DBD, dinkes kota Bandung pelototi rumah sakit dan puskesmas. Tribunjabar. http://jabar.tribunnews.com/2016/02/02/pantau-dbd-dinkes-kota-bandung-pelototi-rumah-sakit-dan-puskesmas. Accessed 8 Aug 2018
Ugarte MD, Goicoa T, Etxeberria J, Militino A (2012) Projections of cancer mortality risks using spatio-temporal P-spline models. Stat Methods Med Res 21(5):545–560
Ugarte MD, Adin A, Goicoa T, Militino AF (2014) On fitting spatio-temporal disease mapping models using approximate Bayesian inference. Stat Methods Med Res 23(6):507–530
Urtasun AA (2017) Hierarchical and spline-based models in space-time disease mapping. Dissertation, Universidad Publica de Navarra, Spain
Utazi C, Afuecheta E, Nnanatu C (2018) A Bayesian latent process spatiotemporal regression model for areal count data. Spat Spatiotemporal Epidemiol 25(1):25–37
Wakefield J (2007) Disease mapping and spatial regression with count data. Biostatistics 8(2):158–183
Waller LA, Carlin BP (2010) Disease mapping. In: Gelfand AE, Diggle PJ, Fuentes M, Guttorp P (eds) Handbook of spatial statistics. Chapman & Hall, Boca Raton, pp 217–243
Waller LA, Carlin BP, Xia H, Gelfand AE (1997) Hierarchical spatio-temporal mapping of disease rates. J Am Stat Assoc 92(438):607–617
Wang X, Yue YR, Faraway JJ (2018) Bayesian regression modeling with INLA. Taylor & Francis, Boca Raton
Watanabe S (2010) Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. J Mach Learn Res 11(1):3571–3594
Watson SC, Liu Y, Lund RB, Gettings JR, Nordone SK, McMahan CS, Yabsley MJ (2017) A Bayesian spatio-temporal model for forecasting the prevalence of antibodies to Borrelia burgdorferi, causative agent of Lyme disease, in domestic dogs within the contiguous United States. PLoS ONE 12(5):1–22
WHO (2009) Dengue guidelines for diagnosis, treatment, prevention, and control. WHO Press, Geneva
WHO (2012) Global strategy for dengue prevention and control 2012–2020. WHO Library Cataloguing-in-Publication Data, New York
WHO (2016) Dengue vaccine research. WHO. http://www.who.int/immunization/research/development/dengue_vaccines/en/. Accessed 2 April 2018
WHO (2018) Dengue and severe dengue. WHO. http://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue. Accessed 2 Feb 2018
Widiyani R (2013) Penyebab DBD. Kompas. http://health.kompas.com/read/2013/04/03/18534298/Empat.Sekawan.Penyebab.DBD. Accessed 25 April 2018

Wijayanti SPM, Thibaud P, Chase-Topping M, Rainey SM, McFarlane M, Schnettler E, Biek R, Kohl A (2016) The importance of socio-economic versus environmental risk factors for reported dengue cases in Java, Indonesia. PLoS Negl Trop Dis 10(9):1–15

Yin P, Mu L, Madden M, Vena J (2014) Hierarchical Bayesian modeling of spatio-temporal patterns of lung cancer incidence risk in Georgia, USA: 2000–2007. J Geogr Syst 16(4):387–407

Zellweger RM, Cano J, Mangeas M, Fo T, Mercier A, Despinoy M, Menkès CE, Dupont-Rouzeyrol M, Nikolay B, Teurlai M (2017) Socioeconomic and environmental determinants of dengue transmission in an urban setting: an ecological study in Noumea, New Caledonia. PLoS Negl Trop Dis 11(4):1–18

Affiliations

I. Gede Nyoman Mindra Jaya[1,2] · Henk Folmer[1,3]

[1] Faculty of Spatial Sciences, University of Groningen, Groningen, The Netherlands

[2] Statistics Department, Padjadjaran University, Bandung, Indonesia

[3] Faculty of Economics and Management, Northwest Agricultural and Forestry University, Yangling, China