



# Spatiotemporal high-resolution prediction and mapping: methodology and application to dengue disease

I. Gede Nyoman Mindra Jaya<sup>1,2</sup> · Henk Folmer<sup>1,2</sup>

Received: 12 May 2021 / Accepted: 8 October 2021 / Published online: 19 February 2022  
© The Author(s) 2022

## Abstract

Dengue disease has become a major public health problem. Accurate and precise identification, prediction and mapping of high-risk areas are crucial elements of an effective and efficient early warning system in countering the spread of dengue disease. In this paper, we present the fusion area-cell spatiotemporal generalized geoad-ditive-Gaussian Markov random field (FGG-GMRF) framework for joint estimation of an area-cell model, involving temporally varying coefficients, spatially and temporally structured and unstructured random effects, and spatiotemporal interaction of the random effects. The spatiotemporal Gaussian field is applied to determine the unobserved relative risk at cell level. It is transformed to a Gaussian Markov random field using the finite element method and the linear stochastic partial differential equation approach to solve the “big  $n$ ” problem. Sub-area relative risk estimates are obtained as block averages of the cell outcomes within each sub-area boundary. The FGG-GMRF model is estimated by applying Bayesian Integrated Nested Laplace Approximation. In the application to Bandung city, Indonesia, we combine low-resolution area level (district) spatiotemporal data on population at risk and incidence and high-resolution cell level data on weather variables to obtain predictions of relative risk at subdistrict level. The predicted dengue relative risk at subdistrict level suggests significant fine-scale heterogeneities which are not apparent when examining the area level. The relative risk varies considerably across subdistricts and time, with the latter showing an increase in the period January–July and a decrease in the period August–December.

**Keywords** Dengue disease · Relative risk · Fusion area-cell generalized geoad-ditive-Gaussian Markov random field model · Bayesian statistics · “Big  $n$  · Problem · Bottom-up approach

---

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

<sup>1</sup> Faculty of Spatial Sciences, University of Groningen, Groningen, The Netherlands

<sup>2</sup> Statistics Department, Padjadjaran University, Bandung, Indonesia

**JEL Classification** C18 · I18**1 Introduction**

Dengue disease is a major challenge to healthcare worldwide, potentially leading to death, especially among the poor in low- and middle-income countries (Ak et al. 2018). In addition, there are latent costs related to stress, productivity loss, school absence, and care taking (Wilastonegoro et al. 2020). To reduce health impacts and treatment costs, there have been substantial efforts aimed at the prevention of dengue disease outbreaks (Kampen et al. 2014). For these purposes, statistical models aimed at identifying the causes, transmission mechanisms, and prediction of outbreaks at a fine spatiotemporal scale are of crucial importance (Messina et al. 2019).

High-resolution information is needed for research on the etiology of the disease and the development of control and prevention strategies (Ak et al. 2018; Jaya et al. 2017; Jaya and Folmer 2020, 2021a, b; Pokharel and Deardon 2016). Dengue disease typically involves a spatiotemporal pattern (Jaya and Folmer 2020, 2021a, b; Phanitchat et al. 2019; Puggioni et al. 2020); therefore, high-resolution spatiotemporal models and maps of the distribution of relative risk are basic elements of an early warning system aimed at identifying where and when an outbreak will occur (Hanigan et al. 2019; Shi et al. 2013; Xu et al. 2019).

In many areas of spatial research including epidemiology, data are often only available at different levels of aggregation (Moraga et al. 2017; Shi et al. 2013; Utazi et al. 2019). If low-resolution information is available whereas high-resolution (cell level) information is needed but lacking, area-to-cell disaggregation through joint area-cell estimation can be applied to obtain the missing information (Moraga et al. 2017; Utazi et al. 2019; Wang et al. 2018a).<sup>1</sup> Data fusion or data assimilation (Banerjee et al. 2015) and Bayesian melding (Fuentes and Raftery 2005; Liu et al. 2011) are terms used to denote integrating multiple data sources of different spatial resolutions.<sup>2</sup> The basic concept involves combining area and cell observations in a single statistical model. Combining data measured at different levels of aggregation can improve parameter estimation and increase prediction accuracy (Wang et al. 2018b). However, it may lead to spatial misalignment (Moraga et al. 2017; Sahu et al. 2010; Truong et al. 2014; Utazi et al. 2019), which induces biased or inconsistent estimators (Liu and Bertazzon 2016; Peng and Bell 2010; Saez and López-Casasnovas 2019).

Several kinds of correction approaches based on regression models have been developed to deal with area-to-cell misalignment problems (Banerjee and Gelfand 2002; Banerjee et al. 2015). Moraga et al. (2017) and Utazi et al. (2019) applied Bayesian geostatistical analysis to deal with misalignment in spatial non-Gaussian

<sup>1</sup> Below we use the notion of “cell” to denote spatial units of higher resolution.

<sup>2</sup> Joint area-cell estimation is based on the assumption that the same factors at area and cell level drive the spatiotemporal process, although usually in various degrees (Banerjee et al. 2015; Utazi et al. 2019). In the present case, the same disease risk factors (weather) drive disease incidence at area and cell levels.

data with linear covariates. However, methods that address area-to-cell misalignment in spatiotemporal non-Gaussian data with nonlinear covariates are less well known. This applies especially to Poisson or Negative Binomial (NB) spatiotemporal data, which are typically applied in dengue and other disease incidence modeling, but also in other kinds of spatial and regional research.

In this paper, we introduce the Fusion Area-Cell Spatiotemporal Generalized Geoadditive (GG)-Gaussian Field (GF) model, abbreviated as FGG-GF model, to generate high-resolution (cell) predictions based on observations at lower (area) resolution of the variable of interest (i.e., the number of dengue incidences) and the population at risk, and high-resolution cell data (i.e., weather variables), while controlling for misalignment. Moraga et al. (2017) and Utazi et al. (2019) have shown that the prediction performance of integrated area-cell models can outperform models that use single level data sources.

Kammann and Wand (2003) introduced the Generalized Geoadditive Model (GGM), which has become popular in disease mapping (among other fields) because of its suitability for making high-resolution maps (Muleia et al. 2020; Wand et al. 2011). The model assumes that there is a spatially continuous variable underlying all observations which can be modeled using a Gaussian process, usually denoted Gaussian Field (GF). A GF is characterized by a first-order autoregressive model with spatially correlated innovations. The GGM combines the Generalized Additive Model (GAM) and the Geostatistical Model (GM). The former was introduced by Hastie and Tibshirani (1986) to provide a flexible means of handling nonlinear and interacting covariates. GAMs are also suitable for handling complex spatial and temporal autocorrelation (French and Wand 2004; Ma et al. 2014). GAMs are non-parametric because they do not require a priori specification of the regression function (Wang et al. 2018a). The GM was introduced by Matheron (1963) to construct high-resolution maps over a particular geographical region based on cell data on (risk) factors associated with a (dependent) variable of interest.

The integrated area-cell observations and the combination of non-Gaussian data, a nonlinear predictor and latent model components, in particular the spatiotemporal GF, make estimation of FGG-GF model, prediction, and mapping computationally complex and time-consuming (Barber et al. 2016) because of the “big  $n$ ” problem. This issue can be handled by transforming a GF with a dense covariance matrix to a Gaussian Markov Random Field (GMRF) with a sparse precision matrix<sup>3</sup> of Matérn covariances (Lindgren et al. 2011).

The objective of this paper is to develop a high-resolution prediction and mapping procedure for spatiotemporal Poisson or Negative Binomial data applying a Fusion Area-Cell Spatiotemporal Generalized Geoadditive-Gaussian Markov Random Field (abbreviated as FGG-GMRF) model. Inference and prediction are handled in a Bayesian framework. The approach will subsequently be applied to dengue

---

<sup>3</sup> Note that although the precision matrix is sparse, the covariance matrix is dense in general. Specifying models in terms of the precision matrix rather than the covariance matrix is useful in many high-dimensional applications because it allows for a sparse representation, typically reducing computational cost and memory usage (Sidén et al. 2018).

disease risk in Bandung city, Indonesia. The purpose is to predict and map the relative dengue risk at subdistrict level, given observations on dengue incidence and population at risk at district level and weather risk factors at cell level.<sup>4</sup> Special attention is paid to high-risk districts and subdistricts requiring public intervention (Aguayo et al. 2020).

The structure of the remainder of this paper is as follows. Section 2 introduces the spatiotemporal GG-GF model. Section 3 presents the FGG-GF and FGG-GMRF models and the Bayesian inference framework. The link between the GF and GMRF models is summarized in Appendix 1. Section 4 applies the methodology to dengue incidence in Bandung city, Indonesia, and Sect. 5 summarizes and concludes the conducted research.

## 2 The spatiotemporal generalized ge additive-Gaussian field model

Consider region  $\mathcal{A} \in \mathbb{R}^2$ , partitioned into  $n_{\mathcal{A}}$  areas (e.g., districts in a city), each measured for  $T$  periods. The areas are labeled  $\{(\mathcal{A}_1, 1), \dots, (\mathcal{A}_1, T), (\mathcal{A}_2, 1), \dots, (\mathcal{A}_i, t), \dots, (\mathcal{A}_{n_{\mathcal{A}}}, T)\}$ , where  $(\mathcal{A}_i, t)$  denotes area  $i$  at time  $t$ , for  $i = 1, \dots, n_{\mathcal{A}}$  and  $t = 1, \dots, T$ . Region  $\mathcal{A}$  is further divided into a finite set of  $n_p$  cells for  $T$  periods. The set of  $n_p$  cells over  $T$  periods is denoted  $\{(s_1, 1), \dots, (s_{n_{\mathcal{A}_1}}, T), (s_{n_{\mathcal{A}_1}+1}, 1), \dots, (s_g, t), \dots, (s_{n_p}, T)\}$ , with  $n_{\mathcal{A}_i}$  denoting the number of cells in area  $\mathcal{A}_i$  for  $g = 1, \dots, n_p$  and  $t = 1, \dots, T$  and  $n_p = \sum_{i=1}^{n_{\mathcal{A}}} n_{\mathcal{A}_i}$ . Note that the notation  $\{(s_{1(\mathcal{A}_1)}, 1), \dots, (s_{n_{\mathcal{A}_1}(\mathcal{A}_1)}, T), (s_{n_{\mathcal{A}_1}+1(\mathcal{A}_2)}, 1), \dots, (s_{g(\mathcal{A}_i)}, t), \dots, (s_{n_p(\mathcal{A}_{n_{\mathcal{A}}})}, T)\}$  will be used to explicitly denote that cell  $s_{g(\mathcal{A}_i)}$  belongs to area  $\mathcal{A}_i$ . If the area is not relevant,  $s_g$  will be used. Moreover, the notation  $g$  for  $s_g$  will be incidentally used if there is no risk of misunderstanding. Finally,  $s_{g,1}$  and  $s_{g,2}$  denote the latitude and longitude coordinates of its centroid, respectively.

Let  $y_{it}$  and  $N_{it}$  denote the number of (dengue) incidences and population at risk in area  $\mathcal{A}_i$  at time  $t$ , respectively, and  $y_{gt}$  and  $N_{gt}$  the number of (dengue) incidences and population at risk in cell  $s_g$  at time  $t$ , respectively. Note that both  $y_{gt}$  and  $N_{gt}$  are unobserved at the cell level.  $y_{it}$  and  $y_{gt}$  are assumed to follow Poisson distributions<sup>5</sup> with means  $\mu_{it} = E_{it}\theta_{it}$  and  $\mu_{gt} = E_{gt}\theta_{gt}$ , respectively, with  $E_{it}$  and  $E_{gt}$  denoting the expected number of (dengue) incidences and  $\theta_{it}$  and  $\theta_{gt}$  the relative (dengue) risk for area  $i$  and cell  $g$  at time  $t$ , respectively (Jaya and Folmer 2020). The expected rate  $E_{it}$  is calculated using external standardization. It is defined based on the overall average across all areas and periods (Abente et al. 2018; Jaya and Folmer 2020, 2021a, b):

<sup>4</sup> Although Dengue disease incidence depends on socioeconomic risk factors such as income, family size, age, education, and weather variables such as humidity, precipitation, and sunshine, the case study presented below only considers the latter, while the socioeconomic risk factors are taken into account by the random effects. The reason is that the latter are only available at city level.

<sup>5</sup> A negative binomial distribution is appropriate when there is large overdispersion of zeros (Berk and MacDonald 2008; Payne et al. 2017).

$$E_{it} = N_{it} \left( \frac{1}{n_A T} \sum_{i=1}^{n_A} \sum_{t=1}^T y_{it} / \frac{1}{n_A T} \sum_{i=1}^{n_A} \sum_{t=1}^T N_{it} \right) \quad \text{for } i = 1, \dots, n_A \text{ and } t = 1, \dots, T. \tag{1}$$

The relative risk is defined as the ratio of the local risk in a spatiotemporal unit relative to the average risk across the whole study region over the entire time period (Yin et al. 2014). It is centered around one, meaning that the total number of incidences is equal to the expected rate. The maximum likelihood (ML) estimator of the relative risk  $\theta_{it}$  is (Jaya et al. 2017; Jaya and Folmer 2020):

$$\hat{\theta}_{it} = \frac{y_{it}}{E_{it}} \quad \text{for } i = 1, \dots, n_A \text{ and } t = 1, \dots, T. \tag{2}$$

This is known as the crude risk or the standardized incidence ratio (SIR).

Following Moraga et al. (2017), Jaya and Folmer (2020, 2021a, b), and Utazi et al. (2019), we model the relative risk as a non-separable Poisson log-linear model as follows<sup>6</sup>:

$$\eta_{it} = \beta_0 + \sum_{k=1}^K f_k(\bar{x}_{k,it}) + \omega_i + v_i + \phi_t + \zeta_t + \delta_{it} + \bar{\Phi}_{it} \quad \text{for } i = 1, \dots, n_A \text{ and } t = 1, \dots, T, \tag{3a}$$

$$\eta_{gt} = \beta_0 + \sum_{k=1}^K f_k(x_{k,gt}) + \omega_{g(A_t)} + v_{g(A_t)} + \phi_t + \zeta_t + \delta_{g(A_t)t} + \Phi_{gt} \tag{3b}$$

for  $g = 1, \dots, n_p$  and  $t = 1, \dots, T$

with  $\eta_{it} = \log(\theta_{it})$  and  $\eta_{gt} = \log(\theta_{gt})$ .

In Eqs. (3a) and (3b),  $\beta_0$  is the overall intercept denoting the average risk across space and time, i.e., across all  $i = 1, \dots, n_A$ ,  $g = 1, \dots, n_p$ , and  $t = 1, \dots, T$ . The latent<sup>7</sup> functions  $f_k(\bar{x}_{k,it})$  and  $f_k(x_{k,gt})$  for  $k = 1, \dots, K$ , represent the (non)linear effects of the metrical area and cell risk factors, respectively. The risk factors at cell level for a given area  $A_i$  and time  $t$  are fixed. However, they vary across areas and times. The latent (non)linear risk factor functions are based on observations at cell level but are predicted at area level. The risk at cell and area levels are assumed to be driven by the same factors; therefore, we adopt joint risk factor functions. For this purpose, we stack the observations on the risk factors such that risk factor  $k$ , at both

<sup>6</sup> Spatiotemporal variation (in disease outcomes) consists of the following three components: a spatial component capturing the overall spatial distribution, a temporal component capturing the overall temporal pattern, and a space–time component capturing space–time interaction. Spatiotemporal models are classified as separable or non-separable (Knorr-Held 2020). A spatiotemporal model is said to be separable if it consists of spatial and temporal components without their interaction, while a non-separable model comprises all three components (Haining and Li 2020; Knorr-Held 2000; Martinez-Beneito and Botella-Rocamora 2019).

<sup>7</sup> The regression functions are latent in that they do not have pre-specified functional forms. This feature makes latent regression functions suitable to accommodate complicated relationships between the risk factors and the dependent variable.

area and cell level, becomes  $\mathbf{z}_k = (\bar{x}_{k,11}, \dots, \bar{x}_{k,n_A T}, x_{k,11}, \dots, x_{k,n_p T})'$  for  $k = 1, \dots, K$  and latent function  $f_k(\mathbf{z}_k)$ . The functions  $f_k(\mathbf{z}_k)$  are commonly centered at the mean, i.e.,  $\mathbb{E}[f_k(\mathbf{z}_k)] = 0$ , for identifiability reasons (Fahrmeir and Lang 2001).

Let  $f(\mathbf{z})$  be the sum of the functions  $f_k(\mathbf{z}_k)$  for  $k = 1, \dots, K$ :

$$f(\mathbf{z}) = \sum_{k=1}^K f_k(\mathbf{z}_k) = f_1(\mathbf{z}_1) + \dots + f_K(\mathbf{z}_K). \tag{4}$$

To account for spatiotemporal variation, Eq. (4) can be extended to a varying coefficients model<sup>8</sup>:

$$f(\mathbf{z}) = f_1(\mathbf{z}_1)\mathbf{v}_1 + \dots + f_K(\mathbf{z}_K)\mathbf{v}_K \tag{5}$$

where the design vector  $\mathbf{v} = (\mathbf{v}_1, \dots, \mathbf{v}_K)'$  contains components of  $\mathbf{z}$  or additional covariates. The vector  $\mathbf{v}_k$  for  $k = 1, \dots, K$ , modifies the relationship between the covariate  $\mathbf{z}_k$  and the log-linear conditional expectation  $\mathbb{E}[\mathbf{y}|\mathbf{z}]$ . If it is identical to the vector  $\mathbf{1}$ , i.e.,  $\mathbf{v}_k = (1, \dots, 1)'$  with dimension  $(n_A + n_p)T \times 1$ , then  $f_k(\mathbf{z}_k)$  presents the overall (main) effect of  $\mathbf{z}_k$ . If it is different from  $\mathbf{1}$ ,  $f_k(\mathbf{z}_k)\mathbf{v}_k$  presents the effect of  $\mathbf{z}_k$  that varies along with  $\mathbf{v}_k$ . In other words,  $f_k(\mathbf{z}_k)\mathbf{v}_k$  models the interaction between  $\mathbf{z}_k$  and  $\mathbf{v}_k$  (Fahrmeir and Lang 2001). According to Martinez-Bello et al. (2017a; b), the varying coefficients model helps refine the association between the regressors (e.g., the weather variables) and the response, thus improving predictions at a fine spatiotemporal scale. For example, if  $\mathbf{v}_k$  denotes the calendar day and  $\mathbf{z}_k$  is the spatiotemporal covariate temperature, then  $f_k(\mathbf{z}_k)\mathbf{v}_k$  represents the temperature effect varying by day. In this study, we apply the temporally varying coefficients model to accommodate the temporally varying nonlinear effects of risk factors on the response. The time-varying effect of, for example, the  $k$ -th covariate can be written as (Franco-Villoria et al. 2019):

$$f_k(\mathbf{z}_{k,t})\mathbf{v}_{k,t} = \beta_k(\mathbf{v}_{k,t})\mathbf{z}_{k,t} \text{ for every } i \text{ and } g, \text{ and for } k = 1, \dots, K \text{ and } t = 1, \dots, T. \tag{6}$$

where  $\beta_k(\mathbf{v}_{k,t})$  for  $t = 1, \dots, T$  is the time-varying regression coefficient, which can be regarded as a stochastic process over  $\mathbf{v}_k$  (Fahrmeir and Lang 2001). For ease of notation, we ignore the term  $\mathbf{v}_k$  and write  $\beta_{k,t}$ .

A time-varying coefficient can be conveniently specified as the sum of a fixed (global mean) effect and a temporal random effect of the risk factor:  $\beta_{k,t} = \beta_k + \zeta_{k,t}$  for  $k = 1, \dots, K$  and  $t = 1, \dots, T$ . The fixed effect ( $\beta_k$ ) presents the effect of the risk factor that remains constant across space or time, while the temporal random effect ( $\zeta_{k,t}$ ) accounts for the time-varying effect of the risk factor (Song et al. 2020). The temporal random effect  $\zeta_{k,t}$  can be conveniently specified as a random walk model

<sup>8</sup> According to Hall et al. (2016) and Martinez-Bello et al. (2017b), temporally varying coefficients models can be seen as distributed lag parameters. They allow the visualization and exploration of changes in the dependent variable as linear or nonlinear functions of time.

of order one (RW1) or two (RW2)<sup>9</sup> (Bernardinelli et al. 1995; Martinez-Bello et al. 2017b; Schrödle and Held 2011):

$$\zeta_{k,t} = \zeta_{k,t-1} + u_{k,t}(\text{RW1}) \text{ for } t = 2, \dots, T \text{ or } \zeta_{k,t} = 2\zeta_{k,t-1} - \zeta_{k,t-2} + u_{k,t}(\text{RW2}) \text{ for } t = 3, \dots, T \tag{7}$$

with  $u_{k,t} \sim \mathcal{N}(0, \sigma_{\zeta_k}^2)$  white noise, and  $\sigma_{\zeta_k}^2$  denoting the variance of the RW process controlling the smoothness of  $\beta_{k,t}$ . A random walk process of order one needs an initial value of  $\zeta_{k,1}$  and a random walk of order two needs initial values of  $\zeta_{k,1}$  and  $\zeta_{k,2}$ .

The components  $\omega_i$  and  $v_i$  are the spatially structured and unstructured main effects at area level, respectively, whereas  $\omega_{g(A_i)}$  and  $v_{g(A_i)}$  denote the spatially structured and unstructured main effects for cell  $s_g$  in the area  $\mathcal{A}_i$  to which it belongs. The components  $\phi_t$  and  $\zeta_t$  are the temporally structured and unstructured main effects. For a given time  $t$ , they are equal for all areas and cells.  $\delta_{it}$  represents the spatiotemporal interaction effect of the unobserved risk factors at area level and  $\delta_{g(A_i)t}$  the impact of the interaction effect  $\delta_{it}$  in cell  $s_g$  in area  $\mathcal{A}_i$  to which it belongs (see Sect. 3.1 for details). We consider four type of space–time interactions (see Table 5 in Appendix 2).

The final component,  $\Phi_{gt}$ , in Eq. (3b) is the spatiotemporal GF in cell  $s_g$  for  $g = 1, \dots, n_p$  at time  $t$ , indicating the true but unobserved relative risk (Cameletti et al. 2013; Godana et al. 2019). Hence,  $\Phi_{gt}$  is the “own” spatiotemporal interaction effect of cell  $s_g$ . Because of the large number of cells, it is continuously indexed (Blangiardo and Cameletti 2015). The component  $\bar{\Phi}_{it}$  in Eq. (3a) denotes the area average of  $\Phi_{gt}$  across the cells within  $\mathcal{A}_i$ . Following Cameletti et al. (2013) and Godana et al. (2019), we assume that  $\Phi(s_g, t)$  changes over time following a first-order autoregressive (AR1) process with coefficient  $\lambda_1$ ,  $|\lambda_1| < 1$ :

$$\Phi(s_g, t) = \lambda_1 \Phi(s_g, t - 1) + \gamma(s_g, t) \quad \text{for } t = 2, \dots, T \text{ and } g = 1, \dots, n_p \tag{8}$$

with  $\Phi(s_g, 1) \sim \mathcal{N}(0, \sigma_{\Phi}^2 / (1 - \lambda_1^2))$  and  $\gamma(s_g, t)$  defined as a mean square differentiable process<sup>10</sup> (Stein 1999) with the temporally independent but spatially correlated innovations following a zero-mean Gaussian distribution with spatiotemporal covariance function:

$$\text{Cov}(\gamma(s_g, t), \gamma(s_h, t')) = \begin{cases} 0 & \text{if } t \neq t' \\ \sigma_{\Phi}^2 R(d) & \text{if } t = t' \end{cases} \tag{9}$$

for  $g \neq h$ .  $\sigma_{\Phi}^2$  is the homogeneous variance of  $\gamma(s_g, t)$ , i.e.,  $\text{Var}(\gamma(s_g, t)) = \sigma_{\Phi}^2$  for every  $s_g$  and  $t$ , and  $R(d)$  the spatial autocorrelation matrix as a function of the distance  $d$  between  $s_g$  and  $s_h$  at time  $t$  (e.g., the Euclidean distance). Under the assumption that the covariance function only depends on  $d$ , it is a Matérn covariance

<sup>9</sup> Random walk models are the Bayesian equivalents of P(enalized) Spline regression models (Wang et al. 2018a).

<sup>10</sup> A GF  $\gamma(s_g, t) = \gamma_t(s_g)$  is mean square differentiable in  $\mathcal{A}$ , if for every  $s_g$  in  $\mathcal{A}$  and for every  $t$ ,  $\gamma'_t(s_g) = \lim_{l \rightarrow 0} \frac{\gamma_t(s_g+l) - \gamma_t(s_g)}{l}$  exists and  $\lim_{l \rightarrow 0} \mathbb{E} \left[ \frac{\gamma_t(s_g+l) - \gamma_t(s_g)}{l} - \gamma'_t(s_g) \right]^2 = 0$  (Stein 1999).

function satisfying the second-order stationarity and isotropy assumptions. Consequently, the mean of the process is constant and only depends on the locations of  $s_g$  and  $s_h$  through the Euclidean distance  $d = \|s_g - s_h\| \in \mathbb{R}$  (Song et al. 2008). The spatial autocorrelation function  $\mathcal{R}(d)$  is defined as:

$$\mathcal{R}(d) = \frac{1}{\Gamma(\nu)2^{\nu-1}}(\kappa d)^\nu K_\nu(\kappa d) \quad \text{for every } t \quad (10)$$

where  $\Gamma(\cdot)$  is the gamma function,  $K_\nu(\cdot)$  the modified Bessel function of the second order (Abramovitz and Stegun 1965) and  $\nu > 0$  the parameter controlling the smoothness of the GF (smoothness parameter). In applications,  $\nu$  is commonly fixed because it is usually poorly identified (Miller et al. 2019; Utazi et al. 2019). In several software packages, including R-INLA (Integrated Nested Laplace Approximation), the default value is one ( $\nu = 1$ ), corresponding to moderate smoothness (Lindgren et al. 2011; Utazi et al. 2019). The scale parameter  $\kappa$ ,  $\kappa > 0$  controls the rate of decay of the correlation and is inversely related to the range parameter  $r$  of the Euclidean distance between  $\gamma(s_g, t)$  and  $\gamma(s_h, t)$ . For large  $r$ ,  $\kappa$  goes to zero. Because of a lack of a simple relationship between  $\kappa$  and  $r$ , Lindgren et al. (2011) proposed the empirically derived relationship  $r = \sqrt{8\nu/\kappa}$  for spatial autocorrelation near 0.1. Substituting Eq. (10) in Eq. (9), the spatiotemporal Matérn covariance function  $\Sigma(d)$  for each time  $t$  is:

$$\Sigma(d) = \frac{\sigma_\Phi^2}{\Gamma(\nu)2^{\nu-1}}(\kappa d)^\nu K_\nu(\kappa d) \quad \text{for every } t. \quad (11)$$

For the joint latent spatiotemporal GF  $\Phi_t = (\Phi_{1t}, \dots, \Phi_{n_t})'$  at cell level, we have:

$$\Phi_t = \lambda_1 \Phi_{t-1} + \gamma_t \text{ and } \gamma_t \sim \mathcal{N}(0, \Sigma) \quad \text{for } t = 2, \dots, T. \quad (12)$$

with  $\gamma_t = (\gamma_{1,t}, \dots, \gamma_{n_t,t})'$  and  $\Sigma = \sigma_\Phi^2 \mathcal{R}$  a Matérn covariance matrix. That is, the joint latent spatiotemporal GF is a second-order stationary, isotropic GF with Matérn covariance function Eq. (11) and initial value distributed as  $\Phi_1 \sim \mathcal{N}\left(0, \frac{\sigma_\Phi^2}{(1-\lambda_1^2)} \mathcal{R}\right)$ .

### 3 Bayesian inference

This section consists of two subsections. In the first, we present the Fusion Area-Cell Spatiotemporal Generalized Geoadditive-Gaussian Field (FGG-GF) model which integrates the sub-models (3a) and (3b) into a single statistical model. The section also presents the Bayesian statistical tools. In the second section, we discuss solving the “big  $n$ ” problem resulting in the Fusion Area-Cell Spatiotemporal Generalized Geoadditive-Gaussian Markov Random Field (FGG-GMRF) model and point out that it can be estimated using the R-INLA package. Details on the link between the GF and the GMRF through the Linear Stochastic Partial Differential Equation (LSPDE) approach are discussed in Appendix 1.

### 3.1 The fusion area-cell spatiotemporal generalized geoadditive-Gaussian field model

As observed above, combining low-resolution and high-resolution data to generate high-resolution predictions entails the risk of misalignment (Moraga et al. 2017; Utazi et al. 2019). To handle misalignment, we first stack the corresponding objects of the area and cell models in Eqs. (3a) and (3b) to give the FGG-GF model<sup>11</sup> which is then estimated as a single model (Blangiardo and Cameletti 2015; Kifle et al. 2017; Utazi et al. 2019). The FGG-GF model reads:

$$\boldsymbol{\eta}_t = \beta_0 \mathbf{1}_{(n_A+n_p)} + \sum_{k=1}^K (\beta_k + \zeta_{k,t}) \mathbf{z}_{k,t} + \ddot{\boldsymbol{\omega}} + \ddot{\mathbf{v}} + \phi_t \mathbf{1}_{(n_A+n_p)} + \zeta_t \mathbf{1}_{(n_A+n_p)} + \ddot{\boldsymbol{\delta}}_t + \ddot{\boldsymbol{\Phi}}_t$$

for  $t = 1, \dots, T$

(13)

where  $\boldsymbol{\eta}_t = (\eta_{1,t}, \dots, \eta_{n_A,t}, \eta_{(n_A+1),t}, \dots, \eta_{(n_A+n_p),t})'$ ,  $\beta_0$  the global mean defined in Eq. (3),  $\mathbf{1}_{(n_A+n_p)}$  a vector of ones of dimension  $(n_A + n_p)$ ,  $\mathbf{z}_{k,t} = (\mathbf{z}_{k,1}, \dots, \mathbf{z}_{k,n_A}, \mathbf{z}_{k,(n_A+1)}, \dots, \mathbf{z}_{k,(n_A+n_p)})'$  the joint  $k$  th risk factor with fixed coefficient  $\beta_k$  and temporal random coefficient  $\zeta_k = (\zeta_{k,1}, \dots, \zeta_{k,T})'$  for  $k = 1, \dots, K$ . Furthermore,  $\ddot{\boldsymbol{\omega}} = (\boldsymbol{\omega}, \boldsymbol{\omega}_A)'$ , with  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_i, \dots, \omega_{n_A})'$  and  $\boldsymbol{\omega}_A = (\omega_{1(A_i)}, \dots, \omega_{g(A_i)}, \dots, \omega_{n_p(A_{n_A})})'$ ,  $\ddot{\mathbf{v}} = (\mathbf{v}, \mathbf{v}_A)'$ , with  $\mathbf{v} = (v_1, \dots, v_i, \dots, v_{n_A})'$  and  $\mathbf{v}_A = (v_{1(A_i)}, \dots, v_{g(A_i)}, \dots, v_{n_p(A_{n_A})})'$ ,  $\ddot{\boldsymbol{\delta}}_t = (\boldsymbol{\delta}_t, \boldsymbol{\delta}_{At})'$ , with  $\boldsymbol{\delta}_t = (\delta_{1t}, \dots, \delta_{it}, \dots, \delta_{n_A t})'$ ,  $\phi_t$  and  $\zeta_t$  defined as in Eq. (3), and  $\delta_{At} = (\delta_{1(A_i)t}, \dots, \delta_{g(A_i)t}, \dots, \delta_{n_p(A_{n_A})t})'$ ,  $\ddot{\boldsymbol{\Phi}}_t = (\overline{\boldsymbol{\Phi}}_t, \boldsymbol{\Phi}_t)'$ , with  $\overline{\boldsymbol{\Phi}}_t = (\overline{\Phi}_{1t}, \dots, \overline{\Phi}_{it}, \dots, \overline{\Phi}_{n_A t})'$  and  $\boldsymbol{\Phi}_t = (\Phi_{1t}, \dots, \Phi_{gt}, \dots, \Phi_{n_p t})'$ . Note that the above vectors are  $((n_A + n_p) \times 1)$  for  $t = 1, \dots, T$ .

The following observations apply. First, the basic components of a high-resolution spatiotemporal relative risk model are the covariates and/or the GF at cell level ( $\Phi(s_g, t)$ ). Either one or both are required for the estimation of the relative risk at cell level. Second, the interaction terms  $\delta_{it}$  and  $\Phi_{gt}$  in the non-separable models in Eqs. (3a) and (3b), respectively, have as covariance matrices the Kronecker products of the spatial and temporal covariance matrices (Blangiardo and Cameletti 2015; Fuentes et al. 2008). See Table 5 in Appendix 2 and Sect. 3.2 for further details. For alternative approaches to handling non-separable models, see among others Bakka et al. (2020), Gneiting (2002), and Sherman (2011). Third, the parameters  $\{\omega_i, v_i, \delta_{it}\}$  are estimated at area level. For the cell level, they are the corresponding area level parameters, implying that they do not vary among cells within a given area  $\mathcal{A}_i$ . Fourth, to control for misalignment, for each area  $\mathcal{A}_i$ , the model component  $\overline{\Phi}_{it}$  and the area values of the risk factors  $\overline{x}_{k,it}$  are taken as the block averages of the cells within  $\mathcal{A}_i$  for a given time point  $t$ , respectively (Banerjee et al. 2015). That is, for  $i = 1, \dots, n_A$  and  $t = 1, \dots, T$ ,  $\overline{x}_{k,it} = |\mathcal{A}_i|^{-1} \int_{\mathcal{A}_i} x_k(s, t) ds$  for  $k = 1, \dots, K$  and  $\overline{\Phi}_{it} = |\mathcal{A}_i|^{-1} \int_{\mathcal{A}_i} \Phi(s, t) ds$

<sup>11</sup> The inla.stack function can be used to define the fusion model framework.

where  $|\mathcal{A}_i| = \int 1 ds$  denotes the size of  $\mathcal{A}_i$ . The simplest procedure to estimate  $\bar{x}_{k,it}$  is to approximate  $|\mathcal{A}_i|^{-1} \int_{\mathcal{A}_i} \mathbf{x}_k(s, t) ds$  for each time  $t$  by taking the average of the values of the cell risk factor  $x_{k,g(\mathcal{A}_i)t}$  in  $\mathcal{A}_i$ :  $\bar{x}_{k,it} \approx \frac{1}{n_{\mathcal{A}_i}} \sum_{s_g \in \mathcal{A}_i} \mathbf{x}_k(s_g, t)$  for  $k = 1, \dots, K$ ,  $i = 1, \dots, n_{\mathcal{A}}$ , and  $t = 1, \dots, T$ , with  $n_{\mathcal{A}_i}$  denoting the number of cells in  $\mathcal{A}_i$  (Lawson et al. 2012; Utazi et al. 2019). Estimation of  $\bar{\Phi}_{it} = |\mathcal{A}_i|^{-1} \int_{\mathcal{A}_i} \Phi(s, t) ds$  is discussed in Sect. 3.2.

Bayesian estimation of the FGG-GF model is initiated by defining the estimated parameter and hyperparameter vectors. Let  $\mathcal{L} = (\beta_0, \beta_1, \dots, \beta_K, \zeta_1, \dots, \zeta_K, \omega, v, \phi, \zeta, \delta, \Phi)$  and  $\Psi = (\sigma_{\beta_0}^2, \sigma_{\beta_1}^2, \dots, \sigma_{\beta_K}^2, \sigma_{\zeta_1}^2, \dots, \sigma_{\zeta_K}^2, \sigma_{\omega}^2, \sigma_v^2, \sigma_{\phi}^2, \sigma_{\zeta}^2, \sigma_{\delta}^2, \sigma_{\Phi}^2, \rho, \lambda_1, \lambda_2, r)'$  denote the parameter and hyperparameter vectors, respectively, of the FGG-GF model in Eq. (13). The joint posterior distribution of the FGG-GF model is:

$$p(\mathcal{L}, \Psi | \mathbf{y}) \propto p(\mathbf{y} | \mathcal{L}, \Psi) p(\mathcal{L} | \Psi) p(\Psi) \tag{14}$$

where  $p(\cdot)$  denotes the probability density function. Below, we first discuss the likelihood function  $p(\mathbf{y} | \mathcal{L}, \Psi)$  and next the joint prior of the GF at cell level. Based on the assumption that  $\mathbf{y}$  follows a Poisson distribution at area and cell levels (see Eq. (3a) and (3b)), the likelihood function  $p(\mathbf{y} | \mathcal{L}, \Psi)$  is given by:

$$\begin{aligned} p(\mathbf{y} | \mathcal{L}, \Psi) &= \prod_{t=1}^T \frac{\exp(-\mathbf{E}_t \exp(\boldsymbol{\eta}_t)) (\mathbf{E}_t \exp(\boldsymbol{\eta}_t))^{y_t}}{y_t!} \\ &= \exp\left(\log\left(\prod_{t=1}^T \frac{\exp(-\mathbf{E}_t \exp(\boldsymbol{\eta}_t)) (\mathbf{E}_t \exp(\boldsymbol{\eta}_t))^{y_t}}{y_t!}\right)\right) \\ &= \exp\left(\sum_{t=1}^T \log\left(\frac{\exp(-\mathbf{E}_t \exp(\boldsymbol{\eta}_t)) (\mathbf{E}_t \exp(\boldsymbol{\eta}_t))^{y_t}}{y_t!}\right)\right) \\ &= \exp\left(\sum_{t=1}^T (y_t (\log(\mathbf{E}_t) + \boldsymbol{\eta}_t) - \mathbf{E}_t \exp(\boldsymbol{\eta}_t) - \log(y_t!))\right). \end{aligned} \tag{15}$$

The joint prior of the GF at cell level is obtained as follows. Since the GF in Eq. (8) at cell level is assumed to follow an AR1 model, the joint prior distribution of  $\Phi = (\Phi_1, \dots, \Phi_T)'$ , i.e.,  $p(\Phi | \lambda_1, \Sigma)$ , is (Godana et al. 2019):

$$\begin{aligned} p(\Phi_1, \dots, \Phi_T | \lambda_1, \Sigma) &= p(\Phi_T | \Phi_{T-1}, \Phi_{T-2}, \dots, \Phi_1, \lambda_1, \Sigma) \\ &\times \dots \times p(\Phi_2 | \Phi_1, \Psi) \times p(\Phi_1 | \lambda_1, \Sigma). \end{aligned} \tag{16}$$

Because of the AR1 process, we have:

$$p(\Phi_T | \Phi_{T-1}, \Phi_{T-2}, \dots, \Phi_1, \lambda_1, \Sigma) = p(\Phi_T | \Phi_{T-1}, \lambda_1, \Sigma). \tag{17}$$

Thus, the joint distribution of the latent spatiotemporal Gaussian process  $\Phi$  is:

$$p(\Phi_1, \dots, \Phi_T | \lambda_1, \Sigma) = p(\Phi_1 | \lambda_1, \Sigma) \prod_{t=2}^T p(\Phi_t | \Phi_{t-1}, \lambda_1, \Sigma). \tag{18}$$

The joint distribution of the GF in Eq. (18) consists of two probability distributions:  $p(\Phi_1 | \lambda_1, \Sigma)$  and  $p(\Phi_t | \Phi_{t-1}, \lambda_1, \Sigma)$  for  $t = 2, \dots, T$ . To obtain the joint distribution of  $\Phi$ , we need the joint distributions of  $p(\Phi_1 | \lambda_1, \Sigma)$  for  $g = 1, \dots, n_p$  and  $t = 1$  and  $p(\Phi_t | \Phi_{t-1}, \lambda_1, \Sigma)$  for  $g = 1, \dots, n_p$  and  $t = 2, \dots, T$ .  $\Phi_1$  is an AR1 stationary process, i.e.,  $\Phi_1 | \lambda_1, \Sigma \sim \mathcal{N}\left(0, \frac{\Sigma}{1-\lambda_1^2}\right)$ . It is called the initial distribution for  $g = 1, \dots, n_p$  and reads as:

$$p(\Phi_1 | \lambda_1, \Sigma) = \left(\frac{1}{\sqrt{2\pi}}\right)^{n_p} \frac{1}{\left|\frac{\Sigma}{1-\lambda_1^2}\right|^{1/2}} \exp\left(-\frac{1}{2}\Phi_1' \left(\frac{\Sigma}{1-\lambda_1^2}\right)^{-1} \Phi_1\right). \tag{19}$$

Because  $\Sigma = \sigma_\Phi^2 \mathcal{R}$ , with  $\mathcal{R}$  defined in Eq. (10), we have:

$$\begin{aligned} p(\Phi_1 | \lambda_1, \Sigma) &= \left(\frac{1}{\sqrt{2\pi}}\right)^{n_p} \frac{1}{\left|\frac{\sigma_\Phi^2}{1-\lambda_1^2} \mathcal{R}\right|^{1/2}} \exp\left(-\frac{1}{2\sigma_\Phi^2} \Phi_1' \left(\frac{\mathcal{R}}{1-\lambda_1^2}\right)^{-1} \Phi_1\right) \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^{n_p} \left(\frac{\sigma_\Phi^2}{1-\lambda_1^2}\right)^{-\frac{n_p}{2}} |\mathcal{R}|^{-\frac{1}{2}} \exp\left(-\frac{(1-\lambda_1^2)}{2\sigma_\Phi^2} \Phi_1' \mathcal{R}^{-1} \Phi_1\right) \\ &\propto \exp\left(-\frac{(1-\lambda_1^2)}{2\sigma_\Phi^2} \Phi_1' \mathcal{R}^{-1} \Phi_1\right). \end{aligned} \tag{20}$$

The joint distribution  $p(\Phi_t | \Phi_{t-1}, \lambda_1, \Sigma)$  for  $g = 1, \dots, n_p$  and  $t = 2, \dots, T$  is given by:

$$\begin{aligned} &\prod_{t=2}^T p(\Phi_t | \Phi_{t-1}, \lambda_1, \Sigma) \\ &= \prod_{t=2}^T \left(\frac{1}{\sqrt{2\pi}}\right)^{n_p} \frac{1}{\left|\sigma_\Phi^2 \mathcal{R}\right|^{1/2}} \exp\left(-\frac{1}{2}((\Phi_t - \lambda_1 \Phi_{t-1})' (\sigma_\Phi^2 \mathcal{R})^{-1} (\Phi_t - \lambda_1 \Phi_{t-1}))\right) \\ &= \prod_{t=2}^T \left(\sqrt{2\pi}\right)^{-\frac{n_p}{2}} (\sigma_\Phi^2)^{-\frac{n_p}{2}} |\mathcal{R}|^{-1/2} \exp\left(-\frac{1}{2}((\Phi_t - \lambda_1 \Phi_{t-1})' (\sigma_\Phi^2 \mathcal{R})^{-1} (\Phi_t - \lambda_1 \Phi_{t-1}))\right) \\ &= \left(\sqrt{2\pi}\right)^{-\frac{n_p(T-1)}{2}} (\sigma_\Phi^2)^{-\frac{n_p(T-1)}{2}} |\mathcal{R}|^{-\frac{(T-1)}{2}} \exp\left(-\frac{1}{2\sigma_\Phi^2} \sum_{t=2}^T ((\Phi_t - \lambda_1 \Phi_{t-1})' \mathcal{R}^{-1} (\Phi_t - \lambda_1 \Phi_{t-1}))\right) \\ &\propto \exp\left(-\frac{1}{2\sigma_\Phi^2} \sum_{t=2}^T ((\Phi_t - \lambda_1 \Phi_{t-1})' \mathcal{R}^{-1} (\Phi_t - \lambda_1 \Phi_{t-1}))\right). \end{aligned} \tag{21}$$

Finally, the joint prior distribution for the AR1 process, denoted as  $p(\Phi|\mathbf{Q}_\Phi^{-1})$ , is given by multiplying Eqs. (20) and (21). It reads:

$$\begin{aligned}
 p(\Phi|\mathbf{Q}_\Phi^{-1}) &\propto \exp\left(-\frac{(1-\lambda_1^2)}{2\sigma_\Phi^2}\Phi_1'\mathcal{R}^{-1}\Phi_1\right) \\
 &\times \exp\left(-\frac{1}{2\sigma_\Phi^2}\sum_{t=2}^T((\Phi_t-\lambda_1\Phi_{t-1})'\mathcal{R}^{-1}(\Phi_t-\lambda_1\Phi_{t-1}))\right) \\
 &\propto \exp\left(-\frac{1}{2}\left(\frac{1}{\sigma_\Phi^2}\left(\Phi_1'(1-\lambda_1^2)\mathcal{R}^{-1}\Phi_1+\sum_{t=2}^T((\Phi_t-\lambda_1\Phi_{t-1})'\mathcal{R}^{-1}(\Phi_t-\lambda_1\Phi_{t-1}))\right)\right)\right) \\
 &\propto \exp\left(-\frac{1}{2}\Phi_1'\mathbf{Q}_\Phi\Phi_1\right) \propto \mathcal{N}(0,\mathbf{Q}_\Phi^{-1})
 \end{aligned}
 \tag{22}$$

with  $\mathbf{Q}_\Phi^{-1} = \Sigma$  denoting the covariance matrix of the GF in Eq. (12).

The priors and joint priors of  $\mathcal{L} = (\beta_0, \beta_1, \dots, \beta_K, \zeta_1, \dots, \zeta_K, \omega, v, \phi, \zeta, \delta, \Phi)$  and the hyperpriors of  $\Psi = (\sigma_{\beta_0}^2, \sigma_{\beta_1}^2, \dots, \sigma_{\beta_K}^2, \sigma_{\zeta_1}^2, \dots, \sigma_{\zeta_K}^2, \sigma_\omega^2, \sigma_v^2, \sigma_\phi^2, \sigma_\zeta^2, \sigma_\delta^2, \sigma_\Phi^2, \rho, \lambda_1, \lambda_2, r)'$  are presented in Table 5 in Appendix 2.<sup>12</sup> Details can be found in Jaya and Folmer (2021a) and the references therein. The prior distributions are assumed to be independent, implying that:

$$\begin{aligned}
 p(\mathcal{L}|\Psi) &= \mathcal{N}(0, \sigma_{\beta_0}^2) \times \mathcal{N}(0; \mathbf{Q}_\beta^{-1}) \times \prod_{k=1}^K \mathcal{N}(0; \mathbf{Q}_{\zeta_k}^{-1}) \times \mathcal{N}(0, \mathbf{Q}_\omega^{-1}) \times \mathcal{N}(0, \mathbf{Q}_v^{-1}) \\
 &\times \mathcal{N}(0, \mathbf{Q}_\phi^{-1}) \times \mathcal{N}(0, \mathbf{Q}_\zeta^{-1}) \times \mathcal{N}(0, \mathbf{Q}_\delta^{-1}) \times \mathcal{N}(0, \mathbf{Q}_\Phi^{-1}).
 \end{aligned}
 \tag{23}$$

The joint hyperparameter distribution is given by<sup>13</sup>:

$$\begin{aligned}
 p(\Psi) &= p(\sigma_{\beta_0}^2)p(\sigma_{\beta_1}^2) \dots p(\sigma_{\beta_K}^2)p(\sigma_{\zeta_1}^2) \dots p(\sigma_{\zeta_K}^2)p(\sigma_\omega^2) \\
 &\quad p(\sigma_v^2)p(\sigma_\phi^2)p(\sigma_\zeta^2)p(\sigma_\Phi^2)p(\sigma_\delta^2)p(\rho)p(\lambda_1)p(\lambda_2)p(\kappa) \\
 &= \prod_{j=1}^{\dim(\Psi)} p(\Psi_j) \quad \text{for } j = 1, \dots, J.
 \end{aligned}$$

Given the likelihood function, the joint prior distributions for the parameter vectors and the joint hyperparameter, the joint posterior distribution in Eq. (14) can be written as:

<sup>12</sup> A joint prior may refer to multiple parameters, to the spatial units (area and cell), to multiple time periods, or to space–time. The specification of the parameter vector or the subscripts of the precision matrices indicate what kind of jointness is in order.

<sup>13</sup> The total number of hyperparameters is  $J = 2K + 11$ , where  $K$  is the total number of risk factors and 11 represents the total of the elements of  $\{\sigma_{\beta_0}^2, \sigma_\omega^2, \sigma_v^2, \sigma_\phi^2, \sigma_\zeta^2, \sigma_\Phi^2, \sigma_\delta^2, \rho, \lambda_1, \lambda_2, r\}$ .

$$\begin{aligned}
 p(\mathcal{E}, \Psi | \mathbf{y}) \propto & \exp\left(\sum_{t=1}^T (y_t \log(E_t \exp(\boldsymbol{\eta}_t)) - E_t \exp(\boldsymbol{\eta}_t))\right) \\
 & \times \exp\left(-\frac{1}{2\sigma_{\beta_0}^2} \beta_0^2\right) \times \exp\left(-\frac{1}{2} \boldsymbol{\beta}' \mathbf{Q}_{\beta} \boldsymbol{\beta}\right) \times \exp\left(-\frac{1}{2} \sum_{k=1}^K \boldsymbol{\zeta}'_k \mathbf{Q}_{\zeta_k} \boldsymbol{\zeta}_k\right) \\
 & \times \exp\left(-\frac{1}{2} \boldsymbol{\omega}' \mathbf{Q}_{\omega} \boldsymbol{\omega}\right) \times \exp\left(-\frac{1}{2} \mathbf{v}' \mathbf{Q}_v \mathbf{v}\right) \times \exp\left(-\frac{1}{2} \boldsymbol{\phi}' \mathbf{Q}_{\phi} \boldsymbol{\phi}\right) \times \exp\left(-\frac{1}{2} \boldsymbol{\zeta}' \mathbf{Q}_{\zeta} \boldsymbol{\zeta}\right) \\
 & \times \exp\left(-\frac{1}{2} \boldsymbol{\Phi}' \mathbf{Q}_{\Phi} \boldsymbol{\Phi}\right) \times \prod_{j=1}^{\dim(\Psi)} p(\Psi_j).
 \end{aligned}
 \tag{24}$$

### 3.2 The fusion area-cell spatiotemporal generalized geoadditive-Gaussian Markov random field model

A continuously indexed GF typically has a dense covariance matrix, such as  $\mathbf{Q}_{\Phi}$  in Eq. (24), leading to complex, time-consuming numerical estimation challenges, commonly referred to as the “big  $n$ ” problem. Lindgren et al. (2011) proposed to solve the big  $n$  problem by substituting a sparse, discretely indexed Gaussian Markov Random Field (GMRF) for the continuously indexed GF.<sup>14</sup> For a GMRF, the full conditional distribution for each component  $\gamma_{g,t}$  for  $g = 1, \dots, n_p$ , only depends on a set of neighbors  $N(g)$  as follows:

$$p(\gamma_{g,t} | \boldsymbol{\gamma}_{-g,t}) = p(\gamma_{g,t} | \boldsymbol{\gamma}_{N(g),t}) \quad \text{for } g = 1, \dots, n_p \text{ and } t = 1, \dots, T \tag{25}$$

where  $\boldsymbol{\gamma}_{-g,t}$  denotes all the elements in  $\boldsymbol{\gamma}$  except  $\gamma_{g,t}$ , and  $\boldsymbol{\gamma}_{N(g),t}$  denotes all the elements of  $\boldsymbol{\gamma}_t$  in the neighborhood  $N(g)$  of  $\gamma_{g,t}$ . The vector of elements of  $\boldsymbol{\gamma}_t$  not in the neighborhood  $N(g)$  of  $\gamma_{g,t}$  is denoted as  $\boldsymbol{\gamma}_{-\{g,N(g)\},t}$ .  $\gamma_{g,t}$  is conditionally independent of the elements of  $\boldsymbol{\gamma}_{-\{g,N(g)\},t}$ . The conditional independence relationship is written as:

$$\gamma_{g,t} \perp \boldsymbol{\gamma}_{-\{g,N(g)\},t} | \boldsymbol{\gamma}_{N(g),t} \quad \text{for } t = 1, \dots, T \text{ and } g = 1, \dots, n_p. \tag{26}$$

If Eq. (26) holds, the precision matrix  $\mathbf{Q} = \boldsymbol{\Sigma}^{-1}$  of  $\boldsymbol{\gamma}_t$  is sparse for each  $t$ . In other words, for a pair  $g$  and  $h$  with  $g \neq h$  and  $h \notin \{N(g)\}$ , we have:

$$\gamma_{g,t} \perp \gamma_{h,t} | \boldsymbol{\gamma}_{-(g,h),t} \iff \mathbf{Q}(g, h) = 0 \quad \text{for } g = 1, \dots, n_p \text{ and } t = 1, \dots, T \tag{27}$$

implying that the nonzero pattern in the precision matrix  $\mathbf{Q}$  is given by the neighborhood structure. Conversely,

$$\mathbf{Q}(g, h) \neq 0, \text{ if } h \in \{N(g)\}. \tag{28}$$

<sup>14</sup> A sparse precision matrix enables the application of computationally efficient numerical methods. For instance, for the factorization of a dense covariance matrix of a spatiotemporal GF, the  $O(n^3)$  computation time is reduced to  $O(n^2)$  for a sparse covariance matrix of a GMRF (Rue and Held 2005).

Lindgren et al. (2011) proposed the Linear Stochastic Partial Differential Equation (LSPDE)<sup>15</sup> approach based on a mesh of the study area, to transform a dense Matérn covariance matrix of a GF, such as in Eq. (11), into a sparse Matérn precision matrix of a GMRF (see Appendix 1). Specifically, for  $t = 1, \dots, T$ , the GF  $\gamma_{g,t}$  in Eq. (12) with Matérn covariance function  $\Sigma = \sigma_{\Phi}^2 \mathcal{R}$ , is transformed into a GMRF,  $\tilde{\gamma}_t(s) \sim \mathcal{N}\left(0, \tilde{\mathbf{Q}}_s^{-1}\right)$ , with sparse spatial precision matrix  $\tilde{\mathbf{Q}}_s$  defined in Eq. (47). Consequently, for  $t = 1, \dots, T$ , the joint latent spatiotemporal GF  $\Phi_t = \left(\Phi_{1t}, \dots, \Phi_{n_p t}\right)'$  at cell level in Eq. (12) is transformed into a GMRF  $\tilde{\Phi}_t$  as:

$$\tilde{\Phi}_t = \lambda_1 \tilde{\Phi}_{t-1} + \tilde{\gamma}_t \text{ and } \tilde{\gamma}_t(s) \sim \mathcal{N}\left(0, \tilde{\mathbf{Q}}_s^{-1}\right) \tag{29}$$

with the initial value distributed as:  $\tilde{\Phi}_1 \sim \mathcal{N}\left(0, \tilde{\mathbf{Q}}_s^{-1} / (1 - \lambda_1^2)\right)$ . The joint distribution of the  $T \times L$ -dimensional cell level GMRF  $\tilde{\Phi} = \left(\tilde{\Phi}'_1, \dots, \tilde{\Phi}'_T\right)'$  is:

$$\tilde{\Phi} \sim \mathcal{N}\left(0, \tilde{\mathbf{Q}}_{\tilde{\Phi}}^{-1}\right) \tag{30}$$

with precision matrix  $\mathbf{Q}_{\tilde{\Phi}} = \mathbf{Q}_T \otimes \tilde{\mathbf{Q}}_s$ , i.e., the Kronecker product of the autoregressive temporal covariance matrix ( $\mathbf{Q}_T$ ) (see Table 5 in Appendix 2) and the Matérn spatial covariance matrix ( $\mathbf{Q}_s$ ), respectively.

To facilitate the estimation of  $\tilde{\Phi}_t = \left(\bar{\Phi}_t, \Phi_t\right)'$ , with  $\bar{\Phi}_t = \left(\bar{\Phi}_{1t}, \dots, \bar{\Phi}_{i t}, \dots, \bar{\Phi}_{n_A t}\right)'$  and  $\Phi_t = \left(\Phi_{1t}, \dots, \Phi_{g t}, \dots, \Phi_{n_p t}\right)'$   $\tilde{\Phi}_t = \left(\Phi_{1t}, \dots, \Phi_{g t}, \dots, \Phi_{n_p t}\right)'$  for  $t = 1, \dots, T$ , (Eq. (13)) as a GMRF  $\tilde{\Phi}$ , we introduce the  $((n_A + n_p) \times L)$ -dimensional partitioned or block matrix  $\mathbf{H} = \left[\mathbf{H}_1 \mathbf{H}_2\right]^{-1}$  that maps the GMFRs associated with the  $L$  triangulation nodes<sup>16</sup> to the  $n_A$  areas and  $n_p$  cells, respectively. The elements of  $\mathbf{H}_1$  correspond to the block average  $|\mathcal{A}_i|^{-1} \int_{\mathcal{A}_i} \Phi(s, t) ds$  for  $i = 1, \dots, n_A, t = 1, \dots, T$  and  $= (s_{1(\mathcal{A}_1)}, \dots, s_{g(\mathcal{A}_i)}, \dots, s_{n_p(\mathcal{A}_{n_A})})'$ . That is,  $\mathbf{H}_1$  is the  $(n_A \times L)$  sparse matrix with  $\mathbf{H}_1(i, l) = 1/V_i$  if vertex  $l$  is in area  $\mathcal{A}_i$  and zero otherwise and  $V_i$  is the number of vertices in the area  $\mathcal{A}_i$ . Hence, matrix  $\mathbf{H}_1$  reads:

<sup>15</sup> The LSPDE approach uses finite basis functions defined by a triangulation of the study region (see Online Resource 1 for application to Bandung city) to transform a GF to a GMRF. Implementation can be conveniently handled using INLA (Rue et al. 2009). The LSPDE approach is immediately applicable to a large class of spatiotemporal models. For instance, it applies to models with complex hierarchical structures or with non-separable covariance functions, as well as non-stationary models with time-varying parameters (Cameletti et al. 2013).

<sup>16</sup> The number of nodes (vertices)  $L$  is larger than the number of cells  $n_p$  because we extend the domain of interest by an outer area to avoid the boundary effect (see Appendix 1 for details).

$$\mathbf{H}_1 = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & \dots & L-u & \dots & L \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ \vdots \\ n_A \end{matrix} & \left( \begin{array}{ccccccc} 1/V_1 & 0 & 0 & \dots & 0 & \dots & 0 \\ 0 & 1/V_2 & 0 & \dots & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 1/V_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 0 & \dots & 1/V_{n_A} \end{array} \right) \end{matrix} \tag{31}$$

Consequently,  $\bar{\Phi}_{it} \approx \sum_{l=1}^L \mathbf{H}_1(i, l) \tilde{\Phi}_{t,l}$  for  $t = 1, \dots, T$  with  $\tilde{\Phi}_{t,l}$  the  $(t, l)$  th element of  $\tilde{\Phi}$ .

$\mathbf{H}_2$  transforms the  $(T \times L)$  elements of  $\tilde{\Phi}$  into  $(n_p \times T)$  elements of  $\Phi$  with the value of the  $(g, t)$  th element  $\Phi$  corresponding to the value of  $(t, l)$  th of  $\tilde{\Phi}$ . That is,  $\mathbf{H}_2$  is an  $(n_p \times L)$  sparse matrix with  $\mathbf{H}_2(g, l) = 1$  if the vertex  $l$  is at location  $s_g$  and zero elsewhere such that for the  $g$ th cell  $\Phi_{g,t} \approx \sum_{l=1}^L \mathbf{H}_2(g, l) \tilde{\Phi}_{t,l}$  for  $i = 1, \dots, n_p$  and  $t = 1, \dots, T$ . Hence, matrix  $\mathbf{H}_2$  reads:

$$\mathbf{H}_2 = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & \dots & L-u & \dots & L \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ \vdots \\ n_p \end{matrix} & \left( \begin{array}{ccccccc} 1 & 0 & 0 & \dots & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 0 & \dots & 1 \end{array} \right) \end{matrix} \tag{32}$$

Given the partitioned matrix  $\mathbf{H}$ , the FGG-GF in Eq. (13) can be written as:

$$\eta_t = \beta_0 1_{(n_A+n_p)} + \sum_{k=1}^K (\beta_k + \zeta_{k,t}) z_{k,t} + \ddot{\omega} + \ddot{v} + \phi_t 1_{(n_A+n_p)} + \zeta_t 1_{(n_A+n_p)} + \ddot{\delta}_t + \mathbf{H} \tilde{\Phi}_t \tag{33}$$

for  $t = 1, \dots, T$ .

Because the FGG-GMRF model in Eq. (33) belongs to the class of the latent Gaussian models, it can be estimated using INLA-LSPDE (Cameletti et al. 2013; Gómez-Rubio et al. 2021). Predictions of the relative risk  $\theta$  for the cells of the triangulated domain can be obtained via the posterior conditional distribution of  $\tilde{\Phi}$ , given  $\mathbf{Q}_{\tilde{\Phi}}$  for all the  $L$  vertices and the posterior distributions of the parameter and hyperparameters in Eq. (24). The FGG-GMRF model setup in Eq. (33) implies that INLA generates predictions for the target cells during model-fitting.

#### 4 Application: relative dengue risk at subdistrict level in Bandung, 2012–2018

Bandung city is divided into 30 districts and 151 subdistricts. The districts are third level administrative units within a province, and the subdistricts are fourth level administrative units. Every district in Bandung city consists of a minimum of four

subdistricts. While the number of dengue incidences in Bandung city is reported at district level, for efficient and effective prevention and control, figures at the subdistrict scale are needed. In Sect. 4.1, we discuss and explore the data; in Sect. 4.2, we estimate the FGG-GMRF model; and in Sect. 4.3, we use the model to predict the relative dengue risk at subdistrict level.

#### 4.1 Data and exploratory data analysis

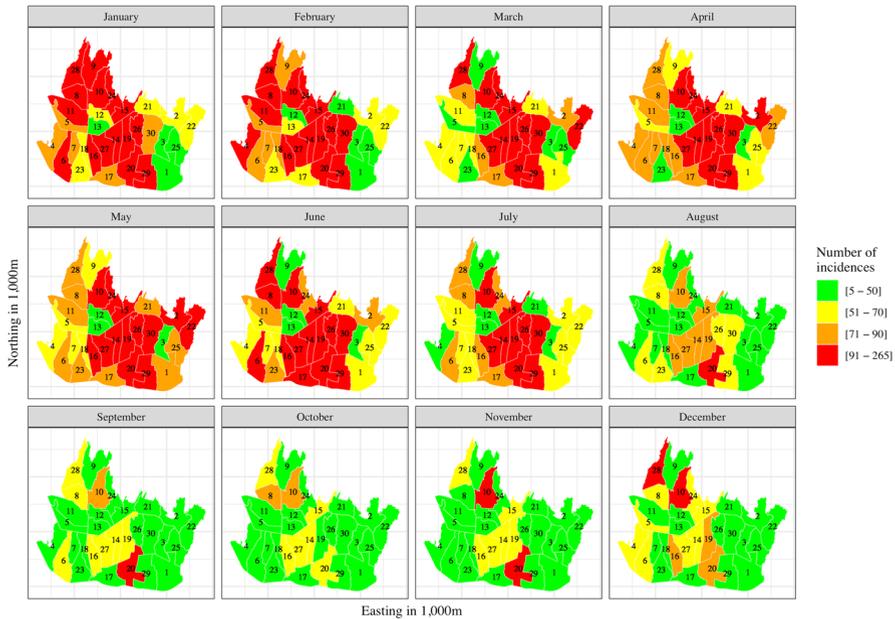
The data were obtained from existing databases. Annual observations at district level on the population at risk (see Online Resource 1) and monthly dengue incidence at district level (see Fig. 1) were obtained from the Bandung Central Statistical Bureau (2012, 2013, 2014, 2015, 2016, 2017, 2018) and the Bandung Health Department (2013, 2014, 2015, 2016, 2017, 2018, 2019), respectively. From January 1, 2012, until December 31, 2018, a total of 26,095 dengue incidences (1,030 per 100,000 inhabitants) were reported. The monthly incidence pattern is highly similar from year to year and is taken as constant. Particularly, the mean Pearson correlation coefficient of monthly dengue incidence for the years 2012–2018 is approximately 0.70. In addition, there were no major shifts across the years in the annual cycle. Hence, as in Jaya and Folmer (2021a), we only consider the monthly cycle for the 30 districts.<sup>17</sup> Figure 1 shows a high number of incidences from January to July, followed by a sharp drop in July and a low level of incidences for the remainder of the year. The monthly incidences range from 5 to 265 (0.197–10.461 per 100,000 inhabitants, respectively).

We derived the crude risk rate (i.e., the standardized incidence ratio, SIR) as the ratio of the observed to the expected number of incidences (see Eq. (2)). Figure 2 presents the monthly dengue SIR per district for the period 2012–2018. It ranges from 0.229 to 3.132. Most districts, primarily those in northern and southern Bandung, have a SIR greater than one from January to July. The districts with the highest SIR are Buah Batu (id = 20), Lengkong (id = 27) and Rancasari (id = 29).

As observed by Ebi and Nealon (2016), Jaya and Folmer (2021a), and Zellweger et al. (2017), among others, socioeconomic and environmental conditions are the main factors influencing dengue disease risk over space and time. However, for Bandung, socioeconomic risk variables such as income, education, occupation and living conditions are unavailable for districts and cells. These factors are accounted for by the random effects, thus controlling for omitted variable bias (Jaya and Folmer 2020, 2021a, b). By contrast, the monthly averages of the weather risk variables of precipitation (mm), temperature (°C), sunshine duration (kJ/m<sup>2</sup>day) and water vapor pressure (kPa) are available at cell level from the WorldClim2.0 database (Fick and Hijmans 2017), obtained from 19 weather stations surrounding Bandung in West Java for 1970–2000. We selected cells of resolution 1 km<sup>2</sup>. Accordingly, Bandung city was divided into 179 cells. Table 1 presents the monthly average weather variables for the period 1970–2000.

Figure 3 shows that precipitation and water vapor pressure are relatively high in the period November–April, temperature is relatively high in the period April–June

<sup>17</sup> The main objective of this research is to find out in which areas and which months Dengue outbreaks will occur annually. For this purpose monthly cycle observations suffice.



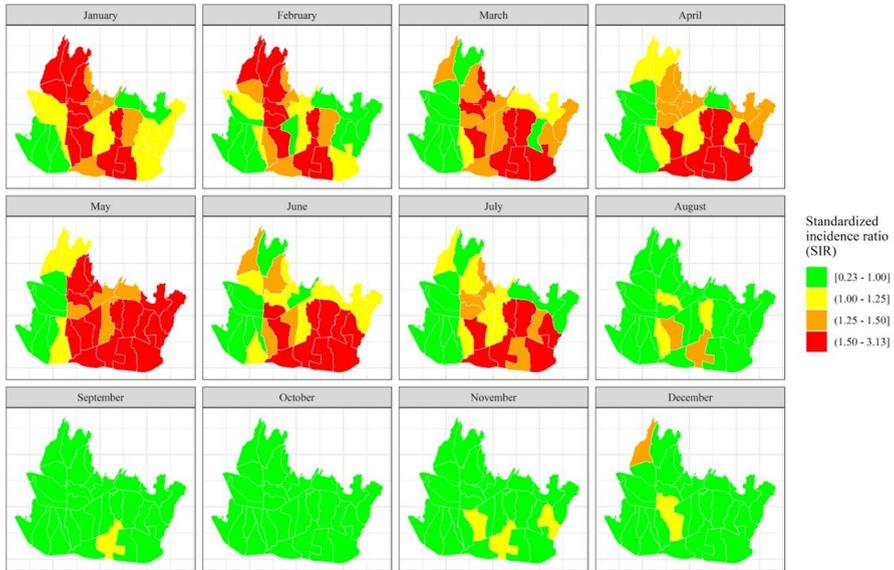
**Fig. 1** Monthly dengue incidences at district level (in 1,000), Bandung City, Indonesia, 2012–2018 (1–30: district codes, see Online Resource 1)

and in October, and solar radiation is relatively high in the period August–November. Figure 4 indicates that the average temperature varies strongly over space. The minimum average temperature occurs in the northern districts, which are mountainous areas at approximately 800 m above sea level. In addition, they are densely covered with forests and have relatively high precipitation. The central districts, where the governmental facilities and businesses are located, also have high precipitation in the period November to April. Moreover, they have higher temperatures than northern Bandung because of differences in forest density and elevation. They also have high population density, high mobility and high air pollution (Jaya and Folmer 2020).

In preparation for estimating the FGG-GMRF model in Eq. (33), we calculated the variance inflation factor (VIF) of the weather variables to check multicollinearity. Table 2 shows that the maximum VIF is 9.312 (for water vapor pressure) which is below the critical (rule of thumb) value of 10, indicating that the correlation among variables is unlikely to affect estimation (Montgomery et al. 2012).

## 4.2 The estimated generalized Geoadditive-Gaussian Markov Random Field model and prediction

The first step in estimating the FGG-GMRF model given by Eq. (33) is the construction of a triangle mesh of the study area for the application of the Finite Element Method (FEM) and LSPDE approach.



**Fig. 2** Monthly dengue standardized incidence ratio (SIR) at district level, Bandung City, Indonesia, 2012–2018

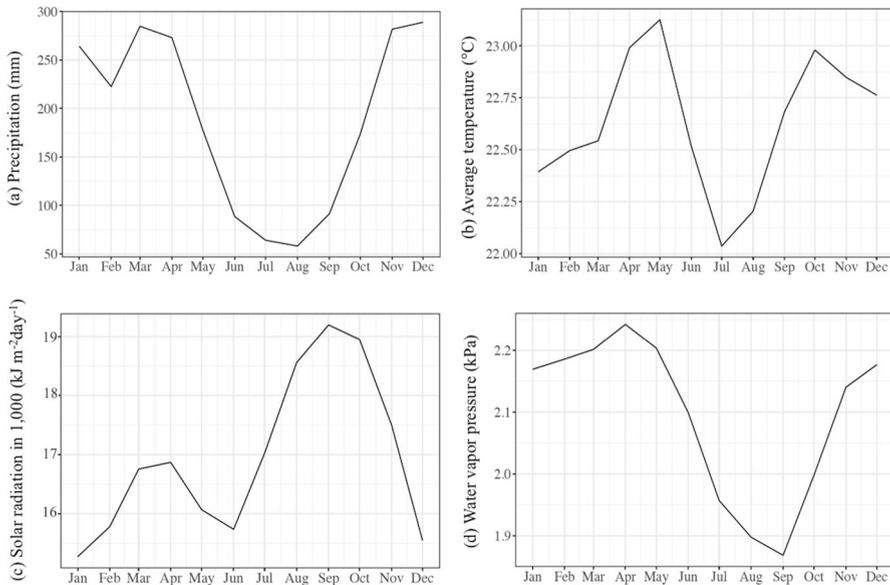
**Table 1** Descriptive statistics for the monthly averages of the weather variables<sup>a</sup>

Description	Mean	SD	Min	Max
Precipitation (mm)	189.024	89.030	52.000	328.000
Average temperature (°C)	22.632	0.896	19.300	24.000
Solar radiation (kJ/m <sup>2</sup> day)	16,938.484	1,317.538	15,041.000	19,924.000
Water vapor pressure (kPa)	2.095	0.140	1.670	2.300

<sup>a</sup>WorldClim (2020) Global climate and weather data, version 2.1. WorldClim: <https://www.worldclim.org/>. Accessed May 2020

As described in Appendix 1, the accuracy of the FEM calculations and the precision of the forecasts, is a function of the number of vertices in the mesh (edge length). Blangiardo and Cameletti (2015) and Utazi et al. (2019) recommended varying the edge length between the minimum distance and approximately 5–8% of the maximum distance of any two cells (18,681 m). Hence, we considered  $G = \{966, 831, 717, 616, 548, 501\}$  vertices, corresponding to edge lengths varying from 1000 to 1500 m, with a difference<sup>18</sup> of 100 m (see Online Resource 1). The data and R code are available in Online Resource 2.

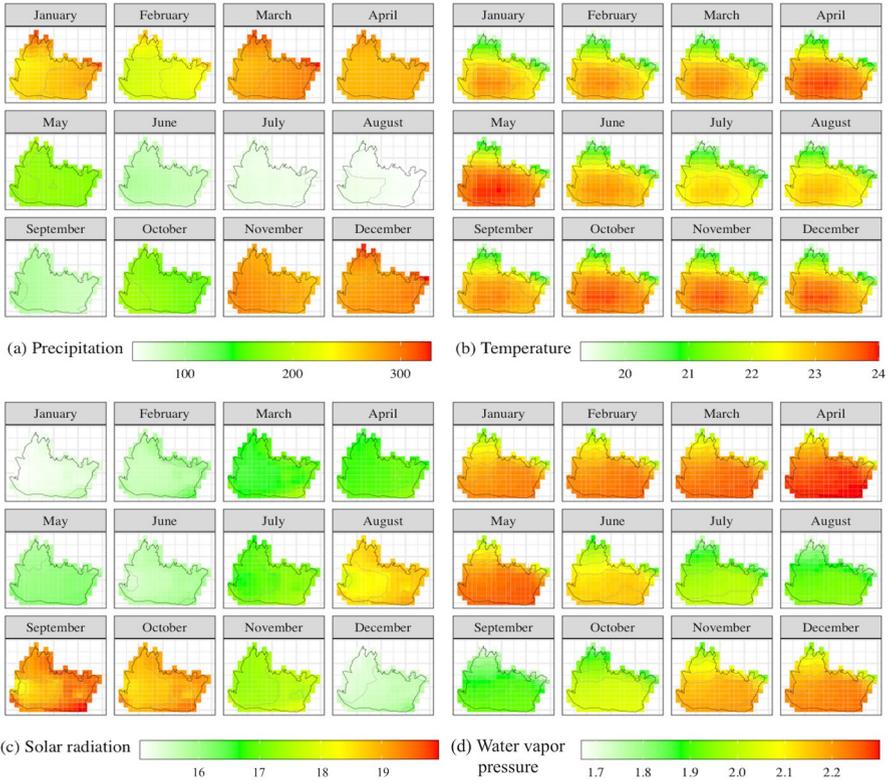
<sup>18</sup> R-INLA contains the routine `inla.mesh.2d()` for meshing of spatial domains. It contains several sub-routines, notably `loc`: for the number and the location of the initial mesh vertices, and `max.edge`: for the maximum edge lengths in the study area and in the outer area.



**Fig. 3** Monthly variation of the mean annual weather variables **a** precipitation (mm), **b** temperature (°C), **c** solar radiation in 1000 (kJ/m<sup>2</sup>day), and **d** water vapor pressure (kPa)

Before turning to the estimations of the spatiotemporal FGG-GMRF model, we make the following observations. First, as explained in Sect. 3, the covariates and/or the state process  $\Phi$  at cell level are required for high-resolution spatiotemporal prediction using the FGG-GMRF model. Hence, either the covariates or  $\Phi$ , or both, are included in the selected model. Second, for every model, we considered Poisson and Negative Binomial model specifications for the number of incidences, a random walk of order one (RW1) and two (RW2) for the time-varying coefficients, structured and unstructured spatial and temporal random effects and their interaction, and six different edge lengths. Third, the best model was selected using the deviance information criterion (DIC), the Watanabe–Akaike information criterion (WAIC) and the marginal predictive likelihood (MPL). As a rule of thumb, the best model is the one with the smallest DIC and WAIC, and the largest MPL. Fourth, we started the estimation with the simplest models with covariates only (M1), and then, we proceeded to the model specifications with covariates and four types of interaction (see Table 5 in Appendix 2) at area and cell levels (M2) and, finally, we estimated the full models with covariates, four types of interaction at area and cell levels, and spatially and temporally structured and unstructured main effects (M3).<sup>19</sup> Finally,

<sup>19</sup> In Jaya and Folmer (2021a), the order of estimation was reverse. The models with the structurally and temporally main random effects were estimated before the models with the interaction effects. The latter outperformed the former. Based on this outcome, we estimated the M2 models before the M3 models to economize on computation time.



**Fig. 4** The spatiotemporal variation of the weather variables: **a** precipitation (mm), **b** temperature (°C), **c** solar radiation in 1000 (kJ/m<sup>2</sup>day), and **d** water vapor pressure (kPa)

**Table 2** Variance inflation factors (VIF) for precipitation, average temperature, solar radiation and water vapor pressure

No	Variable	VIF
1	Precipitation (mm)	3.074
2	Average temperature (°C)	3.234
3	Solar radiation (kJ m <sup>-2</sup> day <sup>-1</sup> )	3.347
4	Water vapor pressure (kPa)	9.312

VIF < 10 indicates that there is no serious multicollinearity

due to the large number of outcomes, we only present the estimates for interaction Type IV (spatially structured  $\otimes$  temporally structured) which, as in Jaya and Folmer (2020), performed best among the models with spatiotemporal interaction.<sup>20</sup>

<sup>20</sup> The estimates of the models with other types of interaction are available from the first author upon request.

The estimations are presented in Table 6 (in Appendix 3). The table shows that the models<sup>21</sup> with covariates only (M1) have the worst fit and predictive performance among the three classes of models. They have the highest DICs, WAICs and smallest MPLs. Given their relatively poor fit and predictive performance, the M1 models were not considered further in the selection procedure. Introduction of the interaction effect type IV (M2) yielded substantially better predictive performance. The full models with covariates, interaction at area and cell levels and spatially and temporally structured and unstructured main effects (M3) had fit and predictive performance similar to the M2 models. Hence, the main spatially and temporally structured and unstructured effects did not improve the model fit and prediction performance, which is consistent with Jaya and Folmer (2020). Based on these observations, and because the M2 models have a simpler structure, we selected the class of M2 models.

Next, we turned to the selection of the best model from the 24 models M2.1.1.1–M2.2.2.6. First, we considered the edge length, finding that the models M2.1.1.1–M2.2.2.6. have similar DIC, WAIC and MPL values. Based on this observation, we selected the edge length of 1500 m for reasons of computational time. Online Resource 1 presents the mesh. Second, among the M2 models with edge length 1500 m, the Poisson model had slightly lower DIC and WAIC and slightly higher MPL than the Negative Binomial model. In addition, the models with temporal trends RW1 and RW2 had similar DIC, WAIC and MPL values. Based on these considerations, we selected model M2 with Poisson distribution, RW1 time-varying effect, and edge length of 1500 m (denoted as model M2.1.1.6 below).

Figure 5a shows that for model M2.1.1.6, the observed and predicted dengue relative risks are strongly correlated (Pearson correlation coefficient = 0.986), indicating that the model fits the data well. Figure 5b shows that the PIT histogram is close to the Uniform distribution, also indicating that model M2.1.1.6 fits the data well.

Table 3 summarizes various components of model M2.1.1.6 which are subsequently used to calculate the posterior means of the monthly relative risk at district and subdistrict levels. Before doing so, we discuss the components separately. To this end, we also make use of Figs. 6 and 7. Before going into detail, we make the following remarks. First, as observed in Sect. 2, p.8, the posterior means of the time-varying coefficients  $\beta_{k,t} = \beta_k + \zeta_{k,t}$  for  $k = 1, \dots, K$  and  $t = 1, \dots, T$ , which are presented in Fig. 6, consist of the fixed effect plus the temporal random effect. The minimum and maximum posterior means of the temporal random effects present the largest negative and largest positive differences of the temporal random effects relative to the global effects. Secondly, the contributions of the weather variables in explaining the spatiotemporal variation of the dengue risk are conveniently summarized by the posterior means of their hyperparameter variances and their percentage contributions as fractions of the total variance (the last column of Table 3). In a similar vein, the posterior means and fractions of the hyperparameter variance of the random components present their variability and strength in explaining the relative risk across space and time.

<sup>21</sup> For a precise definition of the models see the note under Table 6, Appendix 3.

We will begin the discussion of Table 3 with the global mean effects of the weather variables. We only consider the posterior means and disregard credible intervals. Before going into the detail, it is worth noting that an indirect relationship exists between the relative risk of dengue and weather variables via the development and survival of the dengue virus and its vector (Jaya and Folmer 2021a).

Precipitation in general has a negative impact, with a posterior global (overall) mean of  $-0.0041$ , which is consistent with Jaya and Folmer (2021a). For a one mm increase in the global mean, the dengue risk decreases by  $(\exp(-0.0041) - 1)100\% = -0.41\%$ . The explanation is that heavy rainfall disrupts the *Aedes-spp* mosquito's reproductive cycle by washing away breeding sites (Abiodun et al. 2016; Benedum et al. 2018).

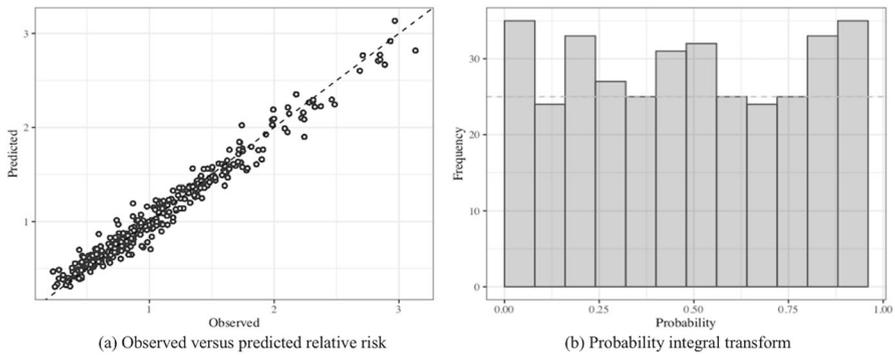
Temperature in general has a positive impact (0.1002), which is consistent with Hurtado-Díaz et al. (2007) and Jaya and Folmer (2021a). The relative risk increases by  $(\exp(0.1002) - 1)100\% = 10.54\%$  for an increase of global mean temperature by  $1^{\circ}\text{C}$ . The explanation is that higher temperatures offer good conditions for mosquito development, particularly feeding (Hales et al. 2002; Lambrechts et al. 2012).

Solar radiation has a negative impact ( $-0.0001$ ), which is consistent with Ekasari et al. (2018), Jaya and Folmer (2021a) and Martínez-Bello et al. (2017b). An increase of  $1 \text{ kJ m}^{-2} \text{ day}^{-1}$  of solar radiation decreases the relative dengue risk by  $(\exp(-0.0001) - 1)100\% = -0.01\%$ . As shown by Rasjid et al. (2019), strong solar radiation negatively influences the breeding and spread of *Aedes-spp* mosquitoes. A longer spell of solar radiation implies a shortened spell of dawn and dusk, during which the *Aedes-spp* mosquito preys on animals and humans, particularly 20 to 30 min after sunset (Ekasari et al. 2018; Jaya and Folmer 2021a).

Water vapor pressure in general has a negative impact, with a posterior global mean of 0.5033. An increase of the global mean by 1% increases the relative dengue risk by  $(\exp(0.5033) - 1)100\% = 65.42\%$  due to an increase in breeding ability (Bambrick et al. 2009).

For the temporal random effects of the weather variables, the greatest negative temporal random effect of precipitation occurred in June ( $-0.0248$ ), while the greatest positive occurred in March (0.0176). For temperature, the largest negative temporal random effect was in February ( $-0.0713$ ), while the largest positive was in June (0.1104). Solar radiation has a temporal random effect close to zero, as indicated by its minimum and maximum posterior means of  $-0.0002$  and  $0.0003$ , respectively. The greatest negative of temporal random effect of water vapor pressure was in January ( $-0.0137$ ), while the greatest positive effect was in June (0.0161).

The variances of the weather variables together account for 31.87% of the total variance of the hyperpriors with solar radiation being the most important ( $\sigma_{\beta_3}^2 = 0.023$ ), explaining 9.88%, while temperature is the smallest, accounting for 6.2%. The variance of the area level interaction effect ( $\sigma_{\delta}^2$ ) accounts for the highest fraction of the total variance (43.47%), followed by the cell level interaction effect ( $\sigma_{\phi}^2$ ), with a fraction of 24.68%. These fractions indicate that the trend of dengue relative risk for each district and subdistrict is strongly affected by neighboring districts and subdistricts, respectively. The relatively low fractions of the total variance for the other effects imply that they are less important. Specifically, the low fraction



**Fig. 5** a Scatterplot for model M2.1.1.6 of the predicted versus the observed dengue relative risk and **b** histogram of the probability integral transform (PIT)

of the total variance for the average temperature indicates that only a small part of the variability of the relative risk of dengue in districts and subdistricts is explained by the average temperature.

The posterior mean of hypermeter of the Leroux CAR spatial autoregressive coefficient ( $\rho$ ), and the posterior means of the hyperparameters of the temporal autoregressive coefficients at area level ( $\lambda_2$ ) and cell level ( $\lambda_1$ ) are substantial (larger than 0.700), indicating strong spatial and temporal dependency. The estimated range is  $r$  equals 13.597 km. This implies that beyond the distance  $r = 13.597$  the spatial correlation among any two cells is smaller than 0.1. Hence, the observations are spatially strongly correlated. Beyond 13.597km it is negligible.

Based on Table 3, we now discuss the estimated parameters for the time-varying effects of the risk factors and, next, the spatiotemporal interaction effects at the area and cell levels.

The posterior means of time-varying effects of the weather variables are presented in Fig. 6. The figure shows that the time-varying effects of precipitation, average temperature and solar radiation vary considerably over the year. The time-varying effect of water vapor pressure, in contrast, is highly constant over time.

The time-varying effect of precipitation is positive for the periods January–April, August–September and December and negative for May–July and October–November. The strongest negative effect was in June ( $-0.0289$ ), and the strongest positive effect was in March ( $0.0135$ ). The negative impact for the period May–July follows after the peak of the rainy season from November–April. The negative impact for October–November is caused by the increase in precipitation after the peak of the dry season in June–July. The positive impacts for the periods January–April and August–September, and the peak in March, correspond to the relatively low rainfall one to two months before these periods.

The time-varying effect of temperature is positive for all months and ranges from 0.0289 (end of August) to 0.2106 (June–July). It increases from January–June, is at its peak in June, decreases from July until mid-September, and then starts increasing up to the global mean, where it remains for the rest of the year. Note that the monthly temperature has a delayed risk effect in that it increases from January–May,

**Table 3** The posterior means and 95% credible intervals of the fixed effects, the minimum and maximum posterior means and 95% credible intervals of the temporal random effects and the posterior means and 95% credible intervals of the hyperparameters of the FGG-GMRF model M2.1.1.6

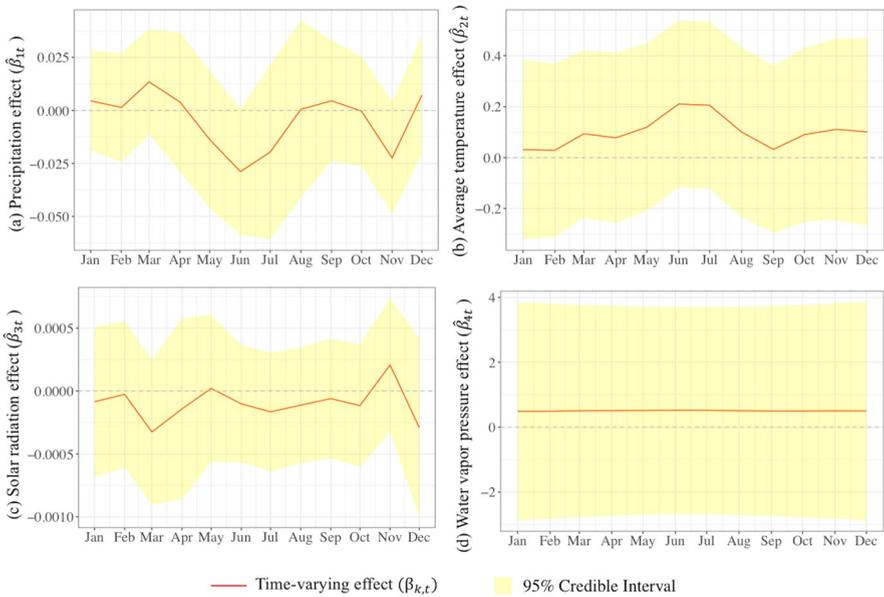
Parameter	Fixed effects /Global effects ( $\beta_k$ )		Random effects		Posterior means of the hyperparameters (95% credible interval)	Fraction of the total variance (%)
	Posterior means of the regression coefficients (95% credible interval) <sup>a</sup>	Percent-age change (%) <sup>b</sup>	Temporal random effects ( $\zeta_{k,t}$ )			
			Minimum posterior means (95% credible interval)	Maximum posterior means (95% credible interval)		
<b>Risk factors</b>						
Precipitation ( $x_1$ )	-0.0041 (-0.0146; 0.0064)	-0.41	-0.0248 (-0.0440; -0.0056)	0.0176 (0.0032; 0.0319)	0.019 (0.0191; 0.0191)	8.17
Temperature ( $x_2$ )	0.1002 (-0.1222; 0.3225)	10.54	-0.0713 (-0.1890; 0.0462)	0.1104 (0.0054; 0.2153)	0.015 (0.0145; 0.0145)	6.20
Solar radiation ( $x_3$ )	-0.0001 (-0.0004; 0.0002)	-0.01	-0.0002 (-0.0005; 0.0000)	0.0003 (0.0001; 0.0005)	0.023 (0.0231; 0.0231)	9.88
Water vapor pressure ( $x_4$ )	0.5033 (-2.4868; 3.4908)	65.42	-0.0137 (-0.3904; 0.3627)	0.0161 (-0.1877; 0.2197)	0.018 (0.0178; 0.0178)	7.62
<b>Interaction effect at area level (<math>\delta</math>)</b>						
Variance of interaction effect ( $\sigma_\delta^2$ )					0.102 (0.1018; 0.1018)	43.47
<b>Leroux CAR spatial autoregressive coefficient (<math>\rho</math>)</b>					0.815 (0.8147; 0.8147)	
<b>Temporal autoregressive coefficient (<math>\lambda_2</math>)</b>					0.962 (0.9619; 0.9619)	
<b>Interaction effect at cell level (<math>\Phi</math>)</b>						
Variance of Interaction effect at cell level ( $\sigma_\Phi^2$ )					0.058 (0.0578; 0.0578)	24.68

**Table 3** (continued)

Parameter	Fixed effects /Global effects ( $\beta_k$ )		Random effects		Fraction of the total variance (%)
	Posterior means of the regression coefficients (95% credible interval) <sup>a</sup>	Percent-age change (%) <sup>b</sup>	Temporal random effects ( $\zeta_{k,t}$ )	Posterior means of the hyperparameters (95% credible interval)	
			Minimum posterior means (95% credible interval)	Maximum posterior means (95% credible interval)	
Spatial range ( $r$ ) in km					13.597 (13.597; 13.598)
Temporal autoregressive coefficient at cell level ( $\lambda_1$ )					0.749 (0.7485; 0.7485)

<sup>a</sup>The 95% credible intervals are obtained from the posterior quantiles, i.e.,  $q(0.025)$  and  $q(0.975)$ . Using INLA, the posterior marginal distributions of the parameters approximately follow a normal distribution due to the Laplace approximation (Blangiardo and Cameletti 2015)

<sup>b</sup>The exact percentage change to  $\theta_{it}$  for a one-unit change in a risk factor with all the other variables in the model held constant, is  $100(e^{\theta} - 1)$



**Fig. 6** Time-varying effect ( $\hat{\beta}_{k,t}$ ) of **a** precipitation (mm), **b** average temperature ( $^{\circ}\text{C}$ ), **c** solar radiation ( $\text{kJ m}^{-2} \text{ day}^{-1}$ ) and **d** water vapor pressure (kPa)

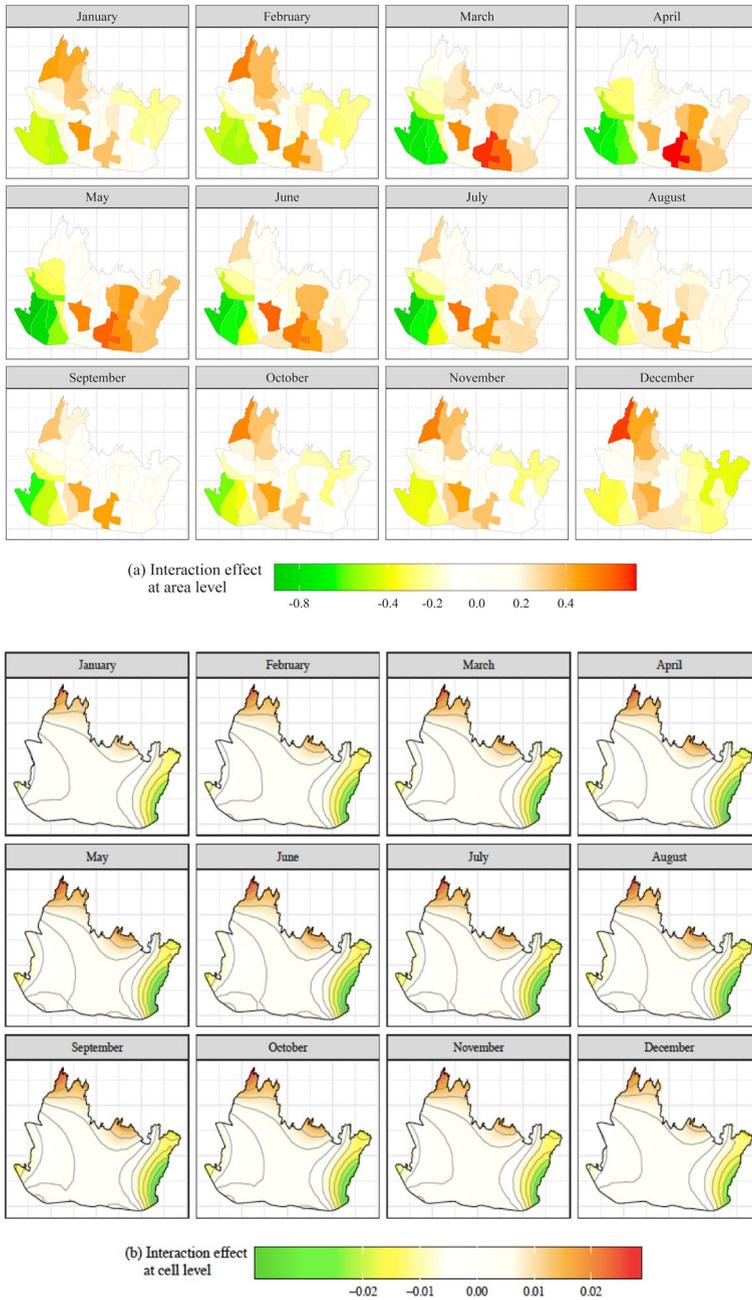
while its impact is largest in June. The delay is due to the mosquito life cycle and incubation period (Jaya and Folmer 2021a).

The time-varying coefficient of solar radiation is below zero for almost all months, except for November, and varies from  $-0.0003$  to  $0.0002$ . This is due to the fact that tropical countries such as Indonesia receive a lot of solar radiation throughout the year (Handayani and Ariyanti 2012). The time-varying effect of water vapor pressure is positive all year round and hardly varies. The effect varies from  $0.4896$  to  $0.5194$ . The strongest effect is in June ( $0.5194$ ).

Figure 7 presents the estimated parameters of the spatiotemporal interaction effects at area and cell level which depend on their hyperparameters in Table 3.

Figure 7a presents the district level spatiotemporal interaction effects (i.e., the residual effect after accounting for the weather effects). The figure shows that the interaction effect varies across districts and time. In the northern districts, it is positive and quite high during January and February, followed by a decrease to around zero during the period March–May. From June–September it is moderately positive followed by a period of high positive interaction for the rest of the year, especially in the most north-western districts. The high interaction effect in the period September–February is related to multiple factors, in particular environmental conditions. The northern areas are ideal breeding habitats because of dense vegetation and high humidity, especially during the rainy season, with low sunshine duration and high humidity.

The central districts have high spatiotemporal interaction effects because of favorable socioeconomic conditions for the spread of the dengue virus,



**Fig. 7** Monthly posterior means of **a** the area level interaction effect and **b** the cell level interaction effect, January–December

including high population density and high density of hotels, hostels and student apartments. For the northern central districts, there is the additional effect of spillover of mosquitoes from the northern districts. As a consequence, the interaction effect of the northern central districts follows a time pattern similar to time patterns of the northern districts, though less intense. The most central districts have high positive interaction effects all year round because they have the highest population density and density of hotels and hostels. The two most central districts have low interaction effects all year round, indicating that the main effects (risk factors) virtually fully explain the dengue risk. These districts have no special socioeconomic or environmental conditions affecting the dengue incidence rate.

In the southern districts, the spatiotemporal interaction effect is similar to that in the most central districts, although for partly different reasons. They have high population density with many residential areas that have unhygienic conditions. See Hsu et al. (2017) for details on the relationship between hygiene and dengue infection.

The situation in the eastern districts differs from that in the northern, southern and central districts. The interaction effect is negative in January and February, highly positive in March–May, slightly positive in June–August and negative for the rest of the year. The negative interaction effect in January–February is probably caused by the interaction of the weather variables and the environmental conditions. The absence of forests, heavy rainfall, and the short spells of sunshine in the period January–February keep the humidity low, which is unfavorable for the presence of dengue mosquitoes. The districts are residential areas with inadequate drainage and sanitation. The heavy rainfall until March combined with inadequate drainage and sanitation leads to large quantities of standing water, which provides favorable breeding habitats, contributing to the positive interaction effect in March.

The western districts have medium to strong negative interaction effects all year round, reducing the effects of the weather variables. The majority of the western districts have good drainage and sanitation, and the lifestyle and health behavior of the population is substantially better than in the other parts of Bandung, reducing dengue infection (Bandung Health Profile 2019). For example, the district with the highest healthy behavior index, Cicendo, is located in the western region. The negative effects also indicate that there is limited spillover of mosquitoes from the other districts.

Figure 7b presents contour maps of the cell level interaction effects. In contrast with Fig. 7a, the cell level interaction effects are almost the same across the months. Hence, after accounting for the weather variables, the cell level residual varies over space but is relatively constant over time, implying that it is related more to a topographical dimension, such as elevation, than to time. Positive cell level interaction effects are found in the northern part of Bandung, which is at 800 m above sea level and has high precipitation and dense vegetation, providing an ideal breeding ground and habitat for the *Aedes-spp* mosquito (Arboleda et al. 2009).

### 4.3 Posterior mean of the relative risk

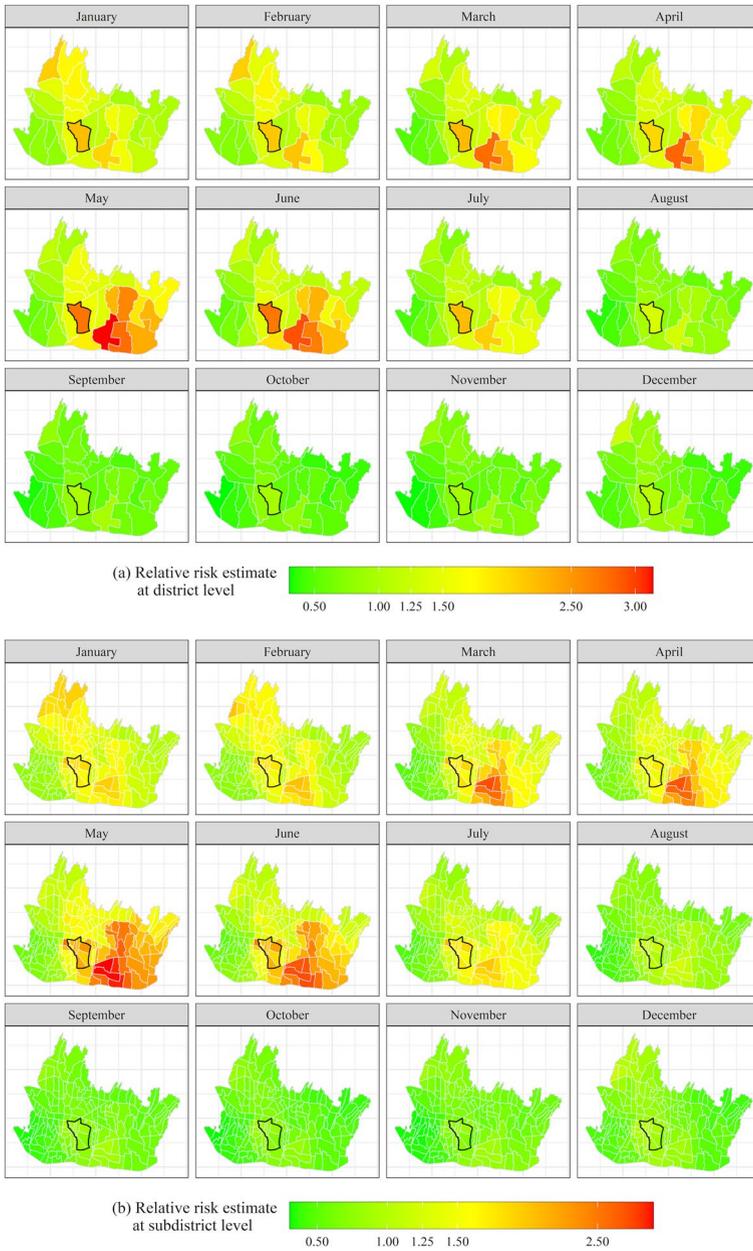
Figure 8a shows the posterior means of the relative risk estimates (based on the posterior means of the time-varying coefficients of the risk factors and the posterior means of the spatiotemporal interaction effects) at district level. The posterior means of the relative risk at subdistrict level (see Fig. 8b) are obtained as the block average of the cell values within each subdistrict boundary and the cell values that are partly outside its boundary.<sup>22</sup> Accordingly, Bandung city is divided into 30 districts, 151 subdistricts and 179 cells.

Comparing Figs. 8a and 8b shows similar temporal trends, in particular, an increase in January–July and a decrease in August–November. This applies especially to the spatial units with the highest risk, notably the districts in southern Bandung. For the spatial dimension, however, we notice substantial differences. For example, according to Fig. 8a, the entire central district of Lengkong (surrounded area in Fig. 8) is categorized as a high-risk area in the period January–July, whereas Fig. 8b shows that this only applies to parts of the district. The explanation is that categorization based on the model given by Eq. (3a) ignores within-district heterogeneity, while this is taken into account when categorization is based on the model given by Eq. (3b). To explore this issue further, we calculated the high-risk and low-risk districts and subdistricts based on the posterior exceedance probability for the two approaches, denoted as the top-down and the bottom-up approaches, respectively. Following Sparks (2015) and Osei and Stein (2017), we fixed the posterior exceedance probability threshold for  $\theta_{it}$  at 1.25. The exceedance probability  $\widehat{Pr}(\theta_{it} > 1.25 | \mathbf{y})$  over space and time is presented in Fig. 9. Table 4 presents the classification of the subdistricts into high and low risk based on the bottom-up and top-down approaches, respectively.

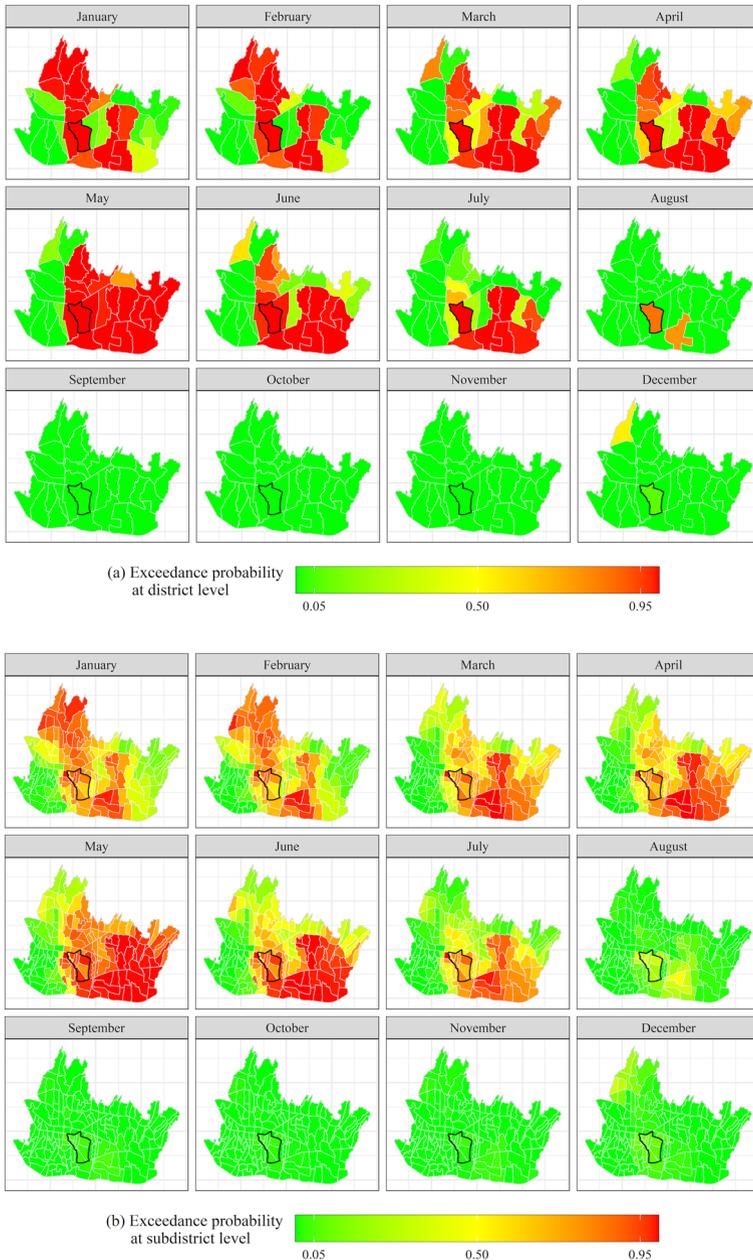
Table 4 shows substantial misclassification for the top-down approach for the period January–July. The overall misclassification rate is 15.6% for all periods (January–December) and 26.7% for the high-risk period (January–July). The table furthermore shows that all the misclassifications occurred in the period January–July, which partly overlaps with the rainy season in November–May, whereas there is no misclassification from August–December. The explanation for the misclassification as such is that the bottom-up approach averages out the differences in risk factors, and consequently the number of incidences, over a set of relatively small number of relatively homogenous cells within a subdistrict. The top-down approach, on the other hand, averages out the differences over a substantially larger number of relatively heterogeneous cells in a district (a district contains at least four subdistricts; see Sect. 4.1).

The misclassification is obviously concentrated in the rainy period January–July with local variation in the risk factors due variation in local characteristics such as elevation or vegetation density. Although the rainy season is from November to May, there are positive rates of misclassification in June–July and unexpected zero rates in November–December. These misclassifications and unexpected rates are due to the

<sup>22</sup> The reason that some cells are partly outside some subdistrict boundaries is that the cells are squares.



**Fig. 8** Monthly posterior means of the relative risk at **a** area (district) level and **b** subdistrict level (surrounded area: Lengkong district)



**Fig. 9** Monthly posterior exceedance probability of the relative risk at **a** district and **b** subdistrict levels (surrounded area: Lengkong district)

delayed responses of mating, breeding and hunting by *Aedes-spp* mosquitos. Mating and breeding occur mainly during the rainy season. Following this, it takes approximately two weeks to one month for the eggs to develop into adult mosquitos and for

**Table 4** Misclassification of the subdistricts based on the bottom-up and top-down approaches

Month	Bottom-up approach	Top-down approach		Misclassification (%)
		Low	High	
January	Low	92	53	35.1
	High	0	6	
February	Low	97	48	31.8
	High	0	6	
March	Low	105	36	24.5
	High	1	9	
April	Low	111	29	19.9
	High	1	10	
May	Low	59	54	35.8
	High	0	38	
June	Low	93	30	19.9
	High	0	28	
July	Low	119	29	19.9
	High	1	2	
August	Low	151	0	0
	High	0	0	
September	Low	151	0	0
	High	0	0	
October	Low	151	0	0
	High	0	0	
November	Low	151	0	0
	High	0	0	
December	Low	151	0	0
	High	0	0	
Average (%)	15.6			

the virus to multiply and reach the salivary glands before it is transmitted to humans. If an individual is infected, the symptoms can be observed approximately four to seven days after being bitten (Ehelepola et al. 2015). Accordingly, there is a delayed infection response with respect to the weather conditions (Jaya and Folmer 2021a).

## 5 Summary and conclusions

Effective and efficient control of a variety of spatial problems, including dengue disease abatement, requires data at a fine spatiotemporal scale. However, data availability at the same (especially fine) spatial scale is quite rare (Moraga et al. 2017; Utazi et al. 2019). A major challenge in spatial sciences, including modeling of infectious diseases such as dengue and COVID-19, is how to align data bases of different resolutions consistently. In this study, we presented the Fusion Area-Cell Spatiotemporal Generalized

Geoadditive-Gaussian Markov Random Field model as a solution to this problem. This model combines observations on the dependent variable and population at risk at the area level and covariates at the cell level to generate predictions of relative risk at the subdistrict level. Special attention was paid to the model setup to generate predictions for the target cells during model-fitting, using Bayesian Integrated Nested Laplace Approximation (INLA). The methodology was applied to monthly dengue disease data for 30 districts in the city of Bandung, Indonesia, for the period January 2012 to December 2018. The risk factors consisted of the monthly averages of precipitation, temperature, solar radiation and water vapor pressure. The analysis showed that the effects of precipitation, temperature and solar radiation varied considerably across space and time, while the effect of water vapor pressure was highly constant over time. Solar radiation was found to be the most important risk factor. The spatiotemporal interaction effect, capturing the effects of omitted variables at area level, also varied across districts and time. In contrast, the cell level interaction effect was almost constant over the months but varied substantially over space, indicating a strong spatial spillover effect.

Based on the posterior means of the relative risk at cell level, we obtained the relative risk estimates at subdistrict level. We found a similar temporal pattern for district and subdistricts. Relative dengue risk was relatively high in the period January–July and relatively low during the period August–December. We further compared the risk estimates per subdistrict based on: (i) the bottom-up approach using the cell level estimates and (ii) the top-down approach assigning the district value to its subdistricts. Using the posterior exceedance probability of the relative risk, we identified high-risk and low-risk districts to find that during the high-risk period of January–July, the top-down approach misclassified 26.4% of the subdistricts as high risk, which according to the bottom-up approach was low risk. The overall misclassification rate was 15.6%.

The main conclusions of the paper are the following. First, effective and efficient policy intervention, such as the control of infectious diseases, requires data at the right level of resolution. In particular, low-resolution maps may misclassify regions. If regions are incorrectly misclassified as high-risk, unnecessary policy intervention with undue financial, social and environmental costs may result. In contrast, if regions are incorrectly misclassified as low-risk, opportunities for policy intervention may be missed which may also have costs of various kinds. Secondly, the proposed FGG-GMRF model adjusts data and maps of different resolutions consistently, and allows more data to be utilized, thus improving the statistical efficiency. Third, application of the FGG-GMRF model to the dengue disease data for Bandung from 2012 to 2018 shows that the relative infection risk is high in various cells, subdistricts and districts from January–July. The strong spatiotemporal interaction indicates that the occurrence of the dengue disease vector is highly contagious and must be detected early in order to prevent its spread. Rapid response measures such as fogging are critical in areas with high dengue incidence. Finally, based on the experiences in the present paper, it is worthwhile to investigate the suitability of the FGG-GMRF model for a variety of other spatiotemporal problems, including other infectious diseases such as COVID-19 (see Jaya and Folmer 2021b), vaccination coverage (Utazi et al. 2019), particulate matter concentration (Cameletti et al. 2013; Lee et al. 2016), and social issues such as unemployment and crime.

## Appendix 1: The linear stochastic partial differential equation approach

A GF with dense covariance matrix can be transformed to a GMRF with sparse covariance matrix by means of a Linear Stochastic Partial Differential Equation (LSPDE) (Lindgren et al. 2011) which reads:

$$(\kappa^2 - \nabla^2)^{\alpha/2} \boldsymbol{\gamma}_t(\mathbf{s}) = \mathcal{W}_t(\mathbf{s}) \quad \text{for } t = 1, \dots, T \text{ and } \mathbf{s} = (s_1, \dots, s_g, \dots, s_n)' \text{ with } s_g = (s_{g,1}, s_{g,2}) \quad (34)$$

where  $\boldsymbol{\gamma}_t(\mathbf{s})$  is a temporally independent GF,  $\mathcal{W}_t(\mathbf{s})$  a Gaussian white noise process,  $\alpha$  a positive integer related to the smoothness parameter of the Matérn covariance function in Eq. (11) by  $\alpha = \nu + 1$ , and  $\nabla^2$  the Laplace operator, i.e.,  $\nabla^2 = \left( \frac{\partial^2}{\partial s_1^2} + \frac{\partial^2}{\partial s_2^2} \right)$ . Using spectral decomposition, Whittle (1954, 1963) showed that for  $\alpha > 1$ , the only exact stationary solution to Eq. (34) is the isotropic Matérn field, i.e., the stationary GF  $\boldsymbol{\gamma}_t(\mathbf{s})$  with the Matérn covariance function in Eq. (11).

A closed-form solution for the LSPDE in Eq. (34) is restricted to regular lattices (Lindgren et al. 2011). For an irregular lattice, it can be approximated through a basis function representation using the Finite Element Method (FEM) defined on the domain  $\mathcal{A} \in \mathbb{R}^2$ . The basis function representation is defined on a mesh, that is, a collection of (i) vertices, (ii) the edges between the vertices, and (iii) the polygons described by the edges. A mesh consists of a minimum of three connected edges conforming to the shape of the domain. It subdivides a continuous geometric space into a finite set of discrete geometric or topological elements such as triangles or rectangles (for a two-dimensional geometric space) or tetrahedral or rectangular prisms (for three-dimensional spaces). It reduces the degrees of freedom from infinite to finite. Because the FEM calculations are based on a finite number of cells and the results are generalized through interpolation for the entire domain, the accuracy of the global solution is a function of the number of elements of the mesh (Bohn and Feischl 2021).

Triangulation is a common FEM meshing scheme (Sloan 1993) because of its flexibility for irregular domains (Lindquist and Gilest 1989) and accuracy (due to the minimization of the discretization error)<sup>23</sup> (Ahmadian et al. 1998). Triangulation divides the domain into a set of non-intersecting triangles, where any two triangles meet in, at most, a common edge or vertex. The popular Delaunay triangulation scheme (Cheng et al. 2013) ensures that the triangulation maximizes the minimum angle of the triangles, thus avoiding sliver (i.e., long and thin) triangles and rendering the transitions between small and large triangles smooth (Lindgren et al. 2011). The restricted Bowyer–Watson algorithm, which is designed to conform to the domain's boundary, is a popular Delaunay triangulation algorithm (Cheng et al. 2013).

A drawback of applying an algorithm with boundary conditions is that the variance near the boundary is inflated by a factor of two (Lindgren et al. 2011). To avoid the boundary effect, Lindgren and Rue (2015) proposed to extend the domain of interest by an outer area at a distance of at least  $r$ , corresponding to a correlation of

<sup>23</sup> The errors are due to the difference between the exact solution of the model equations and the numerical solution (Tu et al. 2018).

approximately 0.1 between two points in the inner and outer areas. The edge length of the triangles of the outer area should be at least equal to the edge length of the triangles of the inner area (Blangiardo and Cameletti 2015). To find the appropriate mesh for the data at hand, meshes of different sizes are usually considered, which can be evaluated using the DIC and WAIC (Righetto et al. 2018).

Given the triangular mesh, the FEM of the solution of the LSPDE in Eq. (34) is:

$$\boldsymbol{\gamma}_t(\mathbf{s}) \approx \sum_{l=1}^L \boldsymbol{\psi}_l(\mathbf{s}) \tilde{\gamma}_{tl} \text{ for } t = 1, \dots, T, \mathbf{s} = (s_1, \dots, s_g, \dots, s_{n_p})' \text{ and } s_g = (s_{g,1}, s_{g,2}) \tag{35}$$

with  $\{\boldsymbol{\psi}_l(\mathbf{s})\}_{l=1}^L$  the set of piecewise linear basis functions,  $L$  the number of vertices in the triangulation and  $l$  the  $l$  th vertex. For location  $s_g$ , the piecewise linear basis function  $\boldsymbol{\psi}_l(s_g)$  is:

$$\boldsymbol{\psi}_l(\mathbf{s}_g) = \begin{cases} 1 & \text{if the vertex } l \text{ is at location } s_g \\ 0 & \text{elsewhere.} \end{cases} \tag{36}$$

The  $\{\tilde{\gamma}_{tl}\}_{l=1}^L$  are zero mean Gaussian-distributed weights determining the value of  $\boldsymbol{\gamma}_t(\mathbf{s})$  for vertex  $l$  at location  $s_g$ . Hence,  $\boldsymbol{\gamma}_t(\mathbf{s})$  is uniquely defined by its values at the vertices  $l, l = 1, 2, \dots, L$  of the mesh. The values in the interior of the triangles are estimated by linear interpolation. Hence, the Gaussian-distributed weights determine the values of the GF at the vertices such that the distribution of  $\boldsymbol{\gamma}_t(\mathbf{s})$  is determined by the joint distribution of the weights  $\tilde{\boldsymbol{\gamma}}_t = (\tilde{\gamma}_{t1}, \dots, \tilde{\gamma}_{tL})'$  with sparse precision matrix  $\mathbf{Q}_s$  (Lindgren et al. 2011).

To show that Eq. (34) approximates the GF  $\boldsymbol{\gamma}_t(\mathbf{s})$ , we take<sup>24</sup>  $\alpha = 2$  corresponding to  $\nu = 1$ , define the LSPDE in Eq. (34) as a variational problem, i.e., multiply it by an arbitrary test function  $\boldsymbol{\varphi}(\mathbf{s})$  for  $\mathbf{s} = (s_1, \dots, s_g, \dots, s_{n_p})'$  with  $s_g = (s_{g,1}, s_{g,2})$ , and integrate it by parts over the domain  $\mathcal{A}$  using Green's first identity theorem<sup>25</sup> (Langtangen and Logg 2016). Multiplying the LSPDE in Eq. (34) by a test function  $\boldsymbol{\varphi}(\mathbf{s})$  gives:

<sup>24</sup> For  $\mathbf{s} \in \mathbb{R}^2$ , two different values of  $\alpha$  are commonly applied, viz.  $\alpha = \{1, 2\}$  corresponding to  $\nu = \{0, 1\}$  (Lindgren et al. 2011). For  $\alpha = 2, (\kappa^2 - \nabla^2)^{\alpha/2} \boldsymbol{\gamma}_t(\mathbf{s})$  is a linear differential operator which is easier to solve than a fractional differential operator when  $\alpha$  is an odd number, e.g.  $\alpha = 1$  (Miller et al. 2019).

<sup>25</sup> Green's first identity theorem is a multidimensional generalization of integration by parts to reduce the order of derivatives (Langtangen and Logg 2016). Applied to Eq. (37), it reads:

$$\int_{\mathcal{A}} \boldsymbol{\varphi}(\mathbf{s}) \nabla^2 \boldsymbol{\gamma}_t(\mathbf{s}) d\mathbf{s} = \int_{\partial \mathcal{A}} \frac{\partial \boldsymbol{\gamma}_t(\mathbf{s})}{\partial \mathbf{n}} \boldsymbol{\varphi}(\mathbf{s}) d\mathbf{s} - \int_{\mathcal{A}} \nabla \boldsymbol{\varphi}(\mathbf{s}) \nabla \boldsymbol{\gamma}_t(\mathbf{s}) d\mathbf{s}$$

with  $\frac{\partial \boldsymbol{\gamma}_t(\mathbf{s})}{\partial \mathbf{n}}$  the directional derivative:

$$\frac{\partial \boldsymbol{\gamma}_t(\mathbf{s})}{\partial \mathbf{n}} = \nabla_{\mathbf{n}} \boldsymbol{\gamma}_t(\mathbf{s}) = \lim_{\epsilon \rightarrow 0} \frac{\boldsymbol{\gamma}_t(\mathbf{s} + \epsilon \mathbf{n}) - \boldsymbol{\gamma}_t(\mathbf{s})}{\epsilon} = \frac{\partial \boldsymbol{\gamma}_t(\mathbf{s})}{\partial \mathbf{s}} = \mathbf{n} \cdot \nabla \boldsymbol{\gamma}_t(\mathbf{s})$$

$$\int_{\mathcal{A}} \varphi(s) (\kappa^2 - \nabla^2) \gamma_t(s) ds = \int_{\mathcal{A}} \varphi(s) \mathcal{W}_t(s) ds$$

$$\int_{\mathcal{A}} \kappa^2 \varphi(s) \gamma_t(s) ds - \int_{\mathcal{A}} \varphi(s) \nabla^2 \gamma_t(s) ds = \int_{\mathcal{A}} \varphi(s) \mathcal{W}_t(s) ds \tag{37}$$

where  $\mathcal{A}$  is the entire spatial domain over which Eq. (37) is to be solved and  $ds$  is shorthand for  $ds_1 ds_2$  of the two-dimensional integral. Partial integration of  $-\int_{\mathcal{A}} \varphi(s) \nabla^2 \gamma_t(s) ds$  gives:

$$-\int_{\mathcal{A}} \varphi(s) \nabla^2 \gamma_t(s) ds = \int_{\mathcal{A}} \nabla \varphi(s) \nabla \gamma_t(s) ds - \int_{\partial \mathcal{A}} \frac{\partial \gamma_t(s)}{\partial \mathbf{n}} \varphi(s) d \tilde{s} \tag{38}$$

where  $\frac{\partial \gamma_t(s)}{\partial \mathbf{n}} = \mathbf{n} \cdot \nabla \gamma_t(s)$  is the directional derivative of  $\gamma_t(s)$  in outward normal direction  $\mathbf{n}$  on the boundary  $\partial \mathcal{A}$  (i.e., the subset of points which can be approached both from  $\mathcal{A}$  and from the outside of  $\mathcal{A}$ ) (Langtangen and Logg 2016) and  $d \tilde{s}$  is shorthand for  $d \tilde{s}_1 d \tilde{s}_2$  of the two-dimensional integral along the boundary  $\partial \mathcal{A}$ . Substituting Eq. (38) in Eq. (37), we have:

$$\int_{\mathcal{A}} \kappa^2 \varphi(s) \gamma_t(s) ds + \int_{\mathcal{A}} \nabla \varphi(s) \nabla \gamma_t(s) ds - \int_{\partial \mathcal{A}} \frac{\partial \gamma_t(s)}{\partial \mathbf{n}} \varphi(s) ds = \int_{\mathcal{A}} \varphi(s) \mathcal{W}_t(s) ds \tag{39}$$

with the Neumann boundary condition:

$$\left. \frac{\partial \gamma_t(s)}{\partial \mathbf{n}} \right|_{\partial \mathcal{A}} = 0,$$

where  $l$  denotes the boundary (Bakka 2019). Hence, Eq. (39) can be written as:

$$\kappa^2 \int_{\mathcal{A}} \varphi(s) \gamma_t(s) ds + \int_{\mathcal{A}} \nabla \varphi(s) \nabla \gamma_t(s) ds = \int_{\mathcal{A}} \varphi(s) \mathcal{W}_t(s) ds. \tag{40}$$

Because the set of the test functions is infinite (Langtangen and Logg 2016), it is not possible to test Eq. (40) for every test function  $\varphi(s)$ . As a solution, the FEM can be applied to construct a finite set of test functions  $\{\varphi_h(s)\}_{h=1}^L$  for  $h = 1, \dots, L$  and tested against Eq. (40). Using Eq. (35) and substituting  $\{\psi_l(s)\}_{l=1}^L$  for  $l = 1, \dots, L$  in Eq. (40), we obtain the system of linear equations:

$$\int_{\mathcal{A}} \kappa^2 \varphi_h(s) \sum_{l=1}^L \psi_l(s) \tilde{\gamma}_{hl} ds + \int_{\mathcal{A}} \nabla \varphi_h(s) \nabla \sum_{l=1}^L \psi_l(s) \tilde{\gamma}_{hl} ds = \int_{\mathcal{A}} \varphi_h(s) \mathcal{W}_t(s) ds,$$

$$\sum_{l=1}^L \left( \kappa^2 \int_{\mathcal{A}} \varphi_h(s) \psi_l(s) ds + \int_{\mathcal{A}} \nabla \varphi_h(s) \nabla \psi_l(s) ds \right) \tilde{\gamma}_{hl} = \int_{\mathcal{A}} \varphi_h(s) \mathcal{W}_t(s) ds. \tag{41}$$

The test functions are commonly taken to be equal to the basis functions, i.e.,  $\varphi_h(s) = \psi_l(s)^{26}$  for  $h = 1, \dots, L$ . Hence:

$$\sum_{l=1}^L \left( \kappa^2 \int_{\mathcal{A}} \psi_l(s)\psi_h(s)ds + \int_{\mathcal{A}} \nabla\psi_h(s)\nabla\psi_l(s)ds \right) \tilde{\gamma}_{tl} = \int_{\mathcal{A}} \psi_l(s)\mathcal{W}_t(s)ds \tag{42}$$

The integral on the right-hand side of Eq. (42) is the Gaussian white noise distribution  $\mathcal{N}(0, \int_{\mathcal{A}} \psi_l^2(s)ds)$ , with mean zero and covariance matrix<sup>27</sup>:

$$Cov\left(\int_{\mathcal{A}} \psi_l(s)\mathcal{W}_t(s)ds, \int_{\mathcal{A}} \psi_h(s)\mathcal{W}_t(s)ds\right) = \int_{\mathcal{A}} \psi_l(s)\psi_h(s)ds. \tag{43}$$

We can write Eq. (42) in matrix form as (Simpson et al. 2012):

$$\mathbf{K}_s \tilde{\boldsymbol{\gamma}}_t \sim \mathcal{N}(0, \mathbf{C}_s) \tag{44}$$

where  $\mathbf{K}_s = \kappa^2 \mathbf{C}_s + \mathbf{M}_s$ ,  $\tilde{\boldsymbol{\gamma}}_t = (\tilde{\gamma}_{t1}, \dots, \tilde{\gamma}_{tL})'$ ,  $C_{s,lh} = \int_{\mathcal{A}} \psi_l(s)\psi_h(s)ds$ , and  $M_{s,lh} = \int_{\mathcal{A}} \nabla\psi_l(s) \cdot \nabla\psi_h(s)ds$ .

Because of the highly local nature of the piecewise linear basis functions,  $\mathbf{C}_s$  and  $\mathbf{M}_s$  are sparse matrices. However,  $\mathbf{C}_s^{-1}$  is a dense matrix and will generally not be a GMRF. Lindgren et al. (2011) proposed to solve this issue by reducing the integration order of the inner product of the piecewise linear basis functions  $\psi_l(s)$  and  $\psi_h(s)$  for vertices  $l = 1, \dots, L$  and  $h = 1, \dots, L$  on the interval  $\mathcal{A}$  (i.e.,  $\langle \psi_l(s), \psi_h(s) \rangle_{\mathcal{A}} = \int_{\mathcal{A}} \psi_l(s)\psi_h(s)ds$ )<sup>28</sup> by taking  $\psi_h(s)$  as the constant function 1 yielding  $\langle \psi_l(s), 1 \rangle_{\mathcal{A}} = \int_{\mathcal{A}} \psi_l(s)ds = \sum_{h=1}^L C_{lh}$ . The result is an approximate diagonal  $\mathbf{C}_s$  matrix,  $\tilde{\mathbf{C}}_s$ , with diagonal elements  $C_{s,ll} = \sum_{h=1}^L C_{lh}$ . Replacing  $\mathbf{C}_s$  with  $\tilde{\mathbf{C}}_s$  yields

$$\mathbf{K}_s \tilde{\boldsymbol{\gamma}}_t \sim \mathcal{N}(0, \tilde{\mathbf{C}}_s) \tag{45}$$

<sup>26</sup> This procedure is known as the Galerkin finite element method (Langtangen and Logg 2016).

<sup>27</sup> By definition, a generalized GF, for example,  $\chi(s)$ , in a domain  $\mathcal{A}$  is a random  $L^2(\mathcal{A})$  generalized function, with  $L^2(\mathcal{A})$  a vector space of square-integrable functions in two-dimensional space  $\mathcal{A}$  (i.e.,  $L^2(\mathcal{A}) = \{\chi(s) : \mathcal{A} \rightarrow \int_{\mathcal{A}} |\chi(s)|^2 ds < \infty\}$ ) such that for every finite set of test functions  $\{\psi_h \in L^2(\mathcal{A}), h = 1, \dots, H\}$ , the inner product  $\int_{\mathcal{A}} \psi_h \chi(s) ds$  for  $h = 1, \dots, H$  is jointly Gaussian. Because Gaussian white noise  $\mathcal{W}_t(s)$  is an  $L^2(\mathcal{A})$ -bounded generalized GF in the domain  $\mathcal{A}$ , the distribution of  $\int_{\mathcal{A}} \psi_l \mathcal{W}_t(s) ds$  is Gaussian, with expectation and covariance given by (Lindgren et al. 2011):

$$\mathbb{E}\left(\int_{\mathcal{A}} \psi_l(s)\mathcal{W}_t(s)ds\right) = 0$$

$$Cov\left(\int_{\mathcal{A}} \psi_l(s)\mathcal{W}_t(s)ds, \int_{\mathcal{A}} \psi_h(s)\mathcal{W}_t(s)ds\right) = \int_{\mathcal{A}} \psi_l(s)\psi_h(s)ds.$$

<sup>28</sup> The inner product of two arbitrary functions  $\psi_l(s)$  and  $\psi_h(s)$  is approximately the same as the summation of the product of pairwise values, with summation replaced by integration (Langtangen and Mardal 2019).

and

$$\tilde{\boldsymbol{\gamma}}_t \sim \mathcal{N}\left(0, \tilde{\mathbf{Q}}_s\right) \quad (46)$$

with  $\tilde{\mathbf{Q}}_s$  the precision matrix

$$\begin{aligned} \tilde{\mathbf{Q}}_s &= \mathbf{K}_s' \tilde{\mathbf{C}}_s^{-1} \mathbf{K}_s = \left(\kappa^2 \tilde{\mathbf{C}}_s + \mathbf{M}_s\right)' \tilde{\mathbf{C}}_s^{-1} \left(\kappa^2 \tilde{\mathbf{C}}_s + \mathbf{M}_s\right) \\ \tilde{\mathbf{Q}}_s &= \left(\kappa^4 \tilde{\mathbf{C}}_s + 2\kappa^2 \mathbf{M}_s + \mathbf{M}_s \tilde{\mathbf{C}}_s^{-1} \mathbf{M}_s\right) \quad \text{for every } t, \end{aligned} \quad (47)$$

where  $\kappa = \sqrt{8v/r}$ . Because of the serial independence assumption in Eq. (9),  $\tilde{\mathbf{Q}}_s$  is constant over time.

Bolin and Lindgren (2009) compared the exact FEM approach (using  $\mathbf{C}_s$  with all the elements of  $\mathbf{C}_s$  calculated as a Hilbert space wavelet model such as a B-spline or a Daubechies wavelet model), and the Markov approximation (replacing  $\mathbf{C}_s$  with  $\mathbf{C}_s$ ) for spatial prediction and found that the differences between both approaches in terms prediction errors are negligible.

## Appendix 2

This appendix consists of three parts. The first part is Table 5. which presents the priors and hyperpriors for the FGG-GMRF model given by Eq. (33). The second part consists of comments on the priors and hyperpriors and the final part deals with identification.

### Priors, joint priors and hyperpriors for the FGG-GMRF model

See Table 5.

### Comments on the priors and hyperpriors

The following observations apply for the parameters and hyperparameters of the FGG-GMRF. Due to the lack of strong prior knowledge, we used a vague Gaussian prior distribution with a zero mean and a very large variance for the parameters  $\beta_0 : \beta_0 \sim \mathcal{N}(0, 10^6)$  and  $\beta_k : \beta_k \sim \mathcal{N}(0, 10^6)$  for  $k = 1, \dots, K$  (Blangiardo and Cameletti 2015; Martinez-Beneito and Botella-Rocamora 2019). A weakly informative prior<sup>29</sup> was assigned to the log-odds of the spatial autoregressive parameter  $\rho : \log(\rho/(1 - \rho)) \sim \mathcal{N}(0, 0.45)$  (Utazi et al. 2019). Note that the transformation is

<sup>29</sup> This prior represents a situation where we are not completely ignorant about the values of a parameter, but have no firm prior information about them (Simpson et al. 2017).

**Table 5** Priors, joint priors and hyperpriors<sup>a</sup>

Component	Prior	Joint prior	Hyperprior	
Intercept	Exchangeable (iid) Gaussian process <sup>b</sup>	$\beta_0   \sigma_{\beta_0}^2 \sim \mathcal{N}(0, \sigma_{\beta_0}^2)$ for every $i, g$ and $t$ . A vague Gaussian prior with zero mean and a large variance $\sigma_{\beta_0}^2 = 10^6$	$p(\beta_0   \sigma_{\beta_0}^2) \propto \exp\left(-\frac{1}{2\sigma_{\beta_0}^2} \beta_0^2\right)$	-
Global effects of the risk factors	Exchangeable (iid) Gaussian process	$\beta_k   \sigma_{\beta_k}^2 \sim \mathcal{N}(0, \sigma_{\beta_k}^2)$ for every $i, g$ and $t$ and $k = 1, \dots, K$ . A vague Gaussian prior with zero mean and a large variance $\sigma_{\beta_k}^2 = 10^6$	$p(\beta   \mathbf{Q}\beta^{-1}) \propto \exp\left(-\frac{1}{2} \beta' \mathbf{Q}\beta\right)$ for every $i, g$ and $t$ where $\mathbf{Q}\beta = \text{diag}\left(\frac{1}{\sigma_{\beta_1}^2}, \dots, \frac{1}{\sigma_{\beta_K}^2}\right)$ with $\beta = (\beta_1, \dots, \beta_K)$ and $\sigma_{\beta_k}^2 = 10^6$ for $k = 1, \dots, K$	-



Table 5 (continued)

Component	Prior	Joint prior	Hyperprior
Spatially structured random effect	<p>Leroux Conditional Autoregressive (CAR)</p> <p><math>\omega_i   (\omega_{-i}, \mathbf{W}) \sim \mathcal{N}(\mu, \sigma^2)</math> for every <math>i</math> and <math>i = 1, \dots, n_A</math>, with</p> $\mu = \frac{\rho \sum_{j=1}^{n_A} w_{ij} \omega_j}{\rho \sum_{j=1}^{n_A} w_{ij} + 1 - \rho}$ $\sigma^2 = \frac{\sigma_{\omega}^2}{\left(\rho \sum_{j=1}^{n_A} w_{ij} + 1 - \rho\right)}$ <p>and <math>\mathbf{W}</math> the <math>(n_A \times n_A)</math> binary spatial weights matrix characterizing the neighborhood structure of the areas, e.g., an inverse distance matrix or a contiguity matrix and <math>\rho</math> the spatial autoregressive coefficient</p>	<p><math>p(\boldsymbol{\omega}   \sigma_{\omega}^2) \propto \exp\left(-\frac{1}{2} \boldsymbol{\omega}' \mathbf{Q}_{\omega} \boldsymbol{\omega}\right) \propto \mathcal{N}(0, \mathbf{Q}_{\omega}^{-1})</math> for every <math>t</math>,</p> <p>with <math>\boldsymbol{\omega} = (\omega_1, \dots, \omega_{n_A})'</math> and <math>\mathbf{Q}_{\omega} = \boldsymbol{\tau}_{\omega} \mathbf{R}_{\omega}</math> the precision matrix with <math>\boldsymbol{\tau}_{\omega} = 1/\sigma_{\omega}^2</math> and <math>\mathbf{R}_{\omega}</math> the <math>(n_A \times n_A)</math> spatial structure matrix, with the <math>(i, j)</math> th element defined as:</p> $R_{\omega}[i, j] = \begin{cases} \rho n_{A_i} + (1 - \rho) & \text{if } i = j \\ -1 & \text{if } i \sim j \\ 0 & \text{otherwise} \end{cases}$ <p>where <math>n_{A_i}</math> is the number of neighbors in region <math>A_i</math>, and <math>i \sim j</math> denoting that areas <math>i</math> and <math>j</math> are neighbors</p>	<p>Half Cauchy (HC) with scale parameter 25 for <math>\sigma_{\omega}</math> i.e., <math>\sigma_{\omega} \sim C^+(0, 25)</math> and <math>\log\left(\frac{\rho}{1-\rho}\right) \sim \mathcal{N}(0, 0.45)</math></p>
Spatially unstructured random effect	<p>Exchangeable (iid) Gaussian process</p> <p><math>v_i   \sigma_v^2 \sim \mathcal{N}(0, \sigma_v^2)</math> for every <math>t</math> and <math>i = 1, \dots, n_A</math>.</p>	<p><math>p(\mathbf{v}   \sigma_v^2) \propto \exp\left(-\frac{1}{2} \mathbf{v}' \mathbf{Q}_v \mathbf{v}\right) \propto \mathcal{N}(0, \mathbf{Q}_v^{-1})</math> for every <math>t</math>, with <math>\mathbf{v} = (v_1, \dots, v_{n_A})'</math> and</p> <p><math>\mathbf{Q}_v = \boldsymbol{\tau}_v \mathbf{I}_{n_A}</math> the precision matrix, <math>\mathbf{I}_{n_A}</math> the <math>(n_A \times n_A)</math> identity matrix and <math>\boldsymbol{\tau}_v = 1/\sigma_v^2</math></p>	<p><math>\sigma_v^2 \sim \text{IG}(1, 0.01)</math></p>

Table 5 (continued)

Component	Prior	Joint prior	Hyperprior
Temporally structured random effect	<p>Autoregressive of order one (AR1)</p> $\phi_t = \lambda_2 \phi_{t-1} + \epsilon_t,$ $\epsilon_t \sim \mathcal{N}(0, \sigma_\phi^2)$ for every $i$ and $g$ and $t = 2, \dots, T$ , with $ \lambda_2  < 1$ the autoregressive parameter and $\phi_1 \sim \mathcal{N}\left(0, \frac{\sigma_\phi^2}{1-\lambda_2^2}\right)$	$p(\boldsymbol{\phi}   \mathbf{Q}_\phi^{-1}) \propto \exp\left(-\frac{1}{2} \boldsymbol{\phi}' \mathbf{Q}_\phi \boldsymbol{\phi}\right)$ $\propto \mathcal{N}(0, \mathbf{Q}_\phi^{-1})$ for every $i$ and $g$ , with $\boldsymbol{\phi} = (\phi_1, \dots, \phi_T)'$ and $\mathbf{Q}_\phi = \tau_\phi \mathbf{R}_{\phi(T \times T)}$ the precision matrix with $\tau_\phi = 1/\sigma_\phi^2$ and $\mathbf{R}_{\phi(T \times T)}$ the $(T \times T)$ temporal structure matrix for the AR1 prior: $\mathbf{R}_{\phi(T \times T)} = \begin{bmatrix} 1 & & & & \\ -\lambda_2 & & & & \\ & (1 + \lambda_2^2) & & & \\ & & \ddots & & \\ & & & -\lambda_2 & \\ & & & & (1 + \lambda_2^2) \\ & & & & & \ddots \\ & & & & & & -\lambda_2 \\ & & & & & & & 1 \end{bmatrix}$	$\sigma_\phi^2 \sim \text{IG}(1, 0.01)$
Temporally unstructured random effect	<p>Exchangeable (iid) Gaussian</p> $\xi_t   \sigma_\xi^2 \sim \mathcal{N}(0, \sigma_\xi^2)$ for every $i$ and $g$ and $t = 1, \dots, T$	$p(\boldsymbol{\xi}   \mathbf{Q}_\xi^{-1}) \propto \exp\left(-\frac{1}{2} \boldsymbol{\xi}' \mathbf{Q}_\xi \boldsymbol{\xi}\right)$ $\propto \mathcal{N}(0, \mathbf{Q}_\xi^{-1})$ for every $i$ and $g$ , with $\boldsymbol{\xi} = (\xi_1, \dots, \xi_T)'$ and $\mathbf{Q}_\xi = \tau_\xi \mathbf{I}_T$ the precision matrix with $\tau_\xi = 1/\sigma_\xi^2$ and $\mathbf{I}_T$ the $(T \times T)$ identity matrix	$\sigma_\xi^2 \sim \text{IG}(1, 0.01)$

Table 5 (continued)

Component	Prior	Joint prior	Hyperprior
Area level Interaction effect	Type I: combines the spatially unstructured ( $\nu_i$ ) and temporally unstructured ( $\zeta_i$ ) random effects. Implies $\rho = 0$ and $\lambda_2 = 0$ .	$p(\boldsymbol{\delta}   \mathbf{Q}_{\delta(t)}^{-1}) \propto \exp\left(-\frac{1}{2} \boldsymbol{\delta}' \mathbf{Q}_{\delta(t)} \boldsymbol{\delta}\right) \propto \mathcal{N}\left(0, \mathbf{Q}_{\delta(t)}^{-1}\right)$ where $\boldsymbol{\delta} = (\delta_{11}, \dots, \delta_{n_A T})'$ with $\delta_{it}$ independent over space and time for $i = 1, \dots, n_A$ , and $t = 1, \dots, T$ . $\mathbf{Q}_{\delta(t)} = \tau_\delta \mathbf{R}_{\delta(t)}$ the precision matrix with $\tau_\delta = 1/\sigma_\delta^2$ and structure matrix $\mathbf{R}_{\delta(t)} = \mathbf{R}_\nu \otimes \mathbf{R}_\zeta = \mathbf{I}_{n_A} \otimes \mathbf{I}_T = \mathbf{I}_{n_A T}$ with $\mathbf{I}_{n_A T}$ the $(n_A T \times n_A T)$ identity matrix and $\otimes$ the Kronecker product	$\sigma_\nu^2 \sim \text{IG}(1, 0.01)$ and $\sigma_\zeta^2 \sim \text{IG}(1, 0.01)$
Type II: combines the spatially unstructured ( $\nu_i$ ) and the temporally structured ( $\phi_i$ ) random effects. Implies $\rho = 0$ and $\lambda_2 \neq 0$ .		$p(\boldsymbol{\delta}   \mathbf{Q}_{\delta(t)}^{-1}) \propto \exp\left(-\frac{1}{2} \boldsymbol{\delta}' \mathbf{Q}_{\delta(t)} \boldsymbol{\delta}\right) \propto \mathcal{N}\left(0, \mathbf{Q}_{\delta(t)}^{-1}\right)$ where $\boldsymbol{\delta} = (\delta_{11}, \dots, \delta_{n_A})'$ with $\delta_i = (\delta_{i1}, \dots, \delta_{iT})'$ for $i = 1, \dots, n_A$ following an AR1, independently distributed for all areas. $\mathbf{Q}_{\delta(t)} = \tau_\delta \mathbf{R}_{\delta(t)}$ the precision matrix with $\tau_\delta = \frac{1}{\sigma_\delta^2}$ and structure matrix $\mathbf{R}_{\delta(t)} = \mathbf{R}_\nu \otimes \mathbf{R}_\phi = \mathbf{I}_{n_A} \otimes \mathbf{R}_\phi$	$\sigma_\nu^2 \sim \text{IG}(1, 0.01)$ and $\sigma_\phi^2 \sim \text{IG}(1, 0.01)$

Table 5 (continued)

Component	Prior	Joint prior	Hyperprior
<p>Type III: combines the temporally unstructured (<math>\varepsilon_t</math>) and the spatially structured (<math>\omega_i</math>) random effects. Implies <math>\rho \neq 0</math> and <math>\lambda_2 = 0</math></p>	<p><math>p(\boldsymbol{\delta}   \mathbf{Q}_{\delta(t)}) \propto \exp\left(-\frac{1}{2} \boldsymbol{\delta}' \mathbf{Q}_{\delta(t)} \boldsymbol{\delta}\right) \propto \mathcal{N}\left(0, \mathbf{Q}_{\delta(t)}^{-1}\right)</math>                      where <math>\boldsymbol{\delta} = (\delta_1, \dots, \delta_T)'</math> with <math>\boldsymbol{\delta}_t = (\delta_{it}, \dots, \delta_{nt})'</math> for <math>t = 1, \dots, T</math> following a Leroux CAR, independently distributed for all periods. <math>\mathbf{Q}_{\delta(t)} = \tau_{\delta} \mathbf{R}_{\delta(t)}</math> the precision matrix with <math>\tau_{\delta} = \frac{1}{\sigma_{\delta}^2}</math> and structure matrix <math>\mathbf{R}_{\delta(t)} = \mathbf{R}_{\xi} \otimes \mathbf{R}_{\omega} = \mathbf{I}_T \otimes \mathbf{R}_{\omega}</math></p>	<p>HC prior with scale parameter 25 for <math>\sigma_{\omega} \sim C^+(0, 25)</math>, <math>\log\left(\frac{\rho}{1-\rho}\right) \sim \mathcal{N}(0, 0.45)</math> and <math>\sigma_v^2 \sim \text{IG}(1, 0.01)</math></p>	
<p>Type IV: combines the spatially (<math>\omega_i</math>) and temporally structured (<math>\phi_t</math>) random effects. Implies <math>\rho \neq 0</math> and <math>\lambda_2 \neq 0</math></p>	<p><math>p(\boldsymbol{\delta}   \mathbf{Q}_{\delta(tV)}) \propto \exp\left(-\frac{1}{2} \boldsymbol{\delta}' \mathbf{Q}_{\delta(tV)} \boldsymbol{\delta}\right) \propto \mathcal{N}\left(0, \mathbf{Q}_{\delta(tV)}^{-1}\right)</math>                      where <math>\boldsymbol{\delta} = (\delta_{11}, \dots, \delta_{n_A T})'</math> with <math>\delta_{it}</math> for <math>i = 1, \dots, n_A</math>, and <math>t = 1, \dots, T</math> dependent over space and time. <math>\mathbf{Q}_{\delta(tV)} = \tau_{\delta} \mathbf{R}_{\delta(tV)}</math> the precision matrix with <math>\tau_{\delta} = \frac{1}{\sigma_{\delta}^2}</math> and structure matrix <math>\mathbf{R}_{\delta(tV)} = \mathbf{R}_{\omega} \otimes \mathbf{R}_{\phi}</math></p>	<p>HC prior with scale parameter 25 for <math>\sigma_{\omega} \sim C^+(0, 25)</math>, <math>\log\left(\frac{\rho}{1-\rho}\right) \sim \mathcal{N}(0, 0.45)</math> and <math>\sigma_{\phi}^2 \sim \text{IG}(1, 0.01)</math></p>	

Table 5 (continued)

Component	Prior	Joint prior	Hyperprior
Cell level Gaussian Field through LSPDE model	$\tilde{\Phi}_t = \lambda_1 \tilde{\Phi}_{t-1} + \tilde{\gamma}_t \text{ with } \tilde{\gamma}_t(s) \sim \mathcal{N}(0, \tilde{\mathbf{Q}}_s)$ <p>for <math>t = 1, \dots, T</math>. <math>\mathbf{I}</math> Initial value</p> $\tilde{\Phi}_1 \sim \mathcal{N}(0, \tilde{\mathbf{Q}}_s / (1 - \lambda_1^2))$	$p(\tilde{\Phi}   \tilde{\mathbf{Q}}_{\Phi}^{-1}) \propto \exp\left(-\frac{1}{2} \tilde{\Phi} \tilde{\mathbf{Q}}_{\Phi} \tilde{\Phi}\right)$ $\propto \mathcal{N}\left(0, \tilde{\mathbf{Q}}_{\Phi}^{-1}\right)$ <p>with <math>\tilde{\Phi} = (\tilde{\Phi}_1, \dots, \tilde{\Phi}_T)'</math> and <math>\tilde{\mathbf{Q}}_{\Phi}^{-1} = \tilde{\mathbf{Q}}_T \otimes \tilde{\mathbf{Q}}_s</math> the <math>(TL \times TL)</math></p>	<p>Penalized complexity (PC) distribution for <math>\sigma_y</math> and <math>r</math>, i.e., <math>\Pr(\sigma_{\Phi} &gt; 1) = 0.01</math> and <math>\Pr(r &lt; 5) = 0.5</math>, respectively</p>
		<p>precision matrix with</p> $\mathbf{Q}_{T \times (T)} = \begin{bmatrix} 1/\sigma_y^2 & & & & \\ & -\lambda_1/\sigma_y^2 & & & \\ & -\lambda_1/\sigma_y^2 & (1 + \lambda_1^2)/\sigma_y^2 & & \\ & & & \ddots & \\ & & & -\lambda_1/\sigma_y^2 & -\lambda_1/\sigma_y^2 \end{bmatrix}$ <p>the <math>T</math>-dimensional precision matrix of the (AR1) process and</p> $\tilde{\mathbf{Q}}_s = \left( \kappa^d \tilde{\mathbf{C}}_s + 2\kappa^2 \mathbf{M}_s + \tilde{\mathbf{M}}_s \tilde{\mathbf{C}}_s \mathbf{M}_s \right)$ the $(L \times L)$ spatial precision matrix of the GMRF	

<sup>a</sup>Table 5 is partly based on Jaya and Folmer (2021a)

<sup>b</sup>An exchangeable (Gaussian) process is a sequence of random variables that are independent and identically (normal) distributed (iid).

<sup>c</sup>Scaling is applied to the temporal structure matrices of RW1 and RW2 to make their marginal variances comparable. Scaling is implemented in R-INLA by using the option scale.model=TRUE

used to ensure that  $\rho$  takes values between 0 and 1 (Bivand et al. 2015; Martinez-Beneito and Botella-Rocamora 2019).

For the scale hyperparameters (variance and standard deviation) of the temporally varying coefficients, spatially and temporally structured and unstructured random effects, spatiotemporal interaction of the random effects, and the Gaussian field (GF), the Inverse Gamma (IG), half-Cauchy (HC), Uniform and Penalized Complexity (PC) are common hyperpriors. The IG distribution is a well-known, easy to use, conjugate hyperprior for the variance (Coly et al. 2021; Hamura et al. 2021).<sup>30</sup> The HC is a Cauchy distribution truncated at zero. Gelman (2006) recommended the HC with scale parameter 25 as an alternative to the IG when the variance hyperparameter is close to zero. The HC is also a frequently used hyperprior for the standard deviation (Gómez-Rubio 2020). When the limit of the scale parameter goes to infinity, the HC converges to the Uniform hyperprior (Gelman 2006; Gómez-Rubio 2020). However, for the Uniform hyperprior, it is difficult to set ranges of values for the standard deviation using R-INLA. The reason is that R-INLA assumes that the standard deviation is unbounded (above) (Gelman et al. 2017; Gómez-Rubio 2020), whereas a fixed upper bound on the standard deviation is frequently needed to avoid overfitting<sup>31</sup> (Simpson et al. 2017). As an alternative to the HC and Uniform hyperpriors for the standard deviation, Simpson et al. (2017) proposed the PC hyperprior. The PC allows using probability statements on values for the parameters such as the standard deviation, autocorrelation and range parameters (Franco-Villoria et al. 2019). The parameters can be restricted using either an upper bound or a lower bound but not both. The PC is frequently used for autoregressive processes (Sørbye and Rue 2017), varying coefficients models and high-resolution prediction models (Franco-Villoria et al. 2019).

In the application, we assigned the IG (1, 0.01) hyperprior to the variance of the autoregressive AR1 model of the temporally structured effects ( $\sigma_\phi^2$ ) and to the variances of the exchangeable (iid) priors of the spatially ( $\sigma_v^2$ ) and temporally ( $\sigma_\zeta^2$ ) unstructured random effects, respectively. Following Lawson et al. (2010), we assigned the HC hyperprior with scale parameter to the standard deviation of the Leroux prior (Leroux et al. 2000) of the spatially structured random effect ( $\sigma_\omega^2$ ).

Regarding the hyperpriors to the variances of the RW1 and RW2 priors the following applies. The marginal variances of the RW1 and RW2 priors reflect the smoothness of the  $k$ th vector of the temporal random effect  $\zeta_k = (\zeta_{k,1}, \dots, \zeta_{k,T})'$  for  $k = 1, \dots, K$ .<sup>32</sup> However, before assigning hyperpriors we need to render the marginal variances as smoothness indicators comparable as they are affected by their temporal structures and the scale of the risk factors which render them inadequate as smoothness indicators (Sørbye and Rue 2014; Blangiardo and Cameletti 2015; Gómez-Rubio 2020; Kang et al. 2015). The temporal structure matrices carry over to the hyperpriors. To overcome the incomparability problem, Sørbye and Rue (2014) proposed to scale the

<sup>30</sup> Note that in the present setting we are dealing with the latent Gaussian model (LGM) for which the IG distribution is a conjugate hyperprior for the variance of the Gaussian prior (Hamura et al. 2021).

<sup>31</sup> The Uniform distribution becomes an increasingly uninformative prior as the upper bound increases (Lemoine 2019). Hence, it is unclear what upper bound is appropriate for a Uniform prior.

<sup>32</sup> The smaller the marginal variance, the smoother the time-varying regression coefficient.

temporal structure matrices by the geometric means of the diagonal elements of their inverses. However, the random walk models of order one (RW1) and two (RW2) are intrinsic Gaussian Markov random fields (IGMRFs).<sup>33</sup> They have zero mean and generalized covariance matrices  $\Sigma_{\zeta_k}^{RW1} = \sigma_{\zeta_k}^2 \mathbf{R}_{\zeta_k}^{RW1-}$  and  $\Sigma_{\zeta_k}^{RW2} = \sigma_{\zeta_k}^2 \mathbf{R}_{\zeta_k}^{RW2-}$ , respectively with  $\mathbf{R}_{\zeta_k}^{RW1-}$  and  $\mathbf{R}_{\zeta_k}^{RW2-}$  the generalized inverses of the temporal structure matrices varying along with RW1 and RW2, respectively.

Sørbye and Rue (2014) proposed to scale the temporal structure matrices of the RW1 and RW2 priors by the geometric mean of the corresponding diagonal elements of their generalized inverses  $\mathbf{R}_{\zeta_k}^-$ :

$$\exp\left(\frac{1}{T} \sum_{t=1}^T \log\left(\text{diag}\left(\mathbf{R}_{\zeta_k}^-\right)\right)\right) = \sigma_{GV}^2$$

with  $\sigma_{GV}^2$  called the generalized variance of the diagonal elements of  $\mathbf{R}_{\zeta_k}^-$  (see Sørbye and Rue (2014) for details). Scaling of  $\mathbf{R}_{\zeta_k}$  gives the scaled temporal structure matrix  $\tilde{\mathbf{R}}_{\zeta_k} = \mathbf{R}_{\zeta_k} \sigma_{GV}^2$  and scaled generalized covariance matrix as  $\tilde{\Sigma}_{\zeta_k} = \sigma_{\zeta_k}^2 \tilde{\mathbf{R}}_{\zeta_k}$  with scaled generalized variance  $\sigma_{GV} = 1$ .

Scaling the RW1 and RW2 priors can be straightforwardly done in R-INLA by using the option `scale.model=TRUE` (Blangiardo and Cameletti 2015; Sørbye 2013). After scaling, the same hyperprior IG (1, 0.01) can be assigned to the variance  $\sigma_{\zeta_k}^2$  of the IGMRF priors RW1 and RW2.<sup>34</sup> We followed this procedure in the application.

To avoid overfitting, we followed Fuglstad et al. (2020) and applied a weakly informative PC hyperprior to the range ( $r$ ) and the standard deviation ( $\sigma_{\phi}$ ) of the LSPDE model. Particularly, we applied  $\Pr(\sigma_{\phi} > \sigma_{\phi 0}) = \alpha_{\sigma_{\phi}}$  with  $\sigma_{\phi 0}$  and  $\alpha_{\sigma_{\phi}}$  denoting the lower limit and lower tail probability of the PC hyperprior for  $\sigma_{\gamma}$ , respectively. The actual value for the standard deviation was 1 and 0.001 for the tail probability, i.e.,  $\Pr(\sigma_{\phi} > 1) = 0.01$ . For the range  $r$  we selected  $\Pr(r < r_0) = \alpha_r$  with  $r_0$  and  $\alpha_r$  denoting the upper limit and upper tail probability, respectively. We took prior information  $r_0 = 5\text{km}$  which is the distance beyond which dengue disease risk is no longer spatially correlated (Sedda et al. 2018). It corresponds to the median distance

<sup>33</sup> An IGMRF has a sparse precision matrix which is not of full rank. According to Rue and Held (2005), the order of an IGMRF is the rank deficiency of its precision matrix, i.e., the number of its zero eigenvalues. Hence, a zero-mean IGMRF of order  $s$  is:

$$p\left(\zeta_k | \mathbf{Q}_{\zeta_k}^-\right) = (2\pi)^{-\frac{T-s}{2}} \left(|\mathbf{Q}_{\zeta_k}^-|\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \zeta_k' \mathbf{Q}_{\zeta_k}^- \zeta_k\right)$$

where  $\mathbf{Q}_{\zeta_k}$  is the  $(T \times T)$  the semi-positive definite precision matrix of the parameters  $\zeta_k$  given by  $\mathbf{Q}_{\zeta_k} = \frac{1}{\sigma_{\zeta_k}^2} \mathbf{R}_{\zeta_k}$  with  $\mathbf{R}_{\zeta_k}$  the  $(T \times T)$  temporal structure matrix.  $|\mathbf{Q}_{\zeta_k}^-|$  denotes the generalized determinant which is equal to the product of the  $T - s$  non-zero eigenvalues of  $\mathbf{Q}_{\zeta_k}$  ( $s = \{1,2\}$  for a RW1 and RW2, respectively). The generalized inverse of the precision matrix ( $\mathbf{Q}_{\zeta_k}^-$ ) is the generalized covariance matrix  $\Sigma_{\zeta_k} = \sigma_{\zeta_k}^2 \mathbf{R}_{\zeta_k}^-$ .

<sup>34</sup> Sørbye and Rue (2014) pointed out that scaling of the RW1 and RW2 variances reduces the sensitivity of the estimated marginal variance to changes in the scale parameter of the IG hyperprior and re-scaling of the covariates.

between the grid cells. For the prior belief we took  $\alpha_r = 0.5$ , i.e.,  $\Pr(r < 5) = 0.5$ . Note that the two preceding decisions imply that we know the true limits ahead of the analysis.

## Identification

Because the ICAR, Leroux, AR1, RW1, and RW2 priors are specified conditionally on neighboring observations and their parameters are unique up to an additive constant, their specifications implicitly include the overall intercept. Consequently, if the model would also include an additional intercept, it would not be identified due to the perfect collinearity (Goicoa et al. 2018). To solve this identification problem, the explicit intercept can be excluded or sum-to-zero constraints on the random effects can be imposed (Goicoa et al. 2018). When sum-to-zero constraints are imposed, the random effects become orthogonal to the explicit overall intercept.

To achieve identifiability, we imposed the following sum-to-zero constraints on the temporal random of effects of the risk factors, the spatial and temporal random effects and their spatiotemporal interaction effects components:

$$\sum_{t=1}^T \zeta_{k,t} = 0 \quad \text{for every } i \text{ and } g \text{ and for } k = 1, \dots, K,$$

$$\sum_{i=1}^{n_A} \omega_i + v_i = 0 \quad \text{and} \quad \sum_{i=1}^{n_A} \delta_{it} = 0 \quad \text{for } t = 1, \dots, T,$$

$$\sum_{t=1}^T \phi_t + \zeta_t = 0 \quad \text{and} \quad \sum_{t=1}^T \delta_{it} = 0 \quad \text{for } i = 1, \dots, n_A.$$

The above-mentioned sum-to-zero constraints are imposed in INLA by setting the `constr = TRUE` argument to the function  $f(\cdot)$  which defines the temporal random effects of the risk factors, the spatial, temporal, and spatiotemporal interaction effect components.

## Appendix 3

The best specification of the FGG-GMRF model in Eq. (33) was selected using the deviance information criterion (DIC), the Watanabe–Akaike information criterion (WAIC), and the marginal predictive likelihood (MPL). The results are presented in Table 6.<sup>35</sup> As a rule of thumb, the best model is the one with the smallest DIC and WAIC, and the largest MPL.

<sup>35</sup> The estimates were obtained using the software package R-INLA (R-version 4.0.3, Integrated Nested Laplace Approximation-INLA). The code is presented in Online Resource 2.

**Table 6** Deviance information criterion (DIC), Watanabe–Akaike information criterion (WAIC) and marginal predictive likelihood (MPL) of a subset of models of the FGG-GMRF model in Eq. (33)<sup>a</sup>

Model specification <sup>b</sup>	Likelihood	Time-varying effect	Edge length (m)	DIC	WAIC	MPL
M1.1.1.0	Poisson	RW1	–	5184.089	5623.433	–2850.213
M1.2.1.0	NB	RW1	–	3281.386	3283.865	–1778.259
M1.1.2.0	Poisson	RW2	–	5229.130	5616.038	–2893.252
M1.2.2.0	NB	RW2	–	3380.390	3381.260	–1941.200
M2.1.1.1	Poisson	RW1	1000	2657.428	2640.887	–1582.563
M2.2.1.1	NB	RW1	1000	2676.950	2671.954	–1583.277
M2.1.1.2	Poisson	RW1	1100	2663.998	2660.371	–1579.446
M2.2.1.2	NB	RW1	1100	2676.937	2671.849	–1582.548
M2.1.1.3	Poisson	RW1	1200	2663.953	2660.405	–1580.599
M2.2.1.3	NB	RW1	1200	2680.719	2676.344	–1582.903
M2.1.1.4	Poisson	RW1	1300	2658.124	2642.776	–1583.341
M2.2.1.4	NB	RW1	1300	2681.375	2675.452	–1582.484
M2.1.1.5	Poisson	RW1	1400	2663.895	2661.152	–1580.897
M2.2.1.5	NB	RW1	1400	2680.761	2676.639	–1582.581
M2.1.1.6	Poisson	RW1	1500	2664.065	2661.030	–1580.177
M2.2.1.6	NB	RW1	1500	2680.761	2676.639	–1582.581
M2.1.2.1	Poisson	RW2	1000	2658.819	2644.598	–1597.562
M2.2.2.1	NB	RW2	1000	2669.423	2653.978	–1599.671
M2.1.2.2	Poisson	RW2	1100	2658.947	2642.814	–1599.163
M2.2.2.2	NB	RW2	1100	2658.947	2642.814	–1599.163
M2.1.2.3	Poisson	RW2	1200	2665.287	2663.565	–1594.933
M2.2.2.3	NB	RW2	1200	2681.804	2677.917	–1597.695
M2.1.2.4	Poisson	RW2	1300	2659.083	2643.993	–1598.382
M2.2.2.4	NB	RW2	1300	2683.923	2679.331	–1596.700
M2.1.2.5	Poisson	RW2	1400	2665.050	2662.933	–1595.102
M2.2.2.5	NB	RW2	1400	2683.923	2679.331	–1596.700
M2.1.2.6	Poisson	RW2	1500	2664.968	2662.321	–1595.516
M2.2.2.6	NB	RW2	1500	2683.821	2679.441	–1596.753
M3.1.1.1	Poisson	RW1	1000	2660.567	2644.342	–1586.289
M3.2.1.1	NB	RW1	1000	2667.208	2659.532	–1588.466
M3.1.1.2	Poisson	RW1	1100	2658.277	2641.471	–1585.928
M3.2.1.2	NB	RW1	1100	2667.213	2659.537	–1588.420
M3.1.1.3	Poisson	RW1	1200	2658.277	2641.472	–1585.928
M3.2.1.3	NB	RW1	1200	2667.213	2659.537	–1588.419
M3.1.1.4	Poisson	RW1	1300	2660.667	2644.078	–1586.228
M3.2.1.4	NB	RW1	1300	2667.213	2659.536	–1588.410
M3.1.1.5	Poisson	RW1	1400	2660.510	2644.422	–1586.429
M3.2.1.5	NB	RW1	1400	2667.191	2659.512	–1588.405
M3.1.1.6	Poisson	RW1	1500	2660.598	2644.207	–1586.461
M3.2.1.6	NB	RW1	1500	2667.191	2659.512	–1588.400

**Table 6** (continued)

Model specification <sup>b</sup>	Likelihood	Time-varying effect	Edge length (m)	DIC	WAIC	MPL
M3.1.2.1	Poisson	RW2	1000	2661.441	2647.045	- 1600.555
M3.2.2.1	NB	RW2	1000	2668.788	2663.003	- 1602.876
M3.1.2.2	Poisson	RW2	1100	2659.836	2645.165	- 1599.883
M3.2.2.2	NB	RW2	1100	2659.836	2645.165	- 1599.883
M3.1.2.3	Poisson	RW2	1200	2659.544	2645.097	- 1599.448
M3.2.2.3	NB	RW2	1200	2668.788	2663.003	- 1602.876
M3.1.2.4	Poisson	RW2	1300	2661.572	2646.658	- 1599.695
M3.2.2.4	NB	RW2	1300	2668.788	2663.003	- 1602.871
M3.1.2.5	Poisson	RW2	1400	2661.619	2647.490	- 1600.489
M3.2.2.5	NB	RW2	1400	2668.803	2663.020	- 1602.851
M3.1.2.6	Poisson	RW2	1500	2661.708	2647.171	- 1600.532
M3.2.2.6	NB	RW2	1500	2668.803	2663.020	- 1602.847

The second number refers to the likelihood: 1: Poisson and 2: Negative Binomial, the third to the time varying effect of the coefficients, 1: random walk of order 1 (RW1) and 2: random walk of order 2 (RW2), the fourth digit defines the edge length, 0: edge length not relevant, 1: 1000 m, 2: 1,100 m, 3: 1,200 m, 4: 1,300 m, 5: 1,400 m, and 6: 1,500 m. Note that there is no edge length for the M1 models because they do not need meshing. M1 contains 4 sub-models, and M2 and M3 contain 24 sub-models each

<sup>a</sup>The FGG-GMRF models are specified as  $E[y_t | E_t, \theta_t]$

<sup>b</sup>The following notation applies. M denotes Model. M1 is the sub-model with covariates only, i.e.,  $\eta_t = \beta_0 1_{(n_A+n_p)} + \sum_{k=1}^4 (\beta_k + \zeta_{k,t}) z_{k,t}$ , M2 the model with covariates plus area and cell level interaction, i.e.,  $\eta_t = \beta_0 1_{(n_A+n_p)} + \sum_{k=1}^K (\beta_k + \zeta_{k,t}) z_{k,t} + \delta_t + \Phi_t$  (Area interaction type IV with precision matrix  $Q_\delta = \tau_\delta R_\omega \otimes R_\phi$ ). M3 the full model, i.e.,  $\eta_t = \beta_0 1_{(n_A+n_p)} + \sum_{k=1}^K (\beta_k + \zeta_{k,t}) z_{k,t} + \dot{\omega} + \ddot{\omega} + \phi_t 1_{(n_A+n_p)} + \zeta_t 1_{(n_A+n_p)} + \delta_t + \Phi_t$

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10109-021-00368-0>.

**Acknowledgements** We thank the Bandung City Health Office for providing the data. This research was funded by ALG Unpad contract: 1427/UN6.3.1/LT/2020.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## 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
- Abiodun G, Maharaj R, Witbooi P, Okosun K (2016) Modelling the influence of temperature and rainfall on the population dynamics of *Anopheles arabiensis*. *Malar J* 15(364):1–15
- Abramovitz M, Stegun I (1965) *Handbook of mathematical functions*. Dover Publications, New York
- Aguayo G, Schritz A, Ruiz-Castell M, Villarroel L, Valdivia G, Fagherazzi G, Valdivia G, Fagherazzi G, Lawson A (2020) Identifying hotspots of cardiometabolic outcomes based on a Bayesian approach: the example of Chile. *PLoS ONE* 15(6):1–16
- Ahmadian H, Friswell M, Mottershead J (1998) Minimization of the discretization error in mass and stiffness formulation by an inverse method. *Int J Numer Methods Eng* 41(2):371–378
- Ak C, Ergonul O, Şencan I, Torunoğlu MA, Gonen M (2018) Spatiotemporal prediction of infectious diseases using structured Gaussian processes with application to Crimean-Congo hemorrhagic fever. *PLoS Negl Trop Dis* 12(8):1–20
- Arboleda S, Jaramillo ON, Peterson A (2009) Mapping environmental dimensions of dengue fever transmission risk in the Aburrá Valley Colombia. *Int J Environ Res Public Health* 6(12):3040–3055
- Bakka H (2019) How to solve the stochastic partial differential equation that gives a Matérn random field using the finite element method, pp 1–17. <https://arxiv.org/abs/1803.03765> [stat.CO]
- Bakka H, Krainski E, Bolin D, Rue H, Lindgren F (2020). The diffusion-based extension of the Matérn field to space-time, pp 1–22. <https://arxiv.org/abs/2006.04917> [stat.ME]
- Bambrick H, Woodruff R, Hanigan I (2009) Climate change could threaten blood supply by altering the distribution of vector-borne disease: An Australian case-study. *Glob Health Action* 2(1):1–11
- Bandung Central Statistical Bureau (2012) Bandung City in Figure 2012. Bandung Government, Bandung
- Bandung Central Statistical Bureau (2013) Bandung City in Figure 2013. Bandung Government, Bandung
- Bandung Central Statistical Bureau (2014) Bandung City in Figure 2014. Bandung Government, Bandung
- Bandung Central Statistical Bureau (2015) Bandung City in Figure 2015. Bandung Government, Bandung
- Bandung Central Statistical Bureau (2016) Bandung City in Figure 2016. Bandung Government, Bandung
- Bandung Central Statistical Bureau (2017) Bandung City in Figure 2017. Bandung Government, Bandung
- Bandung Central Statistical Bureau (2018) Bandung City in Figure 2018. Bandung Government, Bandung
- Bandung Health Department (2013) Health Profile of Bandung Municipality in 2012. Bandung Government, Bandung
- Bandung Health Department (2014) Health Profile of Bandung Municipality in 2013. Bandung Government, Bandung
- Bandung Health Department (2015) Health Profile of Bandung Municipality in 2014. Bandung Government, Bandung
- Bandung Health Department (2016) Health Profile of Bandung Municipality in 2015. Bandung Government, Bandung
- Bandung Health Department (2017) Health Profile of Bandung Municipality in 2016. Bandung Government, Bandung
- Bandung Health Department (2018) Health Profile of Bandung Municipality in 2017. Bandung Government, Bandung
- Bandung Health Department (2019) Health Profile of Bandung Municipality in 2018. Bandung Government, Bandung
- Banerjee S, Gelfand A (2002) Prediction interpolation and regression for spatially misaligned data. *Sankhyā: Indian J Stat* 64(2):227–245
- Banerjee S, Carlin B, Gelfand A (2015) *Hierarchical modeling and analysis for spatial data*, 2nd edn. CRC Press Taylor and Francis Group, Boca Raton
- Barber X, Conesa D, Lladosa S, López-Quiñez A (2016) Modelling the presence of disease under spatial misalignment using Bayesian latent Gaussian models. *Geospat Health* 11(415):11–20

- Benedum C, Seidahmed O, Eltahir E, Markuzon N (2018) Statistical modeling of the effect of rainfall flushing on dengue transmission in Singapore. *PLoS Negl Trop Dis* 12(12):1–18
- Berk R, MacDonald J (2008) Overdispersion and Poisson regression. *J Quant Criminol* 24(3):269–284
- 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
- Bivand R, Gómez-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
- Bohn J, Feischl M (2021) Recurrent neural networks as optimal mesh refinement strategies. *Comput Math Appl* 97:61–76. <https://doi.org/10.1016/j.camwa.2021.05.018>
- Bolin D, Lindgren F (2009) Wavelet Markov models as efficient alternatives to tapering and convolution fields. *Mathematical Sciences Preprint* 13. Lund University, Lund
- Cameletti M, Lindgren F, Simpson D, Rue H (2013) Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *ASTA Adv Stat Anal* 97(2):109–131
- Yb C, Chen Xh, Hi Li, Zy C, Jiang R, Lü J, Fu Hd (2018) Analysis and comparison of Bayesian methods for measurement uncertainty evaluation. *Math Probl Eng*. <https://doi.org/10.1155/2018/7509046>
- Coly S, Garrido I, Abrial D, Yao AF (2021) Bayesian hierarchical models for disease mapping applied to contagious pathologies. *PLoS ONE* 16(1):1–28
- Ebi K, Nealon J (2016) Dengue in a changing climate. *Environ Res* 151(1):115–123
- Ehelepola NDB, Ariyaratne K, Buddhadasa WNMP, Ratnayake S, Wickramasinghe M (2015) A study of the correlation between dengue and weather in Kandy city Sri Lanka (2003–2012) and lessons learned. *Infect Dis Poverty* 4(1):42–55
- Ekasari R, Susanna D, Riskiyani S (2018) Climate factors and dengue fever in Jakarta 2011–2015. *KnE Life Sci* 4(4):151–160
- Fahrmeir L, Lang S (2001) Bayesian inference for generalized additive mixed models based on Markov random field prior. *Appl Stat* 50(2):201–220
- Fick SE, Hijmans RJ (2017) WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *Int J Climatol* 37(2):4302–4315
- Franco-Villoria M, Ventrucci M, Rue H (2019) A unified view on Bayesian varying coefficient models. *Electron J Stat* 13(2):5334–5359
- French J, Wand M (2004) Generalized additive models for cancer mapping with incomplete covariates. *Biostatistics* 5(2):177–191
- Fuentes M, Raftery AE (2005) Model evaluation and spatial interpolation by Bayesian combination of observations with outputs from numerical models. *Biometrics* 61(1):36–45
- Fuentes M, Chen L, Davis JM (2008) A class of nonseparable and nonstationary spatial temporal covariance functions. *Environmetrics* 19(5):487–507
- Fuglstad GA, Hem I, Knight A, Rue H, Riebler A (2020) Intuitive joint priors for variance parameters. *Bayesian Anal* 15(4):1109–1137
- Gelman A (2006) Prior distribution for variance parameters in hierarchical models. *Bayesian Anal* 1(3):515–533
- Gelman A, Simpson D, Betancourt M (2017) The prior can often only be understood in the context of the likelihood. *Entropy* 19(555):1–13
- Gneiting T (2002) Nonseparable, stationary covariance functions for space-time data. *J Am Stat Assoc* 97(458):590–600
- Godana AA, Mwalili SM, Orwa GO (2019) Dynamic spatiotemporal modeling of the infected rate of visceral leishmaniasis in humans in an endemic area of Amhara regional state Ethiopia. *PLoS ONE* 14(3):1–21
- Goicoa T, Adin A, Ugarte M, Hodges J (2018) In spatio-temporal disease mapping models, identifiability constraints affect PQL and INLA results. *Stoch Environ Res Risk Assess* 32(3):749–770
- Gómez-Rubio V (2020) Bayesian inference with INLA. Chapman & Hall/CRC Press, Boca Raton
- Gómez-Rubio V, Bivand R, Rue H (2021) Spatial models using Laplace Approximation methods. In: Fischer MM, Nijkamp P (eds) *Handbook of regional science, second and, extended*. Springer, Berlin, pp 1943–1959
- Haining R, Li G (2020) Modelling spatial and spatial-temporal data a Bayesian approach. CRC Press Taylor & Francis Group, Boca Raton

- Hales S, de Wet N, Maindonald J, Woodward A (2002) Potential effect of population and climate changes on global distribution of dengue fever: an empirical model. *Lancet* 360(9336):830–834
- Hall S, Swamy P, Tavlas G (2016) Time-varying coefficient models: a proposal for selecting the coefficient drivers sets. *Macroecon Dyn* 21(5):1158–1174
- Hamura Y, Irie K, Sugawara S (2021) On global-local shrinkage priors for count data. *Bayesian Anal*. <https://doi.org/10.1214/21-BA1263>
- Handayani N, Ariyanti D (2012) Potency of solar energy applications in Indonesia. *Int J Renew Energy Dev* 1(2):33–38
- Hanigan I, Chaston T, Hinze B, Dennekamp M, Jalaludin B, Kinfu Y, Morgan G (2019) A statistical downscaling approach for generating high spatial resolution health risk maps: A case study of road noise and ischemic heart disease mortality in Melbourne. *Australia Int J Health Geogr* 18(1):20–29
- Hastie T, Tibshirani R (1986) Generalized additive models. *Stat Sci* 1(3):297–318
- Hsu J, Hsieh CL, Lub C (2017) Trend and geographic analysis of the prevalence of dengue in Taiwan 2010–2015. *J Glob Infect Dis* 54:43–49. <https://doi.org/10.1016/j.ijid.2016.11.008>
- Hurtado-Díaz M, Riojas-Rodríguez H, Rothenberg S, Gomez-Dantés H, Cifuentes E (2007) Short communication: impact of climate variability on the incidence of dengue in Mexico. *Trop Med Int Health* 12(11):1327–1337
- 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, vol 2. methodological advances, regional systems modeling and open sciences*. Springer, West Virginia, pp 239–255
- Jaya IGNM, Folmer H (2020) Bayesian spatiotemporal mapping of relative dengue disease risk in Bandung Indonesia. *J Geogr Syst* 22(1):105–142
- Jaya IGNM, Folmer H (2021a) Identifying spatiotemporal clusters by means of agglomerative hierarchical clustering and Bayesian regression analysis with spatiotemporally varying coefficients: methodology and application to dengue disease in Bandung Indonesia. *Geogr Anal* 53(4):767–817
- Jaya IGNM, Folmer H (2021b) Bayesian spatiotemporal forecasting and mapping of COVID-19 risk with application to West Java Province Indonesia. *J Reg Sci* 61(4):849–881
- Kammann EE, Wand MP (2003) Geoadditive models. *Appl Stat* 52(1):1–18
- Kampen GI, Engelfriet P, Pv B (2014) Disease prevention: saving lives or reducing health care costs? *PLoS ONE* 9(8):1–5
- Kang S, McGree J, Baade P, Mengersen K (2015) A case study for modelling cancer incidence using Bayesian spatio-temporal models. *Aust N Z J Stat* 57(3):325–345
- Kifle YW, Hens N, Faes C (2017) Cross-covariance functions for additive and coupled joint spatiotemporal SPDE models in R-INLA. *Environ Ecol Stat* 24(4):551–586
- Knorr-Held L (2000) Bayesian modeling of inseparable space-time variation in disease risk. *Stat Med* 19(17–18):2555–2567
- Lambrechts L, Paaijmans K, Fansiri T, Carrington L (2012) Impact of daily temperature fluctuations on dengue virus transmission by *Aedes aegypti*. *Proc Natl Acad Sci USA* 108(18):7460–7465
- Langtangen HP, Logg A (2016) Solving PDEs in Python. The FEniCS tutorial I. Springer Open, Cham
- Langtangen HP, Mardal KA (2019) Introduction to numerical methods for variational problems. Springer, Cham
- Lawson AB (2010) Hotspot detection and clustering: ways and means. *Environ Ecol Stat* 17(2):231–245
- Lawson AB, Choi J, Cai B, Hossain M, Kirby RS, Liu J (2012) Bayesian 2-stage space-time mixture modeling with spatial misalignment of the exposure in small area health data. *J Agric Biol Environ Stat* 17(3):417–441
- Lee M, Kloog I, Chudnovsky A, Lyapustin A, Wang Y, Melly S, Coull B, Koutrakis P, Schwartz J (2016) Spatiotemporal prediction of fine particulate matter using high resolution satellite images in the southeastern U.S 2003–2011. *J Exp Sci Environ Epidemiol* 26(4):377–384
- Leroux B, Lei X, Breslow N (2000) Estimation of disease rates in small areas: a new mixed model for spatial dependence. In: Halloran M, Berry D (eds) *Statistical models in epidemiology the environment and clinical trials*. Springer, New York, pp 179–191
- Lindgren F, Rue H (2015) Bayesian spatial modelling with R-INLA. *J Stat Softw* 63(19):1–25
- Lindgren F, Rue H, Lindström J (2011) An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *J R Stat Soc B* 73(4):423–498
- Lindquist D, Gilest M (1989) A comparison of numerical schemes on triangular and quadrilateral meshes. In: Dwoyer D, Hussaini M, Voigt R (eds) *11th International conference on numerical methods in fluid dynamics. Lecture notes in physics*. Springer, Berlin, pp 369–373

- Liu X, Bertazzon S (2016) Fine scale spatio-temporal modelling of urban air pollution. In: Miller J, O'Sullivan D, Wiegand N (eds) *Geographic Information Science*. Springer, Cham, pp 210–224
- Liu Z, Le ND, Zidek JV (2011) An empirical assessment of Bayesian melding for mapping ozone pollution. *Environmetrics* 22(3):340–353
- Ma W, Gu S, Wang Y, Zhang X, Wang A, Zhao N, Song Y (2014) The use of mixed generalized additive modeling to assess the effect of temperature on the usage of emergency electrocardiography examination among the elderly in Shanghai. *PLoS ONE* 9(6):1–10
- Martinez-Beneito M, Botella-Rocamora P (2019) *Disease mapping from foundations to multidimensional modeling*. Taylor & Francis Group, Boca Raton
- Martínez Bello DA, López-Quílez A, Torres-Prieto A (2017) Relative risk estimation of dengue disease at small spatial scale. *Int J Health Geogr* 16(31):1–15
- Martinez-Bello DA, Lopez-Quilez A, Torres-Prieto A (2017) Bayesian dynamic modeling of time series of dengue disease case counts. *PLoS Negl Trop Dis* 11(7):1–19
- Matheron G (1963) *Principles of geostatistics*. *Econ Geol* 58(8):1246–1266
- Messina JP, Brady OJ, Golding N, Kraemer MU, Wint GR, Ray SE, Velayudha, (2019) The current and future global distribution and population at risk of dengue. *Nat Microbiol* 4(9):1508–1515
- Miller D, Glennie R, Seaton A (2019) Understanding the stochastic partial differential equation approach to smoothing. *J Agric Biol Environ Stat* 25(1):1–16
- Montgomery D, Peck E, Vining G (2012) *Introduction to linear regression analysis*. Wiley, Hoboken
- Moraga P, Cramb S, Mengersen K, Pagano M (2017) A geostatistical model for combined analysis of point level and area level data using INLA and SPDE. *Spat Stat* 21(Part A):27–41
- Muleia R, Boothe M, Loquiha O, Aerts M, Faes C (2020) Spatial distribution of HIV prevalence among young people in Mozambique. *Int J Environ Res Public Health* 17(3):885–904
- Osei F, Stein A (2017) Diarrhea Morbidity in small areas: accounting for non-stationarity in sociodemographic impacts using Bayesian spatially varying coefficient modelling. *Sci Rep* 7(1):9908–9922
- Payne E, Hardin J, Egede L, Ramakrishnan V, Selassie A, Gebregziabher M (2017) Approaches for dealing with various sources of overdispersion in modeling count data: scale adjustment versus modeling. *Stat Methods Med Res* 26(4):1802–1823
- Peng R, Bell M (2010) Spatial misalignment in time series studies of air pollution and health data. *Biostatistics* 11(4):720–740
- Phanitchat T, Zhao B, Haque U, Pientong C, Ekalaksananan T, Aromseree S, Overgaard H (2019) Spatial and temporal patterns of dengue incidence in northeastern Thailand 2006–2016. *BMC Infect Dis* 19(1):743–755
- Pokharel G, Deardon R (2016) Gaussian process emulators for spatial individual level models of infectious disease. *Can J Stat* 44(4):480–501
- Puggioni G, Couret J, Serman E, Akanda A, Ginsberg H (2020) Spatiotemporal modeling of dengue fever risk in Puerto Rico. *Spat Spatio-Temporal Epidemiol* 35:100375–100383. <https://doi.org/10.1016/j.sste.2020.100375>
- Rasjid A, Yudhastuti R, Notobroto HB, Hartono R (2019) Climate change: An overview of the prevalence of dengue hemorrhagic fever in the South Sulawesi province of Indonesia. *Indian J Public Health Res Dev* 10(8):1982–1986
- Righetto AJ, Faes C, Vandendijck Y, Ribeiro PJ Jr (2018) On the choice of the mesh for the analysis of geostatistical data using R-INLA. *Commun Stat* 49(1):203–220
- Rue H, Held L (2005) *Gaussian Markov random fields: theory and applications*. Chapman and Hall/CRC, Boca Raton
- Rue H, Martino S, Chopin N (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J R Stat Soc Ser B* 7(2):319–392
- Saez M, López-Casasnovas G (2019) Assessing the effects on health inequalities of differential exposure and differential susceptibility of air pollution and environmental noise in Barcelona, 2007–2014. *Int J Environ Res Public Health* 16(18):340–362
- Sahu S, Gelfand A, Holland D (2010) Fusing point and areal level space–time data with application to wet deposition. *J R Stat Soc Ser C Appl Stat* 59(1):77–103
- Schrödle B, Held L (2011) Spatio-temporal disease mapping using INLA. *Environmetrics* 22(6):725–734
- Sedda L, Vilela AP, Aguiá ER, Gaspar CH, Gonçalves AN, Olmo RP, Silva ATS, Silveira LC, Drumond BP, Marques JT (2018) The spatial and temporal scales of local dengue virus transmission in natural settings: A retrospective analysis. *Parasites Vectors* 11(1):79–92
- Sherman M (2011) *Spatial statistics and spatio-temporal data*. Wiley, West Sussex

- Shi X, Miller S, Mwenda K, Onda A, Rees J, Onega T, Moeschler J (2013) Mapping disease at an approximated individual level using aggregate data: a case study of mapping New Hampshire birth defects. *Int J Environ Res Public Health* 10(9):4161–4167
- Sidén P, Lindgren F, Bolin D, Villani M (2018) Efficient covariance approximations for large sparse precision matrices. *J Comput Graph Stat* 27(4):898–909
- Simpson D, Lindgren F, Rue H (2012) Think continuous: Markovian Gaussian models in spatial statistics. *Spat Stat* 1:16–29. <https://doi.org/10.1016/j.spasta.2012.02.003>
- Simpson D, Rue H, Riebler A, Martins T, Sørbye S (2017) Penalising model component complexity: a principled, practical approach to constructing priors. *Stat Sci* 32(1):1–28
- Sloan S (1993) A fast algorithm for generating constrained Delaunay triangulations. *Comput Struct* 47(3):441–450
- Song C, Sh X, Wang J (2020) Spatiotemporally Varying Coefficients (STVC) model: a Bayesian local regression to detect spatial and temporal nonstationarity in variables relationships. *Ann GIS* 26(3):277–291
- Song HR, Fuentes M, Ghosh S (2008) A comparative study of Gaussian geostatistical models and Gaussian Markov random field models. *J Multivar Anal* 99(8):1681–1697
- Sørbye SH (2013) Tutorial: scaling IGMRF-models in R-INLA. University of Tromsø, Tromsø, Department of Mathematics and Statistics
- Sørbye SH, Rue H (2017) Penalised complexity priors for stationary autoregressive processes. *J Time Ser Anal* 38(6):923–935
- Sørbye SH, Rue H (2014) Scaling intrinsic Gaussian Markov random field priors in spatial modelling. *Spat Stat* 8:39–51. <https://doi.org/10.1016/j.spasta.2013.06.004>
- Sparks C (2015) An examination of disparities in cancer incidence in Texas using Bayesian random coefficient models. *PeerJ* 3:1–24. <https://doi.org/10.7717/peerj.1283>
- Stein, (1999) Interpolation of spatial data: some theory for Kriging. Springer, New York
- Truong P, Heuvelink G, Pebesma E (2014) Bayesian area-to-point Kriging using expert knowledge as informative priors. *Int J Appl Earth Obs Geoinform* 30:128–138. <https://doi.org/10.1016/j.jag.2014.01.019>
- Tu J, Yeoh GH, Liu C (2018) Computational fluid dynamics: a practical approach. Butterworth-Heinemann, Oxford
- Utazi C, Thorley J, Alegana V, Ferrari M, Nilsen K, Takahashi S, Tatem A (2019) A spatial regression model for the disaggregation of areal unit-based data to high-resolution grids with application to vaccination coverage mapping. *Stat Methods Med Res* 28(10–11):3226–3241
- Wand H, Whitaker C, Ramjee G (2011) Geoadditive models to assess spatial variation of HIV infections among women in Local communities of Durban South Africa. *In J Health Geogr* 10:28–36. <https://doi.org/10.1186/1476-072X-10-28>
- Wang X, Yue YR, Faraway J (2018a) Bayesian regression modeling with INLA. Taylor and Francis Group LLC, Boca Raton
- Wang C, Puhan M, Furrer R (2018b) Generalized spatial fusion model framework for joint analysis of point and areal data. *Spat Stat* 23:72–90. <https://doi.org/10.1016/j.spasta.2017.11.006>
- Whittle P (1963) Stochastic processes in several dimensions. *Bull Inst Int Stat* 40:974–994
- Whittle P (1954) On stationary processes in the plane. *Biometrika* 41(3–4):434–449
- Wilastonegoro N, Kharisma D, Laksono I, Halasa-Rappel Y, Brady O, Shepard D (2020) Cost of dengue illness in Indonesia across hospital, ambulatory, and not medically attended settings. *Am J Trop Med Hyg* 103(5):2029–2039
- WorldClim (2020) Global climate and weather data, version 2.1. WorldClim: <https://www.worldclim.org/>. Accessed 2 May 2020
- Xu Y, Cancino-Muñoz I, Torres-Puente M, Villamayor L, Borrás R, Borrás-Mañez M, Escribano I (2019) High-resolution mapping of tuberculosis transmission: Whole genome sequencing and phylogenetic modelling of a cohort from Valencia Region. *Spain Plos Med* 6(10):1–20
- 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(1):387–407
- Zellweger R, Cano J, Mangeas M, Taglioni F, Mercier A, Despinoy M, 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