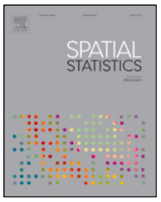
Spatial Statistics 54 (2023) 100744

Contents lists available at ScienceDirect

Spatial Statistics

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



Data fusion in a two-stage spatio-temporal model using the INLA-SPDE approach

Stephen Jun Villejo [a,b,*], Janine B Illian [a], Ben Swallow [a,c]

[a] University of Glasgow, Glasgow, United Kingdom
[b] University of the Philippines Diliman, Quezon City, Philippines
[c] University of St Andrews, St Andrews, United Kingdom

ARTICLE INFO

Article history:
Received 11 November 2022
Received in revised form 27 January 2023
Accepted 16 March 2023
Available online 29 March 2023

Keywords:
Integrated nested Laplace approximation (INLA)
Stochastic partial differential equations (SPDE)
Data fusion
Spatial misalignment

ABSTRACT

This paper proposes a two-stage estimation approach for a spatial misalignment scenario that is motivated by the epidemiological problem of linking pollutant exposures and health outcomes. We use the integrated nested Laplace approximation method to estimate the parameters of a two-stage spatio-temporal model — the first stage models the exposures using data fusion while the second stage links the health outcomes to exposures. The first stage is based on the Bayesian melding model, which assumes a common latent field for the different data sources for the pollutants. The second stage fits a generalized linear mixed model using the spatial averages of the estimated latent field, and additional spatial and temporal random effects. Uncertainty from the first stage is accounted for by simulating repeatedly from the posterior predictive distribution of the latent field. A simulation study was carried out to assess the impact of the sparsity of the data on the monitors, number of time points, and the specification of the priors in terms of the biases, RMSEs, and coverage probabilities of the parameters and the block-level exposure estimates. The results show that the parameters are generally estimated correctly but there is difficulty in estimating the Matèrn field parameters. The effect of exposures on the health outcomes is the primary parameter of interest for spatial epidemiologists and health policy makers, and our results show that the proposed method estimates these very well. The proposed method is applied to measurements of NO_2 concentration and respiratory hospitalizations for year 2007 in

* Corresponding author at: University of Glasgow, Glasgow, United Kingdom.
E-mail addresses: Stephen.Villejo@glasgow.ac.uk (S.J. Villejo), Janine.Illian@glasgow.ac.uk (J.B. Illian), bts3@st-andrews.ac.uk (B. Swallow).

https://doi.org/10.1016/j.spasta.2023.100744
2211-6753/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

England. The results show that an increase in NO2 levels is significantly associated with an increase in the relative risks of the health outcome. Also, there is a strong spatial structure in the risks, a strong temporal autocorrelation, and a significant spatio-temporal interaction effect.

© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Spatially misaligned data, where the data are measured at different spatial scales, are a common issue in spatial modeling (Lawson et al., 2016). In ecological and epidemiological research for example, it is of interest to estimate the effect of pollution exposures on health outcomes such as incidence of certain diseases (Cameletti et al., 2019; Blangiardo et al., 2016; Lee et al., 2017). Measurements of exposures are typically collected from a network of monitoring stations, from satellites, or from outputs of computer simulations using numerical models describing the origins and dispersion of exposures. Data from monitoring stations are typically sparsely located in the spatial domain, while the two latter measurements are high resolution data on fine regular grids. The former are classified as point-referenced or geostatistical data, while the latter are considered as high-resolution areal data on fine regular grids (Bruno et al., 2016). Moreover, data on health outcomes such as the incidence of certain diseases are available as aggregated counts on fixed and irregular areal units. Such data are organized as counts of mortality, morbidity, or hospital admissions by type of disease over a set of regions partitioning the entire study region for each time point (Bruno et al., 2016). Although these data derive from individual-level information, which would enable the analysis of spatial pattern of case event locations and to quantify the effects of long-term exposure, individual-level data at this resolution are usually not available and are also costly to obtain in practice, and hence have been summarized across administrative units (Bruno et al., 2016; Diggle, 2013; Molitor et al., 2006). As a result, time series and areal study designs have been frequently used to model health outcome data. However, this comes at a cost since the summary statistics on the population across areas cannot be used to assess effects at the individual level. Moreover, there is the danger of assuming that the associations observed from the areal level also hold at the individual level, often referred to as the *ecological fallacy* (Bruno et al., 2016).

This paper focuses on a specific spatial misalignment scenario where the outcome of interest is measured and available in areas or blocks, while the explanatory variables include both point-referenced and high-resolution grid/raster data. A sensible classical approach for this kind of spatial misalignment is to compute values of the explanatory variables at the level of the administrative areas or blocks and to then apply standard statistical methods to regress the outcome of interest against the block-level exposure values (Bruno et al., 2016); in other words, the analysis is performed in two stages.

In the context of the problem which motivates this paper, the first stage computes pollution exposure values at the level of the administrative areas or blocks, also termed *upscaling* or *smoothing*, using the points data from the monitors and the high-resolution grid data from the satellite and the dispersion models. Linking pollution exposures and health outcomes therefore necessitates inference on a spatial scale that is different from the scale of the original data, referred to as a change of support problem (COSP). A classical approach to this point-to-area COSP in geostatistics is block kriging (Gotway and Young, 2002). However, kriging approaches require inversion of matrices, which can be computationally expensive especially for large datasets. This paper hence proposes a point-to-area upscaling method that does not require inverting large covariance matrices. If there were no high-resolution data available, a naive approach would be to simply calculate the average of the values from the monitoring stations inside a block, possibly with the use of distance-based or population-based weights, and use this as the block-level value (Bruno et al., 2016). However,

this approach does not work when there are blocks without point-referenced data due to the sparsity of monitoring stations in the entire spatial domain, or if the values exhibit strong spatial heterogeneity (Lee et al., 2015; Krall et al., 2015). This issue may be addressed by incorporating information from satellite data and high-resolution output of numerical models. This process of merging information from different data sources is referred to as *data fusion* (Berrocal et al., 2012). One approach to performing data fusion, called *Bayesian melding*, is to regard the point-referenced data and the high-resolution grid data as outcomes from a common latent continuous random process with some measurement error, and possibly some covariates and spatial and temporal effects (Cameletti et al., 2019; Szpiro and Paciorek, 2013; Bergen and Szpiro, 2015; Fuentes and Raftery, 2005).

In the second stage, the block-level exposure estimates from the first stage are used as an input in a statistical model for the health outcomes. A naive approach is to perform the first stage independently of the second stage. This assumes that the estimated exposure values at the block level are free from estimation error. Since this approach does not account for the uncertainty in the first-stage model, this can potentially lead to biased estimates of the health effects of exposures and can also underestimate the uncertainty in the second stage model estimates (Cameletti et al., 2019; Gryparis et al., 2009). This approach is termed the *plug-in approach* and has been shown to yield biased estimates, especially when the data from the monitors are sparse and the values are heterogeneous in space (Gryparis et al., 2009). An approach to overcome this issue is a two-stage Bayesian approach, in which the first stage estimates the posterior distribution of the parameters and of the latent exposures field, while the second stage estimates the health model by plugging in simulated values from the posterior predictive distribution of the first-stage model (Gryparis et al., 2009; Lee and Shaddick, 2010). Another approach is a fully Bayesian approach which simultaneously estimates the parameters of the exposures model and the health model. However, in this framework, the health outcomes could strongly influence the exposure surface, especially if the number of monitoring stations is relatively small. Nonetheless, the two-stage Bayesian estimator is shown to approximate the fully Bayesian results quite well (Gryparis et al., 2009).

This paper aims to propose and illustrate an approach for a data fusion problem and a change of support problem when estimating the effect of exposures on health outcomes. The proposed method uses the Bayesian melding model to perform data fusion and also to account for the measurement error in the different data sources that we wish to integrate when doing modeling and analysis. The models are latent Gaussian and are estimated using the *integrated nested Laplace approximation* (INLA). Moreover, the *stochastic partial differential equation* approach is used to efficiently estimate the spatial field (Lindgren et al., 2011; Rue et al., 2009).

The use of the INLA approach is motivated by its computational benefits. The INLA method is a deterministic approach for doing Bayesian inference, as opposed to *Markov Chain Monte Carlo* (MCMC) method which is a simulation-based approach (Rue et al., 2009). For particular spatial applications for which the size of the latent Gaussian field is large, the INLA method was able to compute accurate approximations of the posterior marginals which takes longer for MCMC algorithms to compute. In addition, the approximation bias is smaller than the MCMC error for typical examples. Fitting Gaussian random field (GRF) models can be computationally demanding because of the need to factorize and invert typically large covariance matrices. The SPDE approach continuously approximates the GRF with a *Gaussian Markov Random Field* (GMRF) structure, resulting in a sparse *precision matrix* and hence speeding up the factorization of the matrix. In addition to making the computation efficient, the SPDE approach also allows the interpretation of the model in terms of the parameters of the covariance function (Cameletti et al., 2019).

Section 2 provides a discussion of the Bayesian melding model for data fusion. Section 3 discusses the proposed spatio-temporal model and the assumptions of the model. It also discusses in some detail how INLA and the SPDE approach can be used to fit the proposed model. Section 4 discusses the specifics of the simulation study, while Section 5 presents the results of the simulation study. Finally, Section 6 discusses an application of the method to real data.

2. Bayesian melding

Data used to conduct exposure assessment in any environmental health study come from various sources. Most studies use data from a network of monitoring stations set up and maintained by government agencies. However, since the network is typically sparse due to high maintenance costs of the stations, it will be difficult to capture spatial heterogeneity of the true exposure surface. Furthermore, networks are typically located in urban areas where the pollution level is typically high. This leads to biases when fitting models to assess the impact of exposures on health outcomes as the exposure surface will be over estimated (Lawson et al., 2016).

A solution to this problem is the use of additional sources of information of exposures which provide more detailed spatial and temporal information. Two specific sources are satellite images and computer simulations using numerical or deterministic models. Satellite images are remotely sensed data and provide global coverage. On the other hand, numerical models enable the simulation of the creation and dispersion of pollution exposures using information on pollution sources and on chemical and physical processes. However, despite the rich spatial and temporal information from these data, they are prone to error. Remotely sensed data are subject to retrieval errors. Computer simulations using numerical models are sensitive to model misspecification of the underlying process, the input data on pollution sources, and the discretization of the continuous field. These two data sources are also referred to as *proxy data* (Lawson et al., 2016). This problem of combining data from monitoring stations and proxy data from satellite images and computer simulations is referred to as *data fusion* or *data assimilation*.

An approach to data assimilation, called *Bayesian melding*, assumes that both the point-referenced data and the proxy data have a common latent spatial process (Fuentes and Raftery, 2005). The Bayesian melding model in a purely spatial context is based on the following equations:

$$w(\boldsymbol{s}) = x(\boldsymbol{s}) + e(\boldsymbol{s}) \tag{1}$$

$$x(\boldsymbol{s}) = \mu(\boldsymbol{s}) + \xi(\boldsymbol{s}) \tag{2}$$

$$\tilde{x}(\boldsymbol{s}) = \alpha_0(\boldsymbol{s}) + \alpha_1(\boldsymbol{s})x(\boldsymbol{s}) + \delta(\boldsymbol{s}) \tag{3}$$

$$\tilde{x}(B_i) = \frac{1}{|B_i|}\int_{B_i} \tilde{x}(\boldsymbol{s})d\boldsymbol{s}, \tag{4}$$

where $w(\boldsymbol{s})$ denotes the point-referenced random outcome at location $\boldsymbol{s}$, $\tilde{x}(B_i)$ is the outcome of the proxy data in grid cell B_i, and $x(\boldsymbol{s})$ is the latent spatial stochastic process. Eq. (1) states that the observed values at the monitoring stations are error-prone realizations of the latent process with $e(\boldsymbol{s})$ as the random error term. Eq. (2) postulates that the latent spatial stochastic process is decomposed into a mean process $\mu(\boldsymbol{s})$ and a residual process $\xi(\boldsymbol{s})$. The mean of $x(\boldsymbol{s})$ can be a function of covariates, while the residual component may include spatial effects. Eq. (3) models the proxy data at the point-level. The point-level outcome $\tilde{x}(\boldsymbol{s})$ is associated with the latent process as a linear function of $x(\boldsymbol{s})$, and with $\alpha_0(\boldsymbol{s})$ and $\alpha_1(\boldsymbol{s})$ as additive and multiplicative calibration parameters, respectively. In addition to that, there is also an additional random error term, $\delta(\boldsymbol{s})$, for $\tilde{x}(\boldsymbol{s})$. The incorporation of these calibration parameters is due to the assumption that the proxy data have more noise and are less correlated with the true latent field as compared to the observed values from the monitors (Lawson et al., 2016). The calibration parameters $\alpha_0(\boldsymbol{s})$ and $\alpha_1(\boldsymbol{s})$ are assumed to be either constant or varying in space, but they are usually considered as fixed effects to avoid identifiability issues when estimating $x(\boldsymbol{s})$. Finally, Eq. (4) defines the observed value of $\tilde{x}(B_i)$ as a spatial average of $\tilde{x}(\boldsymbol{s})$, where $|B_i|$ denotes the size of block B_i. Studies have shown that Bayesian melding outperforms kriging in predicting pollution fields (Berrocal et al., 2010; Liu et al., 2011). The idea of assuming a common latent spatial process for the observed data at different spatial scales was also used in Wikle and Berliner (2005), McMillan et al. (2010) and Sahu et al. (2010). The proposed method discussed in Section 3 extends these ideas in a spatio-temporal context.

3. Proposed method

There are two main model structures, one for the exposures (referred to as the *first-stage model*) and another one to link the health outcomes with the exposures (*second-stage model*). Section 3.1

discusses the model specification and assumptions. Sections 3.2 and 3.3 discusses the details of model estimation for the first-stage and second-stage model, respectively. When estimating the effect of exposures on the health outcomes using the two-stage modeling framework, it is important to account for the estimation error in the first stage. This issue is discussed in Section 3.3.1.

3.1. Model assumptions

3.1.1. First-stage model

The first-stage model borrows ideas from the Bayesian melding model in Eqs. (1)–(4), extending the original formulation to a spatio-temporal context. An extension of Eq. (2) which specifies the latent spatial process is given by:

$$x(\boldsymbol{s},t) = \beta_0 + \boldsymbol{\beta}^{\top}\boldsymbol{z}(\boldsymbol{s},t) + \xi(\boldsymbol{s},t)$$
$$\xi(\boldsymbol{s},t) = \varsigma\xi(\boldsymbol{s},t-1) + \omega(\boldsymbol{s},t), \quad |\varsigma| < 1, t = 2,\ldots,T.$$

This follows the spatio-temporal model proposed in Cameletti et al. (2013) for particulate matter concentration in the North-Italian region Piemonte. Here, $x(\boldsymbol{s},t)$ is the true exposures value at location $\boldsymbol{s}$ and time t, $\boldsymbol{z}(\boldsymbol{s},t)$ are the covariates, β_0 is the intercept of the model, and $\boldsymbol{\beta}$ is the coefficient vector of the covariates $\boldsymbol{z}(\boldsymbol{s},t)$. The spatio-temporal dependence in the model is induced by $\xi(\boldsymbol{s},t)$ which evolves in time as an autoregressive process of order 1, with ς as the autoregressive parameter and $\xi(\boldsymbol{s},1) \sim N(0, \sigma_\omega^2/(1-\varsigma^2))$. The term $\omega(\boldsymbol{s},t)$ is a temporally-independent Gaussian random field with mean 0 and Matèrn covariance function, i.e.,

$$\text{Cov}(\omega(\boldsymbol{s}_i,t),\omega(\boldsymbol{s}_j,u)) = \begin{cases} 0 & t \neq u \\ \Sigma_{i,j} & t = u, \end{cases}$$

with $\Sigma_{i,j} = \dfrac{\sigma_\omega^2}{2^{\nu-1}\Gamma(\nu)}(\kappa\|\boldsymbol{s}_i-\boldsymbol{s}_j\|)^{\nu}K_\nu(\kappa(\|\boldsymbol{s}_i-\boldsymbol{s}_j\|))$ where $\|\cdot\|$ is the Euclidean distance in $\mathbb{R}^2$ and $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ are two locations. In the Matèrn equation, σ_ω^2 is the marginal variance, K_ν is the modified Bessel function of the second kind and order $\nu > 0$ usually fixed at 1, and $\kappa > 0$ is a scaling parameter. The Matèrn parameters are σ_ω^2 and κ. The interpretation of the spatio-temporal structure of the model is usually in terms of the σ_ω^2 and the range parameter ρ, which is the distance at which the correlation is around 0.1. The empirically derived relationship between ρ and κ is $\rho = \dfrac{\sqrt{8\nu}}{\kappa}$.

Following Eq. (1) of the melding model, the model for the observed values at M monitors is given by:

$$w(\boldsymbol{s}_i,t) = x(\boldsymbol{s}_i,t) + e(\boldsymbol{s}_i,t), \quad e(\boldsymbol{s}_i,t) \overset{\text{iid}}{\sim} N(0,\sigma_e^2), \quad i = 1,\ldots,M, \tag{5}$$

where $w(\boldsymbol{s}_i,t)$ is the observed value at a monitor in location $\boldsymbol{s}_i$ at time t. This equation follows the classical error model in which the observed values at the monitors are error-prone realizations of the true exposures surface and where the error process is modeled as $e(\boldsymbol{s}_i,t) \overset{\text{iid}}{\sim} N(0,\sigma_e^2)$.

With the assumption that the resolution of the proxy data is reasonably fine, the proxy data is treated as geostatistical at the centroids. This provides a simplification of the Bayesian melding model since Eq. (3) is used immediately to model the proxy data instead of using both Eqs. (3) and (4). The proxy data is typically in the form of a raster data; hence, we use $\tilde{x}(B_i,t)$ to denote the observed value at the pixel or grid cell B_i. The model for the proxy data is then given by

$$\tilde{x}(B_i,t) = \tilde{x}(\boldsymbol{g}_i,t) = \alpha_0 + \alpha_1 x(\boldsymbol{g}_i,t) + \delta(\boldsymbol{g}_i,t), \quad \delta(\boldsymbol{g}_i,t) \overset{\text{iid}}{\sim} N(0,\sigma_\delta^2), \tag{6}$$

where $\boldsymbol{g}_i$ is the centroid of B_i. This simplification avoids the joint specification given by

$$\tilde{x}(\boldsymbol{s},t) = \alpha_0 + \alpha_1 x(\boldsymbol{s},t) + \delta(\boldsymbol{s},t), \quad \delta(\boldsymbol{s},t) \overset{\text{iid}}{\sim} N(0,\sigma_\delta^2) \tag{7}$$

$$\tilde{x}(B_i,t) = \frac{1}{|B_i|}\int_{B_i}\tilde{x}(\boldsymbol{s},t)d\boldsymbol{s}. \tag{8}$$

The simplification in Eq. (6) still incorporates an additive bias α_0, a multiplicative bias α_1, and an additional zero-mean random noise term with variance σ_δ^2. The proposed method is not a full Bayesian melding approach since both the proxy data and the observed data from the monitors are considered as geostatistical. The regression calibration method, which is another method to do data fusion, also treats the proxy data as geostatistical; hence this assumption is not new in the literature (Lee et al., 2017). Even with this simplification, the proposed model is still primarily based on the melding model by assuming that the monitors data and the proxy data are a function of a common latent spatial process as shown in Eqs. (5) and (6).

A particular scenario for a joint specification in Eqs. (7) and (8) to be infeasible is when there is no available covariate information whose resolution is finer than the resolution of the proxy data. For instance, if the resolution of the proxy data is the same for the covariate information $\boldsymbol{z}(\boldsymbol{s}, t)$, then it will not be feasible to fit both Eqs. (7) and (8) jointly unless it will be assumed that the value of $\boldsymbol{z}(\boldsymbol{s}, t)$ for any point $\boldsymbol{s}$ inside B is constant. Thus, an implicit assumption in the data generating process other than for the proxy data to have a high resolution is that the covariate information on $\boldsymbol{z}(\boldsymbol{s}, t)$ is collected at the same resolution as the proxy data and also at the level of the monitors. This simplification in the model assumptions also simplifies the estimation procedure for the first-stage model as would be presented in Section 3.2. Also, in doing the simulation study which would be presented in Section 4, the resolution of the simulation grid for the latent field is the same as the resolution of the proxy data but incorporating the biases α_0, α_1 and $\delta(\boldsymbol{s})$. On another important note, the proposed model assumes an *iid* error process for the monitors and the proxy data, as shown in Eqs. (5) and (6). But a more complex structure in the measurement error term could be adopted. Since the proposed method is based on the INLA method, a limitation in the construction of the model for the error process is that it should satisfy the latent Gaussianity assumption in order for the INLA method to be applicable.

Lastly, the true value of the exposures at an area or block B_i at time t, denoted by $x(B_i, t)$, is assumed to be a spatial average of the process $x(\boldsymbol{s}, t)$ over B_i, i.e.,

$$x(B_i, t) = \frac{1}{|B_i|} \int_{B_i} x(\boldsymbol{s}, t) d\boldsymbol{s}. \tag{9}$$

In terms of the notation, $x(B_i, t)$ denotes a spatial average of the latent process, while $\tilde{x}(B_i, t)$ is the observed value of the proxy data at a pixel B_i which is assumed to be equal to $\tilde{x}(\boldsymbol{g}_i, t)$, where $\boldsymbol{g}_i$ is the centroid of B_i, as shown in Eq. (6).

3.1.2. Second-stage model

The second-stage model specifies the model for the health outcomes as the response variable and the block-level spatial averages of $x(\boldsymbol{s}, t)$ defined in Eq. (9) as an input to the model. The observed count at an irregular block B_i at time t, denoted by $Y(B_i, t)$, is assumed to be a Poisson outcome, i.e.,

$$Y(B_i, t) \sim \text{Poisson}\big(P(B_i, t)\lambda(B_i, t)\big), \tag{10}$$

where $P(B_i, t)$ is the expected number of cases in area B_i at time t, and $\lambda(B_i, t)$ is the relative risk modeled as

$$\log(\lambda(B_i, t)) = \gamma_0 + \gamma_1 x(B_i, t) + \varphi_{it},$$

where γ_0 is the intercept, γ_1 is the coefficient of the true block-level exposure $x(B_i, t)$, and φ_{it} is a spatio-temporal random effect. A conditional independence assumption given the effects is assumed for the Poisson outcome, i.e., $Y(B_i, t)|\gamma_0, \gamma_1, \varphi_{it} \overset{ind}{\sim} \text{Poisson}\Big(\mu(B_i, t) = P(B_i, t)\lambda(B_i, t)\Big)$. The spatio-temporal random effect term can take several forms as discussed in Blangiardo and Cameletti (2015). A general form is given as

$$\log(\lambda(B_i, t)) = \gamma_0 + \gamma_1 x(B_i, t) + \phi_i + \psi_i + \zeta_t + \nu_t + \upsilon_{it}, \tag{11}$$

where ϕ_i is an *iid* spatial random effect, ψ_i is a structured spatial random effect, ν_t is an *iid* temporal random effect, ζ_t is a structured temporal random effect, and υ_{it} is a spatio-temporal

interaction effect. The interaction effect has four types, depending on which of the two spatial effects and which of the two time effects interact. Structured random effects include random walk models, autoregressive processes, and correlated random effects of different dimensions. These different ways to specify the spatio-temporal structure of the model do not violate latent Gaussianity, and hence all posterior marginals of the model parameters can be estimated using the INLA method (Blangiardo and Cameletti, 2015).

Fig. 1 shows a simulated data with number of time points $T = 3$: the counts at the irregular blocks $Y(B_i, t)$ (top), the proxy data $\tilde{x}(\boldsymbol{s}, t)$ (middle), and data from a network of monitoring stations $w(\boldsymbol{s}, t)$ (bottom). The true values of the parameters and other details of the simulation study are discussed in Section 4.

3.2. Model fitting for first-stage model

Suppose the data from the monitors at time t are given by $\boldsymbol{w}_t^\top = \begin{pmatrix} w(\boldsymbol{s}_1, t) & w(\boldsymbol{s}_2, t) & \dots & w(\boldsymbol{s}_M, t) \end{pmatrix}$, $t = 1, \dots, T$. We assume that we have T time points and M monitors. Also, the data from the proxy data at time t is given by $\tilde{\boldsymbol{x}}_t^\top = \begin{pmatrix} \tilde{x}(\boldsymbol{g}_1, t) & \tilde{x}(\boldsymbol{g}_2, t) & \dots & \tilde{x}(\boldsymbol{g}_G, t) \end{pmatrix}$, $t = 1, \dots, T$. Here, we denote by G the number of grid cells for the proxy data, and we denote by $\tilde{x}(\boldsymbol{g}_i, t)$ the observed value at the grid cell whose centroid is $\boldsymbol{g}_i$ at time t, $i = 1, \dots, G$. Since both $\boldsymbol{w}_t$ and $\tilde{\boldsymbol{x}}_t$ are error-prone realizations of the true exposure values $\boldsymbol{x}_t$, then we define the vector of true exposures combined for both the monitors and the proxy data at time t as $\boldsymbol{x}_t = \begin{pmatrix} \boldsymbol{x}_{t,M} & \boldsymbol{x}_{t,P} \end{pmatrix}^\top$, where $\boldsymbol{x}_{t,M}$ denotes the vector of true exposures at the monitors at time t, and $\boldsymbol{x}_{t,P}$ denotes the vector of true exposures at the centroids of the grid cells making up the proxy data at time t. Then the vector of true exposures for all $t = 1, \dots, T$ is denoted by $\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_1^\top & \boldsymbol{x}_2^\top & \dots & \boldsymbol{x}_T^\top \end{pmatrix}^\top$. Similarly, we define $\boldsymbol{\xi} = \begin{pmatrix} \boldsymbol{\xi}_1^\top & \boldsymbol{\xi}_2^\top & \dots & \boldsymbol{\xi}_T^\top \end{pmatrix}^\top$, where $\boldsymbol{\xi}_t^\top = \begin{pmatrix} \boldsymbol{\xi}_{t,M}^\top & \boldsymbol{\xi}_{t,P}^\top \end{pmatrix}$, with $\boldsymbol{\xi}_{t,M}$ as the vector of spatio-temporal random effects at the monitors for time t and $\boldsymbol{\xi}_{t,P}$ as the vector of spatio-temporal random effects at the centroids of the grid cells making up the proxy data at time t. And finally $\boldsymbol{\omega} = \begin{pmatrix} \boldsymbol{\omega}_1^\top & \boldsymbol{\omega}_2^\top & \dots & \boldsymbol{\omega}_T^\top \end{pmatrix}^\top$ as the vector of values of the Gaussian random field, where $\boldsymbol{\omega}_t^\top = \begin{pmatrix} \boldsymbol{\omega}_{t,M}^\top & \boldsymbol{\omega}_{t,P}^\top \end{pmatrix}$; and $\boldsymbol{z} = \begin{pmatrix} \boldsymbol{z}_1^\top & \boldsymbol{z}_2^\top & \dots & \boldsymbol{z}_T^\top \end{pmatrix}^\top$, as the vector of a single covariate, where $\boldsymbol{z}_t^\top = \begin{pmatrix} \boldsymbol{z}_{t,M}^\top & \boldsymbol{z}_{t,P}^\top \end{pmatrix}$. This is can easily be generalized to the case of more than one covariate.

The first-stage estimation procedure fits the following joint model:

$$\boldsymbol{w}_t = \boldsymbol{x}_{t,M} + \boldsymbol{e}_t, \quad \boldsymbol{e}_t \sim N(\boldsymbol{0}, \sigma_e^2 \mathbb{I}_M), \quad t = 1, \dots, T \tag{12}$$

$$\tilde{\boldsymbol{x}}_t = \alpha_0 \boldsymbol{1}_G + \alpha_1 \boldsymbol{x}_{t,P} + \boldsymbol{\delta}_t, \quad \boldsymbol{\delta}_t \sim N(\boldsymbol{0}, \sigma_\delta^2 \mathbb{I}_G), \quad t = 1, \dots, T \tag{13}$$

$$\begin{pmatrix} \boldsymbol{x}_{t,M} \\ \boldsymbol{x}_{t,P} \end{pmatrix} = \beta_0 \boldsymbol{1}_{M+G} + \beta_1 \begin{pmatrix} \boldsymbol{z}_{t,M} \\ \boldsymbol{z}_{t,P} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\xi}_{t,M} \\ \boldsymbol{\xi}_{t,P} \end{pmatrix}, \quad t = 1, \dots, T \tag{14}$$

$$\begin{pmatrix} \boldsymbol{\xi}_{t,M} \\ \boldsymbol{\xi}_{t,P} \end{pmatrix} = \varsigma \begin{pmatrix} \boldsymbol{\xi}_{t-1,M} \\ \boldsymbol{\xi}_{t-1,P} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\omega}_{t,M} \\ \boldsymbol{\omega}_{t,P} \end{pmatrix}, \quad \begin{pmatrix} \boldsymbol{\omega}_{t,M} \\ \boldsymbol{\omega}_{t,P} \end{pmatrix} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}), \quad t = 1, \dots, T \tag{15}$$

where $\mathbb{I}_M$ and $\mathbb{I}_G$ are identity matrices of dimension $M \times M$ and $G \times G$, respectively, and $\boldsymbol{1}_G$ and $\boldsymbol{1}_{M+G}$ is a vector of 1's of dimension G and $M + G$, respectively. As discussed in Section 3.1, $\boldsymbol{\omega}_t$ is a temporally-independent Gaussian vector with mean zero and covariance matrix $\boldsymbol{\Sigma}$, whose elements are computed using the Matèrn covariance function with parameters σ_ω^2 and κ.

In the system of equations above, the latent vector $\boldsymbol{x}_{t,M}$ is present in both (12) and (14). Also, $\boldsymbol{x}_{t,P}$ is present in both (13) and (14). In order to make sure that the values of $\boldsymbol{x}_{t,M}$ and $\boldsymbol{x}_{t,P}$ are equivalent for the different equations when fitting the joint model, $\boldsymbol{x}_{t,M}$ in (12) is assumed to be an (almost) identical copy of $\boldsymbol{x}_{t,M}$ in (14). Similarly, $\boldsymbol{x}_{t,P}$ in (13) is assumed to be an (almost) identical 'copy' of $\boldsymbol{x}_{t,P}$ in (14) with α_1 as a scaling parameter. To create these 'copies', the latent field $\boldsymbol{x}_t$ is extended to $\boldsymbol{\chi}_t = \begin{pmatrix} \boldsymbol{x}_t^\top & \boldsymbol{x}_t^{*\top} \end{pmatrix}^\top$, $t = 1, \dots, T$, where $\boldsymbol{x}_t^* = \begin{pmatrix} \boldsymbol{x}_{t,M}^{*\top} & \boldsymbol{x}_{t,P}^{*\top} \end{pmatrix}^\top$ is a copy of $\boldsymbol{x}_t$. The prior specification for the extended latent field at time t, $\pi(\boldsymbol{\chi}_t)$, will ensure that $\boldsymbol{x}_t^*$ is an identical copy of $\boldsymbol{x}_t$. In particular, $\boldsymbol{x}_{t,M}^*$ and $\boldsymbol{x}_{t,P}^*$ will be defined later in such a way that $\mathbb{E}(\boldsymbol{x}_{t,M}^*) = \boldsymbol{x}_{t,M}$ and $\mathbb{E}(\boldsymbol{x}_{t,P}^*) = \alpha_1 \boldsymbol{x}_{t,P}$. This is the same approach as in Martins et al. (2013) and Ruiz-Cárdenas et al. (2012) and is called a *data augmentation approach*.

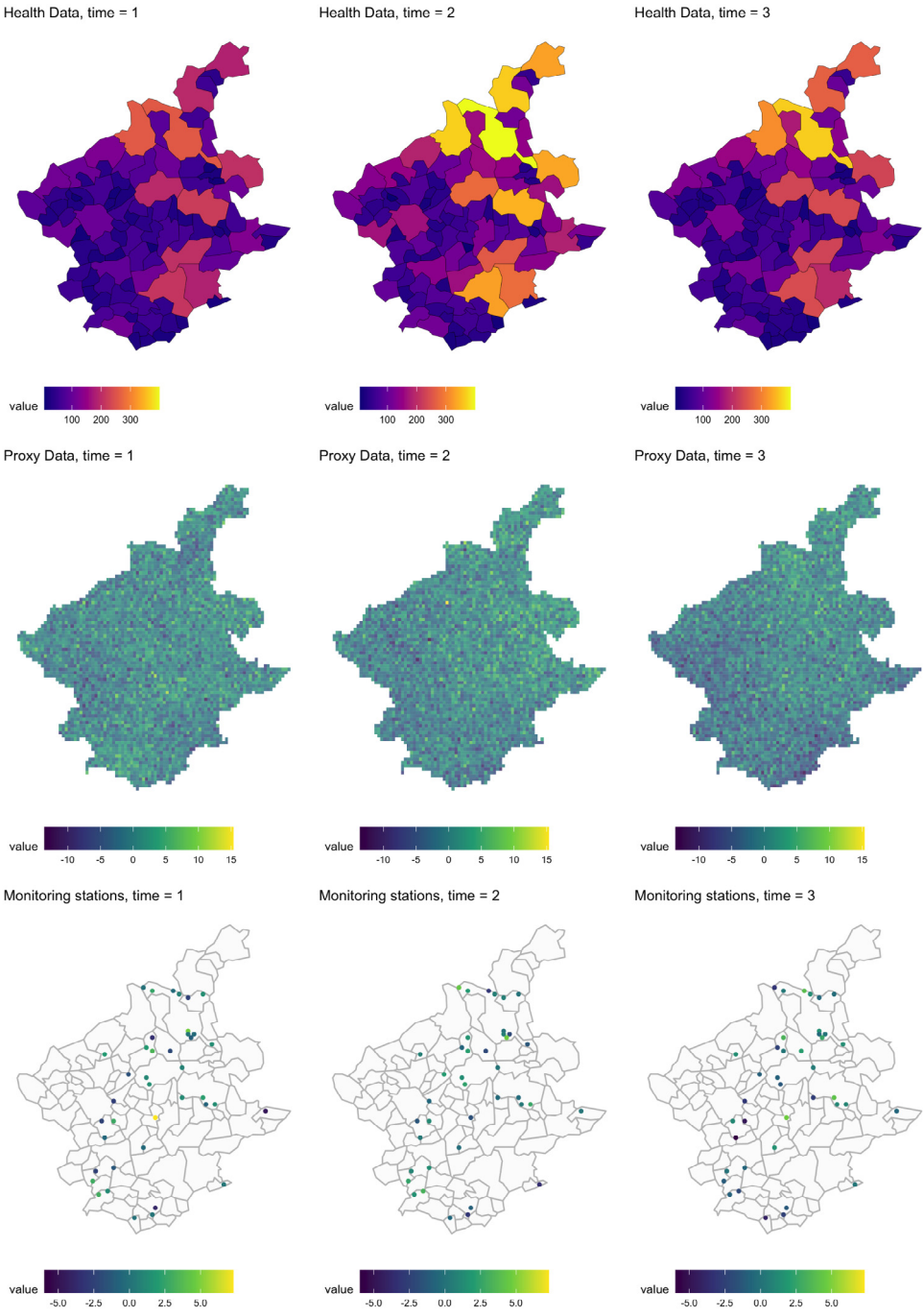


Fig. 1. Sample data for a spatio-temporal analysis: count data (top row), proxy data (middle row), monitoring stations data (bottom).

Since $\boldsymbol{x}_t$ on the left-hand side of (14) is unknown, $\boldsymbol{x}_t$ is transposed on the right-hand side of the equation. The remaining vector of zeros in the left-hand side are referred to as 'pseudo-zeros'. This gives the following re-expression of the joint model:

$$\boldsymbol{w}_t = \boldsymbol{x}^*_{t,M} + \boldsymbol{e}_t, \quad \boldsymbol{e}_t \sim N(\boldsymbol{0}, \sigma_e^2 \mathbb{I}_M), \quad t = 1, \ldots, T \tag{16}$$

$$\tilde{\boldsymbol{x}}_t = \alpha_0 \boldsymbol{1}_G + \boldsymbol{x}^*_{t,P} + \boldsymbol{\delta}_t, \quad \boldsymbol{\delta}_t \sim N(\boldsymbol{0}, \sigma_\delta^2 \mathbb{I}_G), \quad t = 1, \ldots, T \tag{17}$$

$$\boldsymbol{0}_t = -\begin{pmatrix} \boldsymbol{x}_{t,M} \\ \boldsymbol{x}_{t,P} \end{pmatrix} + \beta_0 \boldsymbol{1}_{M+G} + \beta_1 \boldsymbol{z}_t + \boldsymbol{\xi}_t, \quad t = 1, \ldots, T \tag{18}$$

$$\boldsymbol{\xi}_t = \varsigma \boldsymbol{\xi}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}), \quad t = 1, \ldots, T \tag{19}$$

Suppose we have $\boldsymbol{\theta} = \begin{pmatrix} \sigma_e^2 & \sigma_\delta^2 & \alpha_0 & \alpha_1 & \beta_0 & \beta_1 & \varsigma & \sigma_\omega^2 & \kappa \end{pmatrix}^\intercal$. The posterior distribution of interest is $\pi(\boldsymbol{\chi}, \boldsymbol{\xi}, \boldsymbol{\theta} | \boldsymbol{w}, \tilde{\boldsymbol{x}}, \boldsymbol{0})$, given by

$$\pi(\boldsymbol{\chi}, \boldsymbol{\xi}, \boldsymbol{\theta} | \boldsymbol{w}, \tilde{\boldsymbol{x}}, \boldsymbol{0}) \propto \pi(\boldsymbol{w} | \boldsymbol{\chi}, \boldsymbol{\xi}, \boldsymbol{\theta}) \pi(\tilde{\boldsymbol{x}} | \boldsymbol{\chi}, \boldsymbol{\xi}, \boldsymbol{\theta}) \pi(\boldsymbol{0} | \boldsymbol{\chi}, \boldsymbol{\xi}, \boldsymbol{\theta}) \pi(\boldsymbol{\xi} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \pi(\boldsymbol{\chi}).$$

The first two are straightforward since $\boldsymbol{w}_t | \boldsymbol{\chi}, \boldsymbol{\xi}, \boldsymbol{\theta} \overset{\text{ind}}{\sim} N(\boldsymbol{x}^*_{t,M}, \sigma_e^2 \mathbb{I}_M)$ and $\tilde{\boldsymbol{x}}_t | \boldsymbol{\chi}, \boldsymbol{\xi}, \boldsymbol{\theta} \overset{\text{ind}}{\sim} N(\alpha_0 \boldsymbol{1}_G + \boldsymbol{x}^*_{t,P}, \sigma_\delta^2 \mathbb{I}_G)$. For the pseudo-zeros, we have $\boldsymbol{0}_t | \boldsymbol{\chi}, \boldsymbol{\xi}, \boldsymbol{\theta} \overset{\text{ind}}{\sim} N(-\boldsymbol{x}_t + \beta_0 \boldsymbol{1}_{M+G} + \beta_1 \boldsymbol{z}_t + \boldsymbol{\xi}_t, \frac{1}{\tau_0} \mathbb{I}_G)$, where τ_0 is a precision parameter and is fixed at a large value because of the absence of measurement error in the pseudo-zeros. The form of the distribution of $\boldsymbol{\xi}_t | \boldsymbol{\theta}$ uses the fact that $\boldsymbol{\xi}_t | \boldsymbol{\xi}_{t-1} \sim N(\varsigma \boldsymbol{\xi}_{t-1}, \boldsymbol{\Sigma})$, $t = 2, \ldots, T$, and that $\boldsymbol{\xi}_1 \sim N(\boldsymbol{0}, \frac{1}{1-\varsigma^2} \boldsymbol{\Sigma})$.

For the extended latent field $\boldsymbol{\chi}$, we have

$$\pi(\boldsymbol{\chi}) = \prod_{t=1}^{T} \pi(\boldsymbol{\chi}_t) = \prod_{t=1}^{T} \pi(\boldsymbol{x}^*_{t,M} | \boldsymbol{x}_{t,M}) \pi(\boldsymbol{x}^*_{t,P} | \boldsymbol{x}_{t,P}) \pi(\boldsymbol{x}_t).$$

Since $\boldsymbol{x}^*_{t,M}$ and $\boldsymbol{x}^*_{t,P}$ are independent copies of $\boldsymbol{x}_{t,M}$ and $\boldsymbol{x}_{t,P}$, respectively, then both are assumed to be Gaussian centered on $\boldsymbol{x}_{t,M}$ and $\alpha_1 \boldsymbol{x}_{t,P}$ and with very high precision, i.e., $\boldsymbol{x}^*_{t,M} | \boldsymbol{x}_{t,M} \sim N\left(\boldsymbol{x}_{t,M}, \frac{1}{\tau_{x^*}}\right)$ and $\boldsymbol{x}^*_{t,P} | \boldsymbol{x}_{t,P} \sim N\left(\alpha_1 \boldsymbol{x}_{t,P}, \frac{1}{\tau_{x^*}}\right)$, where τ_{x^*} is fixed at some large value. The latent field $\boldsymbol{x}_t$ is assumed to be independent Gaussian centered at zero but with fixed high value for variance (low precision), i.e., $\boldsymbol{x}_t \sim N\left(\boldsymbol{0}, \frac{1}{\tau_x}\right)$, where τ_x is a small value. Although the precision is very low, the pseudo-zeros have very high precision, and so the value of $\boldsymbol{x}_t$ in (18) is forced to be close to its true value.

Finally, the components of $\boldsymbol{\theta}$ are assumed to be independent, i.e., $\pi(\boldsymbol{\theta}) = \prod_{i=1}^{H} \pi(\theta_i)$, where H is the number of parameters in $\boldsymbol{\theta}$.

The joint model specified in Eqs. (16)–(19) is a correct representation of Eqs. (12)–(15) which does not violate latent Gaussianity; hence, allowing the use of INLA to fit the first-stage joint model. This data augmentation approach can be implemented using the `R-INLA` functionalities, some details of which are discussed in Martins et al. (2013) and Ruiz-Cárdenas et al. (2012).

3.2.1. SPDE representation

In the SPDE representation of the model, the latent (Gaussian) field, $\boldsymbol{\omega}_t$, which is continuously defined in space is discretized on a mesh or a triangulation with K vertices. The discretization at an arbitrary location $\boldsymbol{s}$ at time t is given by

$$\omega^D(\boldsymbol{s}, t) = \sum_{k=1}^{K} \psi_k(\boldsymbol{s}) w_{kt}, \tag{20}$$

where $\{\psi_k\}$ are basis functions chosen to be piece-wise linear in each triangle, i.e., $\psi_k = 1$ at vertex k and 0 at other vertices, and $\{w_{kt}\}$ are Gaussian-distributed weights. The basis functions are not indexed by t since we use the same mesh for all time points. The weights $\{w_{kt}\}$ on the other hand do

vary for different time points; hence, the weights are indexed by t. Eq. (20) provides a continuously-indexed but finite-dimensional approximation to the solution of the following stochastic partial differential equation at a fixed time t:

$$(\kappa^2 - \Delta)^{\alpha/2}\omega(\boldsymbol{s}, t) = \mathcal{W}(\boldsymbol{s}, t), \quad \boldsymbol{s} \in \mathbb{R}^d, \quad \alpha = \nu + d/2, \quad \kappa > 0, \quad \nu > 0 \tag{21}$$

where $(\kappa^2 - \Delta)^{\alpha/2}$ is a pseudodifferential operator, $\mathcal{W}(\boldsymbol{s})$ is a Gaussian white noise with unit variance, and Δ is the *Laplacian* defined by $\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial \omega_i^2}$. In the context of the problem at hand, $d = 2$; but this result holds true at higher dimensions. The solution of this SPDE is a Gaussian field with the Matèrn covariance function (Lindgren et al., 2011). The benefit of using the approximation in Eq. (20) is that its full distribution, which is specified by the joint distribution of the weights w_k, has a sparse precision matrix and closely approximates the inverse precision of the Matèrn Gaussian field in some norm (Lindgren et al., 2011). This sparsity speeds up computation and model estimation.

Denoting by $\boldsymbol{\omega}_t^D$ the vector of values at the nodes of the mesh at time t, we have $\boldsymbol{\omega}_t^D \sim N(\mathbf{0}, \mathbf{Q}_s^{-1})$ with $\mathbf{Q}_s$ a sparse precision matrix. The details on how the elements of $\mathbf{Q}_s$ are computed are discussed in Lindgren et al. (2011). Eq. (19) now defined on the nodes of the mesh is then expressed as

$$\boldsymbol{\xi}_t^D = \varsigma \boldsymbol{\xi}_{t-1}^D + \boldsymbol{\omega}_t^D, \quad \boldsymbol{\omega}_t^D \sim N(\mathbf{0}, \mathbf{Q}_s^{-1}), \quad t = 1, \ldots, T,$$

where $\boldsymbol{\xi}_1^D \sim N\left(\mathbf{0}, \frac{1}{1-\varsigma^2}\mathbf{Q}_s^{-1}\right)$ and $\boldsymbol{\xi}_t^D$ is a K-dimensional vector of spatio-temporal random effect at the K nodes of the mesh at time t. The joint distribution of the TK-dimensional vector $\boldsymbol{\xi}^D = \left(\boldsymbol{\xi}_1^{D\intercal} \quad \ldots \quad \boldsymbol{\xi}_T^{D\intercal}\right)^\intercal$ is $\boldsymbol{\xi}^D \sim N(\mathbf{0}, (\mathbf{Q}_s \otimes \mathbf{Q}_T)^{-1})$, where $\mathbf{Q}_T$ is the precision matrix for the autoregressive process of order 1, the form of which is given in Rue and Held (2005).

Since $\boldsymbol{\xi}$ and $\boldsymbol{\omega}$ are estimated in a mesh whose nodes may be different from the observed locations and from the centroids of the grid cells, there needs to be a linear mapping from the nodes to the locations of the observed values of the monitors and proxy data. This is done by incorporating a projection or mapping matrix, say $\boldsymbol{B}$, which is a sparse $(M + G) \times K$ matrix, so that

$$\boldsymbol{x}_t = \beta_0 \mathbf{1}_{G+M} + \beta_1 \boldsymbol{z}_t + \boldsymbol{B}\boldsymbol{\xi}_t^D, t = 1, \ldots, T, \quad \text{or}$$

$$x(\mathbf{s}_i, t) = \beta_0 + \beta_1 z(\mathbf{s}_i, t) + \sum_{k=1}^{K} b_{ik}\xi_{tk},$$

where ξ_{tk} is the kth element of the vector $\boldsymbol{\xi}_t^D$ and b_{ik} is the (i, k)th element of the mapping matrix $\boldsymbol{B}$.

The first-stage hierarchical SPDE model is then given by

$$\boldsymbol{w}_t = \boldsymbol{x}_{t,M}^* + \boldsymbol{e}_t, \quad \boldsymbol{e}_t \sim N(\mathbf{0}, \sigma_e^2 \mathbb{I}_M), \quad t = 1, \ldots, T \tag{22}$$

$$\tilde{\boldsymbol{x}}_t = \alpha_0 \mathbf{1}_G + \boldsymbol{x}_{t,P}^* + \boldsymbol{\delta}_t, \quad \boldsymbol{\delta}_t \sim N(\mathbf{0}, \sigma_\delta^2 \mathbb{I}_G), \quad t = 1, \ldots, T \tag{23}$$

$$\mathbf{0}_t = -\begin{pmatrix} \boldsymbol{x}_{t,M} \\ \boldsymbol{x}_{t,P} \end{pmatrix} + \beta_0 \mathbf{1}_G + \beta_1 \boldsymbol{z}_t + \boldsymbol{B}\boldsymbol{\xi}_t^D \tag{24}$$

$$\boldsymbol{\xi}_t^D = \varsigma \boldsymbol{\xi}_{t-1}^D + \boldsymbol{\omega}_t^D, \quad \boldsymbol{\omega}_t^D \sim N(\mathbf{0}, \mathbf{Q}_s^{-1}), \quad t = 1, \ldots, T \tag{25}$$

The joint model specified in Eqs. (22)–(25) is a latent Gaussian model and hence can be fitted using the INLA method. Using $\boldsymbol{\vartheta}$ to denote the vector of hyperparameters, we then have $\boldsymbol{\vartheta} = \left(\alpha_1 \quad \sigma_e^2 \quad \sigma_\delta^2 \quad \varsigma \quad \sigma_\omega^2 \quad \kappa\right)^\intercal$. The posterior marginals of $\boldsymbol{\vartheta}$ is approximated as

$$\pi(\vartheta_i | \boldsymbol{w}_t, \tilde{\boldsymbol{x}}_t) = \int \pi(\boldsymbol{\vartheta} | \boldsymbol{w}_t, \tilde{\boldsymbol{x}}_t) d\boldsymbol{\vartheta}_{-j}.$$

The latent Gaussian vector is given by $\left(\boldsymbol{x}^\intercal \quad \boldsymbol{x}^{*\intercal} \quad \boldsymbol{\xi}^\intercal \quad \alpha_0 \quad \beta_0 \quad \beta_1\right)^\intercal$, whose posterior marginals are computed as

$$\pi(\cdot | \boldsymbol{w}_t, \tilde{\boldsymbol{x}}_t) = \int \pi(\cdot | \boldsymbol{\theta}, \boldsymbol{w}_t, \tilde{\boldsymbol{x}}_t) \pi(\boldsymbol{\theta} | \boldsymbol{w}_t, \tilde{\boldsymbol{x}}_t) d\boldsymbol{\theta}.$$

The mathematical and computational details of the INLA method are discussed in Rue et al. (2009). Moreover, the R codes to jointly fit Eqs. (22)–(25) is available on https://github.com/StephenVillejo/DataFusionINLASPDE.

3.2.2. Computing block level exposures

The first-stage model is used to predict the values of $x(\boldsymbol{s}, t)$ on a prediction grid. Two methods to compute the block-level exposures discussed in Cameletti et al. (2019) are explored in the simulation study done in this work. Suppose we denote by $\hat{x}(B_i, t)$ the predicted value of exposures at block B_i at time t, so that $\hat{x}(B_i, t) \approx \frac{1}{|B_i|}\int_{B_i} x(\boldsymbol{s}, t)d\boldsymbol{s}$. The first method considers all the prediction grid cells which overlaps with block B_i. Suppose we denote by $\hat{x}(\boldsymbol{g}_j, t)$ the predicted value of $x(\boldsymbol{g}_j, t)$ at the grid cell whose centroid is $\boldsymbol{g}_j$ at time t and $h(\boldsymbol{g}_j, B_i)$ the proportion of the area of block B_i which overlaps with the grid cell. The first method is computed as follows:

$$\textbf{Method 1}: \qquad \hat{x}(B_i, t) = \sum_{\forall j} \hat{x}(\boldsymbol{g}_j, t)h(\boldsymbol{g}_j, B_i). \tag{26}$$

The first method computes the value of $\hat{x}(B_i, t)$ as a weighted mean of the predicted values at the grid cells which overlap with block B_i where the weights are the proportion of block B_i which overlap with the grid cells. On the other hand, the second method uses only the grid cells whose centroids are inside block B_i. The computed value of $\hat{x}(B_i, t)$ using the second method is then a simple mean, i.e., it is computed as

$$\textbf{Method 2}: \qquad \hat{x}(B_i, t) = \frac{1}{\#(\mathbf{g}_j \in B_i)} \sum_{\mathbf{g}_j \in B_i} \hat{x}(\boldsymbol{g}_j, t), \tag{27}$$

where $\#(\mathbf{g}_j \in B_i)$ is the number of grid cells whose centroids are inside block B_i.

3.3. Model fitting for second-stage model

The latent Gaussian vector for the second-stage model is given by $\begin{pmatrix}\gamma_0 & \gamma_1 & \boldsymbol{\varphi}\end{pmatrix}^\intercal$ where $\boldsymbol{\varphi}$ is the vector containing all the spatial effects, temporal effects, and their interaction. The hyperparameter vector, say $\boldsymbol{\theta}_\varphi$, consists of all the model constants used to parameterize the components of $\boldsymbol{\varphi}$. For instance, if a random walk model of order 1 is used for the temporal effect, then the variance parameter for the random walk process is included in the hyperparameter vector. The different model specifications for $\boldsymbol{\varphi}$ are discussed in Blangiardo and Cameletti (2015). The posterior distribution of interest is

$$\pi(\gamma_0, \gamma_1, \boldsymbol{\varphi}, \boldsymbol{\theta}_\varphi|\boldsymbol{y}) \propto \pi(\boldsymbol{y}|\gamma_0, \gamma_1, \boldsymbol{\varphi}, \boldsymbol{\theta}_\varphi)\pi(\boldsymbol{\varphi}|\boldsymbol{\theta}_\varphi)\pi(\gamma_0, \gamma_1, \boldsymbol{\theta}_\varphi),$$

where $\boldsymbol{y}$ is the vector of all observed counts $Y(B_i, t)$. $\pi(\boldsymbol{y}|\gamma_0, \gamma_1, \boldsymbol{\varphi}, \boldsymbol{\theta}_\varphi)$ has a simple form since the elements of $\boldsymbol{y}$ are independent conditional on the latent field and hyperparameters. The form of $\pi(\boldsymbol{\varphi}|\boldsymbol{\theta}_\varphi)$ depends on the model specified for the elements of $\boldsymbol{\varphi}$ and has a straightforward structure (Blangiardo and Cameletti, 2015). Lastly, $\pi(\gamma_0, \gamma_1, \boldsymbol{\theta}_\varphi)$ can be safely assumed as a product of the individual priors.

In the simulation study done in this paper which is discussed in Section 4, a simpler form is assumed for the log risks to illustrate how the model fitting is done and to identify scenarios that could potentially affect model performance. We assume the following form for the log risks:

$$\log(\lambda(B_i, t)) = \gamma_0 + \gamma_1 x(B_i, t) + \phi_i + \nu_t, \tag{28}$$

where $\phi_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma_\phi^2)$ and $\nu_t \overset{\text{iid}}{\sim} \text{N}(0, \sigma_\nu^2)$. Hence, we assume that φ_{it} is equal to the sum of an unstructured spatial effect and an unstructured temporal effect. The latent Gaussian field in this case is given by $\begin{pmatrix}\gamma_0 & \gamma_1 & \boldsymbol{\phi}^\intercal & \boldsymbol{\nu}^\intercal\end{pmatrix}^\intercal$ while the hyperparameter vector is $\begin{pmatrix}\sigma_\phi^2 & \sigma_\nu^2\end{pmatrix}^\intercal$. The posterior distribution of interest is

$$\pi(\gamma_0, \gamma_1, \boldsymbol{\phi}, \boldsymbol{\nu}, \sigma_\phi^2, \sigma_\nu^2|\boldsymbol{y}) \propto \pi(\boldsymbol{y}|\gamma_0, \gamma_1, \boldsymbol{\phi}, \boldsymbol{\nu})\pi(\boldsymbol{\phi}|\sigma_\phi^2)\pi(\boldsymbol{\nu}|\sigma_\nu^2)\pi(\gamma_0, \gamma_1, \sigma_\phi^2, \sigma_\nu^2).$$

For $\pi(\gamma_0, \gamma_1, \sigma_\phi^2, \sigma_\nu^2)$, the prior distributions can be defined independently for each parameter. This

is a latent Gaussian model; hence, all the posterior marginals can be approximated using INLA. In the general case, Eq. (11) provides a more complex form for the spatio-temporal effects.

3.3.1. Propagating uncertainty from the first-stage to the second-stage model

One approach to propagate uncertainty from the first-stage to the second-stage model is to sample several times from the posterior predictive distribution of the latent field $\boldsymbol{x} = \left(\boldsymbol{x}_1^\intercal \quad \boldsymbol{x}_2^\intercal \quad \dots \quad \boldsymbol{x}_T^\intercal\right)^\intercal$, and then to compute all the block-level values of exposures from each sample using Eqs. (26) and (27) which are then used as an input to fit the second-stage model. The final parameter estimates of the second-stage model are then computed using the combined results from all samples (Cameletti et al., 2019; Blangiardo et al., 2016; Lee et al., 2017; Liu et al., 2017). Suppose we simulate from the posterior predictive distribution of $\boldsymbol{x}$ J times, then the steps to propagate uncertainty from the first-stage to the second-stage model are as follows:

Do the following J times:

1. Simulate from the posterior distribution of the latent field $\hat{\pi}(\boldsymbol{x}|\cdot)$ on the prediction grid. This distribution is obtained from the first-stage model.
2. Compute block-level exposures, $\hat{x}(B_i, t)$ for all blocks B_i in the study region, using the two methods in Eqs. (26) and (27).
3. Fit the second-stage model using INLA as described in Section 3.3.
4. For each parameter in the second-stage model, generate samples from its posterior marginal distribution.

After completing all J cycles, all samples from step (4) are combined and used to approximate the posterior distribution of the second-stage model parameters.

4. Simulation study

The performance of the proposed method is investigated using a simulation study. The study region used is the Belo Horizonte region in Brazil which is available in the *spdep* package in R (Bivand and Piras, 2015) and is the same study region used in Cameletti et al. (2019) from which the extensions done in this work are based. Fig. 1 shows the study region map. We used this for the simulation study since it has few areas (98 of them) which makes it more computationally manageable. In the case study presented in Section 6, we used England data which has relatively more areas. The Bayesian computation are done using the functionalities in the `R-INLA` library. All results in the paper can be replicated using the R code which is available on https://github.com/StephenVillejo/DataFusionINLASPDE.

4.1. Simulation of the exposure field

For the Matèrn covariance function parameters, the spatial variance σ_ω^2 is set to 1.5, while the range parameter ρ is 1.89 which corresponds to around 46% of the maximum distance in the 100×100 simulation grid. The autoregressive parameter ς is set to 0.7. The single covariate $z(\boldsymbol{s}, t)$ was generated from $N(\mu = 0, \sigma = 1)$. The fixed effects are $\beta_0 = 0$ and $\beta_1 = 2$. The simulated values of $x(B_i, t)$ is a spatial average of all the points inside block B_i, i.e.,

$$x(B_i, t) = \frac{1}{|B_i|} \int_{B_i} x(\boldsymbol{s}, t) d\boldsymbol{s} \approx \sum_{\forall \boldsymbol{s} \in B_i} \frac{1}{\#(\boldsymbol{s} \in B_i)} x(\boldsymbol{s}, t).$$

4.2. Simulation of the health data

Using the simulated values of $x(B_i, t)$, it is straightforward to simulate the health data. The fixed effects are $\gamma_0 = -3$ and $\log(\gamma_1) = 1.2$. The assumed value of γ_1 implies an expected increase of 20% in the relative risk for a one unit increase in $x(B_i, t)$. The assumed values of the variance parameters of the spatial effect and the temporal effect are $\sigma_\phi^2 = \sigma_\nu^2 = 0.02$. The expected number of cases for each block are generated from a uniform distribution and are made to be proportional to the size of the block so that blocks with bigger surface areas have higher expected number of cases.

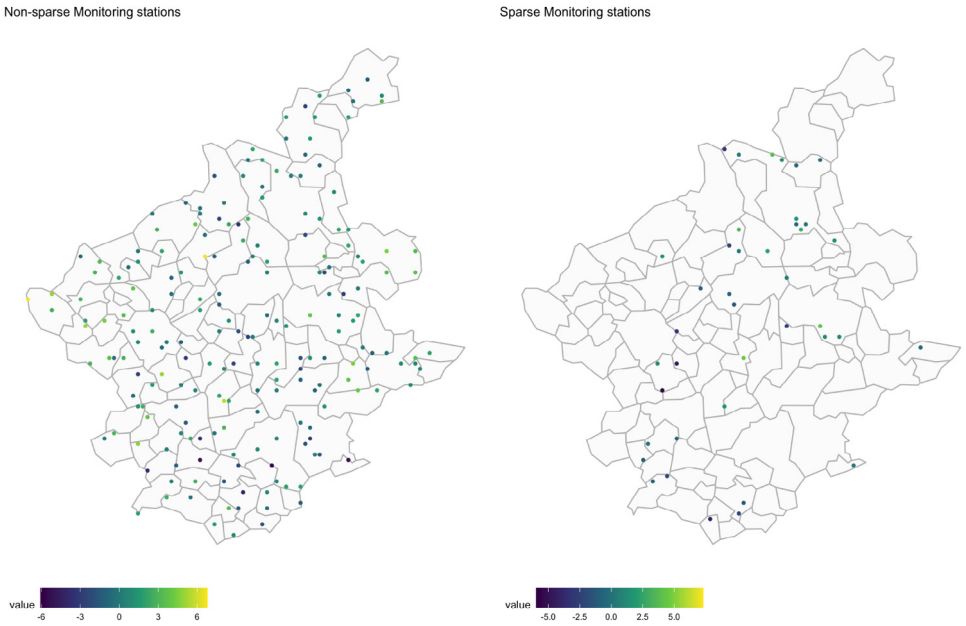


Fig. 2. Non-sparse network of monitors (left) and a sparse network of monitors (right).

4.3. Simulation of the monitors and proxy data

The monitors are simulated by getting a random sample of points from the simulation grid. A non-sparse network was considered and investigated in Cameletti et al. (2019) - either getting 2%, 10%, or 30% of the points from the simulation grid inside each block. Their simulation results showed that block predictions are better in terms of the RMSE and the correlation with the true block-level exposure values when there are more monitoring stations in the data. One contribution of this paper is to look at the case of having a sparse network of monitoring stations, i.e., there are few monitors and several areas or blocks do not have monitors inside. This study considers two scenarios for the monitoring stations data. The first scenario is a non-sparse network similar to that considered in Cameletti et al. (2019). The second scenario is a sparse network and is carefully chosen in such a way that it resembles how actual data of sparse monitoring stations look like. Fig. 2 shows a simulated non-sparse network (left) and a sparse network (right) of monitors. The observed values $w(\boldsymbol{s}_i, t), i = 1, \ldots, M$ follow the classical error model, with the error term assumed as $e(\boldsymbol{s}_i, t) \sim N(0, \sigma_e^2 = 0.1)$. For the proxy data, the bias parameters are $\alpha_0 = -1, \alpha_1 = 1.5$, and $\sigma_\delta^2 = 1$. Fig. 3 shows an example of a simulated exposures field (left) and the corresponding proxy data after adding the biases (right). The figure shows that there is more variability in the values of the proxy data, which is expected because of the scaling in the values through α_1, a shift in the location of the center through α_0, and an additional random noise whose variance is larger than the measurement error in the monitors.

4.4. Prediction grid

The effect of the resolution of the prediction grid on the block-level predictions using Eqs. (26) and (27) has been investigated in Cameletti et al. (2019). Their simulation results have shown that with a finer prediction grid, the block-level exposure predictions are also more accurate. Hence, in the simulation study in this paper, a 100×100 prediction grid is used for all the scenarios, which is the same grid resolution used to simulate the true exposures field and the proxy data.

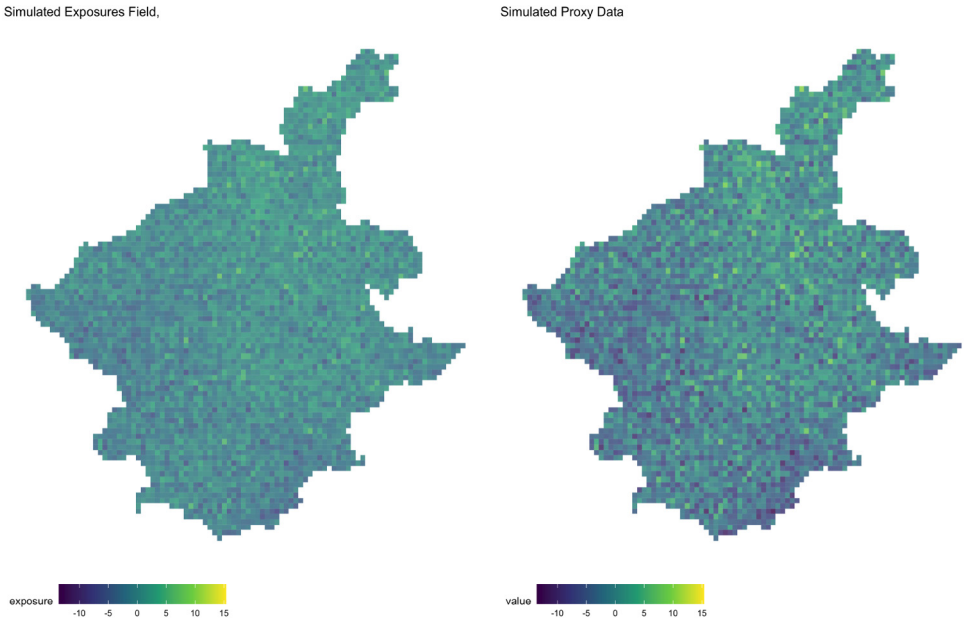


Fig. 3. Simulated true exposures field (left) and corresponding proxy data after adding biases (right).

4.5. Simulation scenarios

There are three simulation settings considered in the study:

1. The sparsity of the monitoring stations: sparse or non-sparse. This is illustrated in Fig. 2.
2. Length of time: $T = 3$, $T = 6$, or $T = 12$. Since this simulation study is a spatio-temporal extension of that in Cameletti et al. (2019), it is important to investigate the effect of the length of time in terms of the model estimates and predictions.
3. Prior specification: use of non-informative priors or informative priors. Since the sensitivity of the posterior estimates to the priors is an important part of Bayesian analysis, the effect of prior specification is therefore investigated in the simulation study. Not all parameters are given informative priors. Only those parameters which are usually difficult to estimate are given informative priors — these include the parameters of the latent Gaussian field σ_ω^2, ρ, and ς; the variance parameters σ_ν^2, σ_ϕ^2, σ_e^2, and σ_δ^2; and α_1 which is the scaling parameter of $\boldsymbol{x}_{t,P}$.

There are a total of 12 simulation scenarios. For each scenario, 500 independent replications are done in order to evaluate the performance of the proposed method. In fitting the second-stage model, the number of simulations from the posterior predictive distribution of the latent field $x(\boldsymbol{s}, t)$ is set at $J = 100$. For each estimated posterior marginal distribution, 200 random values were simulated to compute posterior quantities of interest which include the posterior mean, posterior median, and 95% credible intervals.

Table 1 shows all the simulation scenarios and how they are labeled in the figures in Section 5.

Table 2 shows the priors for the first-stage model parameters for the two cases. Non-informative priors are used in both cases for the parameters β_0, β_1, and α_0. The non-informative priors of the Matèrn parameters are defined in terms of a reparameterization of the parameters of the SPDE in Eq. (21). A detailed discussion of prior specification for Matèrn models using the SPDE approach is discussed in (Lindgren and Rue, 2015). The informative priors for ρ and σ_ω^2 are the so-called penalized-complexity (PC) priors (Fuglstad et al., 2019). The PC prior is a weakly informative prior

Table 1
Simulation scenarios in the simulation study.

Sparsity	Priors specification	Time points		
		3	6	12
No	Informative	A	E	I
	Non-informative	B	F	J
Yes	Informative	C	G	K
	Non-informative	D	H	L

Table 2
Priors specification for first-stage model parameters.

Parameter	Informative prior	Non-informative prior
β_0	$N(0, \infty)$	$N(0, \infty)$
β_1	$N(0, 1000)$	$N(0, 1000)$
α_0	$N(0, \infty)$	$N(0, \infty)$
α_1	$N(1.5, 1)$	$N(0, 1000)$
σ_e^2	Inv Gamma(100, 10)	Inv Gamma(1, 5e−5)
σ_δ^2	Inv Gamma(10, 10)	Inv Gamma(1, 5e−5)
ρ	PC(ρ_0, α)[a]	See Lindgren and Rue (2015)
σ_ω^2	PC($\sigma_{\omega,0}, \alpha$)[a]	See Lindgren and Rue (2015)
ς	$\log\left(\frac{1+\varsigma}{1-\varsigma}\right) \sim N(0.75, 0.05^2)$	$\log\left(\frac{1+\varsigma}{1-\varsigma}\right) \sim N(0, 0.15^2)$

[a]These are called **penalized-complexity priors** which are weakly informative priors (Fuglstad et al., 2019).

Table 3
Priors specification for second-stage model parameters.

Parameter	Informative prior	Non-informative prior
γ_0	$N(0, \infty)$	$N(0, \infty)$
γ_1	$N(0, 1000)$	$N(0, 1000)$
σ_ϕ^2	PC($\sigma_{\phi,0}, \alpha$)[a]	Inv Gamma(1, 5e−5)
σ_v^2	PC($\sigma_{v,0}, \alpha$)[a]	Inv Gamma(1, 5e−5)

[a]These are called **penalized-complexity priors** which are weakly informative priors (Fuglstad et al., 2019).

which penalizes complexity or additional flexibility in the model; hence, the prior tends to prefer the simpler *base* model. It works on the principle that a model further away from the base model should be more strongly penalized. For the Matèrn parameters, the PC prior shrinks the model to the base model with infinite range and zero marginal variance. The informative priors for ρ and σ_ω^2 shown in Table 2 are a joint specification, where $\sigma_{\omega,0}$ and ρ_0 are the upper and lower limit for σ_ω and ρ, respectively, and α is the tail probability. In fitting the models in the simulation study, $\sigma_{\omega,0}$ and ρ_0 are set as equal to the true value, and $\alpha = 0.05$.

There are three precision parameters in the first-stage model that are fixed at an appropriate level. The first one is the precision parameter, τ_0, for the pseudo-zeros which is fixed at a large value. The second one is the precision parameter, τ_x, for the prior of the latent field $\boldsymbol{x}_t$ which is fixed at a very small value. Finally, for the conditional distribution of the copies $\boldsymbol{x}_{t,M}^*$ and $\boldsymbol{x}_{t,P}^*$, the precision parameter τ_x* is also fixed at a large value so that both mimic $\boldsymbol{x}_{t,M}$ and $\boldsymbol{x}_{t,P}$, respectively.

Table 3 shows the priors for the second-stage model parameters for the two cases. The parameters α_0 and α_1 are given non-informative priors for the two cases. The variance parameters σ_ϕ^2 and σ_v^2 are also given the PC priors for the cases of informative priors.

The performance of the proposed method is evaluated in terms of the bias, RMSE, and coverage probabilities for each parameter estimate and block-level predictions $\hat{x}(B_i, t)$. The details are found in Appendix A.

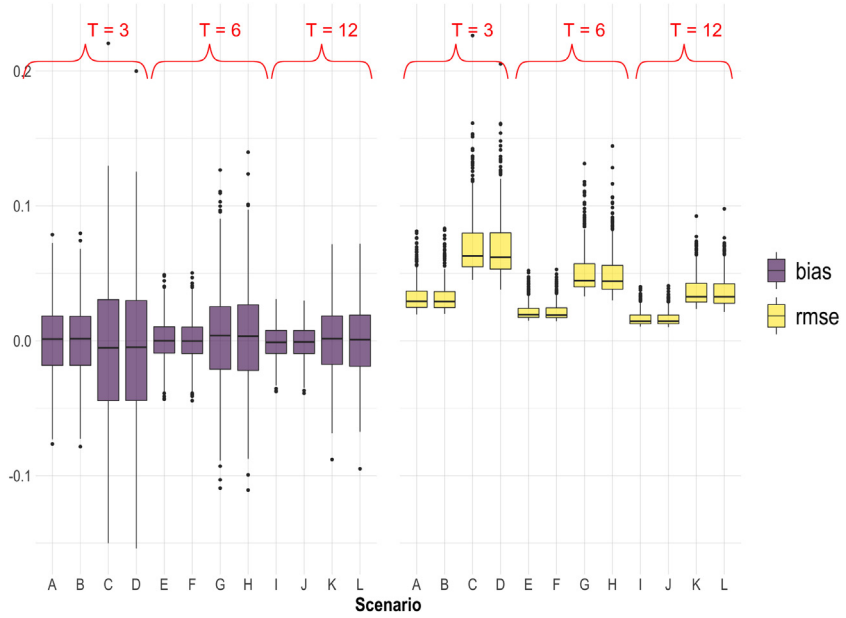


Fig. 4. Plot of bias (purple) and RMSE (yellow) for α_0. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5. Results and discussion

This section presents the results of the simulation study. Section 5.1 presents the results for the first-stage model parameters, Section 5.2 for the block-level exposure estimates, and Section 5.3 for the second-stage model parameters.

5.1. First-stage model parameters

Figs. 4 and 5 show plots of the biases and RMSEs for the additive and multiplicative bias in the proxy data, α_0 and α_1. These parameters are important since both account for the bias in the proxy data. The obtained estimates of these bias parameters will be used to recover the latent exposures field given the observed values of the proxy data. As shown in Figs. 4 and 5, the biases are generally close to zero for all scenarios. However, the RMSEs are higher when the data on the monitors are sparse (scenarios C, D, G, H, K, L); but the values are decreasing with more time points. The proposed method is able to correctly estimate α_1, which is expected to be difficult to estimate since it is a scaling parameter of the latent exposures field $x(\boldsymbol{s}, t)$. There is no difference in the bias and the RMSE for α_1 when using informative or non-informative prior.

Fig. 6 shows the biases and RMSEs for the measurement error at the monitors and at the proxy data, σ_e^2 and σ_δ^2, respectively. Similar to α_0 and α_1, both parameters are of interest since both measure the systematic error in the data. An accurate estimation of these parameters also implies good estimates for the latent exposures field. For σ_e^2, the biases are close to zero for all scenarios but the RMSEs are generally higher when the data on the monitors are sparse and when non-informative priors are used (scenarios D, H, L). This is expected since with few data points from monitors, there is also little information at hand to do the estimation. But for as long as informative priors are used or there are several time points, the RMSEs are generally lower even if the data on the monitors is sparse. For σ_δ^2, the biases are also close to zero for all scenarios. The RMSEs are also generally

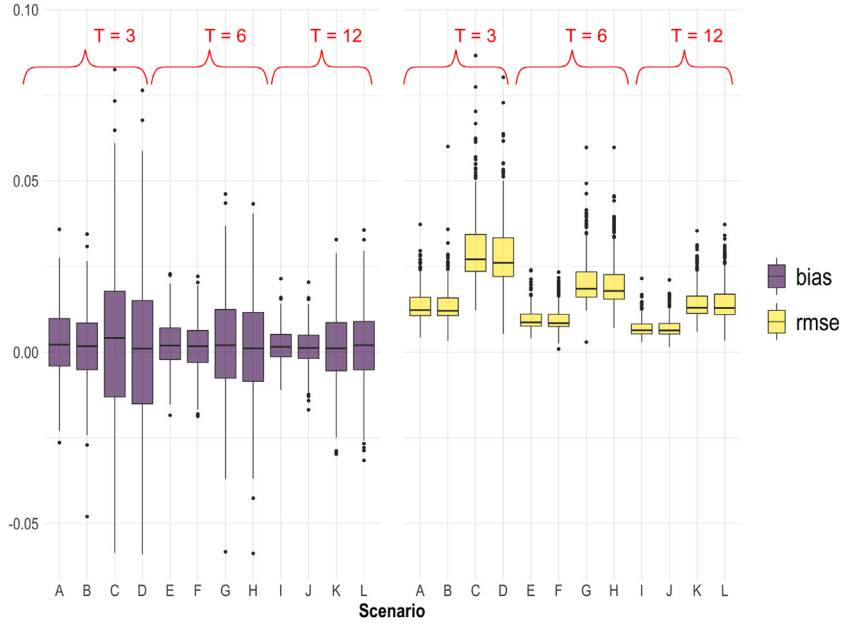


Fig. 5. Plot of bias (purple) and RMSE (yellow) for α_1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

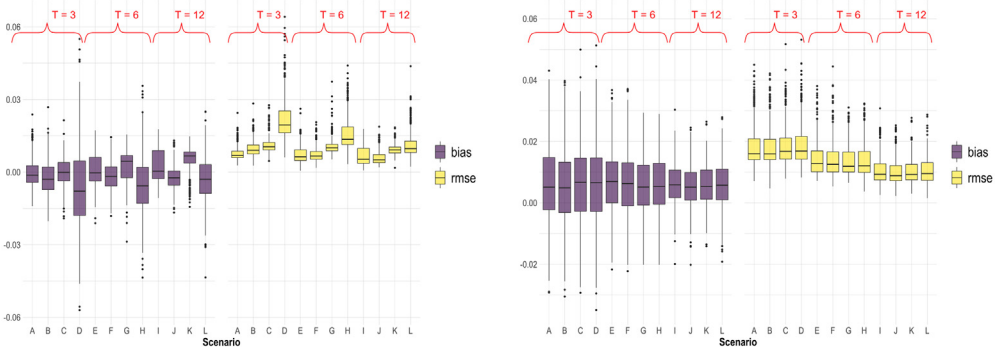


Fig. 6. Plot of bias (purple) and RMSE (yellow) for the measurement error variances σ_e^2 (left) and σ_δ^2 (right). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

small, and is decreasing with more time points. The sparsity in the monitors data and the priors specification does not seem to affect the bias and the RMSE of σ_δ^2.

Fig. B.13(a)–(b) and (d)–(e) in Appendix B shows the biases and RMSEs for the Matèrn parameters σ_ω^2 and ρ. The biases and RMSEs are clearly smaller when informative priors are used which is the expected result since the Matèrn parameters are typically very difficult to estimate. Even if the monitors data is sparse, the biases and RMSEs are still small as long as informative priors are used. This is expected since with few observed values from monitors, we would rely on informative priors to correctly estimate the Matèrn parameters. When non-informative priors are used and data on the monitors are sparse, the bias and RMSEs are expected to be high.

For the temporal parameter of the exposures field, ς, the biases and RMSEs are generally close to zero for all scenarios as shown in Fig. B.13(c) and (f) in Appendix B. The RMSEs are relatively higher when there are fewer time points. This is expected since ς parameterizes the temporal evolution of the spatial field, so that more time points means more information available for estimation. Even with non-informative priors, for as long as there are relatively more time points, the ς parameter is estimated well.

Lastly, shown in Fig. B.14 also in Appendix B are the biases and RMSEs for β_0 and β_1. For both parameters, the biases are close to zero for all scenarios. However, when the data on the monitors is sparse, the RMSE of β_1 is large. The parameter β_1 is the coefficient of the covariate of the latent exposures field. Although this information is also available for the high-resolution proxy data, the proxy data is noisier and less correlated with the true values of the latent field. Hence, this could be the reason for the difficulty in estimating it correctly when the monitors data is sparse. For more time points, the RMSEs for both β_0 and β_1 are generally smaller, since more time points means more information available to do the estimation.

5.2. Block-level exposure estimates

Fig. 7 shows the average of the correlations between the block-level exposure estimates and the corresponding true values for all scenarios. Each point in the plot corresponds to an area or block in the simulation study region. Since there are T time points in the simulated data, the values shown in the figure is the average of the correlations from the T time points which are further averaged from all 500 replicates. All the correlations range from around 0.97 to some value close to 1.0, but using method 2 for computing spatial averages generally gave higher correlations than method 1 for all scenarios. Also, the correlations are generally higher when there are more time points, which is true for both methods.

Fig. B.15 in Appendix B shows the biases and RMSEs in the block-level exposure estimates for all scenarios. Similar to Fig. 7, each point in the plot corresponds to an area or block. The bias for each block at a given time point is computed as the average of the difference between the true block exposure value and its estimated value across all the 500 replicates. The RMSE is computed similarly for each block for a given time point. The final values shown in Fig. B.15 are then the average of the values for all time points T. The bias in the block-level exposures for all 98 areas are generally close to zero; but the spread in the biases are wider for method 1 than method 2. Also, when the data on the monitors are sparse and the priors are non-informative, the biases are generally larger, especially for scenarios with fewer time points. This pattern observed for the biases is also true for the RMSEs. There are a couple of areas with very high RMSEs when using method 1, and this is consistent for all the scenarios. The RMSEs from using method 1 are generally higher compared to method 2 for all the scenarios.

5.3. Second-stage model parameters

The main parameter of interest is γ_1 since it is the coefficient of the exposures in the health model. Fig. 8 shows the biases and the RMSEs for γ_1 for all the scenarios. There is no striking difference in the biases and RMSEs between the two methods for computing the spatial averages of exposures. The figures show that the bias in γ_1 is close to zero for all scenarios. Also, for more time points, the RMSEs become smaller. This is expected since with more time points, there are more information available to estimate the parameter properly. Moreover, the sparsity of the monitors does not affect the quality of estimates for γ_1. This is also expected since as shown in Section 5.2, the obtained block-level exposures estimates are highly correlated and are close to the true values of block-level exposures whether the monitors data is sparse or not. Hence, the obtained estimates for γ_1 will be similar for either case. Lastly, even with non-informative priors on γ_1, the bias and the RMSE is consistently small. Note that in the simulation study, the γ_1 parameter has a non-informative prior for all the scenarios, and so all the values shown in Fig. 8 are computed using non-informative priors for γ_1. The insights for γ_1 also holds true for the intercept γ_0, as shown in Fig. B.16(a) and (b) in Appendix B.

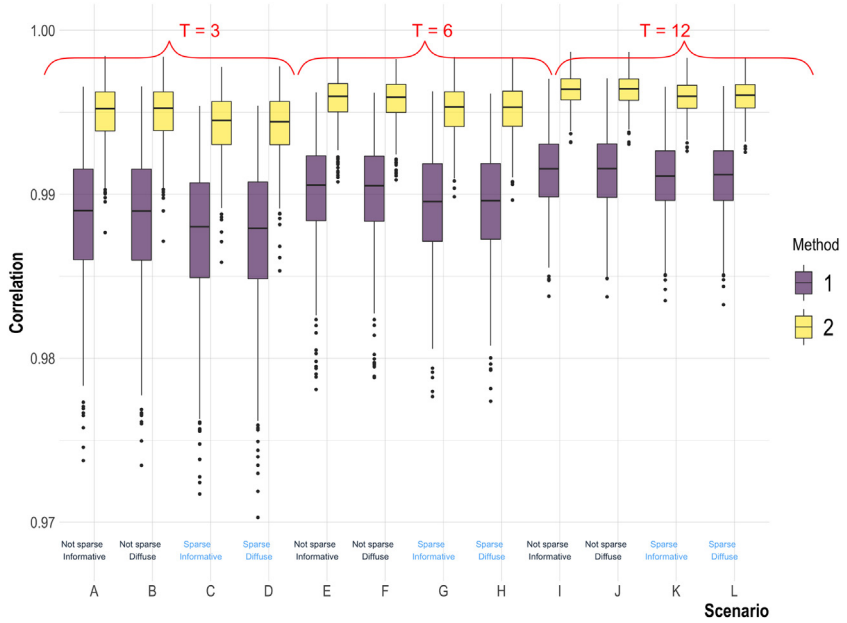


Fig. 7. Plot of correlations with true bock-level exposures for all scenarios.

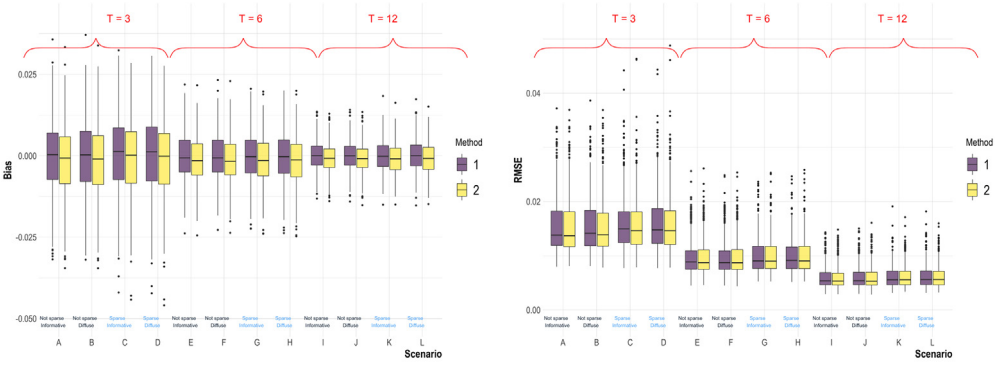
For the variance parameters in the second-stage model, the plots of the biases and RMSEs for all scenarios are shown in Fig. B.16(c)–(f) of Appendix B. For the variance parameter of the time effect σ_ν^2, the biases and RMSEs are generally smaller when there are more time points. This makes sense since σ_ν^2 is a parameter of the temporal random effect, and so it is more accurately estimated when the length of the time series is longer. In addition, the prior specification also affects the precision of the estimates. If the priors used are informative, the RMSEs for σ_ν^2 are generally lower compared the scenarios where the priors are non-informative. However, the difference in the RMSEs for σ_ν^2 for the informative and non-informative priors diminishes with more time points.

For the variance of the block-specific effect σ_ϕ^2, the biases and the RMSEs are generally close to zero for all the scenarios. The number of time points does not seem to unduly affect the biases and RMSEs. This is expected since the model in the simulation study assumes that the block-specific effect is independent with time, and so the number of time points in the data does not potentially affect the accuracy and precision of the estimates of σ_ϕ^2. In addition, this parameter does not seem to be sensitive to the prior specification and the sparsity of the monitors.

Table B.6 in Appendix B shows the coverage probabilities for the health model parameters. For γ_1, all the coverage probabilities are very close to the nominal value of 95%. There is no difference between the two methods of computing spatial averages. For the intercept γ_0, the coverage probabilities are also reasonably close to the nominal value. The coverage probabilities for scenarios with fewer time points is smaller compared to the scenarios with more time points. The coverage probabilities for the variance parameter of the time random effect σ_ν^2 is higher when penalized complexity priors are used or when there are more time points. Finally, for the variance of the block-specific random effect σ_ϕ^2, the coverage probabilities are consistently high and close to the nominal value for all scenarios.

6. Application to real data

The proposed method is applied on a real data in England. The exposures measurements are nitrogen dioxide (NO_2) concentrations at 142 locations from the Automatic Urban and Rural



(a) Plot of bias in γ_1 for all scenarios

(b) Plot of RMSE in γ_1 for all scenarios

Fig. 8. Plot of biases and RMSEs for γ_1.

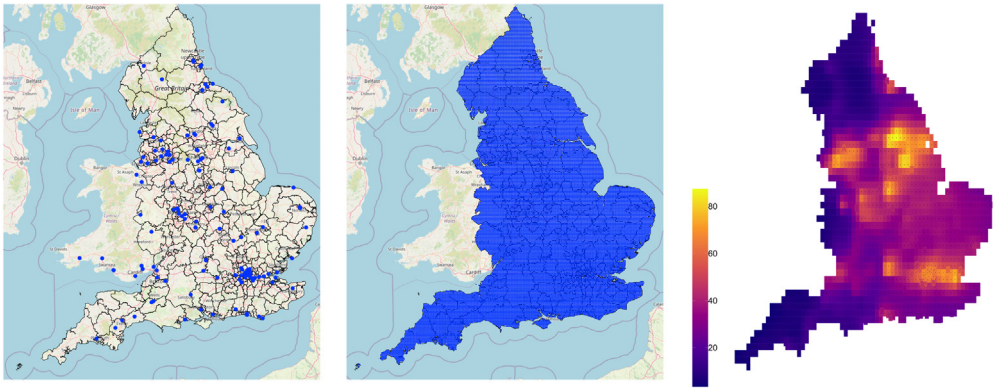
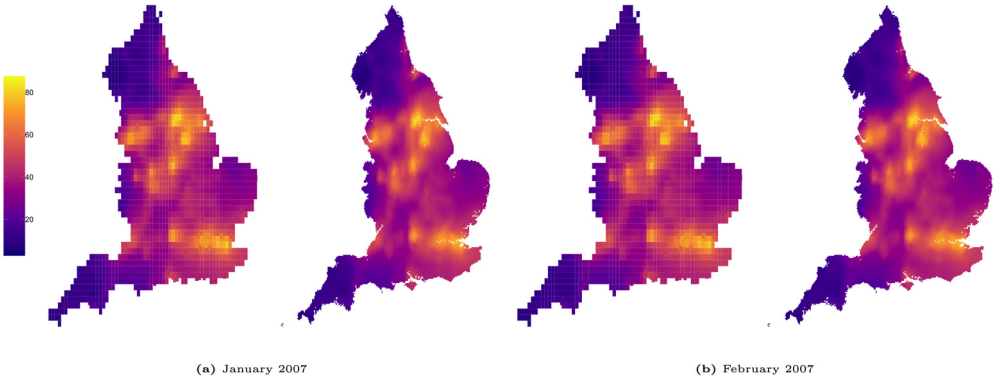


Fig. 9. Plot of the location of monitors in England and Wales (left), the location of the data points for the AQUM data (center), and a plot of NO_2 values from the AQUM for year 2007 and month January (right).

Network (AURN) in England and Wales. Fig. 9 (left) shows the location of the monitors in the study region. The measurements were averaged at a monthly level to be consistent with the temporal resolution of the health data. Data from the Air Quality Unified Model (AQUM) was used as the proxy data for NO_2. The AQUM is a weather and chemical transport model which provides hourly estimates of pollutant concentrations in a 12 km^2 all over England (Lee et al., 2017). Similar to the AURN data, the AQUM data was averaged to monthly level to be consistent with the temporal resolution of the health data. This is the same data used in Lee et al. (2017) where they developed a two-stage model for the same problem of pollution prediction and estimating its long-term impact on health. They used MCMC to do Bayesian inference, as opposed to the proposed method in this paper which uses INLA.

Shown in Fig. 9 (center) are the locations of the data points for the AQUM data in England. Fig. 9 (right) shows the AQUM data for NO_2 for January 2007. The health data used is count of respiratory hospitalizations in England at the level of Local and Unitary Authorities (LUA) which is available at a monthly level. Shown in Fig. 9 (left) is the study map which shows the boundaries of the LUAs in England. We apply the proposed method on a monthly data from January to December 2007. All details of data cleaning and pre-processing are discussed in Lee et al. (2017).

Fig. C.17 in Appendix C shows the mesh for the SPDE component. In fitting the first-stage model, all the priors used were non-informative. Fig. 10 shows the estimated latent NO_2 levels in England



(a) January 2007 (b) February 2007

Fig. 10. Plot of NO_2 from AQUM (left) versus estimated field (right) in England for two time points.

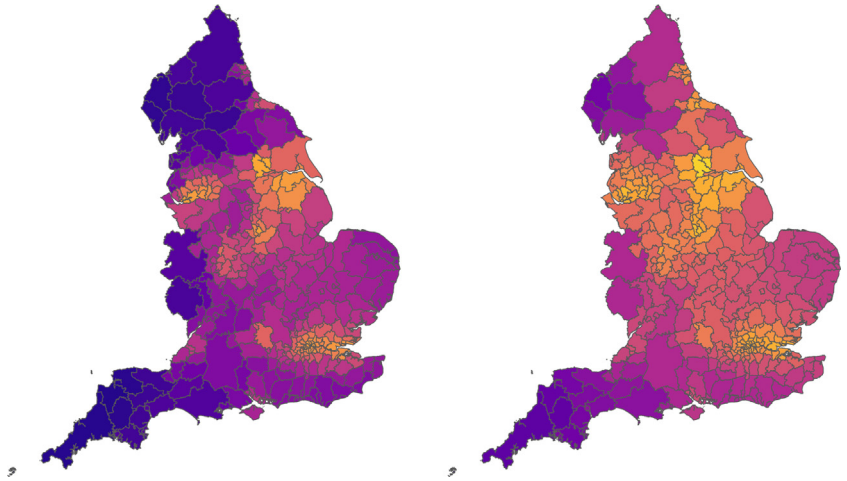


Fig. 11. Block-level estimates of exposures in England for January 2007 (left) and February 2007 (right).

for January 2007 and February 2007 on a very fine prediction grid. Shown also is the AQUM data for the same period. The figure shows that there is a close correspondence between the predicted latent NO_2 concentrations and the AQUM data. Fig. 11 shows the block-level estimates of NO_2 using Method 2 in Eq. (27) using the same scale as in Fig. 10. Method 2 was used to compute the estimates since the simulation study has shown that this method is slightly better than Method 1 in terms of the bias, RMSE, and the correlation with the true exposure values; and the two methods yield almost the same estimates. To account for the uncertainty in the first-stage model, we simulate 100 times from the marginal posterior of the latent exposures field, in Eq. (14), at the prediction grid. For each of the simulated values, block-level estimates are computed which are then used as an input in the second-stage model.

The second-stage model is given in Eqs. (10) and (11). For the general form of the log-risk equation,

$$\log(\lambda(B_i, t)) = \gamma_0 + \gamma_1 \hat{x}(B_i, t) + \phi_i + \psi_i + \zeta_t + \nu_t + \upsilon_{it},$$

it is assumed that the block-specific *iid* effect ϕ_i is $N(0, \sigma_\phi^2)$, the spatially-structured random effect ψ_i follows the intrinsic conditional autoregressive (AR) process which is parameterized by the variance

Table 4
Model choice criteria values.

Model	Marginal likelihood	WAIC	DIC	Failure	PIT	CPO
No interaction	−17 209.0	33 646.12	32 872.71	0%	0.5093	0.0230
Type I	−16 685.1	32 404.37	30 966.22	0.9%	0.5046	0.0204
Type II	−15 831.7	29 366.97	29 284.80	2.51%	0.4972	0.0206
Type III	−17 460.9	29 555.34	29 516.32	11.42%	0.4873	0.0195
Type IV	−18 121.1	29 308.01	29 203.12	2.65%	0.4999	0.0212

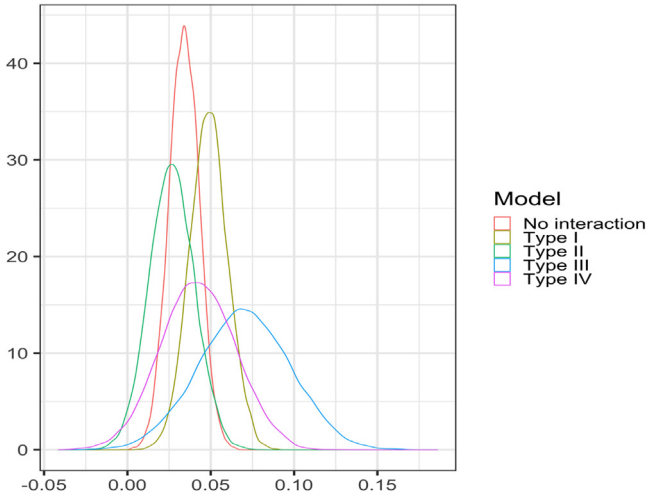


Fig. 12. $\hat{\pi}(\gamma_1|\cdot)$ of the five models.

σ_ψ^2, the time-specific *iid* effect ν_t is $N(0, \sigma_\nu^2)$, and the temporally-structured effect ζ_t is an AR process of order 1 which is parameterized by the lag parameter ρ_ζ and variance σ_ζ^2. The spatio-temporal interaction effect υ_{it} can take several forms. The first case is when the unstructured effects ϕ_i and ν_t interact, which is called *Type I interaction*. When the structured temporal effect ζ_t and the unstructured spatial effect ϕ_i interact, this is *Type II interaction*. *Type III* and *Type IV interaction* are defined as the case when ν_t and ψ_i interact, and when both the structured effects ψ_i and ζ_t interact, respectively. The criteria for model selection are the following: marginal likelihood, widely applicable Bayesian information criterion (WAIC), deviance information criterion (DIC), predictive integral transform (PIT), and the conditional predictive ordinate (CPO). Smaller values for the CPO indicate better model fit. Moreover, if the model fits the data well, the values of the PIT should be close to a uniform distribution.

Table 4 shows the marginal likelihoods, WAIC, DIC, PIT and CPO for the five second-stage models considered. Since several samples ($J = 100$) from the estimated posterior predictive distribution of the latent field $\hat{\pi}(\boldsymbol{x}_t|\cdot)$ were generated and that for each sample, $\hat{x}(B_i, t)$ were computed and were used to fit the second-stage model to properly propagate uncertainty from the first-stage model to the second-stage model, the values shown in Table 4 are the average for all the $J = 100$ simulations.

The model with Type II interaction has the highest marginal likelihood, but the model with Type IV interaction has the lowest WAIC and DIC, although the WAIC and DIC for the Type II model is not too different from the Type IV model. All the mean PIT are close to 0.5 which is the mean of a uniform distribution from 0 to 1, and all the mean CPO are also close to 0. Based on the values, the model with Type II interaction seems to provide the best fit among the five models considered.

Fig. 12 shows the estimated posterior distribution of γ_1, $\hat{\pi}(\gamma_1|\cdot)$, for the five models considered. The γ_1 estimates are quite close to each other except for the model with Type III interaction. As

Table 5
Estimates of the final second-stage model.

Parameter	Mean	P2.5%	P50%	P97.5%
γ_0	−0.5034	−0.8283	−0.5192	−0.0690
γ_1	0.0267	0.0006	0.0266	0.0533
σ^2_ϕ	0.0608	0.0295	0.0592	0.1026
σ^2_ψ	0.3050	0.1800	0.2931	0.4914
σ^2_ζ	0.0625	0.0173	0.0476	0.1954
ρ_ζ	0.7970	0.5054	0.8189	0.9592
σ^2_υ	0.0006	0.0000	0.0001	0.0048
σ^2_υ	0.0225	0.0181	0.0223	0.0281
ρ_υ	0.8105	0.7446	0.8023	0.8540

shown in Table 4, the model with Type III interaction has 11.42% of the observations with predictive measures which are not reliable due to some numerical problems; hence, the γ_1 estimate for this model might not be a reliable one. Table 5 shows the parameter estimates and the 95% credible intervals for the model with Type II interaction. Since the block-level estimates of exposures were log-transformed, for a 10% increase in the NO2 levels for a block, we expect the relative risk of respiratory hospitalization to increase by 2.67%. There seems to be a high spatial correlation in the risks since the estimate for the variance of the spatially structured effect σ^2_ψ (0.3050) is higher than the unstructured effect σ^2_ϕ (0.0608). Also, the temporal correlations is very evident as seen by the estimated AR coefficient ρ_ζ of 0.7970 and the estimate of the variance of the structured temporal effect σ^2_ζ (0.0625) which is bigger than the variance of the unstructured temporal effect σ^2_υ (0.0006). A Type II interaction effect means that ith area/block has its own autoregressive structure which is independent from the other areas. The estimated AR coefficient ρ_υ in the interaction term is 0.8105 which indicates a strong interaction effect.

7. Conclusions

This paper proposes a two-stage spatio-temporal model for the epidemiological problem of estimating the effect of exposures on health outcomes where the data have different spatial supports. The framework of the first-stage model is based on the Bayesian melding model for which a common latent field is assumed for the geostatistical data on the monitors and the high-resolution proxy data, both being error-prone realizations of the latent field. However, instead of treating the proxy data as areal, it is treated as geostatistical data on the centroids of the grid cells. A couple of reasons for this simplification is the assumption that the resolution of the covariate data is not finer than that of the proxy data, that the resolution of the proxy data is fine enough that it can be treated as geostatistical at the centroids, and also for the purpose of simplifying the computation. Nonetheless, the proposed model incorporates a non-spatially varying additive and multiplicative bias, and an additional additive noise in the proxy data since it is noisier and less correlated with the true latent field compared to the measurements from the network of monitoring stations. The SPDE approach is used to model the spatio-temporal structure of the first-stage model in order to speed up computation and spatial interpolation. The second stage fits a Poisson model using the spatial averages of the latent field, and additional spatial and temporal random effects. Both the first-stage and second-stage models, being latent Gaussian, are fitted using the integrated nested Laplace approximation. In estimating the block-level exposure estimates, two methods proposed in Cameletti et al. (2019) were used. In order to account for the estimation error in the first-stage model when fitting the second-stage model, samples from the marginal posterior distribution of the latent field at the prediction grid are generated. For each set of simulated values, all block-level estimates of exposures are computed using the two methods and are then used as input in the second-stage model. This is done several times and then all results are combined and used to approximate the posterior distribution of the second-stage model parameters.

The proposed method worked in the actual data with the expected result that NO_2 is significantly associated with respiratory diseases and with additional insights about the spatial and temporal

structure of the risks. A simulation study was carried out to assess the performance of the proposed method under different scenarios. Considered in the study are the following settings: the sparsity of the data on the monitors, the number of time points, and the prior specification. It is common to work with sparse data on the monitors so it is interesting to look at the effect of the sparsity on the quality of the parameter estimates. Also, it is important in Bayesian analysis to assess the sensitivity of the results to the priors, more so in the context of a complex spatio-temporal process with several model parameters that need to be estimated.

All first-stage model parameters have generally small biases but there is difficulty in estimating the Matèrn field parameters, particularly the spatial variance and the range parameter, especially when non-informative priors are used. As long as informative priors are used, the bias and RSMEs are very small and the coverage probabilities are very high even if the data on the monitors is sparse.

For the main parameter of interest, γ_1, the method provides very good estimates across all scenarios considered in the simulation study. There is no difference between the two methods of computing the spatial averages in terms of the bias, RMSEs, and coverage probabilities. Even with non-informative prior on γ_1 and sparse data on the monitors, the estimates for γ_1 are close to the true value. Finally, with more time points, the RMSEs tend to decrease.

The simulation study done also showed that the sparsity of the data on the monitors can potentially affect the quality of the parameter estimates. When the data on the monitors is sparse, the RMSE of the covariate effect in the latent field is large. This is also true for the measurement error variance in the monitors, but the use of informative priors can be helpful to accurately estimate the parameter. It makes sense for these two parameters to be seriously affected by the sparsity of the monitors data since these are components of the observation model for the monitors; hence, the less monitors in the study map, the less information at our disposal to accurately estimate them. The RMSEs of the Matèrn parameters and the bias parameters of the proxy data are also generally higher when the monitors data is sparse. The parameters in the second-stage model seem to be not affected by the sparsity of the data on the monitors since the proposed method, even with sparse data, was able to estimate well the latent field and that the block-level predicted values are also close to the true value, at least for most of the areas, as shown in the correlations, biases, and RMSEs from the simulation results. However, in practice, this will only be true for as long as the block-level estimates of exposures are close to the unknown block averages of the latent continuous process.

The use of informative priors gave better parameter estimates especially for the latent field parameters which are typically the most difficult parameters to be estimated. When informative priors are used, the estimates of the spatial field variance and the range parameter have lower bias and RMSE. The autoregressive parameter of the latent field also benefits with the use of informative priors, giving smaller errors and higher coverage probabilities. In addition, the measurement error variance at the monitors also have lower RMSEs with informative priors especially for the case when the data is sparse. The rest of the parameters are not too sensitive to the prior specification which includes the variance of the block-specific random effect in the second-stage model.

The number of time points can also potentially affect the quality of the estimates. As already mentioned, if there are more time points, the RMSEs of the fixed effects in the second-stage model are smaller. This is also true for the variance of the time effect in the second-stage model. The bias parameters of the proxy data in the first-stage model also have better estimates with more time points.

The method for computing block-level exposures estimates does not show to have an impact on the parameter estimates. The biases in the block-level estimates for both methods are on the average close to zero, but the first method seem to give higher biases for certain blocks and also higher RMSEs overall. Also, the second method gave slightly higher correlations between the true block-level exposure values and the estimated values. Nonetheless, both methods gave fairly high correlations which are at least 0.97. As far as the parameter estimation is concerned, either of the two methods should work very well.

A possible drawback of the proposed method is that it becomes computationally expensive with a long time series data since it will significantly increase the number of unobserved values of the latent field that need to be estimated. This would primarily depend on the resolution of the proxy data and also the size of the study region.

The use of the full Bayesian melding approach which treats the proxy data as areal would also potentially increase the computational costs. Under the full Bayesian melding model, it is required to have covariate information on $\boldsymbol{z}(\boldsymbol{s}, t)$ at a resolution which is finer than the resolution of the proxy data. Using the data augmentation approach would imply the need for an enlarged extended field, say $\boldsymbol{\chi}_t = \left(\boldsymbol{x}_{t,M}^{\top} \quad \boldsymbol{x}_{t,P*}^{\top} \quad \boldsymbol{x}_{t,M}^{*\top} \quad \boldsymbol{x}_{t,P*}^{*\top}\right)^{\top}, t = 1, \ldots, T$; $\boldsymbol{x}_{t,P*}^{*\top}$ is of larger dimension than $\boldsymbol{x}_{t,P}^{*\top}$ in the original model, where P^* is the resolution of the covariate data $\boldsymbol{z}(\boldsymbol{s}, t)$, which should be greater than the number of grid cells for the proxy data. Hence, when the proxy data has a very high resolution, applying the full Bayesian melding model might amplify the computation effort required. The use of a full Bayesian melding model is a problem that could be investigated in a future work.

Having more than one pollutant (each from multiple data sources) in a single model is a straightforward extension of the current model. In this scenario, the proxy data and the locations of the monitors for each pollutant need not be aligned in space, but this will most likely require the use of multiple but possibly correlated latent processes. This is doable and only requires adding additional equations in the data fusion (first-stage) model. For the case when there are more than two sources of information for a single pollutant – say a satellite data in addition to the monitors data and the data from dispersion models in a single spatio-temporal model – it will also be straightforward to extend the joint (first-stage) model to accommodate this additional information by assuming that this also is an outcome from the same latent process with a different set of calibration parameters. Another extension of the current model is to consider other measurement error process for the monitors data and the proxy data. For as long as the latent Gaussian assumption is satisfied, the current framework of using the INLA-SPDE approach can still be used in such extensions.

The proposed model allows one to obtain estimates of the relative risks at the level of the administrative units where the health outcomes are observed. However, one might be interested to obtain the estimates of the relative risks at a fine scale, say on the same scale as the proxy data since this allows one to look at fine-scale heterogeneity of the risks in space. This is not explored in the current paper but this could be framed in a latent Gaussian model framework. A recent work has been done on this in a so-called fusion area-cell spatio-temporal generalized geoadditive-Gaussian Markov random field (FGG-GMRF) (Jaya and Folmer, 2022).

The proposed model assumes a separable covariance function since the covariance matrix of the SPDE representation of the first-stage model is a Kronecker product of the covariance matrix in space, whose elements are computed from the SPDE model, and the covariance matrix in time, whose elements are computed from the assumed temporal model. The separability assumption is convenient because it simplifies computation; but this assumption could be inadequate. An extension of the model is to assume non-separability which is currently an active area of research. One approach of doing this is to start from an SPDE which yields a Gaussian field with a separability parameter (Lindgren et al., 2020). This parameter determines the type of non-separability of the spatio-temporal covariance. This approach which starts from an SPDE is consistent with the framework adopted in this paper; but there are other approaches for this problem that could be adopted.

The current model assumes that the latent spatio-temporal process is constant through time. But the spatio-temporal process could evolve at some point, i.e., the Matèrn field parameters could change, or the mean structure of the model could evolve as well. This is now a change-point detection problem which is also a promising future work since relatively only few work has been done for change-point detection in spatio-temporal processes. Finally, approaches to improve computational speed and efficiency especially for long time series is a very interesting direction, and also very useful for many practitioners who work with big spatio-temporal data.

CRediT authorship contribution statement

Stephen Jun Villejo: Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Janine B Illian:** Writing – review & editing, Supervision. **Ben Swallow:** Writing – review & editing, Supervision.

Declaration of competing interest

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

Appendix A. Model evaluation

The performance of the method is evaluated by looking at the bias, RMSE, and coverage probabilities. There is a slight difference in the notations for the formulas to compute the model performance metrics between the first-stage and second-stage model parameters. For a first-stage parameter θ, $\hat{\theta}_{ik}$ denotes the kth sampled value from the estimated posterior distribution of θ in the ith replication, $i = 1, \ldots, n_{sim}$ and $k = 1, \ldots, K$. For a second-stage parameter θ, $\hat{\theta}_{ijk}$ denotes the kth sampled value from the estimated posterior distribution of θ using the block-level estimates $\hat{x}(B, t)$ computed using the jth simulated values from the marginal posterior distribution of the latent field, $j = 1, \ldots, J$, $k = 1, \ldots, K$, $i = 1, \ldots, n_{sim}$. In the simulation study, $n_{sim} = 500, J = 100$ and $K = 200$.

1. **Bias** - The bias is computed as

$$\text{bias} = \frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} \left(\frac{1}{K} \sum_{k=1}^{K} \hat{\theta}_{ik} - \theta \right) \quad \text{or} \quad \text{bias} = \frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} \left(\frac{1}{JK} \sum_{j=1}^{J} \sum_{k=1}^{K} \hat{\theta}_{ijk} - \theta \right)$$

   for a first-stage and second-stage model parameter, respectively.

2. **Coverage probability** - For a given data replicate i, suppose $\hat{\theta}_i^{(2.5)}$ is the 2.5th percentile of $\hat{\theta}_{ik}$ for all k for a first-stage parameter or of $\hat{\theta}_{ijk}$ for all j, k for a second-stage parameter. Similarly, suppose $\hat{\theta}_i^{(97.5)}$ is the 97.5th percentile. The 95% coverage probability for θ is given by

$$\text{coverage probability} = \frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} \mathbb{I}_i(\theta), \qquad \mathbb{I}_i(\theta) = \begin{cases} 1 & \hat{\theta}_i^{(2.5)} < \theta < \hat{\theta}_i^{(97.5)} \\ 0 & \text{otherwise,} \end{cases}$$

3. **Root mean square error (RMSE)** - The RMSE is computed as

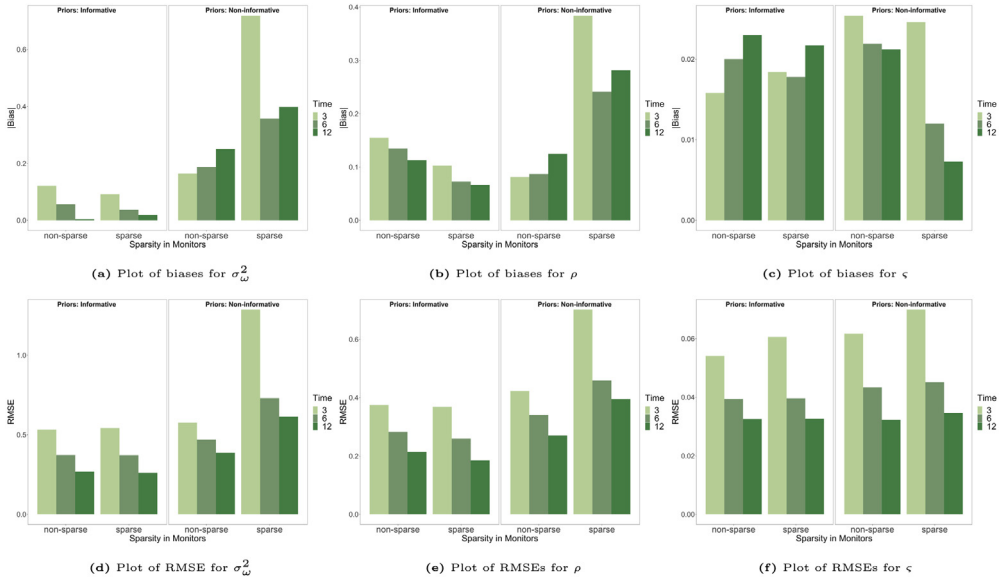
$$\text{RMSE} = \frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} \sqrt{\frac{1}{K} \sum_{k=1}^{K} (\hat{\theta}_{ik} - \theta)^2} \quad \text{or} \quad \text{RMSE} = \frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} \sqrt{\frac{1}{JK} \sum_{j=1}^{J} \sum_{k=1}^{K} (\hat{\theta}_{ijk} - \theta)^2}$$

   for a first-stage parameter and second-stage parameter, respectively.

The same model evaluation metrics are used for the block-level exposure estimates. It should be noted that the true values of the block-level exposures vary for the different replicates.

Appendix B. Simulation results

See Figs. B.13–B.16 and Table B.6.



(a) Plot of biases for σ^2_ω (b) Plot of biases for ρ (c) Plot of biases for ς

(d) Plot of RMSE for σ^2_ω (e) Plot of RMSEs for ρ (f) Plot of RMSEs for ς

Fig. B.13. Plot of biases and RMSEs for the Matèrn field parameters σ^2_ω, ρ, and the temporal parameter ς.

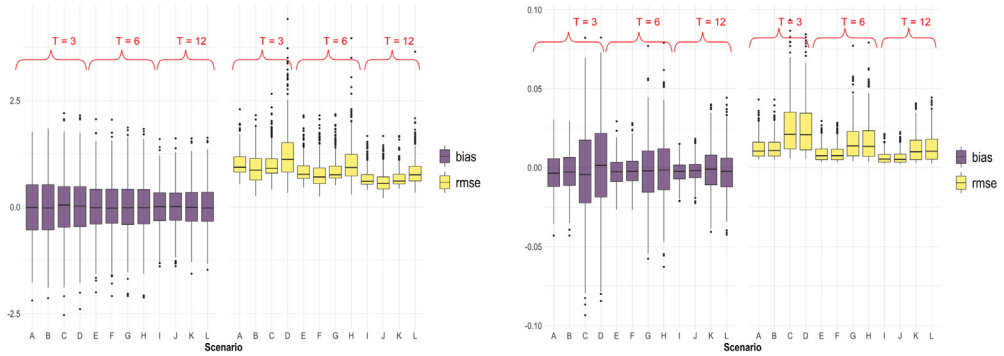


Fig. B.14. Plot of bias (purple) and RMSE (yellow) for β_0 (left) and β_1 (right) for all scenarios. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

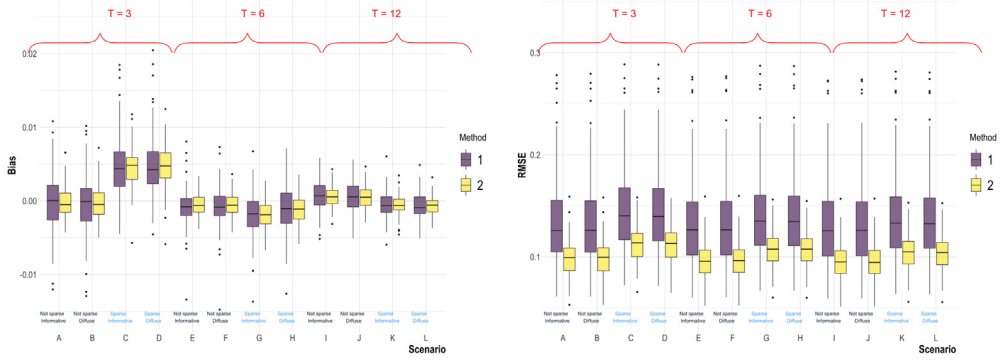
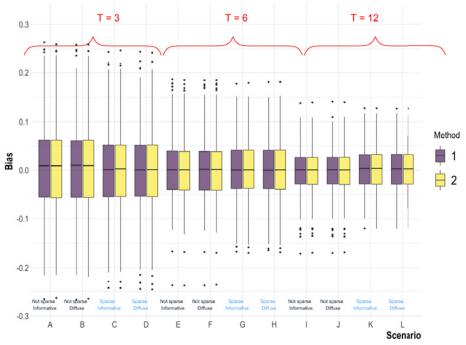
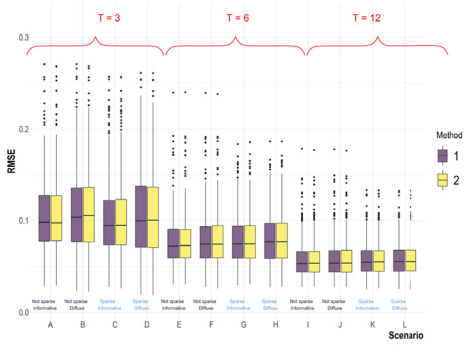


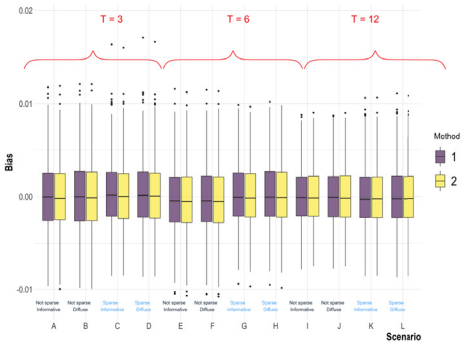
Fig. B.15. Plot of biases (left) and RMSEs (right) in block-level exposure estimates for all scenarios.



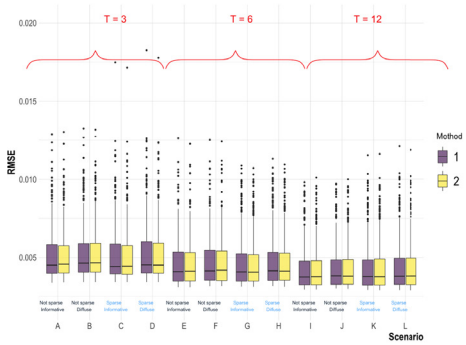
(a) Plot of bias in γ_0 for all scenarios



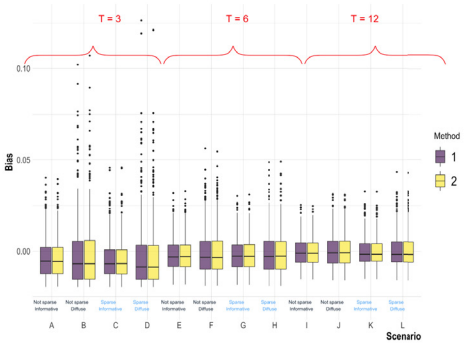
(b) Plot of RMSE in γ_0 for all scenarios



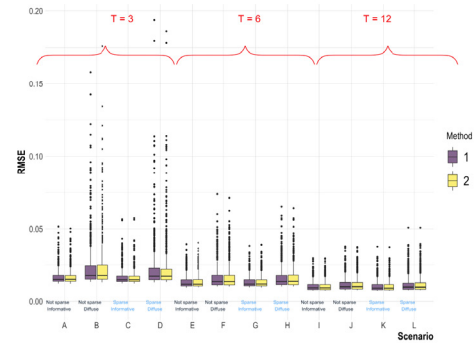
(c) Plot of bias in σ^2_ϕ for all scenarios



(d) Plot of RMSE in σ^2_ϕ for all scenarios



(e) Plot of bias in σ^2_ν for all scenarios



(f) Plot of RMSE in σ^2_ν for all scenarios

Fig. B.16. Plot of biases and RMSEs for second-stage model parameters.

Table B.6
Coverage probabilities of second-stage model parameters for all scenarios.

T	Sparse	Priors	γ_0		γ_1		σ^2_ϕ		σ^2_v	
			M1	M2	M1	M2	M1	M2	M1	M2
3	No	Informative	85.4	85.2	94.2	93	95.2	93.8	88.4	88.6
		Non-informative	81.8	81.2	93.6	93.4	94.8	93.4	71.6	71.6
	Yes	Informative	86.4	85.8	94	95	95	95.4	88	88
		Non-informative	83	82.6	93.2	93.8	94.2	94.6	71.2	71
6	No	Informative	91.8	91.6	93.6	94	94.2	93	89.8	89.4
		Non-informative	91.4	91.2	94	94	94	92.6	84.2	83.4
	Yes	Informative	91	90.6	94.4	94	94.2	94	89.6	90.6
		Non-informative	90	90.4	94.2	94	94.2	93.6	86	85.4
12	No	Informative	92.6	93.2	94	93.8	94.2	95.2	92.4	93
		Non-informative	93.4	92.8	94	94.4	94.2	94.6	90	90
	Yes	Informative	94.8	94.6	95.4	93.8	93.6	93.4	92.8	92
		Non-informative	93.6	94.2	93.4	93.6	93.2	93.4	89.8	89.2

Appendix C. Mesh for the SPDE component

See Fig. C.17.

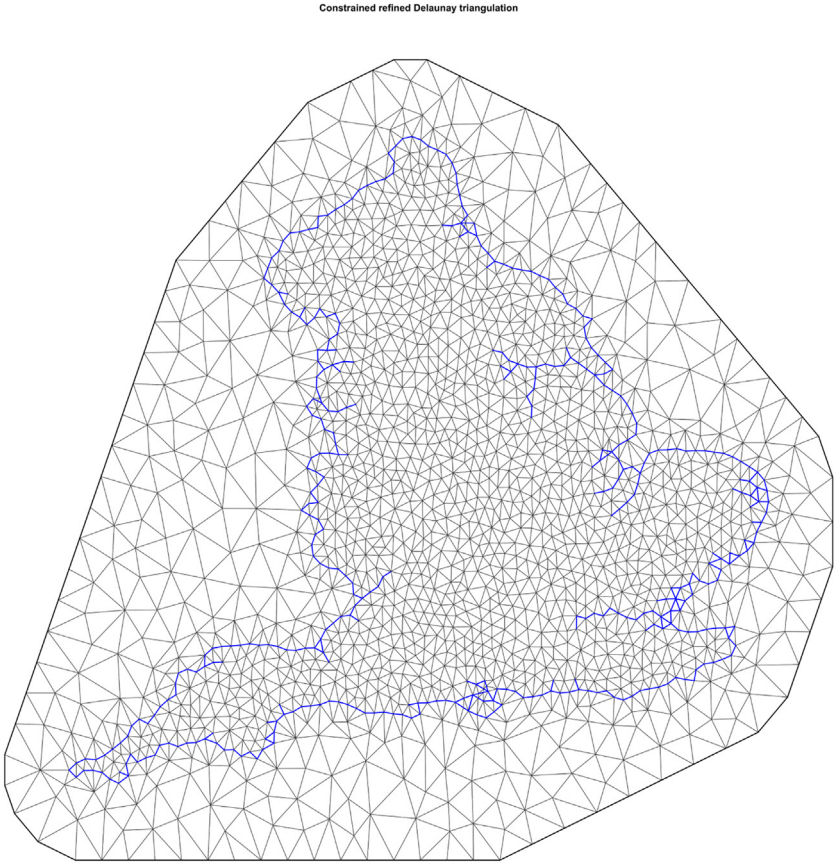


Fig. C.17. Mesh for the SPDE component covering England.

References

Bergen, Silas, Szpiro, Adam A., 2015. Mitigating the impact of measurement error when using penalized regression to model exposure in two-stage air pollution epidemiology studies. Environ. Ecol. Stat. 22 (3), 601–631.

Berrocal, Veronica J., Gelfand, Alan E., Holland, David M., 2010. A spatio-temporal downscaler for output from numerical models. J. Agric. Biol. Environ. Stat. 15 (2), 176–197.

Berrocal, Veronica J., Gelfand, Alan E., Holland, David M., 2012. Space-time data fusion under error in computer model output: An application to modeling air quality. Biometrics 68 (3), 837–848.

Bivand, Roger, Piras, Gianfranco, 2015. Comparing implementations of estimation methods for spatial econometrics. J. Stat. Softw. 63, 1–36.

Blangiardo, Marta, Cameletti, Michela, 2015. Spatial and Spatio-Temporal Bayesian Models with R-INLA. John Wiley & Sons.

Blangiardo, Marta, Finazzi, Francesco, Cameletti, Michela, 2016. Two-stage Bayesian model to evaluate the effect of air pollution on chronic respiratory diseases using drug prescriptions. Spatial Spatio-Tempor. Epidemiol. 18, 1–12.

Bruno, Francesca, Cameletti, Michela, Franco-Villoria, Maria, Greco, Fedele, Ignaccolo, Rosaria, Ippoliti, Luigi, Valentini, Pasquale, Ventrucci, Massimo, 2016. A survey on ecological regression for health hazard associated with air pollution. Spatial Stat. 18, 276–299.

Cameletti, Michela, Gómez-Rubio, Virgilio, Blangiardo, Marta, 2019. Bayesian modelling for spatially misaligned health and air pollution data through the INLA-SPDE approach. Spatial Stat. 31, 100353.

Cameletti, Michela, Lindgren, Finn, Simpson, Daniel, Rue, Håvard, 2013. Spatio-temporal modeling of particulate matter concentration through the SPDE approach. AStA Adv. Stat. Anal. 97 (2), 109–131.

Diggle, Peter J., 2013. Statistical Analysis of Spatial and Spatio-Temporal Point Patterns. CRC Press.

Fuentes, Montserrat, Raftery, Adrian E., 2005. Model evaluation and spatial interpolation by Bayesian combination of observations with outputs from numerical models. Biometrics 61 (1), 36–45.

Fuglstad, Geir-Arne, Simpson, Daniel, Lindgren, Finn, Rue, Håvard, 2019. Constructing priors that penalize the complexity of Gaussian random fields. J. Amer. Statist. Assoc. 114 (525), 445–452.

Gotway, Carol A., Young, Linda J., 2002. Combining incompatible spatial data. J. Amer. Statist. Assoc. 97 (458), 632–648.

Gryparis, Alexandros, Paciorek, Christopher J, Zeka, Ariana, Schwartz, Joel, Coull, Brent A, 2009. Measurement error caused by spatial misalignment in environmental epidemiology. Biostatistics 10 (2), 258–274.

Jaya, IGNM, Folmer, Henk, 2022. Spatiotemporal high-resolution prediction and mapping: methodology and application to dengue disease. J. Geogr. Syst. 1–55.

Krall, Jenna R, Chang, Howard H, Sarnat, Stefanie Ebelt, Peng, Roger D, Waller, Lance A, 2015. Current methods and challenges for epidemiological studies of the associations between chemical constituents of particulate matter and health. Curr. Environ. Health Rep. 2 (4), 388–398.

Lawson, Andrew B, Banerjee, Sudipto, Haining, Robert P, Ugarte, María Dolores, 2016. Handbook of Spatial Epidemiology. CRC Press.

Lee, Duncan, Mukhopadhyay, Sabyasachi, Rushworth, Alastair, Sahu, Sujit K, 2017. A rigorous statistical framework for spatio-temporal pollution prediction and estimation of its long-term impact on health. Biostatistics 18 (2), 370–385.

Lee, Duncan, Shaddick, Gavin, 2010. Spatial modeling of air pollution in studies of its short-term health effects. Biometrics 66 (4), 1238–1246.

Lee, A., Szpiro, A., Kim, S.Y., Sheppard, L., 2015. Impact of preferential sampling on exposure prediction and health effect inference in the context of air pollution epidemiology. Environmetrics 26 (4), 255–267.

Lindgren, Finn, Bakka, Haakon, Bolin, David, Krainski, Elias, Rue, Håvard, 2020. A diffusion-based spatio-temporal extension of Gaussian Matérn fields. arXiv e-prints arXiv–2006.

Lindgren, Finn, Rue, Håvard, 2015. Bayesian spatial modelling with R-INLA. J. Stat. Softw. 63, 1–25.

Lindgren, Finn, Rue, Håvard, Lindström, Johan, 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J. R. Stat. Soc. Ser. B Stat. Methodol. 73 (4), 423–498.

Liu, Zhong, Le, Nhu D., Zidek, James V., 2011. An empirical assessment of Bayesian melding for mapping ozone pollution. Environmetrics 22 (3), 340–353.

Liu, Yi, Shaddick, Gavin, Zidek, James V., 2017. Incorporating high-dimensional exposure modelling into studies of air pollution and health. Stat. Biosci. 9 (2), 559–581.

Martins, Thiago G, Simpson, Daniel, Lindgren, Finn, Rue, Håvard, 2013. Bayesian computing with INLA: new features. Comput. Statist. Data Anal. 67, 68–83.

McMillan, Nancy J, Holland, David M, Morara, Michele, Feng, Jingyu, 2010. Combining numerical model output and particulate data using Bayesian space–time modeling. Environmetrics: Official J. Int. Environmetrics Soc. 21 (1), 48–65.

Molitor, John, Molitor, Nuoo-Ting, Jerrett, Michael, McConnell, Rob, Gauderman, Jim, Berhane, Kiros, Thomas, Duncan, 2006. Bayesian modeling of air pollution health effects with missing exposure data. Am. J. Epidemiol. 164 (1), 69–76.

Rue, Havard, Held, Leonhard, 2005. Gaussian Markov Random Fields: Theory and Applications. CRC Press.

Rue, Håvard, Martino, Sara, Chopin, Nicolas, 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. R. Stat. Soc. Ser. B Stat. Methodol. 71 (2), 319–392.

Ruiz-Cárdenas, Ramiro, Krainski, Elias T., Rue, Håvard, 2012. Direct fitting of dynamic models using integrated nested Laplace approximations—INLA. Comput. Statist. Data Anal. 56 (6), 1808–1828.

Sahu, Sujit K., Gelfand, Alan E., Holland, David M., 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.

Szpiro, Adam A., Paciorek, Christopher J., 2013. Measurement error in two-stage analyses, with application to air pollution epidemiology. Environmetrics 24 (8), 501–517.

Wikle, Christopher K., Berliner, L. Mark, 2005. Combining information across spatial scales. Technometrics 47 (1), 80–91.