A low-rank Bayesian approach for geoadditive modeling

A low-rank Bayesian approach for geoadditive modeling

Bryan Sumalinab*[1,2], Oswaldo Gressani[1], Niel Hens[1,3], Christel Faes[1]

[1]Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat), Data Science Institute (DSI), Hasselt University, Hasselt, Belgium

[2]Department of Mathematics and Statistics, College of Science and Mathematics, Mindanao State University - Iligan Institute of Technology, Iligan City, Philippines

[3]Centre for Health Economic Research and Modelling Infectious Diseases (CHERMID), Vaccine & Infectious Disease Institute, Antwerp University, Antwerp, Belgium

*Corresponding author. *E-mail address*: bryan.sumalinab@uhasselt.be

Abstract

Kriging is an established methodology for predicting spatial data in geostatistics. Current kriging techniques can handle linear dependencies on spatially referenced covariates. Although splines have shown promise in capturing nonlinear dependencies of covariates, their combination with kriging, especially in handling count data, remains underexplored. This paper proposes a new Bayesian approach to the low-rank representation of geoadditive models, which integrates splines and kriging to account for both spatial correlations and nonlinear dependencies of covariates. The proposed method accommodates Gaussian and count data inherent in many geospatial datasets. Additionally, Laplace approximations to selected posterior distributions enhances computational efficiency, resulting in faster computation times compared to Markov chain Monte Carlo techniques commonly used for Bayesian inference. The performance of our novel method is assessed through a simulation study, demonstrating the effectiveness of the proposed approach. The methodology is applied to the analysis of heavy metal concentrations in the Meuse river and vulnerability to coronavirus disease 2019 in Belgium.

Keywords: Kriging, Geoadditive models, Bayesian P-splines, Laplace approximations, Low-rank model.

1 Introduction

Observations characterized by spatial locations often exhibit inherent correlations, with closer observations demonstrating stronger dependencies than those farther apart. This spatial correlation phenomenon is a fundamental aspect of spatial data analysis, especially in disciplines such as geostatistics, where the spatial arrangement of data points provides valuable insights into underlying processes where proximity often implies similarity. Cressie (1993) categorized spatial data into three main types: areal (or lattice) data, geostatistical (continuous) data, and point patterns. Our primary focus here is on the analysis of geostatistical data. Geostatistics is a field dedicated to studying phenomena that are continuously distributed over spatial domains. For example, in environmental studies, nearby soil samples are likely to have similar characteristics due to shared environmental conditions and geological processes. The principles and methodologies of geostatistics can also be applied to phenomena that are not strictly continuous (e.g. areal data), including those related to epidemic modeling and public health outcomes. For instance, geostatistical methods can be used to assess the impact of spatially referenced exposure/covariates on health outcomes, model the spatial variation in disease incidence rates, identify high-risk areas, and assess the spatial dependence of disease transmission (see e.g. Waller and Gotway, 2004; Diggle and Giorgi, 2019). In addition, spatial interpolation techniques, similar to those used in geostatistics, can also help to estimate disease prevalence in regions with sparse or missing data.

One of the primary applications in geostatistics is spatial prediction/interpolation, where missing or unobserved values at specific (unsampled) locations within a spatial domain are estimated. Kriging, a widely used geostatistical technique, relies on spatial correlations to interpolate and predict values across a spatial domain. Kriging methods estimate the spatial variability by considering the spatial arrangement and correlation between observations, resulting in predictions that minimize estimation errors. The strength of kriging lies in its ability to incorporate both the spatial trend and the spatial correlation structure of the data, making it a powerful tool

for spatial data analysis. Although theories about kriging are well established, dealing with larger sample sizes presents a significant computational burden, which is particularly evident in the inversion of the covariance matrix. The computational cost increases as the dimension of the covariance matrix grows with the sample size. A promising solution comes in the form of low-rank representations of spatial models using basis functions, which substantially improve computation time. One such approach is the fixed rank kriging (FRK) proposed by Cressie and Johannesson (2008), for which the spatial component is modeled as a linear combination of basis functions and spatially correlated random coefficients. This methodology was further extended to accommodate non-Gaussian spatial and spatio-temporal data (Sainsbury-Dale et al., 2024). For low-rank Bayesian approaches, one example is the predictive process model (Banerjee et al., 2008; Eidsvik et al., 2012). Another approach that motivates the method proposed here is that of Kammann and Wand (2003), which involves reducing the spatial locations to a subset, called "knots", using a space-filling algorithm (Johnson et al., 1990; Nychka and Saltzman, 1998). They adopted a spline-basis approach and relied upon the commonly used stationary covariance matrix in kriging as the basis functions. Their approach offers not only computational advantages for handling big data but also ease of implementation through standard mixed models software. Another basis function method that gained significant attention in recent years is the Stochastic Partial Differential Equation (SPDE) framework introduced by Lindgren et al. (2011). Their method offers a computationally efficient representation of Gaussian Random Fields (GRFs) by linking Gaussian spatial processes with the Matérn covariance function as a solution to a specific SPDE. This connection allows dense covariance matrices to be transformed into sparse precision matrices, enabling more efficient analysis of large spatial datasets and is implemented using Integrated Nested Laplace Approximations (INLA) (Rue et al., 2009). For a review of low-rank or basis function approaches, see Wikle (2010b) and Cressie et al. (2022). Stein (2014) noted the limitations of low-rank covariance approximations for Gaussian models and demonstrated that an independent-block approach can yield more accurate approximations in certain scenarios, mainly when the observations are sufficiently dense with a small enough nugget variance. Nevertheless, Stein (2014) also mentioned that the conclusions primarily consider cases when the true model is stationary, and it remains an open question whether these limitations extend directly to nonstationary spatial processes. Another alternative to low-rank approaches is covariance

tapering, which approximates the covariance matrix with a sparse matrix, achieving computational efficiency by using sparse matrix techniques (Furrer et al., 2006; Kaufman et al., 2008). Sang and Huang (2012) combined covariance tapering with low-rank covariance approximations, thereby leveraging the strengths of both methods to improve modeling accuracy and addressing limitations inherent in either approach when used independently. Another approach that induces sparsity in the precision matrix is to work with Vecchia approximations (Vecchia, 1988) generalized by Katzfuss and Guinness (2021), which include several special cases. One example is the nearest-neighbor Gaussian process (Datta et al., 2016; Finley et al., 2019) implemented within a Bayesian hierarchical modeling framework and leveraging sparsity for computational efficiency.

While the methods mentioned above offer practical solutions for modeling large spatial datasets, their explicit integration with nonlinear additive covariate effects, especially in the context of non-Gaussian observations, has remained underexplored. Kammann and Wand (2003) modeled nonlinear covariates using mixed-model representations of additive models (Wand, 2003; Ruppert et al., 2003), combining penalized splines to model nonlinear effects with spatial smoothing, a method called "geoadditive models". On the other hand, the method of Lindgren et al. (2011) handles the nonlinear effect of covariates using random walk or autoregressive processes, integrated into the R-INLA software. Vandendijck et al. (2017) extended the method of Kammann and Wand (2003) by proposing to estimate the spatial decay parameter rather than assuming it fixed. Kammann and Wand (2003) and Vandendijck et al. (2017) used likelihood-based estimation methods and implemented their approach in the context of Gaussian data.

This paper proposes a Bayesian methodology for geoadditive models where the spatial component is modeled in a similar way as in Kammann and Wand (2003). The Laplace approximation is used to approximate the posterior distribution of regression parameters, so as to significantly reduce the computational time to carry out inference as compared to traditional Markov chain Monte Carlo (MCMC) algorithms. P-splines (Eilers and Marx, 1996) are used to model the nonlinear effects of covariates. This smoother offers the advantage of a penalty matrix that can be easily constructed and naturally translated into a Bayesian framework (Lang and Brezger, 2004). The combination of Laplace approximations and P-splines (Laplacian-P-splines) in generalized

additive models developed by Gressani and Lambert (2021) offers a computationally efficient alternative to classic MCMC approaches and serves as a backbone to the methodology developed here. We extend the Laplacian-P-splines methodology to a geoadditive model. Additionally, a novel approach is proposed for handling count data in combination with linear and nonlinear dependencies on covariates, an aspect not well explored in the literature on geostatistical modeling. Typically, a Poisson distribution is assumed for count data. However, this assumption is inadequate for handling overdispersion, as the Poisson model implies equality between mean and variance. In contrast, the negative binomial distribution, although more complex, accounts for overdispersion and permits more sophisticated modeling of count data. This paper implements both Poisson and negative binomial distributions, providing a robust and flexible approach for handling spatial count data.

The article is organized as follows. Section 2 introduces the geoadditive model, explaining how the smooth covariates and spatial components are modeled. It also presents the Bayesian geoadditive model and discusses the use of Laplace approximations. Section 3 conducts a simulation study to assess the proposed methodology using various performance measures, including bias, relative bias, and credible and prediction interval coverage. Additionally, we compare our method to the approach of Lindgren et al. (2011). In Section 4, the proposed model is applied to the analysis of two real-world datasets: the Meuse river data and the coronavirus disease 2019 (COVID-19) vulnerability data in Belgium. Finally, Section 5 concludes the paper. Code to reproduce the results of this article is available on the following repository https://github.com/bryansumalinab/Geoadditive-Modeling.

2 Important concepts and methodology

2.1 Geoadditive model

Consider spatially indexed observations denoted by $y_i(\boldsymbol{w}_i)$, where $\boldsymbol{w}_i = (w_{1i}, w_{2i})^\top \in \mathbb{R}^2$ denotes the spatial coordinates for observations $i = 1, ..., n$. The observations $y_i(\boldsymbol{w}_i)$, conditional on the covariates, are typically assumed to have a distribution from an exponential family as in generalized linear models. The geoadditive model can be written as:

$$g(\mu_i) = \underbrace{\beta_0 + \sum_{r=1}^{p} \beta_r x_r(\boldsymbol{w}_i)}_{\text{Linear predictors}} + \underbrace{\sum_{j=1}^{q} f(h_j(\boldsymbol{w}_i))}_{\text{Smooth terms}} + \underbrace{s(\boldsymbol{w}_i)}_{\text{Spatial component}}, \tag{1}$$

where $g(\cdot)$ is the link function and $\mathbb{E}(y_i(\boldsymbol{w}_i)) = \mu_i$. Model (1) consists of three different components. The first component contains the linear predictors $(1, x_1(\boldsymbol{w}_i), \ldots, x_p(\boldsymbol{w}_i))$, with corresponding parameters $(\beta_0, \beta_1, \ldots, \beta_p)$. The second component captures nonlinear dependencies of $g(\mu_i)$ on the covariates $h_j(\boldsymbol{w}_i)$ through the smooth function $f(\cdot)$, for $j = 1, \ldots, q$. The third component, $s(\boldsymbol{w}_i)$, accounts for unobserved or unmeasured spatial structures that are not fully explained by the covariates. In practice, it is also often difficult to obtain all relevant spatial covariates, and thus bias may be present due to unmeasured factors. Note that the classical kriging model can be equivalently represented using basis functions derived from the spatial covariance function (see supplementary material S1). By constructing the basis functions from commonly used spatial covariance functions in kriging, our proposed model also accounts for spatial dependency. For ease of notation, we let $x_{ir} := x_r(\boldsymbol{w}_i)$ and $h_{ij} := h_j(\boldsymbol{w}_i)$ for $r = 1, \ldots, p$ and $j = 1, \ldots, q$. Each smooth covariate h_{ij} can be modeled as:

$$f(h_{ij}) = \sum_{k=1}^{K} \theta_{jk} b_{jk}(h_{ij}), \qquad j = 1, \ldots, q,$$

where $b_{jk}(\cdot)$ is a basis function and θ_{jk} is the coefficient for $k = 1, \ldots, K$. In our case, B-spline basis functions are used with a discrete difference penalty on successive B-spline coefficients proposed by Eilers and Marx (1996). The penalty controls the roughness of the fit and can be naturally extended to the Bayesian framework in formulating the joint prior distribution of the B-spline coefficients (Lang and Brezger, 2004). A Gaussian prior is assumed on $\boldsymbol{\theta}_j = (\theta_{j1}, \theta_{j2}, \ldots, \theta_{jK})^\top$, i.e. $(\boldsymbol{\theta}_j | \lambda_j) \sim \mathcal{N}_K(\mathbf{0}, (\lambda_j P)^{-1})$ where $\lambda_j > 0$ is the penalty parameter for the jth smooth model component and $P = D_m^\top D_m$ where D_m denotes the mth order difference matrix ($m = 2$ in this paper). To ensure that the penalty matrix is full rank, a diagonal matrix is added to P with small entries on the main diagonal (i.e., 10^{-12}).

The spatial term, $s(\boldsymbol{w}_i)$, can be modeled in several ways depending on the type of spatial data.

In kriging or classical geostatistics, the observations $y_i(\boldsymbol{w}_i)$ are assumed to be continuous in the spatial domain $\boldsymbol{w}_i$, and $s(\boldsymbol{w}_i)$ is assumed to be a Gaussian process with mean 0 and variance σ_w^2. An important assumption of kriging is that $s(\boldsymbol{w}_i)$, for $i = 1, \ldots, n$, are correlated, such that $\text{Cov}(s(\boldsymbol{w}_i), s(\boldsymbol{w}_j)) = R(\boldsymbol{w}_i - \boldsymbol{w}_j)$, which satisfies the stationarity assumption, meaning that the covariance function $R(\cdot)$ does not depend on spatial locations but only on the distance between locations (refer below for possible choices of $R(\cdot)$). The main goal of kriging is to predict observations at a given location. For Gaussian data, the best linear unbiased prediction for an arbitrary location is analytically available along with the corresponding prediction uncertainty (Zimmerman and Stein, 2010). For non-Gaussian data, one can rely, for example, on a Bayesian approach (Wikle, 2010a) or a maximum likelihood approach (Zimmerman, 2010).

Kriging predictions require to compute the inverse of the covariance matrix $R(\cdot)$ which is of dimension $n \times n$. With increasing sample size, the computational burden associated to these predictions becomes non-negligible. One way to address this problem is to write the spatial component in terms of the basis function model as follows:

$$s(\boldsymbol{w}_i) = \sum_{l=1}^{L} \alpha_l b_l(\boldsymbol{w}_i) + \varepsilon(\boldsymbol{w}_i), \tag{2}$$

where $b_l(\boldsymbol{w}_i)$ is a known basis function evaluated at location $\boldsymbol{w}_i$ and α_l are the coefficients for $l = 1, \ldots, L$ with $L < n$. The vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_L)^\top$ is assumed to have a multivariate Gaussian distribution, $\boldsymbol{\alpha} \sim \mathcal{N}_L(0, \Sigma_{\boldsymbol{\alpha}})$, and the error term $\varepsilon(\boldsymbol{w}_i)$ is usually assumed to have a Gaussian distribution with mean 0 and constant variance. The addition of the error term $\varepsilon(\boldsymbol{w}_i)$ accounts for the errors stemming from approximating the underlying spatial process with a finite set of basis functions or its low-dimensional representation (Wikle, 2010b; Cressie et al., 2022). Several choices can be made for the basis functions. One such choice is the method proposed by Kammann and Wand (2003) using the stationary covariance function in kriging, defined earlier, as the basis function. By replacing the coordinates with a set of points, called two-dimensional knots ($\boldsymbol{\kappa}_s$), this approach allows for a low-rank representation of the covariance function. Due to the evaluation thereof in a finite and limited set of knots, the computational cost is significantly reduced. The spline approach is used for a smoothing perspective regarding

the stationary covariance function. Therefore, this approach does not scale with sample size. The spatial component $s(\boldsymbol{w}_i)$ is modeled as follows:

$$s(\boldsymbol{w}_i) = \beta_{w_1} w_{1i} + \beta_{w_2} w_{2i} + \sum_{s=1}^{S} z_{is}(\rho) u_s, \tag{3}$$

where w_{1i} and w_{2i} are the spatial coordinates with corresponding coefficients β_{w_1} and β_{w_2}, $z_{is}(\rho) = R_\rho(\boldsymbol{w}_i - \boldsymbol{\kappa}_s)$ and $\boldsymbol{u} = (u_1, u_2, \ldots, u_S)^\top$ are assumed to be normally distributed such that $(\boldsymbol{u}|\lambda_{spat}, \rho) \sim \mathcal{N}_S(0, (\lambda_{spat}\Omega_\rho)^{-1})$, where $\lambda_{spat} > 0$ and $\Omega_\rho = R_\rho(\boldsymbol{\kappa}_s - \boldsymbol{\kappa}_{s'})$ is an $S \times S$ matrix for all $s, s' \in 1, \ldots, S$. Here, $R_\rho(\cdot)$ is the covariance function used in kriging and $\boldsymbol{\kappa}_s$ $(s = 1, \ldots, S)$ is a subset of the spatial coordinates. One way to efficiently choose these two-dimensional knots $(\boldsymbol{\kappa}_s)$ is through the use of a space-filling algorithm (Johnson et al., 1990; Nychka and Saltzman, 1998). In addition, ρ represents the range parameter in kriging. In this paper, we consider covariance models commonly used for stationary processes in kriging (Fahrmeir et al., 2013, pp. 453–456):

$$\text{Exponential} : R_\rho(\boldsymbol{d}) = \lambda_{spat}^{-1} \exp(-\rho||\boldsymbol{d}||),$$

$$\text{Matérn} : R_\rho(\boldsymbol{d}) = \lambda_{spat}^{-1} \exp(-\rho||\boldsymbol{d}||)(1 + \rho||\boldsymbol{d}||),$$

$$\text{Spherical} : R_\rho(\boldsymbol{d}) = \lambda_{spat}^{-1}(1 - 1.5\rho||\boldsymbol{d}|| + 0.5\rho^3||\boldsymbol{d}||^3)\, \mathbb{I}(||\boldsymbol{d}|| \leq \rho^{-1}),$$

$$\text{Circular} : R_\rho(\boldsymbol{d}) = \lambda_{spat}^{-1} \exp(-\rho^2||\boldsymbol{d}||^2),$$

where $||\cdot||$ refers to the Euclidean norm and $\mathbb{I}(\cdot)$ is the indicator function. In the Gaussian scenario, the error term in (2) is not identifiable because of the presence of another error component for the response conditional on the mean. Nevertheless, even without the inclusion of the additional error term in (2), model (3) is capable of generating accurate predictions for $y_i(\boldsymbol{w}_i)$. This is primarily due to the fact that any extra variance introduced by $\varepsilon(\boldsymbol{w}_i)$ tends to be absorbed by the measurement error variance inherent in the Gaussian model. However, the same cannot be said for count data when the Poisson distribution is assumed. To address this, we propose the use of a negative binomial model. In this way, the error term $\varepsilon(\boldsymbol{w}_i)$ is regarded as excess variability, which can be effectively managed by the overdispersion parameter in the negative binomial model. This overdispersion parameter quantifies the additional variance not captured by the Poisson model and can be interpreted as arising from the error variance in

the outcome distribution. This link is evident when considering the general formulation of the exponential dispersion family (EDF) (see Section 2.2 below), where the variance function includes a dispersion parameter that corresponds to the overdispersion in the negative binomial model. This allows for a more accurate estimation of the prediction uncertainty. However, if this excess variability is small, then the Poisson assumption may be sufficient. Furthermore, it is also important to note that equation (3) represents a model for the average spatial process.

2.2 Bayesian model formulation

For ease of notation let $y_i := y_i(\boldsymbol{w}_i)$. To generalize the derivations, we can write the probability distributions of y_i as an exponential dispersion family, $y_i \sim \text{EDF}(\gamma_i, \phi)$, with probability distribution given by $p(y_i; \gamma, \phi) = \exp\{[y_i\gamma_i - b(\gamma_i)]/a(\phi) + c(y_i, \phi)\}$ where ϕ is a dispersion parameter and γ_i is a natural parameter with mean $\mathbb{E}(y_i) = b'(\gamma_i) = \mu_i$ and variance $\mathbb{V}(y_i) = a(\phi)b''(\gamma_i)$.

The log-link function is used for the Poisson and negative binomial model, such that $g(\mu_i) = \log(\mu_i)$. Note that an offset term N_i (e.g. population size for each location) may be added so that $\mu_i = \exp(g(\mu_i)) \times N_i$. In matrix form, (1) can be written as:

$$\log(\boldsymbol{\mu}) = X\boldsymbol{\beta} + \sum_{j=1}^{q} B_j\boldsymbol{\theta}_j + Z(\rho)\boldsymbol{u}, \tag{4}$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$, X is a matrix of dimension $n \times (p + 3)$ with ith row vector $\boldsymbol{x}_i = (1, x_{i1}, x_{i2}, \ldots, x_{ip}, w_{1i}, w_{2i})^\top$ and coefficient vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \cdots, \beta_p, \beta_{w_1}, \beta_{w_2})^\top$, B_j is an $n \times K$ matrix with ith row $\boldsymbol{b}_{ij} = (b_{j1}(h_{ij}), b_{j2}(h_{ij}), \ldots, b_{jK}(h_{ij}))^\top$ and coefficients $\boldsymbol{\theta}_j = (\theta_{j1}, \theta_{j2}, \ldots, \theta_{jK})^\top$, $Z(\rho)$ is an $n \times S$ matrix with ith row $\boldsymbol{z}_i(\rho) = (z_{i1}(\rho), z_{i2}(\rho), \ldots, z_{iS}(\rho))^\top$ and coefficients $\boldsymbol{u} = (u_1, u_2, \ldots, u_S)^\top$ for $i = 1, \ldots, n$ and $j = 1, \ldots, q$.

A Gaussian prior is assumed for $\boldsymbol{\beta}$, i.e., $\boldsymbol{\beta} \sim \mathcal{N}_{\dim(\boldsymbol{\beta})}(0, V_{\boldsymbol{\beta}}^{-1})$ where $V_{\boldsymbol{\beta}} = \zeta I$ with small ζ (e.g. $\zeta = 10^{-5}$), implying a large prior variance and thus a noninformative Gaussian prior. Denote the global design matrix by $C_\rho = [X : B_1 : B_2 : \cdots : B_q : Z(\rho)]$ and the corresponding parameter vector $\boldsymbol{\xi} = (\boldsymbol{\beta}^\top, \boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_q^\top, \boldsymbol{u}^\top)^\top$ such that equation (4) becomes $\log(\boldsymbol{\mu}) = C_\rho\boldsymbol{\xi}$. Moreover, let $\lambda_{spat} := \lambda_{q+1}$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_q, \lambda_{q+1})^\top$. Denote the precision of $\boldsymbol{\xi}$ by $Q_{\boldsymbol{\xi}}^{\boldsymbol{\lambda}} =$

blkdiag(V_β, $\lambda_1 P$, $\ldots$, $\lambda_q P$, $\lambda_{q+1}\Omega_\rho$), where blkdiag$(\cdot)$ is a block diagonal matrix. The full Bayesian model is given by:

$$
\begin{aligned}
&(y_i|\boldsymbol{\xi}) \sim \text{EDF}(\gamma_i, \phi), \quad i = 1, \ldots, n, \\
&(\boldsymbol{\xi}|\boldsymbol{\lambda}, \rho) \sim \mathcal{N}_{dim(\boldsymbol{\xi})}(\mathbf{0}, (Q_{\boldsymbol{\xi}}^{\boldsymbol{\lambda}})^{-1}), \\
&(\lambda_j|\delta_j) \sim \mathcal{G}\left(\frac{\nu}{2}, \frac{\nu\delta_j}{2}\right), \quad j = 1, \ldots, q+1, \\
&\delta_j \sim \mathcal{G}(a_\delta, b_\delta), \quad j = 1, \ldots, q+1, \\
&p(\rho) \propto \rho^{-1}, \\
&p(\phi) \propto \phi^{-1},
\end{aligned}
$$

where $\mathcal{G}(a, b)$ denotes a Gamma distribution with shape parameter a and rate parameter b, which corresponds to a mean of a/b and variance a/b^2. This robust prior specification on the penalty parameters follows from Jullion and Lambert (2007) where $a_\delta = b_\delta$ are chosen to be small enough, say 10^{-5}, with fixed ν (e.g. $\nu = 3$ in this paper). Furthermore, to address identifiability issues between the smooth terms and the spatial component, we imposed centering by constraining the basis matrix for smooth and spatial terms to have columns with zero mean.

2.3 Laplace approximation

The Laplace approximation is used to approximate the conditional posterior $p(\boldsymbol{\xi}|\boldsymbol{\lambda}, \rho, \phi, \mathcal{D})$ by a Gaussian distribution. This posterior approximation is particularly advantageous for its computational efficiency, significantly reducing computational time, as it eliminates the need for sampling compared to MCMC methods. In the case of a Gaussian response, the derived conditional posterior is exactly Gaussian, and the derivations are provided in supplementary material S3. Denote the (Poisson or negative binomial) likelihood function by $\mathcal{L}(\boldsymbol{\xi}, \rho, \phi; \mathcal{D})$ where $\mathcal{D}$ is the observed data. Using Bayes' rule, the conditional posterior of $\boldsymbol{\xi}$ can be written as $p(\boldsymbol{\xi}|\boldsymbol{\lambda}, \rho, \phi, \mathcal{D}) \propto \mathcal{L}(\boldsymbol{\xi}, \rho, \phi; \mathcal{D}) p(\boldsymbol{\xi}|\boldsymbol{\lambda}, \rho)$. The gradient and Hessian of the log-conditional posterior, $\log p(\boldsymbol{\xi}|\boldsymbol{\lambda}, \rho, \phi, \mathcal{D})$, with respect to $\boldsymbol{\xi}$ are analytically derived and used in a Newton-Raphson algorithm to approximate the mode of the conditional posterior of $\boldsymbol{\xi}$. The availability of the derived analytic gradient and Hessian further enhances the computational speed. After

convergence, the Laplace approximation of $p(\boldsymbol{\xi}|\boldsymbol{\lambda}, \rho, \phi, \mathcal{D})$ is a multivariate Gaussian density denoted by $\widetilde{p}_G(\boldsymbol{\xi}|\boldsymbol{\lambda}, \rho, \phi, \mathcal{D}) = \mathcal{N}_{\dim(\boldsymbol{\xi})}(\widehat{\boldsymbol{\xi}}_{\boldsymbol{\lambda}}, \widehat{\Sigma}_{\boldsymbol{\lambda}})$ where $\widehat{\boldsymbol{\xi}}_{\boldsymbol{\lambda}}$ is the mode and $\widehat{\Sigma}_{\boldsymbol{\lambda}}$ is the inverse of the negative Hessian matrix evaluated at the posterior mode. The derivations of the approximate posterior distributions of the hyperparameters are presented in supplementary material S2.

3 Simulation study

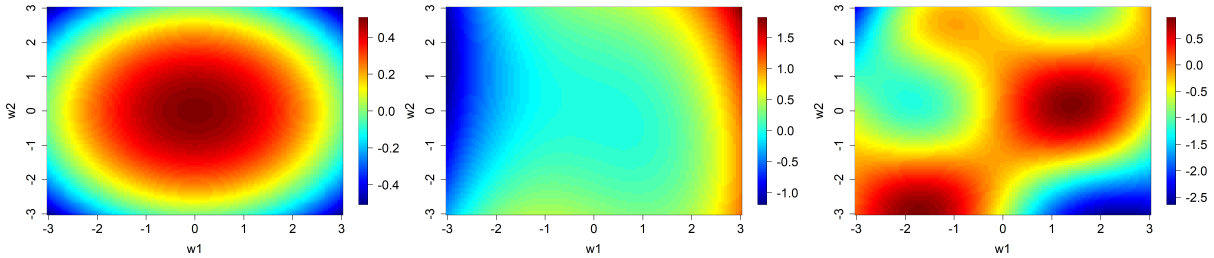
A simulation study is conducted to evaluate the performance of our proposed methodology. For the count data, samples are generated from a Poisson distribution with rate parameter $\mu \cdot \exp(\varepsilon)$. The inclusion of $\exp(\varepsilon)$ preserves the stochasticity of the true spatial process in agreement with equation (2). This is equivalent to simulating a spatial component from a two-dimensional smooth function plus an error term $(s(w_1, w_2) + \varepsilon)$. For the Gaussian data, ε represents the error term of the Gaussian distribution, such that observations y are generated from a normal distribution with mean μ and variance σ^2. In the simulation, the following mean structure is assumed:

$$\mu = \beta_0 + \beta_1 x_1 + f(x_2) + s(w_1, w_2),$$

where $\beta_0 = 3$, $\beta_1 = -0.5$, and $f(x_2) = \cos(2\pi x_2)$. The covariates x_1 and x_2 are randomly simulated from a uniform distribution over the unit interval, and ε is drawn from a zero mean Gaussian distribution with standard deviation $\sigma = 0.25$ for the count data and $\sigma = \sqrt{0.10}$ for the Gaussian data. The spatial component $s(w_1, w_2)$ is a two-dimensional smooth function, for which the following three different forms are considered:

$$\begin{aligned}
s_1(w_1, w_2) &= 0.5 - \frac{w_1^2 + w_2^2}{18}, \\
s_2(w_1, w_2) &= \frac{w_1^3 + w_1 w_2 + w_2^2}{25}, \\
s_3(w_1, w_2) &= -\frac{(w_1 - w_2)^2}{15} + \sin(w_1)\cos(w_2).
\end{aligned}$$

Here, the spatial coordinates w_1 and w_2 are simulated from a uniform distribution on the interval $(-3, 3)$. The plot for the two-dimensional smooth functions considered in the simulation is shown in Figure 1.



(a) Function s_1 (b) Function s_2 (c) Function s_3

Figure 1: Two-dimensional smooth functions.

The performance measures used in our study are bias, relative bias (%Bias), and credible interval (CI) coverage. These quantities are computed for μ, the smooth term $f(x_2)$, and the spatial term $s(w_1, w_2)$. Additionally, the prediction interval (PI) coverage for the samples y is calculated. The details of obtaining the prediction and prediction interval are provided in supplementary material S4. For smooth and spatial terms, a grid of values for x_2 and a two-dimensional grid for (w_1, w_2) are created. Using these grids, $f(x_2)$ and $s(w_1, w_2)$ are computed, and the corresponding relative bias and PI coverage are obtained. To avoid identifiability issues, column mean centering is imposed for the basis functions associated with the smooth and spatial components and the performance measures are computed on the mean centered $f(x_2)$ and $s(w_1, w_2)$. Due to this centering, the bias is only computed for μ. If the quantity of interest is denoted by ω with corresponding estimate $\widehat{\omega}$, then the bias and relative bias are computed as follows:

$$\text{Bias} = \frac{1}{C \times M} \sum_{c=1}^{C} \sum_{m=1}^{M} (\omega_{cm} - \widehat{\omega}_{cm}),$$

$$\%\text{Bias} = \frac{1}{C \times M} \sum_{c=1}^{C} \sum_{m=1}^{M} \left| \frac{\omega_{cm} - \widehat{\omega}_{cm}}{\omega_{cm}} \right| \times 100\%,$$

where $M = 100$ is the length of the grid created for the covariates, and $C = 500$ is the number of simulations or generated realizations, each based on a generated dataset of size 1000.. Moreover, CI or PI coverage is determined by calculating the percentage of ω values that fall within the computed credible/prediction interval.

Table 1 presents the results for function s_3 across various covariance structures (circular, exponential, Matérn, and spherical) and different model distributions (negative binomial, Poisson, and Gaussian). The results for functions s_1 and s_2 are presented in supplementary material S6. The percentage bias for the smooth term is around 4% to 7% for all scenarios, demonstrating low bias and suggesting robust performance in estimating smooth effects using our proposed method. For all functions (s_1, s_2 and s_3), the negative binomial model has coverage rates associated to the smooth term that are generally closer to the 95% nominal level as compared to the Poisson model which shows undercoverage, with rates ranging from 73% to 80%. The Gaussian models, on the other hand, also demonstrate good coverage, ranging from 91% to 94%, which is slightly lower than the nominal level. Note that the relative bias for Gaussian data is very low, around 4% to 5%.

Regarding the spatial term, the Gaussian model generally exhibits the lowest relative bias, followed by the negative binomial model, while the Poisson model has the highest bias. Notably, all covariance functions show similar relative biases across each model for the spatial term, except for the Matérn covariance function. For the Poisson model, the Matérn covariance consistently shows lower bias. In the negative binomial and Gaussian models, the Matérn covariance has a relatively low bias for function s_1 but a higher bias for functions s_2 and s_3. The coverage for the spatial term is highest for the negative binomial model, ranging from 97% to 99%, except for the Matérn covariance, which has lower coverage for functions s_2 and s_3, with rates of 83% and 69%, respectively. This pattern is similar to the Gaussian data, with coverage mostly close to the 95% nominal level, although there is undercoverage for the Matérn covariance for functions s_2 and s_3. Thus, for the spatial term, the Matérn covariance perform better for function s_1. In contrast, the other covariance functions demonstrate robust performance across all three shapes of the underlying spatial structure. However, it is important to note that the Matérn covariance considered here is just a special case where we fixed the smoothness parameter $\nu = 1.5$. Whereas, when $\nu = 0.5$, it reduces to the exponential covariance, which works well across various scenarios. In practice, the choice among these models could be guided by model selection criteria, such as the Bayesian information criterion (BIC) (see supplementary material S5).

For the mean response, biases are generally close to zero for all scenarios, with the percentage bias being lowest for the Gaussian data, ranging between 1% to 3%, except for the Matérn covariance with function s_3, where the relative bias is around 6%. The relative bias for μ is mostly lower for the negative binomial model compared to the Poisson model, but their values are very similar, ranging around 4% to 9%, except for function s_3 with the Matérn covariance in the negative binomial model, which has a relative bias of 13%. We repeated this scenario with different seeds to investigate the observed difference between the Poisson and negative binomial models for function s_3. However, the results seem to be consistent for different samples. We believe this might be due to function s_3, which is the most complex of the functions considered. In addition, when looking at the simulation results for INLA-SPDE in Table S7.1 (supplementary material S7), the function s_3 has the highest percentage bias for the mean response μ. Finally, for the response y, the Gaussian model has prediction interval coverage very close to the nominal level. For count data, the Poisson model shows undercoverage, while the negative binomial model has high prediction interval coverage. Overall, the negative binomial and Gaussian models demonstrate robust performance, characterized by low bias and high coverage rates across all quantities of interest. In contrast, although the Poisson model exhibits low bias, it suffers from lower coverage rates due to the presence of an additional error term, treated as an overdispersion parameter, that is unaccounted for when working with a Poisson model.

The simulation results using INLA-SPDE (Lindgren et al., 2011) are presented in Table S7.1 in supplementary material S7. For discussion purposes, we call our approach LPS (Laplacian-P-splines). For the Gaussian case, it can be observed that both LPS and INLA-SPDE demonstrate comparable relative bias for the smooth term, with minimal differences across functions. INLA-SPDE exhibits a tendency for over-coverage for the smooth term, while LPS shows under-coverage. For the parameter μ, the bias is similar between the two methods, but LPS provides a lower percentage bias. For the prediction interval coverage of y, LPS tends to have higher coverage closer to the nominal level, although the differences are minor. For the negative binomial case, the point estimates show similar patterns to those observed in

the Gaussian case, with comparable results between the two methods. LPS exhibits slightly lower bias percentages for μ compared to INLA-SPDE. In terms of coverage, INLA-SPDE demonstrates a tendency for over-coverage in the smooth term, while LPS shows slight under-coverage. When it comes to prediction interval coverage for y, INLA-SPDE appears to perform better, maintaining coverage closer to the nominal level, while LPS tends to exhibit over-coverage. The spatial component is not centered in INLA-SPDE, therefore there is no common ground for comparison. Furthermore, the computation time is compared using the `microbenchmark` package in R with 10 functions evaluations. The analysis is implemented on a device with an Intel(R) Core(TM) i5-1135G7 CPU running at a base frequency of 2.40GHz, and having 4 cores with 16GB of RAM. Our proposed method demonstrates significantly faster performance compared to INLA-SPDE, with average computation times of approximately 14 seconds for the Gaussian case and 15 seconds for the negative binomial case. In contrast, INLA-SPDE requires an average of around 85 seconds for the Gaussian case and 109 seconds for the negative binomial case (see Table S7.2 in supplementary material S7). These results highlight the computational benefit of our methodology. It is worth noting that INLA-SPDE involves code written in the C-family programming language, which is generally optimized for computational efficiency. This highlights the computational advantage of our methodology even more as our code is entirely based on the R programming language.

Additional simulation results without covariates are presented in supplementary material S8, comparing the proposed model with the classical kriging approach for the Gaussian data using the `geoR` package, yielding comparable results, although the low-rank approach has a slightly larger percentage bias in terms of the mean. The average real elapsed time for the proposed Bayesian approach is around 1 second, while the classical kriging approach takes around 110 seconds on average (see Table S8.2 in supplementary material S8).

Table 1: Simulation results for function s_3. Model - distributional assumption for the response. Covariance - covariance functions used in modeling the spatial component. CP(%) - indicates 95% credible/prediction interval coverage.

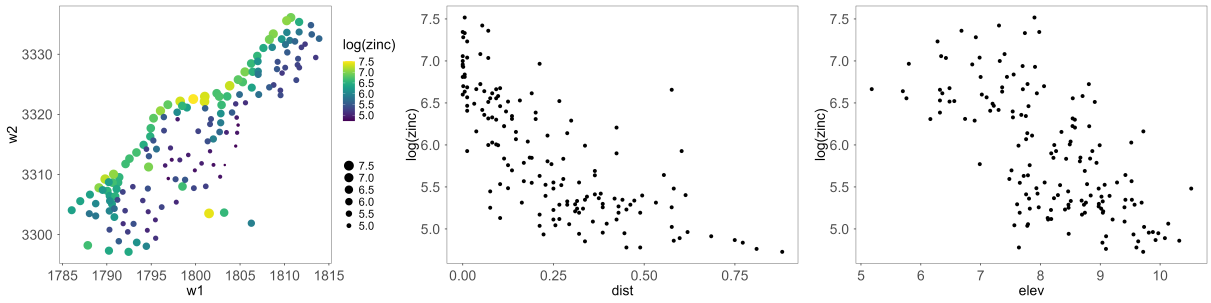
Model	Covariance	Smooth term $f(x_2)$		Spatial term $s(w_1, w_2)$		μ		y
		Bias(%)	CP(%)	Bias(%)	CP(%)	Bias	Bias (%)	CP(%)
Negative binomial	Circular	6.06	93.95	13.44	97.16	-0.34	8.05	98.15
	Exponential	5.83	94.62	13.60	97.17	-0.33	8.03	98.25
	Matérn	7.31	91.99	19.86	69.01	-0.20	13.09	99.15
	Spherical	5.95	94.93	13.28	97.60	-0.34	7.97	98.26
Poisson	Circular	6.51	79.43	15.82	83.74	-0.53	8.92	87.19
	Exponential	6.46	79.82	15.84	82.86	-0.51	9.03	87.29
	Matérn	6.22	78.44	12.75	70.28	-0.64	8.26	86.65
	Spherical	6.28	79.42	15.85	83.77	-0.45	8.98	87.36
Gaussian	Circular	5.01	92.66	11.10	95.28	-0.01	2.85	94.70
	Exponential	4.97	92.41	11.24	94.76	-0.02	2.87	94.64
	Matérn	5.32	94.14	22.74	44.49	-0.02	5.91	96.31
	Spherical	5.00	93.25	11.23	95.17	-0.02	2.87	94.55

4 Data application

The proposed methodology is applied to the analysis of two datasets: (1) Meuse river data using the Gaussian model and (2) vulnerability to coronavirus 2019 disease (COVID-19) in Belgium using the negative binomial model.

4.1 Meuse river data

The Meuse dataset contains measurements of heavy metal concentrations in topsoil collected from the flood plain of the Meuse River near the village of Stein, Netherlands. It also includes the geographic coordinates of each sampling location and is commonly used to demonstrate kriging and other geostatistical techniques. This analysis focuses on zinc concentrations in the topsoil, using two covariates: (1) distance to the Meuse river (dist), and (2) relative elevation above the local riverbed (elev). The dataset is available in the R package `sp`.



(a) (b) (c)

Figure 2: (a) Observed log-zinc values over the sampling locations (w_1, w_2); (b) Scatterplot of observed log-zinc with the covariate distance; (c) Scatterplot of observed log-zinc with the covariate elevation.

Figure 2 displays the observed logarithm of zinc (log-zinc) values at sampling locations (w_1, w_2) and scatterplots illustrating the relationship between log-zinc and the variables, distance, and elevation. Based on the graphical exploration, a nonlinear relation between distance and the log-transformed zinc concentrations is observed, while for the relation between log-zinc and elevation, linearity seems a plausible assumption. Classical geostatistical methods typically accommodate only the linear effects of covariates. Therefore, transformation is commonly applied (e.g. a square root transformation of distance) in order to fit a linear geostatistical model. One benefit of our method is the direct incorporation of nonlinear covariates without the need for any transformation. The distance and elevation are included as smooth covariates, using a Gaussian model for the log-zinc, with various covariance structures. Table 2 shows that the BIC values are similar for different model covariances, with the circular covariance having the lowest BIC. Consequently, the model using circular covariance is examined further. Both tests for the significance of the two smooth covariates yield a p-value < 0.0001, indicating a statistically significant relationship between the covariates and log-zinc. The details for testing the nonlinear effect of covariates are presented in supplementary material S5. Since Figure 2c indicates a linear effect of elevation, a model is fitted with elevation as a linear covariate, resulting in a BIC of -126.89, slightly higher than the BIC (-128.94) obtained when elevation is included as a smooth covariate. Therefore, the final model includes both distance and elevation as smooth covariates given by:

$$\log(\text{zinc}_i) = \beta_0 + f_1(\text{dist}_i) + f_2(\text{elev}_i) + s(w_{1i}, w_{2i}) + \epsilon_i, \tag{5}$$

for $i = 1, \ldots, 155$. Figure 3 presents the estimated effects of the covariates. Figure 4 shows the estimated spatial surface and the comparison between fitted and observed log-zinc values based on model (5).

Table 2: Results for the Gaussian model fitted on Meuse river data.

Covariance	**BIC**
Circular	-128.94
Exponential	-128.12
Matérn	-126.61
Spherical	-127.14

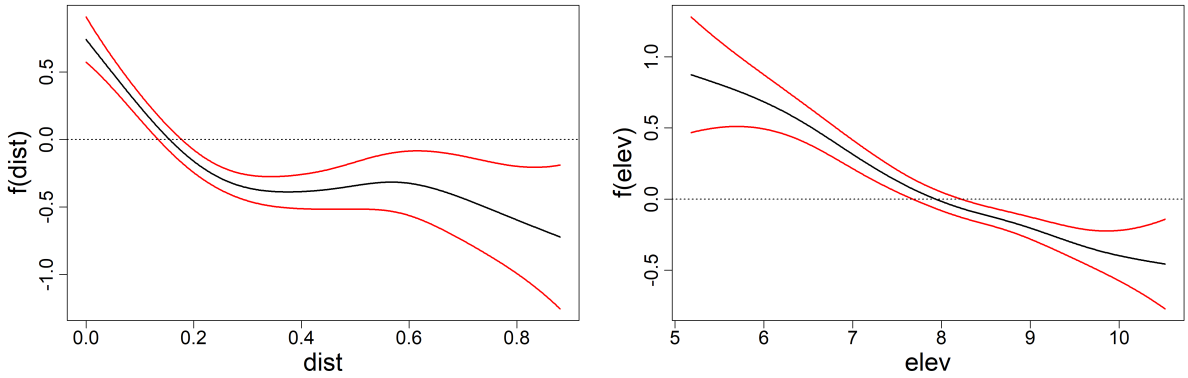
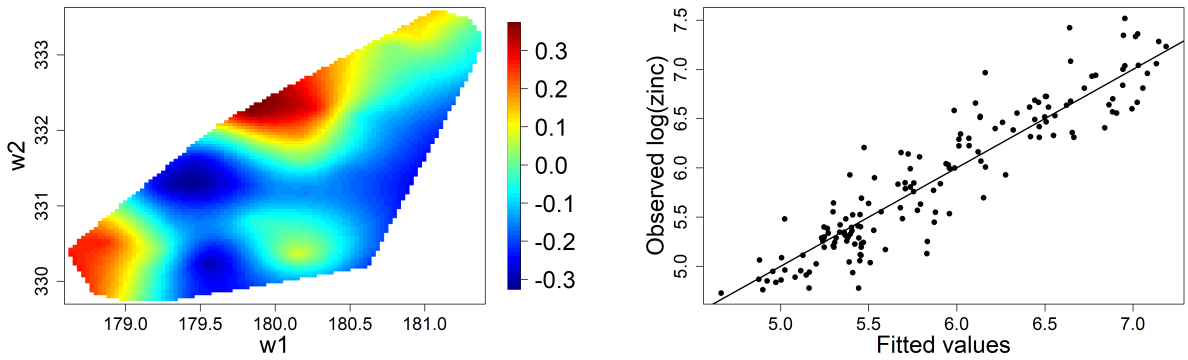


Figure 3: Estimated effects of smooth covariates on log-zinc. 95% credible interval is shown by the red lines.



(a) (b)

Figure 4: Results for the Meuse data using model (5). (a) Estimated continuous surface for the spatial term $s(w_1, w_2)$; (b) Estimated mean vs observed log-zinc values.

4.2 COVID-19 vulnerability data

The proposed negative binomial model is applied to the analysis of the COVID-19 incidence data with respect to confirmed cases from Flanders and Brussels regions in Belgium from September 1, 2020, to December 31, 2020. The study area is divided into 9627 statistical sectors, each with a population ranging from a minimum of 7 to a maximum of 6082, with an average population of approximately 740 inhabitants. The population per 100 inhabitants of each statistical sector is used as an offset term in the model. The centroid of each statistical sector serves as the coordinate for the spatial analysis. The dataset includes the total number of positive cases for the considered period and various risk factors for each statistical sector. Variables identified in the literature as risk factors for vulnerability to COVID-19 are considered. Correlations between these variables are then examined, with only one variable retained from each pair that had an absolute Pearson correlation greater than 0.5. The final factors include median net income (*med_inc*) (Rozenfeld et al., 2020; Wachtler et al., 2020), the proportion of retired people (*pensinr*) (Rozenfeld et al., 2020; Pijls et al., 2021), the proportion of non-Belgian residents (*nonBel*) (Hayward et al., 2021), the proportion of single parents (Sung, 2021) (*snglprn*), the yearly average black carbon level (*bc*) (Rozenfeld et al., 2020; Wu et al., 2020), and the proportion of females (*female*) (Wu and Qian, 2022). All these variables are standardized.

Initially, all factors are considered as smooth covariates. Various covariance functions are fitted with the spherical covariance having the lowest BIC value, as shown in Table 3. Therefore, further data analysis is conducted using spherical covariance, showing that all factors have statistically significant effects (p-value < 0.0001). The plot illustrating the estimated smooth effects is shown in Figure S9.2 in supplementary material S9. To investigate the linear effects, each smooth effect for a given covariate is subsequently replace by a linear effect, and the BIC value is compared to that of the full model (BIC = 1221406), where all factors are smooth covariates. As shown in Table 4, the factors *med_inc* and *bc* resulted in lower BIC when included as linear covariates rather than smooth covariates.

Table 3: BIC results for the negative binomial model fitted on COVID-19 data using different covariance functions.

Covariance	**BIC**
Circular	-1221386
Exponential	-1221397
Matérn	-1221187
Spherical	-1221406

Table 4: BIC comparison for the COVID-19 data using spherical covariance when the variables are added as linear covariates.

Variables	**BIC with linear covariate**	**Difference from BIC with smooth covariate**
med_inc	-1221420	-14
pensinr	-1221388	18
nonBel	-1221403	3
snglprn	-1221351	55
bc	-1221422	-16
female	-1221388	18

Therefore, the final model is given by:

$$\begin{aligned}\log(\mu_i) =&\beta_0 + \beta_1 med\_inc_i + \beta_2 bc_i + f(pensinr_i) + f(nonBel_i) + \\ &f(snglprn_i) + f(female_i) + s(w_{1i}, w_{2i}) + \log(N_i),\end{aligned} \tag{6}$$

where N_i is the offset term per 100 inhabitants for $i = 1, \ldots, 9627$. The estimated coefficients for the linear covariates in Table 6 indicate that areas with higher median net income are negatively correlated with the number of COVID-19 cases, while areas with high levels of black carbon are positively correlated with the number of cases. Table 5 shows the results for the nonlinear covariates in model (6) and the plot of the estimated smooth effects is shown in Figure 5. The latter figure shows that areas with higher proportions of pensioners and single parents exhibit higher vulnerability to COVID-19 compared to the average. However, areas with average or lower proportions of these groups correspond to average vulnerability. For the covariate *pensinr*, the smooth trend for higher values shows a wider 95% credible interval, indicating higher uncertainty in the trend direction, which could potentially go up or down. This is due to the limited number of observations (only five) for *pensinr* values greater than 5.

Other covariates also show higher uncertainty at the right tail because of few observations, with the empirical distributions exhibiting right skewness (see Figure S9.1 in supplementary material S9). Areas with higher proportions of non-Belgians and lower proportions of females appear to be less vulnerable to COVID-19. Finally, the plot for the estimated spatial surface and the comparison between fitted and observed cases is shown in Figure 6.

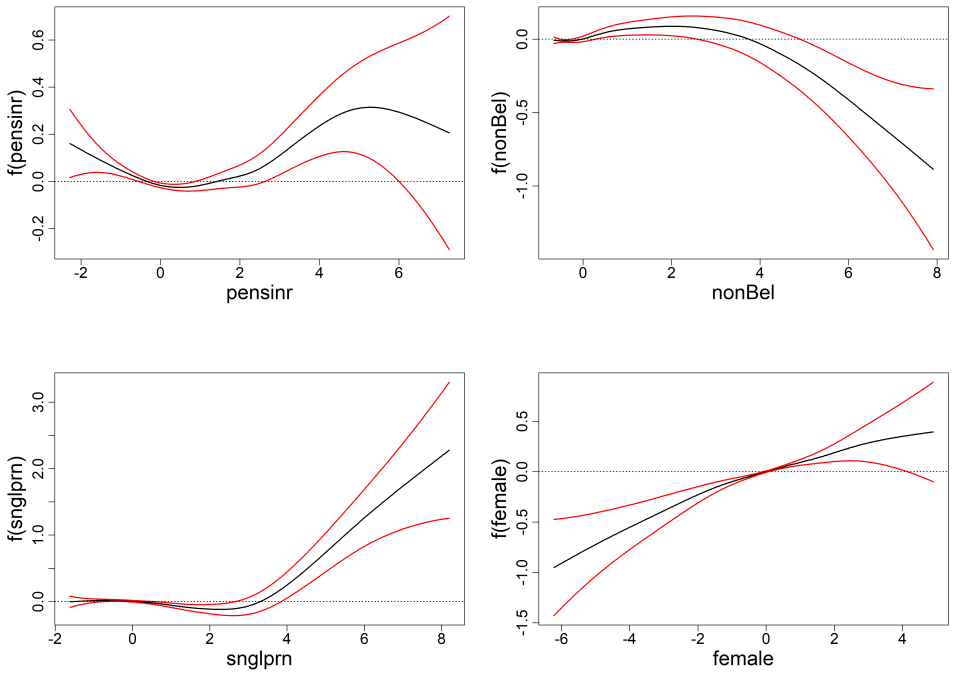


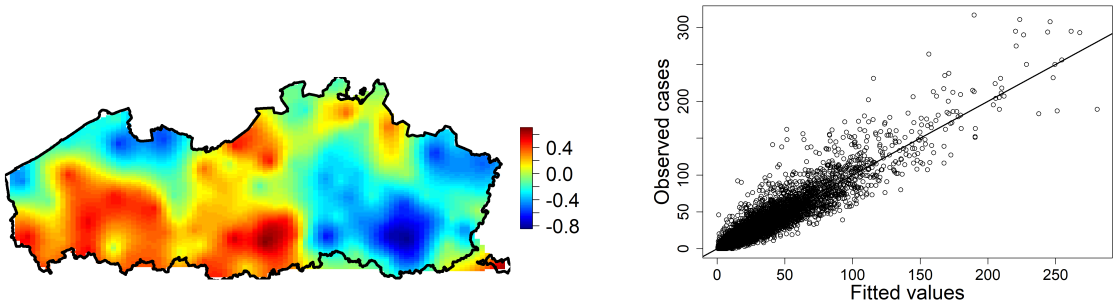
Figure 5: Estimated effects of smooth covariates for COVID-19 data using spherical covariance. 95% credible interval is shown by the red lines.

Table 5: Results for the smooth covariates in model 6.
ED - effective degrees of freedom; T_r - test statistic.

Variables	**ED**	$\boldsymbol{T_r}$	**p-value**
pensinr	5.10	37.95	< 0.0001
nonBel	4.92	26.27	< 0.0001
snglprn	5.87	45.68	< 0.0001
female	3.43	63.86	< 0.0001

Table 6: Results for the linear covariates in model 6. Estimate - estimated coefficient; SE - standard error; CI lower - 95% lower credible interval; CI upper - 95% upper credible interval.

Variables	**Estimate**	**SE**	**CI lower**	**CI upper**
(Intercept)	1.180	0.005	1.170	1.190
med_inc	-0.073	0.007	-0.088	-0.059
bc	0.047	0.010	0.028	0.066



(a) (b)

Figure 6: Results for the COVID-19 data using model (6). (a) Estimated continuous surface for the spatial term $s(w_1, w_2)$ overlaying the map of the study region; (b) Estimated mean vs observed number of COVID-19 positive cases.

5 Discussion

This paper presents a novel Bayesian method for geostatistical modeling that combines Laplace approximations, P-splines, and low-rank representations for spatial processes. The proposed model offers several advantages, including the ability to perform spatial interpolation and investigate the effects of both linear and nonlinear covariates. Its computational efficiency is substantially improved by using Laplace approximations and a low-rank representation of the spatial process, leading to faster computation times. Additionally, the model extends beyond Gaussian responses to accommodate Poisson and negative binomial models, thereby providing robust options for handling count data. Thus, the proposed approach not only enhances the flexibility and speed of geostatistical analyses but also broadens their applicability across different types of data.

While the model in equation (1) may initially appear to be a simple extension of the generalized additive model proposed by Gressani and Lambert (2021), there are several important aspects that add to its complexity. First, we extend their framework by adopting a negative binomial distribution for the count response, which introduces additional challenges beyond the Poisson setting. Second, the incorporation of the spatial component adds another layer of complexity. Third, unlike the standard approach where basis functions are fixed, the basis function for the spatial effect in our model depends on a spatial range parameter that must be estimated. These factors together require a re-derivation, especially the gradient and Hessian matrices, and make the implementation less straightforward.

Simulation studies demonstrate that the proposed model exhibits low relative bias and has credible intervals close to the nominal coverage with respect to the underlying target smooth function and spatial process. These results are consistent across all covariance functions, except for the Matérn covariance, which did not perform well in terms of credible interval coverage for the spatial component in non-symmetric two-dimensional functions. Additionally, the proposed model demonstrates good predictive interval coverage across all covariance functions, highlighting its robust predictive ability. In the presence of overdispersion, the Poisson model, as expected, results in undercoverage, but the availability of the negative binomial model within our proposed framework effectively addresses this issue.

The proposed model has been applied to the Meuse river data and COVID-19 vulnerability data in Belgium, demonstrating its practical applicability. The analysis of COVID-19 data reveals that areas with lower median incomes are more vulnerable to the virus, highlighting the economic disparities exacerbated by the pandemic (Rozenfeld et al., 2020; Wachtler et al., 2020). Additionally, areas with a higher proportion of pensioners who are at increased risk for severe COVID-19 outcomes show higher vulnerability (Rozenfeld et al., 2020; Pijls et al., 2021). Environmental factors, such as higher levels of black carbon, further contribute to increased susceptibility, underscoring how air pollution contributes to health disparities (Rozenfeld et al., 2020; Wu et al., 2020). Social demographics also play a significant role, for instance, areas with higher proportions of females (Wu and Qian, 2022) and single parents (Sung, 2021) are

more vulnerable to the virus. Interestingly, while studies indicate that migrants are generally at high risk for COVID-19 (Hayward et al., 2021), our findings suggest that areas with a high proportion of non-Belgians are less vulnerable to the disease.

Our method shares similarity with the approach of Lindgren et al. (2011) in modeling the spatial field based on a basis function approach. While the method of Lindgren et al. (2011) uses triangulation to define basis functions with a sparse precision matrix for the random field, our approach derives both the basis functions and the precision matrix directly from the covariance function. Although the precision matrix in our method is not sparse, it still achieves significant computational gains. Our method is also conceptually simpler and can accommodate other covariance functions, such as circular and spherical covariance functions, whereas the approach of Lindgren et al. (2011) is restricted to the Matérn covariance. On the other hand, the method of Lindgren et al. (2011), through the use of the R-INLA software, can accommodate a wider range of advanced models, such as those involving spatio-temporal data, point processes, and multivariate responses (Krainski et al., 2018). While the INLA-SPDE approach is well-established and widely used, we have demonstrated that our method similarly produces accurate estimates and predictions with notable computational advantages. Therefore, our approach serves as a complementary or an alternative framework to existing methods, such as the one of Lindgren et al. (2011) and other low-rank methodologies in geostatistical modeling for which very few include nonlinear effects of covariates.

A common limitation in low-rank spatial modeling approaches is selecting the number of basis functions. However, this issue is not unique to our method or to other low-rank approaches, such as those of Cressie and Johannesson (2008). For example, the SPDE approach of Lindgren et al. (2011) requires mesh construction, which determines the number of basis functions. Similarly, Vecchia approximations involve choosing the size of the conditioning set (i.e., the number of neighbors), a choice that also affects both computational efficiency and model accuracy. Nevertheless, we have shown in the simulation study that our proposed approach generally yields accurate results. Moreover, Cressie et al. (2022) reviewed basis-function approaches in spatial statistics and highlighted their significance in modeling nonstationary processes, with

illustrations in Gaussian, non-Gaussian, multivariate, and spatio-temporal settings.

Finally, an interesting extension of this work is to assume a binomial distribution for the count response, considering the population size. This could enhance model performance in scenarios such as disease prevalence modeling, where count data naturally follow a binomial distribution due to the presence of binary outcomes (e.g. disease vs. no disease) within a given population size.

Competing Interest Statement

The authors have declared no competing interest.

References

Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 70(4):825–848.

Cressie, N. (1993). *Statistics for Spatial Data*. John Wiley & Sons.

Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 70(1):209–226.

Cressie, N., Sainsbury-Dale, M., and Zammit-Mangion, A. (2022). Basis-function models in spatial statistics. *Annual Review of Statistics and Its Application*, 9:373–400.

Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E. (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, 111(514):800–812.

Diggle, P. J. and Giorgi, E. (2019). *Model-Based Geostatistics for Global Public Health: Methods and Applications*. Chapman and Hall/CRC.

Eidsvik, J., Finley, A. O., Banerjee, S., and Rue, H. (2012). Approximate Bayesian inference for large spatial datasets using predictive process models. *Computational Statistics & Data Analysis*, 56(6):1362–1380.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2):89–121.

Fahrmeir, L., Kneib, T., Lang, S., and Marx, B. (2013). *Regression Models*. Springer.

Finley, A. O., Datta, A., Cook, B. D., Morton, D. C., Andersen, H. E., and Banerjee, S. (2019). Efficient algorithms for Bayesian nearest neighbor Gaussian processes. *Journal of Computational and Graphical Statistics*, 28(2):401–414.

Furrer, R., Genton, M. G., and Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 15(3):502–523.

Gressani, O. and Lambert, P. (2021). Laplace approximations for fast Bayesian inference in generalized additive models based on P-splines. *Computational Statistics & Data Analysis*, 154:107088.

Hayward, S. E., Deal, A., Cheng, C., Crawshaw, A., Orcutt, M., Vandrevala, T. F., Norredam, M., Carballo, M., Ciftci, Y., Requena-Méndez, A., et al. (2021). Clinical outcomes and risk factors for COVID-19 among migrant populations in high-income countries: a systematic review. *Journal of Migration and Health*, 3:100041.

Johnson, M. E., Moore, L. M., and Ylvisaker, D. (1990). Minimax and maximin distance designs. *Journal of Statistical Planning and Inference*, 26(2):131–148.

Jullion, A. and Lambert, P. (2007). Robust specification of the roughness penalty prior distribution in spatially adaptive Bayesian P-splines models. *Computational Statistics & Data Analysis*, 51(5):2542–2558.

Kammann, E. and Wand, M. P. (2003). Geoadditive models. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 52(1):1–18.

Katzfuss, M. and Guinness, J. (2021). A general framework for vecchia approximations of Gaussian processes. *Statistical Science*, 36(1):124–141.

Kaufman, C. G., Schervish, M. J., and Nychka, D. W. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association*, 103(484):1545–1555.

Krainski, E., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., and Rue, H. (2018). *Advanced spatial modeling with stochastic partial differential equations using R and INLA*. Chapman and Hall/CRC.

Lang, S. and Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics*, 13(1):183–212.

Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73(4):423–498.

Nychka, D. and Saltzman, N. (1998). Design of air-quality monitoring networks. In *Case studies in environmental statistics*, volume 132, pages 51–76. Springer.

Pijls, B. G., Jolani, S., Atherley, A., Derckx, R. T., Dijkstra, J. I. R., Franssen, G. H. L., Hendriks, S., Richters, A., Venemans-Jellema, A., Zalpuri, S., et al. (2021). Demographic risk factors for COVID-19 infection, severity, ICU admission and death: a meta-analysis of 59 studies. *BMJ Open*, 11(1):e044640.

Rozenfeld, Y., Beam, J., Maier, H., Haggerson, W., Boudreau, K., Carlson, J., and Medows, R. (2020). A model of disparities: risk factors associated with COVID-19 infection. *International Journal for Equity in Health*, 19(1):126.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric regression*. Number 12. Cambridge University Press.

Sainsbury-Dale, M., Zammit-Mangion, A., and Cressie, N. (2024). Modeling big, heterogeneous, non-Gaussian spatial and spatio-temporal data using FRK. *Journal of Statistical Software*, 108:1–39.

Sang, H. and Huang, J. Z. (2012). A full scale approximation of covariance functions for large spatial data sets. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 74(1):111–132.

Stein, M. L. (2014). Limitations on low rank approximations for covariance matrices of spatial data. *Spatial Statistics*, 8:1–19.

Sung, B. (2021). A spatial analysis of the association between social vulnerability and the cumulative number of confirmed deaths from COVID-19 in United States counties through November 14, 2020. *Osong Public Health and Research Perspectives*, 12(3):149.

Vandendijck, Y., Faes, C., and Hens, N. (2017). Estimating the spatial covariance structure using the geoadditive model. *Environmental and Ecological Statistics*, 24:341–361.

Vecchia, A. V. (1988). Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 50(2):297–312.

Wachtler, B., Michalski, N., Nowossadeck, E., Diercke, M., Wahrendorf, M., Santos-Hövener, C., Lampert, T., and Hoebel, J. (2020). Socioeconomic inequalities and COVID-19–A review of the current international literature. *Journal of Health Monitoring*, 5(Suppl 7):3.

Waller, L. A. and Gotway, C. A. (2004). *Applied Spatial Statistics for Public Health Data*. John Wiley & Sons.

Wand, M. P. (2003). Smoothing and mixed models. *Computational Statistics*, 18:223–249.

Wikle, C. K. (2010a). Hierarchical modeling with spatial data. *Handbook of Spatial Statistics*, pages 89–106.

Wikle, C. K. (2010b). Low-rank representations for spatial processes. *Handbook of Spatial Statistics*, 107:118.

Wu, C. and Qian, Y. (2022). The gender peak effect: Women are most vulnerable to infections during COVID-19 peaks. *Frontiers in Public Health*, 10:937179.

Wu, X., Nethery, R. C., Sabath, M. B., Braun, D., and Dominici, F. (2020). Exposure to air pollution and covid-19 mortality in the United States: A nationwide cross-sectional study. *MedRxiv*, pages 2020–04.

Zimmerman, D. L. (2010). Likelihood-based methods. *Handbook of Spatial Statistics*, pages 45–56.

Zimmerman, D. L. and Stein, M. (2010). Classical geostatistical methods. *Handbook of Spatial Statistics*, pages 29–44.