Received: 8 April 2021 | Revised: 24 March 2022 | Accepted: 26 March 2022

DOI: 10.1002/env.2727

RESEARCH ARTICLE

WILEY

Mitigating spatial confounding by explicitly correlating Gaussian random fields

Isa Marques[1] | **Thomas Kneib**[1] | **Nadja Klein**[2]

[1]Georg-August-University of Göttingen, Chairs of Statistics and Econometrics, Göttingen, Germany

[2]Humboldt-Universität zu Berlin, Chair of Statistics and Data Science, Berlin, Germany

Correspondence
Isa Marques, Georg-August-University of Göttingen, Chairs of Statistics and Econometrics, Humboldtallee 3, 37073 Göttingen, Germany.
Email: imarques@uni-goettingen.de

Funding information
Deutsche Forschungsgemeinschaft

Abstract

Spatial models are used in a variety of research areas, such as environmental sciences, epidemiology, or physics. A common phenomenon in such spatial regression models is spatial confounding. This phenomenon is observed when spatially indexed covariates modeling the mean of the response are correlated with a spatial random effect included in the model, for example, as a proxy of unobserved spatial confounders. As a result, estimates for regression coefficients of the covariates can be severely biased and interpretation of these is no longer valid. Recent literature has shown that typical solutions for reducing spatial confounding can lead to misleading and counterintuitive results. In this article, we develop a computationally efficient spatial model that explicitly correlates a Gaussian random field for the covariate of interest with the Gaussian random field in the main model equation and integrates novel prior structures to reduce spatial confounding. Starting from the univariate case, we extend our prior structure also to the case of multiple spatially confounded covariates. In simulation studies, we show that our novel model flexibly detects and reduces spatial confounding in spatial datasets, and it performs better than typically used methods such as restricted spatial regression. These results are promising for any applied researcher who wishes to interpret covariate effects in spatial regression models. As a real data illustration, we study the effect of elevation and temperature on the mean of monthly precipitation in Germany.

KEYWORDS

Bayesian inference, penalized complexity prior, spatial statistics, stochastic partial differential equation

1 | INTRODUCTION

Spatial regression data are regression data $\{(y(\boldsymbol{s}), \boldsymbol{z}(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S}\}$, where both the response variable $y(\boldsymbol{s})$ and the explanatory variables $\boldsymbol{z}(\boldsymbol{s})$ are indexed by a spatial variable $\boldsymbol{s}$ representing the location of the corresponding observational unit in a spatial domain $\mathcal{S}$. Based on observations collected at locations $\boldsymbol{s}_i \in \mathcal{S}$, $i = 1, \ldots, n$, a standard regression model would then be of the form

$$y(\boldsymbol{s}_i) = \beta_0 + \boldsymbol{z}(\boldsymbol{s}_i)^T\boldsymbol{\beta} + \varepsilon_i \quad (1)$$

with i.i.d. error terms $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$. However, this model could only be applied if—after adjusting for the covariates $\boldsymbol{z}(\boldsymbol{s}_i)$—no spatial dependence remains such that the residuals can indeed be assumed to be independent. This assumption is questionable for typical, observational spatial data where spatial dependence is likely to arise, for example, due to direct interaction between observational units of interest with spatial proximity determining the intensity of mutual interactions (e.g., when studying outcomes relating to animal behavior) or due to omitted variables that themselves obey spatial dependence. In this article, we are considering situations of the second type, that is, model (1) is assumed to represent the true data generating mechanism, yet we only have access to a subset of covariates instead of the complete vector.

More formally, we assume that $\boldsymbol{z}(\boldsymbol{s}) = (\boldsymbol{z}_{\text{obs}}(\boldsymbol{s})^T, \boldsymbol{z}_{\text{unobs}}(\boldsymbol{s})^T)^T$ and $\boldsymbol{z}_{\text{obs}}(\boldsymbol{s})$ is a subset of covariates that are observed, while the remainder $\boldsymbol{z}_{\text{unobs}}(\boldsymbol{s})$ is unobserved. This refers to the classical setup of unobserved confounders where

- naive estimation of the model relying exclusively on observed data will be biased unless observed and unobserved covariates are uncorrelated, and
- estimation uncertainty will usually be underestimated in the naive model based on observed data only,

see Greene (1990) or Wooldridge (2015) for details. In spatial data, it is particularly unlikely that $\boldsymbol{z}_{\text{obs}}(\boldsymbol{s})$ and $\boldsymbol{z}_{\text{unobs}}(\boldsymbol{s})$ are uncorrelated when both obey spatial dependence themselves. A common approach that is considered to adjust for unobserved spatial covariates is to approximate their overall effect as $\gamma(\boldsymbol{s}) \approx \boldsymbol{z}_{\text{unobs}}(\boldsymbol{s})^T\boldsymbol{\beta}_{\text{unobs}}$ and to model $\gamma(\boldsymbol{s})$ as a spatially correlated random effect in the model

$$y(\boldsymbol{s}_i) = \beta_0 + \boldsymbol{z}_{\text{obs}}(\boldsymbol{s}_i)^T\boldsymbol{\beta}_{\text{obs}} + \gamma(\boldsymbol{s}_i) + \varepsilon_i. \quad (2)$$

Unfortunately, this modeling strategy does not completely solve the problem since the correlation between the observed and unobserved part of the covariates carries over to dependence between $\boldsymbol{z}_{\text{obs}}(\boldsymbol{s})$ and $\gamma(\boldsymbol{s})$, which is then usually referred to as spatial confounding. As a consequence, the problem of unobserved confounding remains but, unlike the standard case of omitted variables, the possibility to estimate the spatial effect $\gamma(\boldsymbol{s})$ as a part of model (2) allows one to device dedicated estimation approaches.

Initially, most research on spatial confounding was motivated by the empirical observation that including a correlated spatial effect in model (2) could have severe (and often detrimental) impact on estimating the regression effects $\boldsymbol{\beta}_{\text{obs}}$ (Hodges & Reich, 2010; Schnell & Papadogeorgou, 2020). Early approaches were notably agnostic about the underlying model and reasons for this behavior, and simply tried to device improved estimation approaches that allowed to recover the true effects $\boldsymbol{\beta}_{\text{obs}}$. Whether or not spatial confounding is a relevant concern, depends both on the specific properties of the omitted variables $\boldsymbol{z}_{\text{unobs}}(\boldsymbol{s})$ and the purpose of the analysis. For the former, the scale at which spatial variability is observed and how this scale relates to the spatial variability in $\boldsymbol{z}_{\text{obs}}(\boldsymbol{s})$ is of major relevance (Hefley et al., 2017; Paciorek, 2010; Page et al., 2017).

With respect to the aim of the analysis, one can distinguish at least four situations. First, if the main goal is to obtain point predictions for the response only, it may be sufficient to just estimate the spatial model (2) since the potential bias in estimating $\boldsymbol{\beta}_{\text{obs}}$ will usually be compensated by estimating $\gamma(\boldsymbol{s})$ (Page et al., 2017). This does not necessarily carry over to the quantification of prediction uncertainty, but still spatial confounding may only be a second order problem in this case. Second, one may be interested in evaluating whether the overall spatial variability observed in the response of interest can be sensibly attributed to spatial risk factors by including candidate risk factors in $\boldsymbol{z}_{\text{obs}}(\boldsymbol{s})$ and evaluating how much spatial variation remains in $\gamma(\boldsymbol{s})$ after adjusting for these risk factors (Adin et al., 2021; Paciorek, 2010). In this case, spatial confounding is a major concern since in general one will not be able to tell apart spatial variation in $\boldsymbol{z}_{\text{obs}}(\boldsymbol{s})$ and $\gamma(\boldsymbol{s})$ without further structural model assumptions. One may also be interested in evaluating the effect of a specific treatment variable on the response while simultaneously adjusting for a number of additional confounding variables which, together, form the vector $\boldsymbol{z}_{\text{obs}}(\boldsymbol{s})$ (Davis et al., 2019; Guan et al., 2020; Schnell & Papadogeorgou, 2020). Here, dealing with spatial confounding is particularly important to identify the causal effect of the treatment. When it comes to the fourth situation, we distinguish a very significant part of the literature that is agnostic to the specific aims of the study (Dupont et al., 2021; Hanks et al., 2015; Hefley et al., 2017; Hodges & Reich, 2010; Hughes & Haran, 2013). In this article, we pragmatically consider dealing with spatial confounding as a way of getting a more realistic approximation of the underlying data generating process that enables meaningful interpretation of $\boldsymbol{\beta}_{\text{obs}}$, positioning our contribution close

to the evaluation of spatial risk factors. While our modeling approach may serve as a useful tool for exploring the presence of spatial confounding in a first place, the prior structures can also be useful as a building block in more complex spatial models.

The most common solution for spatial confounding is restricted spatial regression (RSR), which has been utilized both when evaluating the effect of a specific treatment (e.g., Reich et al., 2006) or risk factors (e.g., Adin et al., 2021). This method alleviates confounding by projecting the spatial effects into the orthogonal space spanned by the covariates and, thus, carries strong assumptions for the fixed effects. Recent studies show that RSR has counterintuitive consequences for causal inference and can perform worse than a nonspatial (NS) model such as the one in (1) (Khan & Calder, 2020; Zimmerman & Ver Hoef, 2021). Several alternatives to RSR have been brought forward. In the causal inference front, Thaden and Kneib (2018) use a three-stage structural equation model to extract the spatial structure in $\boldsymbol{z}_{\text{obs}}(\boldsymbol{s})$ and $\gamma(\boldsymbol{s})$, while Schnell and Papadogeorgou (2020) consider a joint model for $\boldsymbol{z}_{\text{obs}}(\boldsymbol{s})$ and $\gamma(\boldsymbol{s})$ based on Gaussian Markov random fields. Guan et al. (2020) explicitly address the importance of spatial scale, develop a joint model for $\boldsymbol{z}_{\text{obs}}(\boldsymbol{s})$ and $\gamma(\boldsymbol{s})$ in the spectral domain and study their correlation at different scales. On the more agnostic side, Dupont et al. (2021) use a two-stage approach to remove the spatial structure in $\boldsymbol{z}_{\text{obs}}(\boldsymbol{s})$ and, thus, avoid spatial confounding in the main equation (2).

Some literature has focused on simultaneously alleviating the computational bottlenecks of dealing with high dimensional latent spatial effects and spatial confounding (Hughes & Haran, 2013). To address this in our setting, we develop a Bayesian framework that allows to deal with spatial confounding in continuously indexed spatial models using a novel prior structure. We model $\gamma(\boldsymbol{s})$ and $\boldsymbol{z}_{\text{obs}}(\boldsymbol{s})$ jointly using a multivariate Gaussian random field (MGRF) distribution. For this, we use the stochastic partial differential equation (SPDE) approach (Lindgren et al., 2011) which utilizes a Gaussian Markov random field (GMRF) for computations. Moreover, we explore a penalized complexity (PC) prior for the correlation parameter that controls the shrinkage toward a base model, that is, the case of no spatial confounding (Simpson et al., 2017). Finally, thus far, the literature on spatial confounding does not discuss in-depth the case of multiple spatially confounded covariates in a spatial model. In this article, we address this gap, while explicitly accounting for spatial scale and computational complexity.

In summary, with this article, we

1. develop a new prior structure to deal with spatial confounding,
2. increase computational efficiency of the developed prior by exploiting the sparsity of the precision matrix of GMRFs in the SPDE-approach, and
3. extend the prior structure to the case of multiple covariates being correlated with the spatial random effect.

The remainder of this article is organized as follows. In Section 2, we introduce the novel prior structure. In Section 3, we extend the prior structure to the case of multiple spatial covariates. In Section 4, we present the details relating to the MCMC sampling. In Section 5, we present a simulation study that investigates the sources and consequences of spatial confounding. This simulation study compares our model with a spatial model that does not account for spatial confounding (called base model), RSR and a NS model. Additionally, we present a simulation study for the multiple covariates case. In Section 6, we demonstrate the useful properties of our prior structure in an application to precipitation data from Germany. Section 7 discusses the results and proposes future steps to further improve upon the present work.

2 | A NOVEL MGRF PRIOR FOR DEALING WITH SPATIAL CONFOUNDING IN ONE COVARIATE

A spatial process $\{\gamma(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S}\}$ is a continuously indexed GRF if all finite-dimensional distributions of the process are Gaussian, that is, for all $n \in \mathbb{N}$ and all $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n \in \mathcal{S}$, the vector $(\gamma(\boldsymbol{s}_1), \ldots, \gamma(\boldsymbol{s}_n))^T$ is multivariate Gaussian distributed. A GRF is specified by using a mean function $\mu_{\boldsymbol{\gamma}}(\boldsymbol{s})$ and a covariance function $\Sigma_{\boldsymbol{\gamma}}(\boldsymbol{s}, \boldsymbol{s}')$, for $\boldsymbol{s}, \boldsymbol{s}' \in \mathcal{S}$. We assume $\mu_{\boldsymbol{\gamma}}(\boldsymbol{s}) = 0$ for all $\boldsymbol{s}$.

2.1 | A MGRF model specification

Consider the model in (2). For notational simplicity, we will refer to $\boldsymbol{z}_{obs}$ as $\boldsymbol{z}$ and omit the spatial index $\boldsymbol{s}$. Let $\boldsymbol{z}$ be a covariate having a spatial structure that can be represented by a GRF with mean $\boldsymbol{\mu_Z}$ and covariance matrix $\boldsymbol{\Sigma_Z}$, that is,

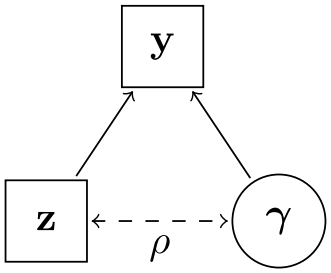

FIGURE 1 Visual representation of the prior structure. Squares represent observable variables and circles represent nonobservables

$\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z)$. Let $\boldsymbol{\gamma} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_\gamma)$. We assume that $\boldsymbol{\gamma}$ and $\boldsymbol{z}$ are jointly Gaussian distributed such that

$$\begin{pmatrix} \boldsymbol{\gamma} \\ \boldsymbol{z} \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{\mu}_z \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_\gamma & \rho \boldsymbol{\Sigma}_\gamma^{1/2} (\boldsymbol{\Sigma}_z^{1/2})^T \\ \rho \boldsymbol{\Sigma}_z^{1/2} (\boldsymbol{\Sigma}_\gamma^{1/2})^T & \boldsymbol{\Sigma}_z \end{pmatrix} \right], \tag{3}$$

where $\rho \boldsymbol{\Sigma}_\gamma^{1/2} (\boldsymbol{\Sigma}_z^{1/2})^T = \boldsymbol{\Sigma}_{\gamma\, z}$, $\boldsymbol{\Sigma}_{z\, \gamma} = \boldsymbol{\Sigma}_{\gamma\, z}^T$, and $\rho \in [-1, 1]$. The matrix roots $\boldsymbol{\Sigma}_\gamma^{1/2}$ and $\boldsymbol{\Sigma}_z^{1/2}$ satisfy $\boldsymbol{\Sigma}_\gamma = \boldsymbol{\Sigma}_\gamma^{1/2} (\boldsymbol{\Sigma}_\gamma^{1/2})^T$ and $\boldsymbol{\Sigma}_z = \boldsymbol{\Sigma}_z^{1/2} (\boldsymbol{\Sigma}_z^{1/2})^T$, and can be derived from a Cholesky decomposition or an eigenvalue decomposition. The covariance structure is constructed such that $\text{Corr}(\gamma_m, Z_m) = \rho$, for all $m = 1, \ldots, M$.

The conditional distribution of $\boldsymbol{\gamma}$ given $\boldsymbol{z}$ is a Gaussian distribution with mean and covariance

$$\boldsymbol{\mu}_{\gamma|z} = \boldsymbol{0} + \boldsymbol{\Sigma}_{\gamma\, z} \boldsymbol{\Sigma}_z^{-1} (\boldsymbol{z} - \boldsymbol{\mu}_z), \tag{4}$$

$$\boldsymbol{\Sigma}_{\gamma|z} = \boldsymbol{\Sigma}_\gamma - \boldsymbol{\Sigma}_{\gamma\, z} \boldsymbol{\Sigma}_z^{-1} \boldsymbol{\Sigma}_{z\, \gamma} \tag{5}$$

leading to

$$\boldsymbol{\gamma}|\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{\mu}_{\gamma|z}, \boldsymbol{\Sigma}_{\gamma|z}) \quad \text{and} \quad \boldsymbol{z} \sim \mathcal{N}(\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z). \tag{6}$$

If $\rho = 0$, then $\boldsymbol{\gamma}|\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_\gamma)$, that is, we get the typically used prior in continuously indexed spatial models, which does not account for spatial confounding. We refer to a prior that assumes $\rho = 0$ as the base prior, and the model associated with it as the base model. The more general case with $\rho \in [-1, 1]$ in (6) is denoted here a multivariate Gaussian random field (MGRF) prior, and the model associated with it is a MGRF model. The MGRF prior structure is illustrated graphically in Figure 1.

2.2 | GMRF reformulation

GMRFs are multivariate Gaussian distributions satisfying conditional independence assumptions, that is, they have a sparse precision matrix, $\boldsymbol{Q}_\gamma = \boldsymbol{\Sigma}_\gamma^{-1}$. While GMRFs are less attractive than GRFs from a theoretical viewpoint, the sparseness of the precision matrix, and potentially also of the corresponding Cholesky factorization, makes GMRFs attractive from a computational viewpoint.

Rue and Tjelmeland (2002) demonstrated empirically that GMRFs can closely approximate most of the commonly used covariance functions in geostatistics and proposed using them as computational replacements for GRFs for computational reasons. The stochastic partial differential equation (SPDE) approach (Lindgren et al., 2011) exploits this empirical equivalence and obtains a model that benefits from both the computational benefits of the GMRF and a well-defined continuous GRF counterpart.

Recently, the nearest neighbors Gaussian process (NNGP) model (Datta et al., 2016) has been presented as a computational competitor for the SPDE-approach. However, while NNGPs benefit from the construction of a sparse Cholesky factorization of the precision matrix (and thus a GMRF specification), they do not have a well-defined continuous counterpart. As a consequence, the interpretability of the resulting model, for example, in terms of concepts such as spatial

scale, is hindered and the model is dependent on the order in which the observations are included in the model, which in turn can lead to large discrepancies between the original covariance and the NNGP covariance (Lindgren et al., 2022). In contrast, for full Cholesky factorization such as in the SPDE-approach, the ordering of the nodes only affects the sparsity of the computed factorization (Lindgren et al., 2022; Rue & Held, 2005).

All spatial models in this article use the SPDE-approach (Lindgren et al., 2011). In this approach, a GRF $\gamma(\boldsymbol{s})$ (similarly for $\boldsymbol{z}(\boldsymbol{s})$) is expanded into a piecewise linear basis through

$$\gamma(\boldsymbol{s}) = \sum_{m=1}^{M} \psi_m(\boldsymbol{s})\gamma_m, \tag{7}$$

where the joint distribution of the weight vector $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_M)^T$ are normally distributed weights $\boldsymbol{\gamma}$ with mean zero and sparse precision matrix $\boldsymbol{Q}_\gamma$ such that $\boldsymbol{\gamma} \sim N(\boldsymbol{0}, \boldsymbol{Q}_\gamma^{-1})$, that is, the GMRF approximation to the GRF. Each basis function $\psi_m(\cdot)$ is piecewise linear. Note that at this point $\boldsymbol{\gamma}$ gains a new meaning, specifically as the GMRF approximation of the GRF $\gamma(\boldsymbol{s})$ via (7), that is, omitting $\boldsymbol{s}$ no longer represents a simplification of notation.

The sparse precision matrix $\boldsymbol{Q}_\gamma$ will typically reduce the cost of factorizations—such as the Cholesky decomposition—to $O(n^{3/2})$ in spatial models, instead of $O(n^3)$ for GRFs with dense covariance matrices. Moreover, the Cholesky decomposition of the precision $\boldsymbol{Q}_\gamma$ inherits its sparse structure in the SPDE-approach (Rue & Held, 2005). Thus, in this article, we assume $\boldsymbol{\Sigma}_\gamma^{1/2}$ (similarly for $\boldsymbol{\Sigma}_z^{1/2}$) represents the Cholesky decomposition of $\boldsymbol{\Sigma}_\gamma$. This allows us to exploit the sparsity of the Cholesky factor to reduce the computational complexity of the resulting prior structure.

In the SPDE-approach, the precision of the resulting GMRF depends on $\tau_\gamma > 0$, a precision parameter, $\kappa_\gamma > 0$, a parameter related to the spatial range, and $\nu_\gamma > 0$ which is a smoothness parameter. Throughout this article, we assume $S \subseteq \mathbb{R}^2$. The marginal variance σ_γ^2 of the GRF can be derived from the parameters τ_γ and κ_γ, via

$$\sigma_\gamma^2 = \frac{\Gamma(\nu_\gamma)}{\Gamma(\alpha_\gamma)4\pi\kappa_\gamma^{2\nu_\gamma}\tau_\gamma^2}.$$

Furthermore, by defining the range as the distance at which the spatial correlation falls to 0.05, the range parameter r_γ can be written as

$$r_\gamma = \frac{\sqrt{8\nu_\gamma}}{\kappa_\gamma}.$$

2.2.1 | Reformulation I

For positive definite $\boldsymbol{Q}_\gamma$, there is a unique Cholesky factor $\boldsymbol{L}_\gamma$ such that $\boldsymbol{L}_\gamma$ is a lower triangular matrix with $L_{\gamma(ii)} > 0\ \forall i$ and $\boldsymbol{Q}_\gamma = \boldsymbol{L}_\gamma \boldsymbol{L}_\gamma^T$. By exploiting the Markov graph structure of the GMRF, the Cholesky factor of the precision matrix can be guaranteed to be sparse (Rue & Held, 2005). The same logic applies to $\boldsymbol{z}$ and we can rewrite (4) and (5) such that

$$\boldsymbol{\mu}_{\gamma|z} = \boldsymbol{0} + \rho\, \boldsymbol{L}_\gamma^{-T}\boldsymbol{L}_z^T(\boldsymbol{z} - \boldsymbol{\mu}_Z), \tag{8}$$

$$\boldsymbol{Q}_{\gamma|z} = \boldsymbol{\Sigma}_{\gamma|z}^{-1} = \frac{1}{1-\rho^2}\boldsymbol{Q}_\gamma, \tag{9}$$

as derived in detail in Supplement A.

2.2.2 | Reformulation II

Reformulation I of the first and second order structure of the prior takes advantage of the sparse structure of $\boldsymbol{Q}_\gamma$ and $\boldsymbol{Q}_z$, and it can lead to a considerable decrease in the computational cost of factorizations. However, to guarantee positive definiteness, we need to avoid boundary cases $|\rho| \approx 1$.

To address this, one can consider the unconstrained GMRF, $\boldsymbol{\gamma} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q}_{\gamma}^{-1})$, and then compute

$$\boldsymbol{\gamma}^{*} = \boldsymbol{\gamma} + \rho\, \boldsymbol{L}_{\gamma}^{-T} \boldsymbol{L}_{z}^{T} (\boldsymbol{z} - \boldsymbol{\gamma}_{z}), \quad \text{where} \quad \boldsymbol{\gamma}_{z} \sim \mathcal{N}(\boldsymbol{\mu}_{z}, \boldsymbol{\Sigma}_{z}).$$

The resulting $\boldsymbol{\gamma}^{*}$ has the correct conditional distribution (see Supplement B). This procedure is similar to "conditioning by kriging" in Rue and Held (2005) and it essentially removes from the linear predictor the part of the spatial effect that is correlated with the covariate of interest.

In comparison to reformulation I, this reformulation requires the sampling of one additional parameter, $\boldsymbol{\gamma}_{z}$, of size M. This can increase the computational costs associated with MCMC, as $\boldsymbol{\gamma}_{z}$ is sampled every iteration.

2.3 | Prior for ρ

Estimating the correlation between $\boldsymbol{z}$ and $\boldsymbol{\gamma}$ is a complex task, namely because $\boldsymbol{\gamma}$ is an unobservable random effect. In this article, we estimate ρ within the Bayesian hierarchical model presented in more detail in Section 4, and set a PC-prior for ρ.

The concept of a PC-prior was first developed by Simpson et al. (2017) and an in-depth introduction can be found there. In short, as the name penalized complexity indicates, this prior invokes the principle of parsimony, for which a simpler *base* model should be preferred until there is enough support for a more complex model. The prior density decays at a constant decay-rate as a function of a measure of the increased complexity between the more flexible model and the base model. The resulting prior allows for user-defined scaling: an upper bound for the parameter of interest can be defined, which describes what is considered as "tail event" and what is the weight put on this event. Hence, the user can describe how informative the resulting PC-prior is.

Simpson et al. (2017) developed a PC-prior for correlation parameters. Here, we consider a base model with $\rho_0 = 0$, that is, a base model that assumes no spatial confounding between fixed effects and spatial random effects. Let $\frac{-1}{c-1} < \rho < 1$ with $c \geq 2$. Simpson et al. (2017) show that the PC-prior for ρ has density

$$p(\rho) = \frac{c-1}{2} \left(\frac{1}{1-\rho} - \frac{1}{1+(c-1)\rho} \right) \frac{\lambda}{\sqrt{-lR(\rho)}} \exp(-\lambda \sqrt{-lR(\rho)}),$$

where $lR(\rho) = \log(R(\rho))$ and $R(\rho) = (1+(c-1)\rho)(1-\rho)^{c-1}$.

The decay-rate λ can be chosen by sampling PC-priors from various values of λ and choosing a λ satisfying

$$\text{Prob}(|\rho| > U) = a. \tag{10}$$

The resulting PC-prior is shown in Figure 2, for $c = 2$ and different values of U and a. As it can be observed, the prior can be made less informative by increasing U or a, that is, by reducing shrinkage toward the base model. For different values of c, the left boundary for the correlation can be shifted toward zero, resulting in an asymmetric prior distribution for ρ. However, throughout this article, we assume there is no information concerning the type of correlation between the unobservable spatial random field and the covariate. Thus, we only consider the case $c = 2$. Moreover, we set $a = 0.05$ and simply change U to get the desired prior density. We refer to the resulting PC-prior as $\mathcal{PC}(U)$.

2.4 | Full hierarchical model

In what follows, we present the full hierarchical structures of the models for reformulations I and II, respectively. The priors on θ_{γ}^{τ}, θ_{z}^{τ}, θ_{γ}^{κ} and θ_{z}^{κ} for $\mathcal{S} = [0,1] \times [0,1]$ are chosen such that the spatial range and marginal variance vary uniformly in the intervals $[0.01, 1]$ and $[0.01, 5]$, respectively, as this should cover all the variability of the considered (scaled) responses. For σ_{ϵ}^{2}, we specify an inverse-gamma distribution $\mathcal{IG}(c = 0.001, d = 0.001)$. Cases $c = d$, with both values approaching zero, are widely used as a weakly informative choice for σ_{ϵ}^{2} (see section 4.4 of Fahrmeir et al., 2013). Recall that $\gamma(\boldsymbol{s}) = \sum_{m=1}^{M} \psi_{m}(\boldsymbol{s}) \gamma_{m}$ and $\boldsymbol{z}(\boldsymbol{s}) = \sum_{m=1}^{M} \psi_{m}(\boldsymbol{s}) \boldsymbol{z}_{m}$. Moreover, $\boldsymbol{\theta}_{\boldsymbol{z}} = (\theta_{z}^{\tau}, \theta_{z}^{\kappa})^{T}$ and $\boldsymbol{\theta}_{\boldsymbol{\gamma}} = (\theta_{\gamma}^{\tau}, \theta_{\gamma}^{\kappa})^{T}$.

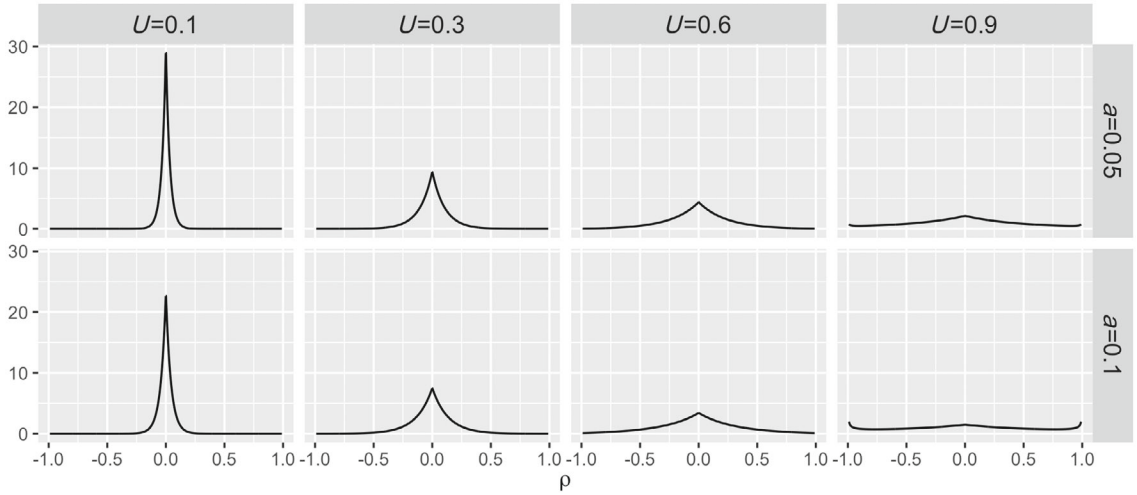

FIGURE 2 Prior densities for ρ using a PC-prior with $\rho_0 = 0$, $c = 2$ and different values for U and a

The hierarchical structure for $B = 1$ and reformulation I in Section 2.2.1 is

$$
\begin{aligned}
y(\boldsymbol{s}_i) = \eta(\boldsymbol{s}_i) + \varepsilon_i &= \beta_0 + z(\boldsymbol{s}_i)^T\boldsymbol{\beta} + \gamma(\boldsymbol{s}_i) + \varepsilon_i \text{ with} \\
\boldsymbol{\gamma}|\boldsymbol{\theta}_{\boldsymbol{z}}, \boldsymbol{\theta}_{\boldsymbol{\gamma}}, \mu_{\boldsymbol{z}} &\sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\gamma}|\boldsymbol{z}}, \boldsymbol{Q}_{\boldsymbol{\gamma}|\boldsymbol{z}}(\boldsymbol{\theta}_{\boldsymbol{z}}, \boldsymbol{\theta}_{\boldsymbol{\gamma}})^{-1}), \\
\beta_0 &\sim \mathcal{N}(\mu_{\beta_0}, \sigma^2_{\beta_0}), \\
\boldsymbol{\beta} &\sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \text{diag}(\boldsymbol{\sigma}^{\boldsymbol{2}}_{\boldsymbol{\beta}})), \\
\boldsymbol{\varepsilon}|\sigma^2_\varepsilon &\sim \mathcal{N}(\boldsymbol{0}, \sigma^2_\varepsilon \boldsymbol{I}), \\
\sigma^2_\varepsilon &\sim \mathcal{IG}(0.001, 0.001), \\
\theta^\tau_{\boldsymbol{\gamma}}, \theta^\tau_{\boldsymbol{z}} &\sim \mathcal{U}(-10, 0), \\
\theta^\kappa_{\boldsymbol{\gamma}}, \theta^\kappa_{\boldsymbol{z}} &\sim \mathcal{U}(1, 5), \\
\rho &\sim \mathcal{N}(U, 0.05), \\
\mu_{\boldsymbol{z}} &\sim \mathcal{N}(\mu_{\mu_{\boldsymbol{z}}}, \sigma^2_{\mu_{\boldsymbol{z}}}),
\end{aligned}
$$

while the hierarchical structure for $B = 1$ and reformulation II in Section 2.2.2 is

$$
\begin{aligned}
y(\boldsymbol{s}_i) = \eta(\boldsymbol{s}_i) + \varepsilon_i &= \beta_0 + z(\boldsymbol{s}_i)^T\boldsymbol{\beta} + \gamma^*(\boldsymbol{s}_i) + \varepsilon_i \\
&= \beta_0 + z(\boldsymbol{s}_i)^T\boldsymbol{\beta} + \sum_{m=1}^{M} \psi_m(\boldsymbol{s}_i)\gamma^*_m + \varepsilon_i \text{ with} \\
\boldsymbol{\gamma}^* &= \boldsymbol{\gamma} + \rho\, \boldsymbol{L}_{\boldsymbol{\gamma}}^{-T}\boldsymbol{L}_{\boldsymbol{z}}^T(\boldsymbol{z} - \boldsymbol{\gamma}_{\boldsymbol{z}}), \\
\boldsymbol{\gamma}|\boldsymbol{\theta}_{\boldsymbol{\gamma}} &\sim N(\boldsymbol{0}, \boldsymbol{Q}_{\boldsymbol{\gamma}}(\boldsymbol{\theta}_{\boldsymbol{\gamma}})^{-1}), \\
\beta_0 &\sim \mathcal{N}(\mu_{\beta_0}, \sigma^2_{\beta_0}), \\
\boldsymbol{\beta} &\sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \text{diag}(\boldsymbol{\sigma}^{\boldsymbol{2}}_{\boldsymbol{\beta}})), \\
\boldsymbol{\varepsilon}|\sigma^2_\varepsilon &\sim \mathcal{N}(\boldsymbol{0}, \sigma^2_\varepsilon \boldsymbol{I}), \\
\sigma^2_\varepsilon &\sim \mathcal{IG}(0.001, 0.001), \\
\theta^\tau_{\boldsymbol{\gamma}}, \theta^\tau_{\boldsymbol{z}} &\sim \mathcal{U}(-10, 0), \\
\theta^\kappa_{\boldsymbol{\gamma}}, \theta^\kappa_{\boldsymbol{z}} &\sim \mathcal{U}(1, 5), \\
\rho &\sim \mathcal{PC}(U, 0.05), \\
\mu_{\boldsymbol{z}} &\sim \mathcal{N}(\mu_{\mu_{\boldsymbol{z}}}, \sigma^2_{\mu_{\boldsymbol{z}}}), \\
\boldsymbol{\gamma}_{\boldsymbol{z}}|\boldsymbol{\theta}_{\boldsymbol{z}}, \mu_{\boldsymbol{z}} &\sim N(\mu_{\boldsymbol{z}}\boldsymbol{1}_M, \boldsymbol{Q}_{\boldsymbol{z}}(\boldsymbol{\theta}_{\boldsymbol{z}})^{-1}).
\end{aligned}
$$

3 | SPATIAL CONFOUNDING IN MULTIPLE COVARIATES

In many situations, we are interested in accounting for multiple spatially varying risk factors and, consequently, we need to tell apart spatial variation in multiple elements of $z_{\text{obs}}(\boldsymbol{s})$ and the spatial effect $\gamma(\boldsymbol{s})$. In the spatial confounding literature, the case of multiple spatially confounded covariates in not well explored, and the usual recommendation is to extend the single-covariate strategy to all spatial covariates included in the model (Dupont et al., 2021; Thaden & Kneib, 2018). This can lead the resulting model to be quite complex and, potentially, hard to estimate.

The prior structure in Section 2 exploits the sparsity of precision matrices of GMRFs for computational efficiency purposes. In this section, we develop a novel prior structure for the case of multiple spatially confounded covariates that explores an additional source of sparsity. Namely, we investigate the efficacy of an approach based on the preprocessing of the matrix of spatial varying covariates $\boldsymbol{z} = (\boldsymbol{z}_1, \dots, \boldsymbol{z}_b, \dots, \boldsymbol{z}_B)^T$ using principal component analysis (PCA; Abdi & Williams, 2010; Pearson, 1901).

Thinking in terms of GMRFs, consider that the model includes B spatially varying covariates $\boldsymbol{z}_1, \dots, \boldsymbol{z}_b, \dots, \boldsymbol{z}_B$ and that we need to disentangle the spatial variation between these and $\boldsymbol{\gamma}$. Moreover, for now, ignore potential collinearity between the different covariates. If we follow Section 2, the most direct way to disentangle spatial variation in $\boldsymbol{\gamma}$ and $\boldsymbol{z}_1, \dots, \boldsymbol{z}_b, \dots, \boldsymbol{z}_B$ is to model the joint $(B+1)$-variate spatial distribution of $(\boldsymbol{\gamma}, \boldsymbol{z}_1, \dots, \boldsymbol{z}_b, \dots, \boldsymbol{z}_B)^T$ in the prior for $\boldsymbol{\gamma}$. Depending on how large B is, this strategy could imply a great computational burden, as well as the estimation of $\frac{B(B+1)}{2}$ correlation parameters.

The PCA methodology reduces the dimensionality of multidimensional datasets for analysis purposes. This methodology allows us to represent the data in a new coordinate system in which basis vectors follow the direction of greatest variation in the data. Consider the eigendecomposition of $\boldsymbol{z}^T\boldsymbol{z}$ utilized in PCA. We choose the first B^* (largest) eigenvalues $\lambda_1, \dots, \lambda_{B^*}$ and their respective eigenvectors, such that

$$\frac{\sum_{b=1}^{B^*}\lambda_b}{\sum_{b=1}^{B}\lambda_b} > \psi, \tag{11}$$

where $\psi \in (0,1)$ is the required threshold (typically, $\psi = 0.95$). We denote the vector of variables resulting from PCA by $\boldsymbol{w} = (\boldsymbol{w}_1, \dots, \boldsymbol{w}_{b^*}, \dots, \boldsymbol{w}_{B^*})^T$. The model follows

$$y(\boldsymbol{s}_i) = \beta_0 + \boldsymbol{z}(\boldsymbol{s}_i)^T\boldsymbol{\beta} + \gamma(\boldsymbol{s}_i) + \varepsilon_i$$

but now the spatial effect has the prior $\boldsymbol{\gamma}|\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\gamma}|\boldsymbol{w}}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}|\boldsymbol{w}})$ that follows

$$\begin{pmatrix} \boldsymbol{\gamma} \\ \boldsymbol{w}_1 \\ \dots \\ \boldsymbol{w}_{B^*} \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \boldsymbol{0} \\ \mu_{w_1}\boldsymbol{I} \\ \dots \\ \mu_{w_{B^*}}\boldsymbol{I} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}} & \rho_1\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{1/2}(\boldsymbol{\Sigma}_{\boldsymbol{w}_1}^{1/2})^T & \dots & \rho_{B^*}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{1/2}(\boldsymbol{\Sigma}_{\boldsymbol{w}_{B^*}}^{1/2})^T \\ \rho_1\boldsymbol{\Sigma}_{\boldsymbol{w}_1}^{1/2}(\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{1/2})^T & \boldsymbol{\Sigma}_{\boldsymbol{w}_1} & \dots & \boldsymbol{0} \\ \dots & \dots & \dots & \dots \\ \rho_{B^*}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{1/2}(\boldsymbol{\Sigma}_{\boldsymbol{w}_{B^*}}^{1/2})^T & \boldsymbol{0} & \dots & \boldsymbol{\Sigma}_{\boldsymbol{w}_{B^*}} \end{pmatrix}\right].$$

Besides dimensionality reduction, choosing the B^* largest eigenvalues has an intuitive spatial interpretation. Spatial confounding typically biases regression coefficients when $\boldsymbol{z}$ has a larger spatial scale than $\boldsymbol{\gamma}$ (Adin et al., 2021;Paciorek, 2010; Page et al., 2017), since when the spatial effect operates on a smaller scale than the covariate, it is likely to explain the data better than the covariate. Also from a spectral analysis perspective, "unconfoundness" is typically associated with high-frequencies (small scale) of the covariate (Guan et al., 2020; Hefley et al., 2017). By choosing the largest eigenvalues in the PCA of the (scaled) $\boldsymbol{z}$, we can extract the large scale behavior associated with the covariate vector. This is illustrated in Figure 3 which shows $\boldsymbol{w}_{b^*}$ arising from the PCA-analysis of 20 GRF covariates $\boldsymbol{z}_1, \dots, \boldsymbol{z}_{20}$ with decreasing spatial range. We observe that the first eigenvalues, that is, the largest, are associated with larger spatial scales.

Our strategy does not correct for correlation between the covariates. Nonetheless, if the covariates in $(\boldsymbol{z}_1, \dots, \boldsymbol{z}_B)^T$ are correlated among themselves, the user can choose common solutions against multicollinearity in linear regression models, for example, by omitting some of the affected covariates or constructing a single combined and easily interpretable covariate from the covariates in question. Another possibility is to use PCA also in this case (Fahrmeir et al., 2013).

The full hierarchical structure in the multiple covariates case is close to the structure in Section 2.4 and can be found in the Supplement C.

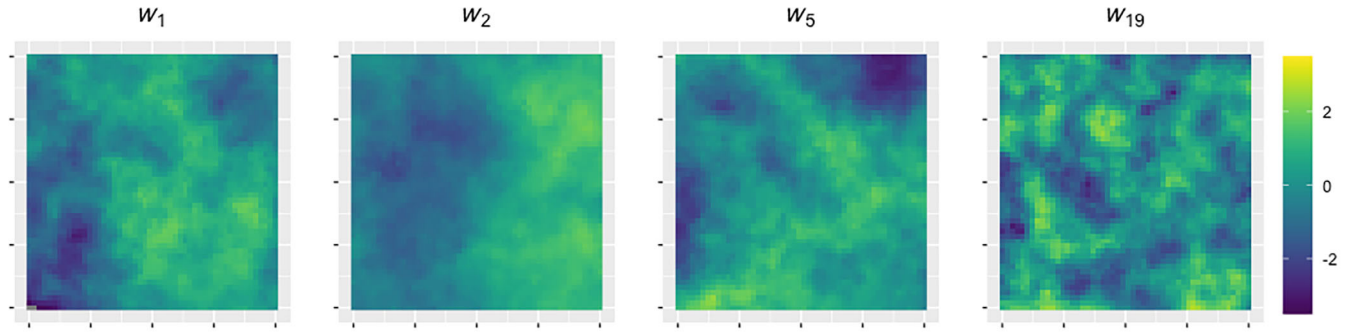

FIGURE 3 The PCA variables $\boldsymbol{w}_1$, $\boldsymbol{w}_2$, $\boldsymbol{w}_5$, and $\boldsymbol{w}_{19}$, associated with eigenvalues index 1, 2, 5, and 19

4 | POSTERIOR EVALUATION

To perform posterior estimation for the full vector of model parameters $\boldsymbol{\vartheta}$, we develop a fully Bayesian approach using MCMC, and separate $\boldsymbol{\vartheta}$ into smaller blocks. Of particular relevance is the separation between variables for which we know the corresponding (full conditional) posterior distributions, and those for which the posterior distribution is not available. We will use a Gibbs sampler (Gelfand, 2000) for the former and derive the corresponding posterior distributions in Section 4.1. For the latter, we use Metropolis Hastings (MH, Metropolis et al., 1953) steps, which are explained in more detail in Section 4.2. Section 4.3 summarizes the final structure of the sampler.

4.1 | Posterior distributions for Gibbs sampler

The priors for β_0, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, σ^2_ϵ, $\mu_{\boldsymbol{z}}$, and additionally $\boldsymbol{\gamma}_{\boldsymbol{z}}$ in reformulation II, are conjugate priors such that Gibbs steps are possible. For all parameters that have normal distributed priors, we can use the properties of the product of two normal densities to derive their posterior distribution (see supplement B of Fahrmeir et al., 2013). In what follows, we start by describing the posterior distributions that are identical for both reformulation I and reformulation II, from Sections 2.2.1 and 2.2.2, respectively. Subsequently, we derive the posteriors for the remaining parameters.

4.1.1 | Posterior distributions identical for reformulations I and II

Let $\tilde{\boldsymbol{z}}$ be matrix $\boldsymbol{z}$ with first column $\mathbf{1}_M$. Consider the $(B+1)\times(B+1)$ matrices $\boldsymbol{A} = \frac{1}{\sigma^2}\tilde{\boldsymbol{z}}\boldsymbol{\psi}^T\boldsymbol{\psi}\tilde{\boldsymbol{z}}^T$ and $\boldsymbol{B} = \text{diag}\left(\sigma^2_{\beta_0}, \boldsymbol{\Sigma}^{2\,T}_{\boldsymbol{\beta}}\right)^{-1}$, where diag denotes a diagonal matrix. Moreover, consider the $(B+1)$-dimensional vectors $\boldsymbol{a} = \tilde{\boldsymbol{z}}\boldsymbol{\psi}^T(\boldsymbol{\psi}\tilde{\boldsymbol{z}}^T\tilde{\boldsymbol{z}}\boldsymbol{\psi}^T)^{-1}(\boldsymbol{y}-\boldsymbol{\psi}\boldsymbol{\gamma})$, and $\boldsymbol{b} = (\mu_{\beta_0}, \boldsymbol{\mu}^T_{\boldsymbol{\beta}})^T$, where $\boldsymbol{y} = (y(\boldsymbol{s}_1), \ldots, y(\boldsymbol{s}_n))^T$. Then,

$$(\beta_0, \boldsymbol{\beta}^T)^T|\cdot \sim \mathcal{N}\left((\boldsymbol{A}+\boldsymbol{B})^{-1}(\boldsymbol{A}\boldsymbol{a}+\boldsymbol{B}\boldsymbol{b}), (\boldsymbol{A}+\boldsymbol{B})^{-1}\right). \tag{12}$$

Consider the single covariate case and Section 2.4. The posterior for $\mu_{\boldsymbol{z}}|\cdot$ follows the same structure as (12), but we instead have the scalars $A = \mathbf{1}^T_M\boldsymbol{Q}_{\boldsymbol{Z}}\mathbf{1}_M$, $B = \frac{1}{\sigma^2_{\mu_{\boldsymbol{z}}}}$, and the M-size vectors $a = \mathbf{1}^T_M\boldsymbol{I}_M\boldsymbol{z}$ and $b = \mu_{\mu_{\boldsymbol{z}}}$, with $\boldsymbol{z}$ replaced by $\boldsymbol{w}$ in the multiple covariates case.

In the case of σ^2_ϵ, the prior is inverse-gamma distributed, which combined with the normal distributed response leads to a posterior distribution (see p. 229 of Fahrmeir et al. (2013))

$$\sigma^2_\epsilon|\cdot \sim \mathcal{IG}\left(0.001 + \frac{n}{2}, 0.001 + \frac{1}{2}(\boldsymbol{y}-\boldsymbol{\eta})^T(\boldsymbol{y}-\boldsymbol{\eta})\right),$$

where $\boldsymbol{\eta} = (\eta(\boldsymbol{s}_1), \ldots, \eta(\boldsymbol{s}_n))^T$ and, according to Section 2.4 and Supplement C.

4.1.2 | Posterior distributions not identical for reformulations I and II

In reformulation I, $\boldsymbol{\gamma}|\cdot$ once again follows the structure in (12), but with $M \times M$ matrices $\boldsymbol{A} = \frac{1}{\sigma^2}\boldsymbol{\psi}^T\boldsymbol{\psi}$ and $\boldsymbol{B} = \boldsymbol{Q}_{\boldsymbol{\gamma}|\boldsymbol{z}}^{-1}$ and M-vectors $\boldsymbol{a} = \boldsymbol{\psi}^T(\boldsymbol{\psi}\boldsymbol{\psi}^T)^{-1}(\boldsymbol{y} - \beta_0\mathbf{1}_n) - \boldsymbol{z}^T\boldsymbol{\beta}$ and $\boldsymbol{b} = \boldsymbol{\mu}_{\boldsymbol{\gamma}|\boldsymbol{z}}$. In the multiple covariates case, $\boldsymbol{B} = \boldsymbol{Q}_{\boldsymbol{\gamma}|\boldsymbol{w}}^{-1}$ and $\boldsymbol{b} = \boldsymbol{\mu}_{\boldsymbol{\gamma}|\boldsymbol{w}}$.

In reformulation II, $\boldsymbol{\gamma}|\cdot$ is sampled according to $M \times M$ matrices $\boldsymbol{A} = \frac{1}{\sigma^2}\boldsymbol{\psi}^T\boldsymbol{\psi}$ and $\boldsymbol{B} = \boldsymbol{Q}_{\boldsymbol{\gamma}}^{-1}$ and M-sized vectors $\boldsymbol{a} = \boldsymbol{\psi}^T(\boldsymbol{\psi}\boldsymbol{\psi}^T)^{-1}(\boldsymbol{y} - \beta_0\mathbf{1}_n) - \boldsymbol{z}^T\boldsymbol{\beta} - \rho\, \boldsymbol{L}_{\boldsymbol{\gamma}}^{-T}\boldsymbol{L}_{\boldsymbol{z}}^T(\boldsymbol{z} - \boldsymbol{\gamma}_{\boldsymbol{z}})$ and $\boldsymbol{b} = \mathbf{0}$. In the multiple covariates case, $\boldsymbol{a} = \boldsymbol{\psi}^T(\boldsymbol{\psi}\boldsymbol{\psi}^T)^{-1}(\boldsymbol{y} - \beta_0\mathbf{1}_n) - \boldsymbol{z}^T\boldsymbol{\beta} - \rho\, \boldsymbol{L}_{\boldsymbol{\gamma}}^{-T}\boldsymbol{L}_{\boldsymbol{w}}^T(\boldsymbol{w} - \boldsymbol{\gamma}_{\boldsymbol{w}})$.

Finally, in reformulation II, $\boldsymbol{\gamma}_{\boldsymbol{z}}|\cdot$ is sampled according to the $M \times M$ matrices $\boldsymbol{A} = (\rho\, \boldsymbol{L}_{\boldsymbol{\gamma}}^{-T}\boldsymbol{L}_{\boldsymbol{z}}^T)^T\boldsymbol{Q}_{\boldsymbol{\gamma}}\rho\, \boldsymbol{L}_{\boldsymbol{\gamma}}^{-T}\boldsymbol{L}_{\boldsymbol{z}}^T$, $\boldsymbol{B} = \boldsymbol{Q}_{\boldsymbol{z}}$, $\boldsymbol{a} = \boldsymbol{z} - (\rho\, \boldsymbol{L}_{\boldsymbol{\gamma}}^{-T}\boldsymbol{L}_{\boldsymbol{z}}^T)^{-1}\boldsymbol{\gamma}$ and $\boldsymbol{b} = \mu_{\boldsymbol{z}}\mathbf{1}_K$, with all $\boldsymbol{z}$ and $\boldsymbol{\gamma}_{\boldsymbol{z}}$ replaced by $\boldsymbol{w}$ and $\boldsymbol{\gamma}_{\boldsymbol{w}}$, respectively, in the multiple covariates case.

4.2 | Posterior estimation for MH steps

Since parameters $\boldsymbol{\theta}_{\boldsymbol{\gamma}}$, $\boldsymbol{\theta}_{\boldsymbol{z}}$, and ρ^* are highly correlated, we generally sample them in one block in a Metropolis–Hastings (MH) step using the robust adaptive MH method from Vihola (2012) with Student t-distributed proposal densities. This algorithm estimates the shape of the target distribution and simultaneously coerces the acceptance rate. Here, we use the typically desired acceptance rate of 23.4% for multidimensional settings (Roberts et al., 1997). However, in the multiple covariates case, ρ^* is updated separately as the block update led to updates that were negligible on practical terms. In this cases, for ρ^*, we use RAM with the typical acceptance rate of 44% for unidimensional settings (Roberts et al., 1997).

The t-distributed proposal densities might lead to invalid proposals for ρ. Consequently, in our MCMC sampler ρ is reparameterized using the general Fisher's z-transformation (Fisher, 1958) such that

$$\rho = \frac{\exp(\rho^*) - 1}{\exp(\rho^*) + 1} \quad \text{and} \quad \rho^* = \log\left(\frac{1 + \rho}{1 - \rho}\right).$$

We use the change of variable theorem to correct the full conditional of ρ for this reparameterization (see theorem B.1 on supplement B.1 of Fahrmeir et al., 2013).

4.3 | Sampler

With a focus in the one covariate case, the final MCMC scheme is outlined in Algorithm 1. It follows similarly for the multiple covariates case.

Algorithm 1. Outline of the MCMC sampler for the one covariate case

Data: Starting values $\boldsymbol{\vartheta}_0$. Number of iterations T.
t = 0
while $t<T$ **do**
  Generate move from $p(\mu_{\boldsymbol{z}}|\cdot)$
  Perform MH step for $\boldsymbol{\theta}_{\boldsymbol{z}}$
  (In reformulation II, generate move from $p(\boldsymbol{\gamma}_{\boldsymbol{z}}|\cdot)$
  Generate move from $p(\sigma_\epsilon^2|\cdot)$
  Generate move from $p((\beta_0, \boldsymbol{\beta}^T)^T|\cdot)$
  Perform MH step for $\{\boldsymbol{\theta}_{\boldsymbol{\gamma}}, \rho\}$
  Generate move from $p(\boldsymbol{\gamma}|\cdot)$
  t += 1
end

5 | SIMULATION STUDY

In the introduction to this article, we positioned our methodology close to the goal of evaluating risk factors in spatial regression models. In this context, we introduced model equation (2), for which spatial confounding is a major concern

since, in general, one will not be able to tell apart spatial variation in the observed risk factors $\boldsymbol{z}_{\text{obs}}(\boldsymbol{s})$ and the remaining spatial variation $\gamma(\boldsymbol{s})$ without further structural model assumptions. Ultimately, we wish to obtain more realistic approximations of the underlying data generating process that enable meaningful interpretation of the coefficients $\boldsymbol{\beta}_{\text{obs}}$ associated with the (observed) risk factors (without delving into causality aspects). Following Sections 2 and 3, we can now evaluate how well the priors we developed perform with respect to this goal.

Thus, the aim of this simulation section is two-fold. First, we want to understand the sources and consequences of spatial confounding in spatial models. Second, we want to investigate to which extent the priors developed in Sections 2 and 3 reduce the bias of regression coefficients when spatial confounding is present. For this, we conduct three simulation studies: (1) a proof of concept simulation where the data generating process exactly matches with our model assumptions, (2) a simulation study in the univariate case where we explicitly construct spatial confounding via an unobserved risk factor, (3) a simulation study in the multiple covariate case where the multiple covariates follow our model assumptions but we use the PCA-based dimension reduction to correct for spatial confounding. These simulation scenarios correspond to Sections 5.1,5.2, and 5.3, respectively.

In the data analysis model (DAM), we compare four different specifications of (2): (i) a nonspatial model (NS), that is, one that excludes $\gamma(\boldsymbol{s})$, (ii) a spatial model with the base prior, that is, not correcting for spatial confounding, (iii) restricted spatial regression (RSR), and (iv) a spatial model with our novel prior structure. We sample $\gamma(\boldsymbol{s})$ for RSR using conditioning kriging (see Supplement D). All $\gamma(\boldsymbol{s})$ are estimated using the SPDE-approach. The two-dimensional domain considered is $S = [0, 1] \times [0, 1]$.

In all studies, we report the bias of regression coefficients and the posterior mean of ρ for simulation study 1 is provided in Supplement E. The results for each scenario are based on $N = 50$ replicates. Inference is based on 10,000 MCMC samples with burn-in of 5000 and thinning of 5 in simulation study 1 and 3, and on 25,000 MCMC with burn-in of 10,000 and thinning of 15 in simulation study 2.

5.1 | Simulation study 1: Proof of concept

In simulation study 1, we provide a proof of concept, by investigating the performance of the MGRF prior when the data-generating model (DGM) and data-analysis model (DAM) are the same. This proof of concept also provides insights on the sources of spatial confounding in DGMs based on GRFs. We assume there is only one observed risk factor $z_{\text{obs}}(\boldsymbol{s})$, and test how well the prior structure in Section 2 can tell apart the spatial variation in $\boldsymbol{z}_{\text{obs}}(\boldsymbol{s})$ and the remaining unobserved spatial variation in the response.

5.1.1 | Data generation model

Consider Equation (2), where we assume $y(\boldsymbol{s})$ is spatially varying with two sources of spatial variability: $\boldsymbol{z}_{\text{obs}}(\boldsymbol{s})$ and $\gamma(\boldsymbol{s})$. We further assume that only one spatial risk factor $\boldsymbol{z}_{\text{obs}}(\boldsymbol{s})$ can be observed and the remaining unobserved spatial variability in $y(\boldsymbol{s})$ - represented $\gamma(\boldsymbol{s})$ - is unobserved. Unobserved and observed spatial risk factors may be correlated, making it hard to tell apart their spatial variability. This has been shown to be problematic when the spatial range of the observed risk factors is larger than the unobserved counterpart, since it leads to spatial confounding bias in β_{obs} (Paciorek, 2010; Page et al., 2017). Consequently, besides their correlation, the spatial range of $\boldsymbol{z}_{\text{obs}}(\boldsymbol{s})$ and $\gamma(\boldsymbol{s})$ is fundamental when trying to tell apart spatial variation in these, in order to achieve a meaningful interpretation of β_{obs}.

The data are generated additively based on (i) an observed spatial risk factor $\boldsymbol{z}_{\text{obs}}(\boldsymbol{s})$, and (ii) unobserved spatial risk factors summarized in $\gamma(\boldsymbol{s})$ such that

$$y(\boldsymbol{s}_i) = -1.5 + 1 \times z_{\text{obs}}(\boldsymbol{s}_i)^T + \gamma(\boldsymbol{s}_i) + \varepsilon_i,$$

where the ε_i's are i.i.d. and follow $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ with $\sigma_\varepsilon^2 = 0.1$. Moreover, $z_{\text{obs}}(\boldsymbol{s})$ and $\gamma(\boldsymbol{s})$ are GRFs generated using the SPDE-approach with mean zero and marginal variance and smoothness equal to one, that is, $\sigma^2_{z_{\text{obs}}} = \sigma^2_\gamma = 1$ and $\nu^2_{z_{\text{obs}}} = \nu^2_\gamma = 1$, respectively. We consider nine different pairs of spatial ranges where $r_\gamma, r_{z_{\text{obs}}} \in \{0.1, 0.5, 0.9\}$. Parameters $z_{\text{obs}}(\boldsymbol{s})$ and $\gamma(\boldsymbol{s})$ are correlated and have correlation parameter ρ (see Equation (3)). We consider five different correlation values, corresponding to five different scenarios: (1) $\rho = 0$; (2) $\rho = 0.3$; (3) $\rho = 0.7$; (4) $\rho = -0.3$; (5) $\rho = -0.7$. Moreover, we consider sample sizes $n \in \{200, 1000\}$.

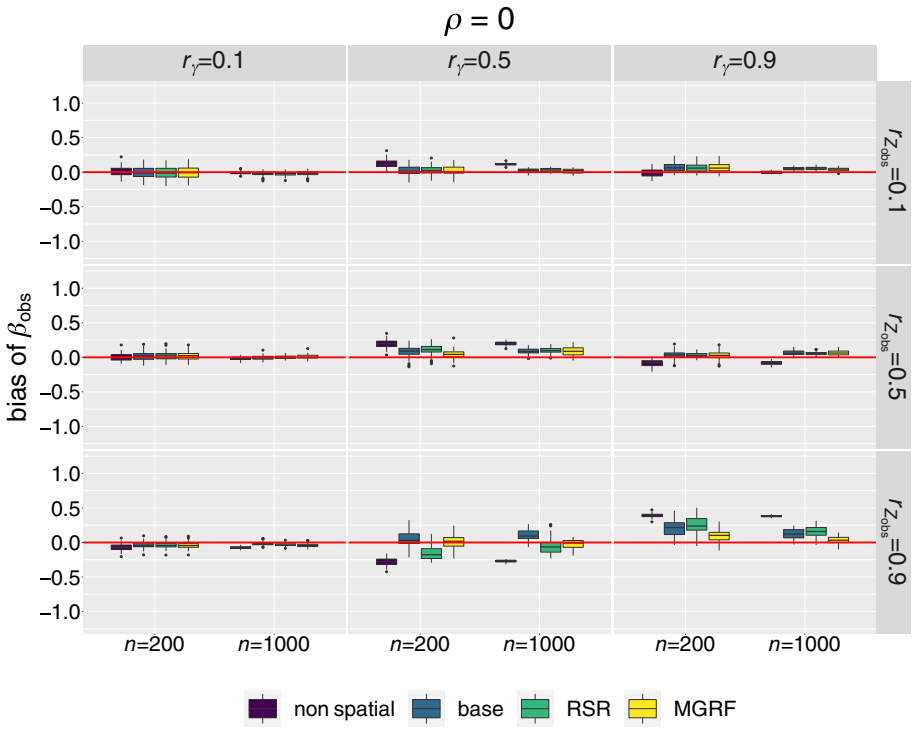

FIGURE 4 Bias of posterior mean of β_{obs} for the four models considered in Scenario 1, and based on 50 replicates. The red line indicates bias of zero

In what follows, we refrain from presenting the results for Scenarios 4 and 5, since the behavior for the negative correlations is identical to that of the positive correlations (just with opposite sign for the bias and, thus, adds no additional information).

5.1.2 | Prior elicitation

We use the priors $\mu_{\boldsymbol{z}_{\text{obs}}} \sim \mathcal{N}(0, 0.1^2)$ and $(\beta_0, \beta_{\text{obs}})^T \sim \mathcal{N}(\mathbf{0}, \text{diag}(100^4, 100^2))$, where β_0 is the model's intercept. The values of U in the shape of the PC-prior (see Equation (10)) vary between scenarios, and represent a trade-off between the accuracy of the posterior mean of β_{obs} over 50 replications and the dispersion of these posterior means. Concretely, we set U high and reduce it in case we get high dispersion for the posterior mean of β_{obs}.

5.1.3 | Results

In Figure 4, our prior prevents overfitting by keeping the results close to those of the base model, except in scenarios where this reduces the bias of β_{obs}. In Figure 5, where $\boldsymbol{z}_{\text{obs}}$ and $\boldsymbol{\gamma}$ are not independent, spatial confounding leads to especially biased regression coefficients when $r_{\boldsymbol{z}_{\text{obs}}} \geq r_{\boldsymbol{\gamma}}$. These results are in line with Paciorek (2010) and Page et al. (2017), and they make sense because if the spatial effect operates on a smaller scale than the covariate, it is likely to explain the data better than the covariate, and thus lead to confounding between the two. When $r_{\boldsymbol{\gamma}} = r_{\boldsymbol{z}_{\text{obs}}}$, as expected, spatial variability cannot be told apart, as both $\boldsymbol{z}_{\text{obs}}$ and $\boldsymbol{\gamma}$ have the same spatial structure. When $r_{\boldsymbol{z}_{\text{obs}}} > r_{\boldsymbol{\gamma}}$, our prior performs considerably better than the competitors. Moreover, there is a tendency to overestimate β_{obs} for $\rho > 0$ (and to underestimate β_{obs} for $\rho < 0$) and the bias increases for higher (absolute) values of ρ. Finally, we can observe a bias/uncertainty trade-off in several range combinations, such that bias reduction might be coupled with increased uncertainty.

All in all, our novel prior either performs better or similarly well to the competitors in terms of the bias of the posterior mean of β_{obs}. All spatial models reach similar values of the continuous ranked probability score (CRPS) and the CRPS for

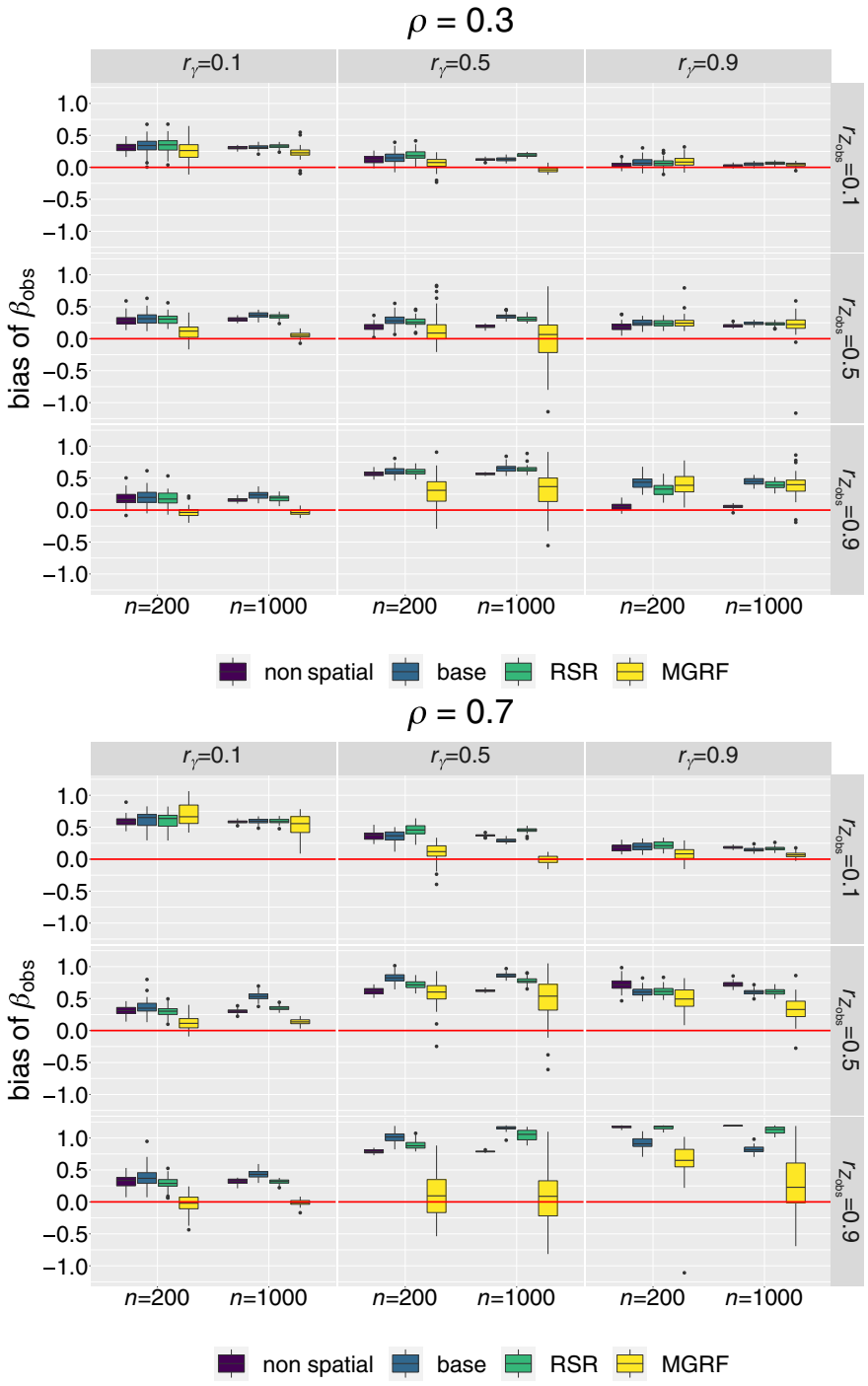

FIGURE 5 Bias of posterior mean of β_{obs} for the four models considered in Scenario 2 and 3, and based on 50 replicates. The red line indicates bias of zero

the NS model is noticeably higher (see Supplement F). A brief report on the efficiency of the MCMC sampler for RSR and our novel prior is provided in Supplement D.

5.2 | Simulation study 2: One single observed risk factor

In simulation study 2, we consider a different DGM in which the unobserved risk factor $\gamma(\boldsymbol{s})$ is composed of a spatially structured variable which affects both the observed risk factor $z_{\text{obs}}(\boldsymbol{s})$ and the response $y(\boldsymbol{s})$ linearly, as well an

additional independent spatial process $\zeta(\boldsymbol{s})$ that represents the remaining spatial dependence structure in the response. This is inspired by the simulation study in Dupont et al. (2021) and follows closely the standard set-up of an omitted variable bias problem. The DAM remains the same as in simulation study 1. An additional scenario that does not include $\zeta(\boldsymbol{s})$ (based on Thaden & Kneib, 2018) is presented in Supplement G.

5.2.1 | Data generation model

The data are generated additively based on (i) an observed spatial risk factor $z_{\text{obs}}(\boldsymbol{s})$, and (ii) unobserved spatial risk factors summarized in $\gamma(\boldsymbol{s})$. The unobserved risk factor affects both the observed risk factor and the response linearly such that

$$\begin{aligned} y(\boldsymbol{s}_i) &= -1.5 + 1 \times z_{\text{obs}}(\boldsymbol{s}_i)^T + \gamma(\boldsymbol{s}_i) + \varepsilon_i, \\ z_{\text{obs}}(\boldsymbol{s}_i) &= 1 \times x(\boldsymbol{s}_i) + \varepsilon_i^z, \\ \gamma(\boldsymbol{s}_i) &= 1 \times x(\boldsymbol{s}_i) + 1 \times \zeta(\boldsymbol{s}_i), \end{aligned}$$

where $x(\boldsymbol{s}_i) = s_{1,i}^2 + s_{2,i}^2$ for $\boldsymbol{s}_i = (s_{1,i}, s_{2,i})^T$ and $\zeta(\boldsymbol{s})$ is a GRF generated using the SPDE-approach with mean zero, marginal variance and smoothness equal to one and range equal to 0.5, that is, $\sigma_\zeta^2 = 1$, $\nu_\zeta^2 = 1$, and $r_\zeta = 0.5$. The variables $x(\boldsymbol{s})$ and $\zeta(\boldsymbol{s})$ are independent. The ε_i^z's and ε_i's are i.i.d. and follow $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ and $\varepsilon_i^z \sim N(0, \sigma_z^2)$ with $\sigma_z^2 = 0.1$, and $\sigma_\varepsilon^2 = 0.5$. The variance parameter σ_z is relatively small such that the model matrix for the spatial model has nearly collinear columns and σ_ε is relatively large to encourage smoothing (Dupont et al., 2021; Thaden & Kneib, 2018).

5.2.2 | Prior elicitation

The choice of priors is identical to Section 5.1.

5.2.3 | Results

The results in Figure 6 show that the MGRF model performs better than all the remaining models. Namely, for $n = 1000$ it is able to reach a bias of zero. This demonstrates the flexibility of the prior structure, beyond the proof of concept in Section 5.1. The results are equally positive for the additional simulation study in Supplement G. All spatial models reach similar CRPS and the CRPS of the NS model is noticeably higher (see Supplement F).

5.3 | Simulation study 3: Multiple risk factors

In simulation study 3, we assume that multiple risk factors are observed and test the prior structure suggested in Section 3 in terms of the bias of the resulting coefficients $\boldsymbol{\beta}_{\text{obs}}$. The DAM follows the model from Section 3 and, consequently, the DGM and DAM are different.

5.3.1 | Data generation model

The data is based on: (1) two observed spatial risk factor $z_1^{\text{obs}}(\boldsymbol{s})$ and $z_2^{\text{obs}}(\boldsymbol{s})$, and (2) an unobserved spatial risk factor $\gamma(\boldsymbol{s})$ such that

$$y(\boldsymbol{s}) = -1.5 + 1 \times z_1^{\text{obs}}(\boldsymbol{s}) - 0.5 \times z_2^{\text{obs}}(\boldsymbol{s}) + \gamma(\boldsymbol{s}) + \varepsilon,$$

where $\sigma_\varepsilon^2 = 0.1$ and $z_1^{\text{obs}}(\boldsymbol{s})$, $z_2^{\text{obs}}(\boldsymbol{s})$, and $\gamma(\boldsymbol{s})$ are GRFs generated using the SPDE-approach with mean zero and marginal variance and smoothness equal to one, that is, $\sigma^2_{z_1^{\text{obs}}} = \sigma^2_{z_2^{\text{obs}}} = \sigma^2_\gamma = 1$ and $\nu^2_{z_1^{\text{obs}}} = \nu^2_{z_2^{\text{obs}}} = \nu^2_\gamma = 1$. When generating the data, we consider nine pairs of spatial ranges where $r_\gamma, r^2_{z_{\text{obs}}} \in \{0.1, 0.5, 0.9\}$ and $r^1_{z_{\text{obs}}} = 0.5$. Covariates $z_1^{\text{obs}}(\boldsymbol{s})$

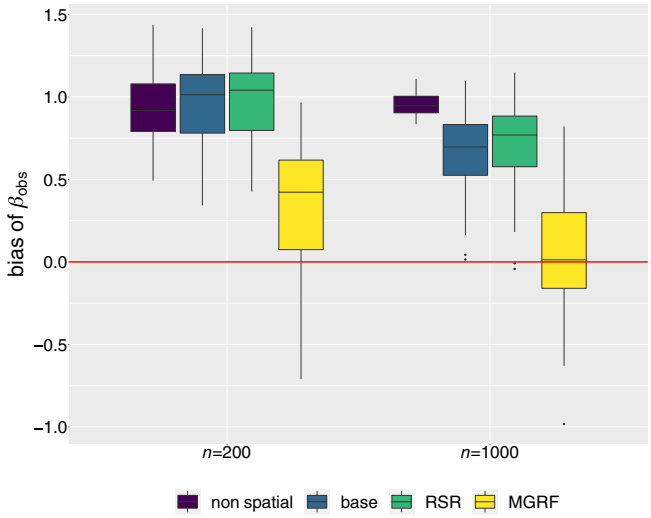

FIGURE 6 Bias of posterior mean of β_{obs} for the four models considered in Scenario 2, and based on 50 replicates. The red line indicates bias of zero

and $z_2^{obs}(\boldsymbol{s})$ are independent, $z_1^{obs}(\boldsymbol{s})$ and $\gamma(\boldsymbol{s})$ are correlated and have correlation parameter ρ_1 (corresponding to ρ in Equation (3)), and $z_2^{obs}(\boldsymbol{s})$ and $\gamma(\boldsymbol{s})$ are correlated and have correlation parameter ρ_2. We consider two scenarios: (1) $(\rho_1, \rho_2) = (0.5, 0.2)$; (2) $(\rho_1, \rho_2) = (-0.5, 0.2)$. Moreover, $n \in \{200, 1000\}$. We use a threshold $\psi = 0.95$ in the PCA (see Equation (11)).

5.3.2 | Prior elicitation

The choice of U follows the same strategy as in Section 5.1. We additionally consider the priors $\mu_{\boldsymbol{z}}^* \sim \mathcal{N}(0, 0.1^2)$ and $(\beta_0, \beta_1^{obs}, \beta_2^{obs})^T \sim \mathcal{N}(\mathbf{0}, \text{diag}(100^4, 100^2, 100^2))$.

5.3.3 | Results

The results are shown in Figures 7 and 8, corresponding to Scenarios 1 and 2, respectively. The benefits of using RSR in reducing the bias of the coefficients associated with (observed) risks factors becomes clearer in the multiple covariates case. In Scenario 1, RSR never performs worse than the NS or the base models. When $r_{\gamma} = 0.9$, RSR outperforms the latter two. Nonetheless, MGRF outperforms the competitors for almost all range combinations. Thus, although MGRF requires more user-defined input compared to RSR, it can lead to larger reductions in bias. While RSR is a good and stable indicator of the direction of the bias, MGRF is more successful in (further) reducing this bias.

6 | APPLICATION TO MONTHLY PRECIPITATION IN GERMANY

In this section, we analyze average monthly precipitation in Germany using open-access data from the German Meteorological Institute (Deutscher Wetterdienst, DWD)[1]. The website also contains detailed information on elevation and climatic variation at the 280 DWD weather stations, which we can use as covariates in our models. We study the month of October 2015 as this example clearly demonstrates the effects our prior can have on the posterior distribution of $\boldsymbol{\beta}_{obs}$.

[1]https://www.dwd.de.

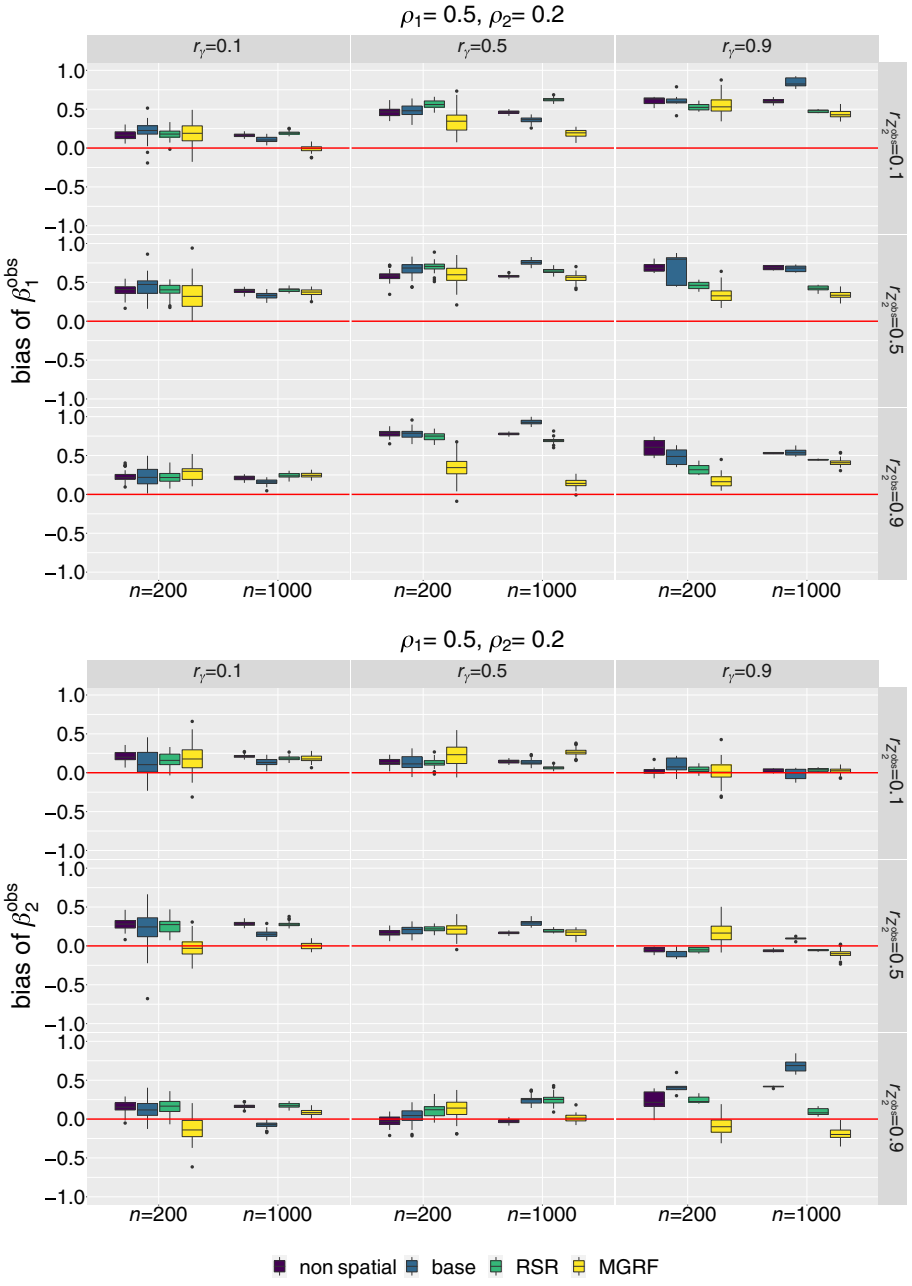

FIGURE 7 Bias of posterior mean of β_1^{obs} (top) and β_2^{obs} (bottom) for the four models considered in Scenario 1, and based on 50 replicates. The red line indicates bias of zero. The bias of β_1^{obs} and β_2^{obs} should be analyzed together, for each spatial range combination

6.1 | Model

We consider model (2), with $y(\boldsymbol{s}_i)$ being the *standardized* amount of precipitation in milliliters at weather station $\boldsymbol{s}_i \in \mathcal{S}$, where $\mathcal{S}$ represents Germany. The covariate $z_1(\boldsymbol{s}_i)$ is the standardized elevation in meters and $z_2(\boldsymbol{s}_i)$ is the standardized average monthly minimum temperature in degrees Celsius. We standardize all variables because the effects of different marginal variances of the GRFs was not analyzed in Section 5.

We use the priors $(\beta_0, \beta_1^{\text{obs}}, \beta_2^{\text{obs}})^T \sim \mathcal{N}(\mathbf{0}, \text{diag}(100^4, 100^2, 100^2))$ and $\mu_{\boldsymbol{z}} \sim \mathcal{N}(0, 1^2)$. We test values of $U \in \{0.3, 0.5, 0.7\}$ and choose the lowest value of U for which the results for the posterior distribution of β_1^{obs} and β_2^{obs} seem to stabilize. The remaining priors are the ones referred to in Section 3.

We test four models with different specifications for $\gamma(\boldsymbol{s})$, as done in the simulation studies from Section 5. Concretely: (i) $\gamma(\boldsymbol{s})$ is excluded from the model, that is, NS, (ii) $\gamma(\boldsymbol{s})$ has a base prior, (iii) $\gamma(\boldsymbol{s})$ has our novel MGRF prior, (iv) $\gamma(\boldsymbol{s})$ is the

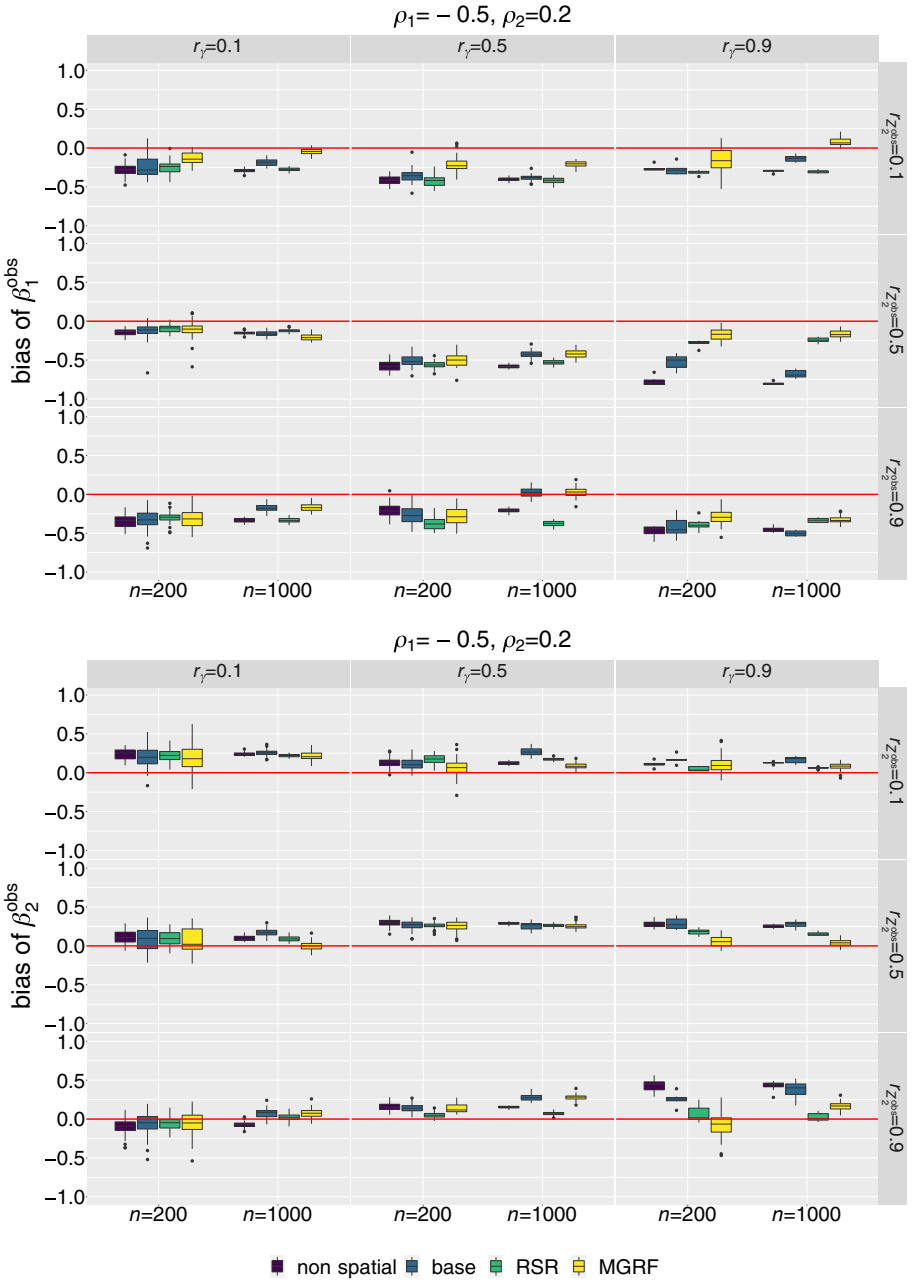

FIGURE 8 Bias of posterior mean of β_1^{obs} (top) and β_2^{obs} (bottom) for the four models considered in Scenario 2, and based on 50 replicates. The red line indicates bias of zero

spatial effect not in the span of the fixed effects, that is, RSR. We use reformulation II for sampling the GRFs. Inference is based on 60,000 MCMC samples with burn-in of 30,000 and thinning of 10.

6.2 | Results

Although linearity is a strong assumption, one would generally expect a positive association between elevation and monthly precipitation ($\beta_1^{\text{obs}} > 0$) and a negative association between minimum temperature and monthly precipitation ($\beta_2^{\text{obs}} < 0$).

Table 1 shows the posterior summaries for β_1^{obs} and β_2^{obs} and it clearly demonstrates the discrepancies in the posterior distribution of β_1^{obs} and β_2^{obs} for different models. For NS, the association between precipitation and elevation

TABLE 1 Mean and equal-tailed 95% credible interval (CI) for the posterior of β_1^{obs} and β_2^{obs} in the five models considered

	β_1^{obs} (elevation)		β_2^{obs} (temperature)	
Model	**Mean**	**95% CI**	**Mean**	**95% CI**
NS	0.024	[−0.024, 0.071]	−0.381	[−0.429, −0.332]
Base model	0.641	[0.493, 0.802]	−0.035	[−0.140, 0.066]
RSR	0.023	[−0.04, 0.088]	−0.170	[−0.228, −0.013]
MGRF	0.165	[0.068, 0.258]	−0.143	[−0.202, −0.086]

is very small and the 95% credible intervals (CI) covers zero. For temperature, we get the expected negative association and the CI does not cover zero. When using the base model, the results are quite different: elevation has the expected positive association and the CI does not cover zero, while temperature has a small effect and the CI covers zero. When using RSR, similarly to NS, we get an association between elevation and precipitation that is small and the CI covers zero, and temperature has the expected association and the CI does not cover zero. The similarity of the behavior of NS and RSR in terms of the estimated β_1^{obs} and β_2^{obs} is in-line with recent literature (Adin et al., 2021; Khan & Calder, 2020).

By using our prior, we go further into the expected direction of the association between the two covariates and monthly precipitation: none of the CIs cover zero and both coefficients show the expected association. As a matter of example, for our prior the interpretation would be that a standard deviation increase in the minimum temperature is associated with, on average, a 0.143 standard deviation decrease in monthly precipitation. The posterior mean of the correlation parameter is −0.266 and the associated 95% CI is [−0.392, −0.147]. Thus, similar to the simulation studies, the RSR seems to pull the coefficients in the direction of the expected association, but the MGRF seems to go further. This indicates that our prior can outperform RSR.

7 | DISCUSSION

In spatial statistics, spatial confounding refers to the phenomenon where spatially varying covariates that model the mean of a response are correlated with the spatial random field that models spatial correlations. We use the Bayesian framework to develop a prior structure that allows to deal with spatial confounding in continuously indexed spatial models. The MGRF prior exploits sparsity in two ways: (i) by using sparse precision matrices of GMRFs within the SPDE-approach, and (ii) by introducing a light parameterization in the case of multiple covariates that accounts for the correlation between a reduced-dimension covariate matrix $\boldsymbol{w}$ and the spatial random effect. Our prior can serve a building block in various types of spatial regression models and, in particular, it can be extended to different response distributions, by using an appropriate link function.

In a simulation study, we investigate the sources of spatial confounding with a particular focus on the spatial range. Future simulation studies should also address the effect of different marginal variances. In both the proof-of-concept and the simulation studies in which DGM and DAM differ, our prior outperforms the competitors in the majority of scenarios and, at worst, it performs equally well.

In an application on monthly precipitation data in Germany, we explore the effect of using our prior on the estimated regression coefficients of average monthly temperature and elevation. This application gives evidence that regression coefficients can be biased if we do not account for spatial confounding and that our prior pulls coefficients toward the expected direction of association.

Future research steps might follow two directions: (i) the model developed is used as a testing tool, or (ii) the model stays in its current format as standalone estimation method for spatial models. In the former, the prior is used to test for the presence of spatial confounding. Then, one could additionally use posterior evidence on the distance to the base model to decide on the presence of spatial confounding.

In future work, since we estimate potentially multiple correlation parameters in the multiple covariates case, one might investigate a hierarchical PC-prior structure that induces more shrinkage. Finally, one should reflect on how to additionally address correlation between spatial covariates.

ACKNOWLEDGMENT

The authors gratefully acknowledge the Deutsche Forschungsgemeinschaft for funding the project within the Research Training Group 2300.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Open Data Server at https://opendata.dwd.de.

REFERENCES

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(4), 433–459.

Adin, A., Goicoa, T., Hodges, J. S., Schnell, P. M., & Ugarte, M. D. (2021). Alleviating confounding in spatio-temporal areal models with an application on crimes against women in India. *Statistical Modelling*, 1471082X211015452.

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

Davis, M. L., Neelon, B., Nietert, P. J., Hunt, K. J., Burgette, L. F., Lawson, A. B., & Egede, L. E. (2019). Addressing geographic confounding through spatial propensity scores: A study of racial disparities in diabetes. *Statistical Methods in Medical Research*, *28*(3), 734–748.

Dupont, E., Wood, S. N., & Augustin, N. (2021). Spatial+: A novel approach to spatial confounding. *Biometrics*.

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

Fisher, R. A. (1958). *Statistical methods for research workers*. *Biological Monographs and Manuals* (13th ed.). Edinburgh and London: Oliver and Boyd.

Gelfand, A. E. (2000). Gibbs sampling. *Journal of the American statistical Association*, *95*(452), 1300–1304.

Greene, W. H. (1990). *Econometric Analysis*. Macmillan.

Guan, Y., Page, G. L., Reich, B. J., Ventrucci, M., & Yang, S. (2020). A spectral adjustment for spatial confounding. *arXiv preprint arXiv:2012.11767*.

Hanks, E. M., Schliep, E. M., Hooten, M. B., & Hoeting, J. A. (2015). Restricted spatial regression in practice: Geostatistical models, confounding, and robustness under model misspecification. *Environmetrics*, *26*(4), 243–254.

Hefley, T. J., Hooten, M. B., Hanks, E. M., Russell, R. E., & Walsh, D. P. (2017). The Bayesian group lasso for confounded spatial data. *Journal of Agricultural, Biological and Environmental Statistics*, *22*(1), 42–59.

Hodges, J. S., & Reich, B. J. (2010). Adding spatially-correlated errors can mess up the fixed effect you love. *The American Statistician*, *64*(4), 325–334.

Hughes, J., & Haran, M. (2013). Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *75*(1), 139–159.

Khan, K., & Calder, C. A. (2020). Restricted spatial regression methods: Implications for inference. *Journal of the American Statistical Association*. *117*(537), 482–494.

Lindgren, F., Bolin, D., & Rue, H. (2022). The SPDE approach for Gaussian and non-Gaussian fields: 10years and still running. *Spatial Statistics*, 100599.

Lindgren, F., Rue, H., & 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.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, *21*(6), 1087–1092.

Paciorek, C. J. (2010). The importance of scale for spatial-confounding bias and precision of spatial regression estimators. *Statistical Science*, *25*(1), 107–125.

Page, G. L., Liu, Y., He, Z., & Sun, D. (2017). Estimation and prediction in the presence of spatial confounding for spatial linear models. *Scandinavian Journal of Statistics*, *44*(3), 780–797.

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *2*(11), 559–572.

Reich, B. J., Hodges, J. S., & Zadnik, V. (2006). Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. *Biometrics*, *62*(4), 1197–1206.

Roberts, G. O., Gelman, A., & Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk metropolis algorithms. *The Annals of Applied Probability*, *7*(1), 110–120. https://doi.org/10.1214/aoap/1034625254

Rue, H., & Held, L. (2005). *Gaussian Markov random fields: Theory and applications*. Chapman & Hall/CRC Press.

Rue, H., & Tjelmeland, H. (2002). Fitting Gaussian Markov random fields to Gaussian fields. *Scandinavian Journal of Statistics*, *29*(1), 31–49. https://doi.org/10.1111/1467-9469.00058

Schnell, P. M., & Papadogeorgou, G. (2020). Mitigating unobserved spatial confounding when estimating the effect of supermarket access on cardiovascular disease deaths. *The Annals of Applied Statistics*, *14*(4), 2069–2095.

Simpson, D., Rue, H., Riebler, A., Martins, T. G., & Sørbye, S. H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, *32*(1), 1–28. https://doi.org/10.1214/16-sts576

Thaden, H., & Kneib, T. (2018). Structural equation models for dealing with spatial confounding. *The American Statistician*, *72*(3), 239–252.

Vihola, M. (2012). Robust adaptive Metropolis algorithm with coerced acceptance rate. *Statistics and Computing*, *22*(5), 997–1008.

Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage Learning.
Zimmerman, D. L., & Ver Hoef, J. M. (2021). On deconfounding spatial confounding in linear models. *The American Statistician*, 1–9.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Marques, I., Kneib, T., & Klein, N. (2022). Mitigating spatial confounding by explicitly correlating Gaussian random fields. *Environmetrics*, *33*(5), e2727. https://doi.org/10.1002/env.2727