Alleviating Spatial Confounding for Areal Data Problems by Displacing the Geographical Centroids

Marcos Oliveira Prates[*], Renato Martins Assunção[†], and Erica Castilho Rodrigues[‡]

Abstract. Spatial confounding between the spatial random effects and fixed effects covariates has been recently discovered and showed that it may bring misleading interpretation to the model results. Techniques to alleviate this problem are based on decomposing the spatial random effect and fitting a restricted spatial regression. In this paper, we propose a different approach: a transformation of the geographic space to ensure that the unobserved spatial random effect added to the regression is orthogonal to the fixed effects covariates. Our approach, named SPOCK, has the additional benefit of providing a fast and simple computational method to estimate the parameters. Also, it does not constrain the distribution class assumed for the spatial error term. A simulation study and real data analyses are presented to better understand the advantages of the new method in comparison with the existing ones.

Keywords: Areal Data, Bayesian Statistics, Spatial Confounding, Spatial Regression.

1 Introduction

Spatial generalized linear mixed models (SGLMM) for areal data analysis have become a common tool for data analysis in recent years with the availability of spatially referenced data sets. Besag et al. (1991) introduced a hierarchical modeling adding random spatial effects to a generalized linear regression model. The spatial dependence is captured by a latent Gaussian Markov Random Field (GMRF). One important advantage of this GMRF approach is to induce a sparse precision matrix that allows for intuitive conditional interpretation and fast Bayesian computation (Rue and Held, 2005). In the past 20 years, Besag et al. (1991) model (hereafter ICAR) has become the most popular areal model for its flexibility and due to its implementation in the `WinBUGS` software (Lunn et al., 2000), which is freely available.

Most spatial data are observational rather than experimental and we commonly observe correlation or multicollinearity between the covariates, also called explanatory or fixed-effect variables. As a consequence of this multicollinearity, in non-spatial regression problems, the estimated coefficients are affected by the presence of the other covariates. This is called confounding and it can lead to implausible estimates. It also affects the

[*]Departmento de Estatística, UFMG, Belo Horizonte, Brasil, marcosop@est.ufmg.br
[†]Departmento de Ciência da Computação, UFMG, Belo Horizonte, Brasil, assuncao@dcc.ufmg.br
[‡]Departmento de Estatística, UFOP, Ouro Preto, Brasil, ericacastirodrigues@gmail.com

variance of the covariates' coefficients estimators which are inflated with respect to what one would have in case the covariates were orthogonal to each other.

In the spatial context, Clayton et al. (1993) and Reich et al. (2006) identified the existence of confounding between the fixed and random effects in SGLMM. In their work, Reich et al. (2006) show that explanatory variables having a spatial pattern may be confounded with the spatial random effects, resulting in fixed effects estimates that are counterintuitive. Thus, they propose an alternative model, called hereafter RHZ model, to alleviate this confounding problem. The RHZ model projects the spatial effects into the orthogonal space spanned by the explanatory variables.

Another well known shortcoming of SGLMM is the computational burden when dealing with high dimensional latent effects. Recently, Hughes and Haran (2013) introduce an alternative model (hereafter, HH model) that alleviates spatial confounding and, at the same time, requires less computational effort. They also consider an orthogonal projection of the spatial effects in a way that takes into account the explanatory variables and the spatial structure. Properties of the HH model was further studied by Murakami and Griffith (2015).

The main idea behind the Reich et al. (2006) and Hughes and Haran (2013) methods is to project the unobserved spatial effects vector in the linear space orthogonal to the explanatory variables. As a consequence, they end up with a precision matrix that is far from sparse losing one of the main advantages of the Markov random fields. Another drawback from RHZ and HH methods is that them become computationally expensive when the spatial effects have parameters. Therefore, it is harder to use them with more flexible spatial structures (Besag, 1974; Leroux et al., 1999; Rodrigues and Assunção, 2012).

From a geostatistical perspective, Paciorek (2010) dealt with the effects of confounding between the spatial random effects and possible explanatory variables. He showed that it can lead to bias in the parameter estimation. Hanks et al. (2015) also studied the effects of spatial confounding in the continuous support situation. They proposed a posterior predictive approach to correct Type-S error on the credibility intervals produced by RHZ. Recently, Hefley et al. (2017) proposed a regularized spatial regression that can also deal with spatial confounding and improve the predictive power the model.

In this paper, we adopt a different approach to deal with spatial confounding in lattice data. Rather than removing the explanatory variables effect from the spatial random effects, we alter the spatial pattern of the map by purging the covariates from the geography. We do this by projecting the spatial coordinates of the areas into the orthogonal space of the covariates, producing a new set of geographical coordinates. These new coordinates induce a different neighborhood structure for the areas that are used to define a different precision matrix. The transformed geography retains the spatial neighborhood that is orthogonal to the covariates. One important consequence of our approach is that we maintain the sparsity of the spatial effects precision matrix allowing for very fast Bayesian computation while controlling for spatial confounding.

Our new approach is called SPOCK, an acronym for *SPatial Orthogonal Centroid "K"orrection*. SPOCK led us to understand better the role of spatial confounding and

its effect on parameter estimation. It gives a more clear understanding of the conceptual differences between the parameters in models that do and do not alleviate spatial confounding. In a striking example, Reich et al. (2006) showed that spatial confounding may provide counter intuitive results in some situations. Sign and relevance of covariates can change drastically after one adds a spatial random effect in a lattice-data regression model. Some important questions arise. When does that occur? Is it always necessary to correct for spatial confounding? If not, when to do so? There is an on-going discussion among the researchers about the need for spatial confounding correction and the meaning of the resulting fixed-effect parameters (Paciorek, 2010; Hanks et al., 2015; Hefley et al., 2017). With our model as a framework, in the simulation study of Section 5, we revisit these questions and try to better understand what are the consequences of correcting for spatial confounding and when it is adequate to do so.

The contributions of this paper are the following:

- a new conceptual approach, SPOCK, to handle spatial confounding;

- because our approach retains the Markovian property, it is extremely efficient for Bayesian computation even in high dimension problems generated by maps with very large number of areas;

- in contrast with the present alternatives, SPOCK deals easily with other spatial models for the random effects;

- SPOCK is very simple to implement and it can be run in any Bayesian spatial software such as `WinBUGS`, `CARBayes` (Lee, 2013) and `R-INLA` (Rue et al., 2009).

The paper proceeds as follows. In Section 2 we review the traditional SGLMM model and present a summary of the RHZ and HH methods. Section 3 introduces SPOCK and its properties are discussed. Section 4 proposes a diagnostic tool to test the need to correct for spatial confounding. In Section 5, a simulation study is performed to compare the proposed method against RHZ and HH methods in terms of precision of estimates and time. It also provides insights on when to correct for spatial confounding and what are the consequences of it. In Section 6.1, we revisit three classical real spatial datasets, the Slovenia dataset (Zadnik and Reich, 2006), the North Carolina SIDS (Cressie, 1991), and the Scotland lips (Breslow and Clayton, 1993). These examples illustrate that all three models lead to very similar estimates but SPOCK is one or two orders of magnitude faster than the other methods. Finally, the paper concludes with a discussion in Section 7.

2 Existing methods

2.1 SGLMM

Spatial Generalized Linear Mixed Models (SGLMM) is a wide class of models that accommodates spatial dependence through a random effect term. Let Y_i be the observation of an area i, $i = 1, \ldots, n$ with distribution given by $Y_i \sim \pi(y|\mu_i, \boldsymbol{\delta}, \boldsymbol{X}_i, \beta)$ with

$g(\mu_i) = \boldsymbol{X}_i\boldsymbol{\beta} + \theta_i$, where $\boldsymbol{\beta}$ is the fixed effects coefficients vector and $\boldsymbol{X}$ is a $n \times q$ full-rank design matrix with the covariates. Typically, $\boldsymbol{X}$ includes a first column of constant values equal to 1. We let $\boldsymbol{X}_i$ be its i-th row, g is an appropriate link function, and $\mu_i = E(Y_i|\boldsymbol{X}_i)$. The vector of hyperparameters of the distribution is $\boldsymbol{\delta}$. Finally, $\boldsymbol{\theta}$ represents a vector with spatially structured random variables capturing the spatial patterns shared by the areas in study.

The most simple instance of this model is the Gaussian model where

$$Y_i|\mu_i, \tau_\epsilon \sim N(\mu_i, \tau_\epsilon) \quad \text{with} \quad \mu_i = \boldsymbol{X}_i\boldsymbol{\beta} + \theta_i. \tag{1}$$

Traditionally the spatial random effect $\boldsymbol{\theta}^\top = (\theta_1, \ldots, \theta_n)$ is defined as an intrinsic conditional autoregressive (ICAR) models introduced by Besag et al. (1991). The prior specification of the ICAR model is given by

$$\pi(\boldsymbol{\theta}|\tau_\theta) \propto \tau_\theta^{n-1} \exp\left(-\frac{\tau_\theta}{2}\boldsymbol{\theta}^\top\boldsymbol{Q}\boldsymbol{\theta}\right), \tag{2}$$

where $\boldsymbol{Q}$ is the precision matrix given by $\boldsymbol{Q} = \boldsymbol{D} - \boldsymbol{A}$ where $\boldsymbol{D}$ a diagonal matrix with the i-th entry equal to the number of neighbors of region i and $\boldsymbol{A}$ is the graph zero/one adjacency matrix. The parameter τ_θ is the spatial precision.

2.2 Non-Confounding SGLMM

Clayton et al. (1993) introduced the concept of spatial confounding between the fixed effects estimates and the spatial random effects in SGLMM. However, only recently, Reich et al. (2006) revisited the problem motivated by a striking case study where the credible interval for a certain fixed effect coefficient changes drastically after introducing the spatial random effects. They proposed an alternative method to alleviate this confounding problem. The idea proposed is to include in the model a random effect that belongs to the orthogonal space of the fixed effects predictors. A new $\mathbb{R}^n$ basis can be used to re-express

$$\boldsymbol{\theta} = \boldsymbol{\theta}^X + \boldsymbol{\theta}^\perp = \boldsymbol{K}\boldsymbol{\theta}_1 + \boldsymbol{L}\boldsymbol{\theta}_2, \tag{3}$$

where $\boldsymbol{K}$ is a $n \times q$ matrix that has the same span as $\boldsymbol{X}$, $\boldsymbol{L}$ is a $n \times (n-q)$ matrix whose columns lie in the orthogonal space of $\boldsymbol{X}$, and $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are vectors with dimensions q and $n-q$, respectively. Equation (1) can be represented as $Y_i|\mu_i, \tau_\epsilon \sim N(\mu_i, \tau_\epsilon)$ with $\mu_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{K}_i\boldsymbol{\theta}_1 + \boldsymbol{L}_i\boldsymbol{\theta}_2$, where $\boldsymbol{K}_i$ and $\boldsymbol{L}_i$ are the i-th rows of $\boldsymbol{K}$ and $\boldsymbol{L}$, respectively. The vector $\boldsymbol{\theta}$ follows the ICAR distribution in (2). Using this parametrization, it was shown that $\boldsymbol{K}$ causes a confounding in the estimates of $\boldsymbol{\beta}$. Reich et al. (2006) suggested the removal of the $\boldsymbol{K}$ component leading to the RHZ model:

$$Y_i|\mu_i, \tau_\epsilon \sim N(\mu_i, \tau_\epsilon) \text{ with } \mu_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{L}_i\boldsymbol{\theta}_2, \quad \boldsymbol{\theta}_2 \sim N_{n-q}\left(0, \boldsymbol{L}^\top\boldsymbol{Q}\boldsymbol{L}\right). \tag{4}$$

Although it alleviates the confounding between the spatial random effects and the estimates of the fixed effects, Hughes and Haran (2013) noticed that this method is computationally inefficient. The reason is that the new precision matrix generated by

(4) is not sparse and has dimension $n - q \approx n$. To reduce the computational demand of the RHZ model they proposed a low-rank alternative model for the $\boldsymbol{L}$ matrix. This new model uses the so-called Moran operator $\boldsymbol{P}^{\perp}\boldsymbol{A}\boldsymbol{P}^{\perp}$ and

$$\boldsymbol{P}^{\perp} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^{\top}\boldsymbol{X})^{-1}\boldsymbol{X}^{\top} \tag{5}$$

is the projection matrix into the orthogonal space of the span of $\boldsymbol{X}$. They showed that the Moran operator retains the spatial patterns of the data and it is only necessary to select the $h \ll n$ higher positive eigenvalues of the spectrum of the Moran operator, **called attractive eigenvalues**. In this way, they were able to reduce the dimension of the random effect by using a low-rank approximation of $\boldsymbol{P}$ maintaining the spatial information necessary in model estimation. The HH model is defined as $Y_i | \mu_i, \tau_\epsilon \sim N(\mu_i, \tau_\epsilon \boldsymbol{I})$ with $\mu_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{M}_i\boldsymbol{\theta}_2$ and $\boldsymbol{\theta}_2 \sim N_h(0, \boldsymbol{M}^{\top}\boldsymbol{Q}\boldsymbol{M})$, where $\boldsymbol{M}$ contains the first h eingenvectors of the Moran operator.

3 Purging the covariates from the geography

Although Reich et al. (2006) and Hughes and Haran (2013) were successful in alleviating the confounding and in reducing the dimension of the spatial random effects as presented in Section 2.2, their approaches have two main drawbacks: (1) the RHZ and HH models lost the original SGLMM model sparsity of the precision matrix and hence do not take advantage of the Markov property in the Bayesian calculations; (2) compounding with (1), their computational cost increases sharply if parameters appear in the precision matrix $\boldsymbol{Q}$ (e.g., Besag, 1974; Leroux et al., 1999; Rodrigues and Assunção, 2012). In this case, one needs to recalculate eigenvalues and eigenvectors of $\boldsymbol{Q}$ each time its parameters are updated.

In order to propose a fast SGLMM that alleviates the confounding problem, we introduce SPOCK, a novel approach to the problem capable of maintaining the Markov properties of the precision matrix as well as allowing for unknown parameters in the matrix $\boldsymbol{Q}$. SPOCK not only produces a fast alternative to RHZ and HH but, more importantly, it also represents a new and rather different conceptual perspective on how to deal with the problem. We will show that this new way of seeing the spatial confounding can help in clarifying the discussion about the need for and the consequences of the confounding alleviation.

Instead of reparametrizing the random effects vector by decomposing it into two orthogonal subspaces, we project the neighborhood graph vertices into the orthogonal space of the covariate matrix $\boldsymbol{X}$. In geostatistics, Sampson and Guttorp (1992) also suggested a deformation of the space to transform a non stationary field into stationary one. Our method requires a coordinate system representing each area by a centroid point and then we transform these geographical centroids inducing a new neighborhood structure that is not influenced by the covariates.

SPOCK depends crucially on the fact that the distance between the centroids represents approximately the graph structure. For example, in the usual maps of counties in the US and Brazilian states, this assumption is satisfied, as we will show in Table 1. However, it is possible to imagine situations in which this assumption is not true. Consider
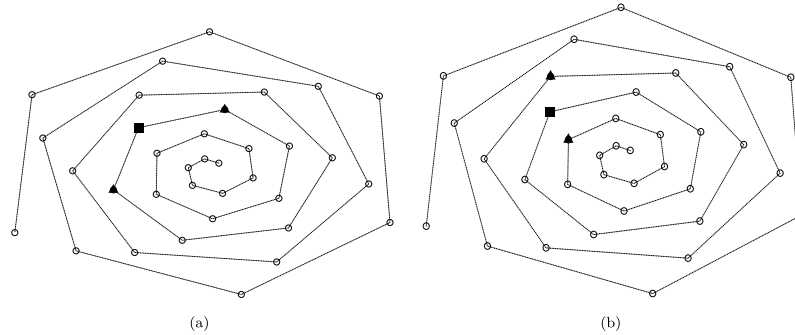
Figure 1: (a) Original centroids of a pseudo-map where the adjacency neighborhood structure is represented by a spiral. One area is highlighted with a square symbol as well as its two neighbors with respect to this graph represented by triangles. (b) The same graph with the square area is shown with its two nearest neighbors (as triangles) with respect to the Euclidian distance between the centroids.

a map in which the areas are arranged in such a way that the neighborhood structure is represented by a spiral graph, as in Figure 1(a). We mark an arbitrary area in this map with a square. Its neighbors in the spiral neighborhood structure are marked with triangles. In Figure 1(b), we show the same graph with the same square reference area but its two neighboring triangles are now those two nearest centroids with respect to the Euclidean distance between the centroids. Comparing the two graphs, we see that the nearest neighbors in the spiral neighborhood structure do not coincide with the nearest neighbors in the structure induced by the distance between the centroids. This type of situation is adverse for SPOCK. However, in usual geographical maps, it is large the correspondence between nearest neighbors by the two criteria, as we shall see in Table 1.

Spatial effects defined on this transformed geography will attenuate the confounding with the predictors. The projected image of the original graph (hereafter, projected graph) allows us to keep the sparsity of the precision matrix Q and the Markov properties of random effects, making the fitting of our model more efficient than RHZ and HH (Rue and Held, 2005). Since SPOCK works only by redefining the non-zero pattern of Q by means of the projected graph, it allows the user to adjust a variety of parameter-based spatial structure such as Besag (1974); Besag et al. (1991); Leroux et al. (1999); Rodrigues and Assunção (2012).

To motivate our method, suppose that we can partition the neighborhood graph as we did with the random effects in Section 2.2. That is, suppose that the neighborhood graph could also be partitioned into two parts, one related with the covariates and another one orthogonal to them. It is not clear how to directly partition the graph in this way. When the distance between the areas centroids reflects the neighborhood structure of the graph, a projection of the numerical coordinates of these centroids can produce a partition of the graph into the two desired components: one correlated with X and another one uncorrelated.

In a more formal way to provide intuition about spatial confounding and about our method, we interpolate our lattice data with a smooth surface. Let $\boldsymbol{s} = [\boldsymbol{s}_1, \boldsymbol{s}_2]$ be the $n \times 2$ matrix with the areas' centroids coordinates. We associate each entry of the random effect $\boldsymbol{\theta}$ with a specific location, its area centroid. Conceptually, interpolate the θ_i values with a smooth surface $\psi(s_1, s_2)$ defined for any arbitrary position (s_1, s_2) in the continuous map region such that $\boldsymbol{\theta} = (\psi_1, \ldots, \psi_n)$ with $\psi_i = \psi(s_{i1}, s_{i2})$.

We want to emphasize the purely mathematical nature of this interpolating function. We are aware that statistical lattice data are not continuous and that each θ_i is associated with the entire i-th region rather than with one single location. We do not attach a substantive interpretation for such function $\psi(s_1, s_2)$. We are not suggesting any kind of asymptotic process, where the areas shrink in size until they become a point. Our interpolating function is simply a convenient mathematical abstraction useful to explain our method and it does not play any role in our method itself.

Let (s_{10}, s_{20}) be a reference location such as a central position in the map and $\nabla\psi = (\gamma_1, \gamma_2)'$ be the gradient of ψ evaluated at (s_{10}, s_{20}). When ψ is smooth enough, a Taylor expansion allows the value $\psi(s_1, s_2)$ in an arbitrary map position (s_1, s_2) to be written as

$$
\begin{aligned}
\psi(s_1, s_2) &= \psi(s_{10}, s_{20}) + (s_1 - s_{10}, s_2 - s_{20}) \nabla\psi + R(s_1, s_2, s_{10}, s_{20}) \\
&= \gamma_0 + \gamma_1(s_1 - s_{10}) + \gamma_2(s_2 - s_{20}) + R(s_1, s_2, s_{10}, s_{20}) .
\end{aligned}
\tag{6}
$$

where $R(s_1, s_2, s_{10}, s_{20})/||\boldsymbol{h}||$ goes to zero as $\boldsymbol{h} = (s_1 - s_{10}, s_2 - s_{20})' \to \boldsymbol{0}$. The remainder $R(s_1, s_2, s_{10}, s_{20})$ is a quadratic form given by $\boldsymbol{h}'\boldsymbol{H}(\boldsymbol{r})\boldsymbol{h}$ where $\boldsymbol{H}(\boldsymbol{r})$ is the ψ Hessian matrix evaluated at a point $\boldsymbol{r}$ that lies somewhere between (s_1, s_2) and (s_{10}, s_{20}). The value of $R(s_1, s_2, s_{10}, s_{20})$ is bounded by the largest eigenvalue modulus of the Hessian matrix $\boldsymbol{H}(\boldsymbol{r})$ as we vary (s_1, s_2) (and hence, $\boldsymbol{r}$). For smooth surfaces $\psi(s_1, s_2)$, the gradient $\nabla\psi$ does not change abruptly and hence the Hessian is not expected to vary substantially and should be small. We stress that these assumptions are not required to apply our method. Their role is only to provide a motivation to understand the rationale behind SPOCK.

Evaluating the expression (6) at the centroids $\boldsymbol{s}$ and organizing the result as a vector, we have

$$
\begin{aligned}
\boldsymbol{\theta} = \psi(\boldsymbol{s}) &= \gamma_0 \boldsymbol{1} + \gamma_1(\boldsymbol{s}_1 - s_{10}\boldsymbol{1}) + \gamma_2(\boldsymbol{s}_2 - s_{20}\boldsymbol{1}) + \boldsymbol{R}(\boldsymbol{s}_1, \boldsymbol{s}_2, s_{10}, s_{20}) \\
&= (\gamma_0 - s_{10} - s_{20})\boldsymbol{1} + \gamma_1 \boldsymbol{s}_1 + \gamma_2 \boldsymbol{s}_2 + \boldsymbol{R}(\boldsymbol{s}_1, \boldsymbol{s}_2, s_{10}, s_{20}) \\
&= [\boldsymbol{1}, \boldsymbol{s}_1, \boldsymbol{s}_2]\boldsymbol{\gamma} + \boldsymbol{R} .
\end{aligned}
$$

In this way, we expressed the spatial random effects with a linear combination of the coordinates $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ based on the assumption that $\boldsymbol{\theta}$ can be cast on a smooth surface. Returning to the SGLMM model, we can rewrite (1) as

$$
\begin{aligned}
\boldsymbol{Y} &= \boldsymbol{X}\boldsymbol{\beta} + [\boldsymbol{1}, \boldsymbol{s}_1, \boldsymbol{s}_2]\boldsymbol{\gamma} + \boldsymbol{R} \\
&= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{P}[\boldsymbol{1}, \boldsymbol{s}_1, \boldsymbol{s}_2]\boldsymbol{\gamma} + \boldsymbol{P}^{\perp}[\boldsymbol{1}, \boldsymbol{s}_1, \boldsymbol{s}_2]\boldsymbol{\gamma} + \boldsymbol{R} \\
&= \boldsymbol{X}\left(\boldsymbol{\beta} + \left(\boldsymbol{X}^{\top}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\top}[\boldsymbol{1}, \boldsymbol{s}_1, \boldsymbol{s}_2]\boldsymbol{\gamma}\right) + \boldsymbol{P}^{\perp}[\boldsymbol{1}, \boldsymbol{s}_1, \boldsymbol{s}_2]\boldsymbol{\gamma} + \boldsymbol{R}
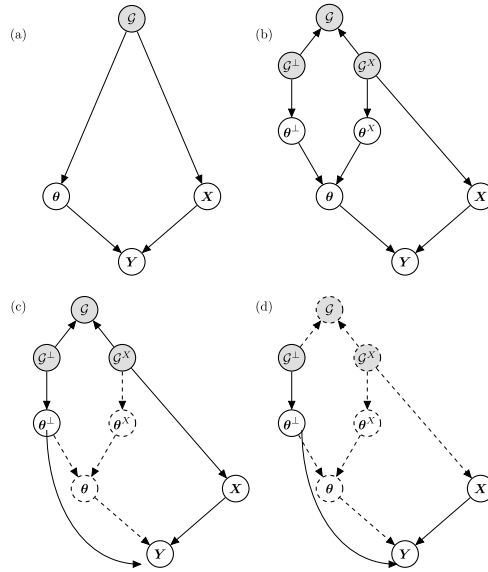\end{aligned}
$$

Figure 2: Graphical model representation of the models: (a) usual ICAR model; (b) ICAR with some nodes decomposed; (c) RHZ and HH models; (d) SPOCK model.

$$= \quad \boldsymbol{X}\boldsymbol{\beta}^* + \boldsymbol{P}^\perp[\mathbf{1}, \boldsymbol{s}_1, \boldsymbol{s}_2]\boldsymbol{\gamma} + \boldsymbol{R}\,, \tag{7}$$

where we used that $\boldsymbol{I} = \boldsymbol{P} + \boldsymbol{P}^\perp$ with $\boldsymbol{P}^\perp$ defined in (5). One of the main advantages of expression (7) is to clearly and intuitively answer to Hodges and Reich (2011) when they ask how "adding spatially correlated errors can mess up the fixed effect you love". The $\boldsymbol{X}$ fixed effect is messed up by the spatial random effect $\boldsymbol{\theta}$ when two conditions are met: (a) the covariates in $\boldsymbol{X}$ have a linear association with $\boldsymbol{s}$ so $\boldsymbol{P}[\mathbf{1}, \boldsymbol{s}_1, \boldsymbol{s}_2]$ is not close to the zero matrix and (b) $\boldsymbol{\gamma}$ is not small. Under these two conditions, the difference between $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^*$ will be large. This motivates our diagnostic tool proposed in Section 4 to verify the need to correct for spatial confounding.

Expression (7) also justifies SPOCK as a method to deal with spatial confounding in spatial regression. We split the linear component of the spatial random effect $\boldsymbol{\theta}$ into two pieces and remove the component $\boldsymbol{P}[\mathbf{1}, \boldsymbol{s}_1, \boldsymbol{s}_2]\boldsymbol{\gamma}$, which can be confounded with $\boldsymbol{X}$. In Figure 2, we show a graphical model representation that explains SPOCK and how it differs from the RHZ and HH concepts. In (a), we have the usual ICAR model. The node $\mathcal{G}$ represents the neighborhood graph, which affects both, the covariates $\boldsymbol{X}$ and the spatial effects $\boldsymbol{\theta}$. The response node $\boldsymbol{Y}$ is formed according to (1). In (b), we have a new representation of the same model in (a) but now we decompose some of the nodes to contrast the different approaches. The node $\boldsymbol{\theta}$ is split into two pieces, $\boldsymbol{\theta}^X$ and $\boldsymbol{\theta}^\perp$, following the RHZ solution as in (3). We also introduce a conceptual decomposition of the graph $\mathcal{G}$ into two pieces. The first one is $\mathcal{G}^X$, and it carries the shared information between the graph and $\boldsymbol{X}$. The second one is $\mathcal{G}^\perp$, and it contains the orthogonal information in $\mathcal{G}$, after extracting $\mathcal{G}^X$. We explain in the next paragraph
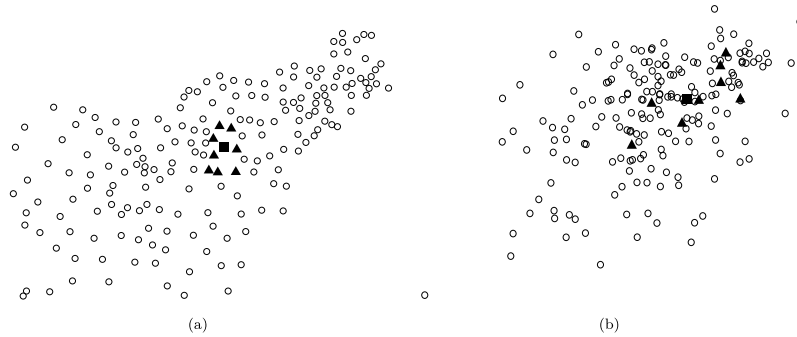
Figure 3: (a) Original centroids of the Slovenia. (b) New centroids after projecting the original centroids coordinates into the vector space orthogonal to the span of $\boldsymbol{X}$.

what we mean by this decomposition of the neighborhood graph and how it is carried out in practice. Node $\mathcal{G}^{\perp}$ links directly only to the $\boldsymbol{\theta}^{\perp}$ component of the spatial random effect, while $\mathcal{G}^{X}$ links only to $\boldsymbol{\theta}^{X}$. The RHZ and HH models are represented in (c) and they are obtained by removing the node $\boldsymbol{\theta}^{X}$ and keeping only the $\boldsymbol{\theta}^{\perp}$, as shown in (4). The nodes and edges removed are shown with dashed lines in the figure. Our SPOCK approach is represented in (d). Differently from RHZ and HH, our geography transformation means the removal of $\mathcal{G}^{X}$ and its children. One important additional difference is that $\boldsymbol{\theta}^{\perp}$ in our model is not the same as that defined by RHZ and HH. In our case, $\mathcal{G}^{\perp}$ directly induces a new $\boldsymbol{\theta}^{\perp}$ with a sparse precision matrix $\boldsymbol{Q}^{\star}$.

To explain how we construct $\mathcal{G}^{\perp}$, we use the same spatial dataset from Slovenia that motivated Reich et al. (2006) in their work. Figure 3(a) shows the Slovenia map represented by its centroids (grey dots). The single predictor variable is a social-economic status measure (SES). This predictor has a strong spatial pattern, with a gradient crossing from the SouthWest to the NorthEast in the map. Our new geography is given by the projected centroids $\boldsymbol{s}^{\star} = \boldsymbol{P}^{\perp}\boldsymbol{s}$ shown in Figure 3(b). These new set of coordinates represent the node with the projected graph $\mathcal{G}^{\perp}$ in Figure 2. If there is no spatial confounding, we expect $\boldsymbol{s}^{\star}$ and $\boldsymbol{s}$ to be approximately the same. In the original map in the left hand side, we selected one area (black square) and its original neighbors (black triangles) to show where they are located in the new projected map. We can see that the neighboring areas in the original map are separated out and may become isolated from one another in this new geography.

After projecting the original coordinates with $\boldsymbol{P}^{\perp}$ to obtain the new spatial coordinates $\boldsymbol{s}^{\star}$, we need to specify the neighborhood structure. To define the edges of the $\mathcal{G}^{\perp}$ neighborhood structure, we consider two alternatives: 1) to use the number of neighbors k_i of each area in the original graph (Knn); 2) to use the Delaunay triangulation to automatically define the number of neighbors. In the first method, we add edges between areas i and j if area j is among the k_i nearest neighbors of area i in the new geography. In the second method, a Delaunay triangulation is performed over the new centroids and the vertices connected by the triangles' edges are considered neighbors.

	avg. Sensitivity		avg. Recall	
	Delaunay	Knn	Delaunay	Knn
48 U.S. state maps	91.9	96.1	80.7	85.3
26 Brazilian state maps	91.9	96.7	80.5	86.6

Table 1: Percentage agreement in the zero-one pattern between the original graph precision matrix and the projected precision matrices.

To choose between these two methods, we analyzed their ability to reproduce the true and original map adjacency structure when there is no spatial confounding. After all, when X is not associated with θ, we expect to see only small changes in the spatial relationship between the areas. Table 1 shows the result of using the two methods in two collections of maps totaling 76 widely different geographical structures. The first set is composed by the 48 maps of counties from each one of the US states but Alaska and Hawaii. The size and shape of the US states varies widely and hence this collection gives a good idea of how the method works in different situations. The second set is composed by 26 maps of counties (indeed, municipalities) of individual Brazilian states. For each one of these 76 maps, we calculate the *sensitivity* (or precision) of each method: the proportion of neighboring pairs of counties in the original state map that is reproduced into either the Delaunay or Knn neighborhood map in the new geography. Table 1 shows the average sensitivity over all 76 maps. We also calculate the *recall* of each method: the proportion of neighboring pairs of counties in the reconstructed adjacency structure that is also a neighboring pair in the original map. For both measures, sensitivity and recall, the higher value, the better. It is clear from the table that the Knn method is preferred according to this criterion and, from now on, all analysis in this manuscript is performed using the Knn method to reconstruct the neighborhood structure after projecting the coordinates.

The new adjacency matrix generated by the Knn method replaces the original adjacency matrix used in the SGLMM model. After that, the user can select his preferred algorithm to fit the model and carry out inference. For example, he can use the INLA (Rue et al., 2009) algorithm in order to take advantages of the Markov properties of the new matrix Q generated by the projected graph and to avoid traditional Markov chain Monte Carlo (MCMC) convergence problems. For this reason we use R-INLA to perform our analysis in this paper but since there is no software restriction to fit our method we also used a SPOCK MCMC version through CARBayes.

To better visualize the impact of different spatial trends in the projected graph we revisit the Slovenia map example in Figure 4(a) where the map is now centered at the origin $(0,0)$ and presents its original neighboring structure. For all areas, we keep the k_i number of neighbors in the original map to reconstruct the Knn neighboring structure of the projected geography. Different spatial trends are created for one explanatory variable X. Figure 4(b) is associated with a single covariate given by $X_i = s_{1i} + s_{2i}$, a

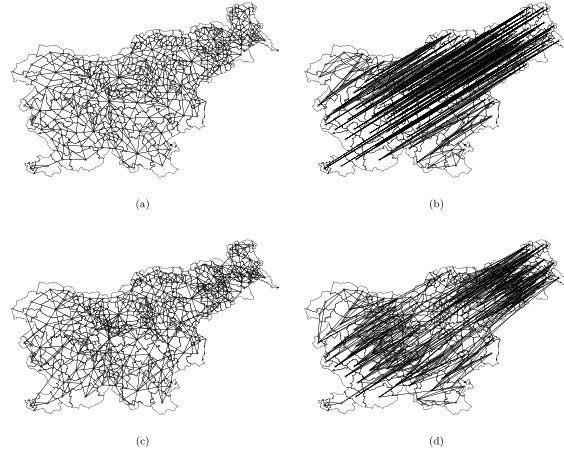Figure 4: (a) Original Slovenia map. (b) Reconstructed adjacency matrix when $\boldsymbol{X}$ is a linear trend of the centroids coordinates. (c) Reconstructed adjacency matrix when $\boldsymbol{X}$ is a quadratic trend of the centroids coordinates (d) Reconstructed adjacency matrix when $\boldsymbol{X}$ is a cubic trend of the centroids coordinates.

linear trend based on the centroids' coordinates. Figure 4(c) has $X_i = s_{1i}^2 + s_{2i}^2$, while Figure 4(d) has $X_i = s_{1i}^3 + s_{2i}^3$.

The original centroids are projected into the space generated by $\boldsymbol{P}^\perp$. We do not show this new geography but rather, in Figure 4, we show the original map with each i-th centroid connected to its new k_i neighbors in the projected space. Although there is some overlap between segments connecting the new neighbors, we can see that each area in Figure 4(b) has all its k_i neighbors approximately in the $(1,1)$ direction. As $X_i = (1,1)(s_{i1}, s_{i2})^\top$, it explains the Y variation in the $(1,1)$ direction and so, for the spatial effect, it remains to explain the Y variation in the orthogonal $(-1,1)$ direction. In the new geography, far away areas (and therefore, with practically independent spatial effects) are those distant in the $(-1,1)$ direction. Hence, the new neighborhood structure must be the k_i nearest neighbors in the $(1,1)$ direction. When $\boldsymbol{X}$ has a quadratic spatial trend, Figure 4(c), there is almost no effect on the projected centroids. This happens because the orthogonal component of $\boldsymbol{s}$ in the span of $\boldsymbol{X}$. is negligible and thus no confounding correction is necessary. This pattern is very similar to the one we obtain when $\boldsymbol{X}$ is randomly generated (without any spatial pattern). Finally, Figure 4(d) presents a cubic pattern for $\boldsymbol{X}$. In this case, although not as perfect as in Figure 4(b), the orthogonal projection of $\boldsymbol{s}$ in the span of $\boldsymbol{X}$ is significant and therefore some correction is needed. From Figure 4, it is clear that different spatial trends generates different effects over the projected graph. This is expected since the projected graph should be orthogonal to the information in the span of $\boldsymbol{X}$.

4   When do we need to correct for spatial confounding?

An additional advantage of the SPOCK approach is to provide a diagnostic tool for detecting the need for spatial confounding alleviation. Hefley et al. (2017) provide a way to measure spatial multicollinearity through model fitting. However, practitioners might prefer to use the traditional SGLMM rather than the set of new methods to deal with spatial confounding, unless it is really necessary. The following diagnostic tool can be used in an exploratory analysis providing guidance to which model is appropriate prior to any model fit. As we saw in Figure 4, when the spatial coordinates have no linear association with the covariates, our methodology has no effect on the geography. This could lead one to think that we should always correct for spatial confounding as an insurance policy. However, there is a clear price in using the methods that alleviate spatial confounding. They are more complex and hence epidemiologists and public health agents have a harder time to interpret them (Hanks et al., 2015). Next, we provide a tool based on canonical correlation suggesting either one should do the confounding correction or not.

To verify the degree of linear association between the centroids s and the covariates X, we apply the canonical correlation technique which measures how much variability the two sets share and if they have some common linear dimensions.

The main idea of the technique is to find two linear combinations $U = sa$ and $V = Xb$, where a and b are 2×1 and $q \times 1$, respectively, such that the correlation between U and V is maximized. The solution leads to a maximum correlation given by ρ, the largest eigenvalue of the matrix $S_{ss}^{-1/2} S_{sx} S_{xx}^{-1} S_{xs} S_{ss}^{-1/2}$, where $S = [S_{ss} \; S_{sx}; S_{xs} \; S_{xx}]$ is the empirical covariance matrix of $[s, X]$. The need for spatial confounding correction is based on a statistical test of the null hypothesis that $\rho = 0$. There are two possible approaches, one based on an asymptotic test and another based on a Monte Carlo test.

The asymptotic test is based on the Wilks' Lambda statistic Λ, which asymptotically has a distribution F (Wilks, 1935). This statistic is a multivariate generalization of the coefficient of determination. The basic idea is to fit a multivariate regression model where the multivariate response variable is s and the covariates are composed by X. The test statistic measures the proportion of the s variability that can not be explained by the predictors in X. If this value is small, we have evidence that there is a linear relationship between the two data sets and a correction for spatial confounding should be considered. Under the normality assumption for the two matrices, s and X, or if we have a large number of observations, we have Λ following approximately an F distribution. In cases where these assumptions are not valid, we carry out a randomization test by randomly permuting the rows of either s or X.

5   Simulation

In this section, we present three simulation studies. The first one aims to give insights on the implications of the model assumptions to alleviate spatial confounding when using SPOCK. It also compares the results obtained by the SPOCK methodology with the competitor methods and the computational time of the available implementations.

The second study compares the effect on the spatial confounding problem obtained by SPOCK and the alternative methods when the spatial coordinates $\boldsymbol{s}$ have a non-linear dependence on the covariates $\boldsymbol{X}$. The third, study the Type-S error rate (Gelman and Tuerlinckx, 2000) and coverage of the proposed methods from an areal data perspective.

5.1 Linear confounding

Using the Slovenia map, we carry out a simulation study with a linear dependence between the covariates and the space with two main goals: 1) to understand and discuss what are the implications of the models assumptions and when it is necessary to use some spatial confounding alleviation method; 2) to compare the results of the SPOCK methodology with the existing RHZ and HH alternative models. The following model is selected to generate the data $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \theta_i + e_i$ with $e_i \overset{ind}{\sim} N(0, \tau_e)$ and $i = 1, \ldots, n$ where $\boldsymbol{\theta}$ is the spatial effect, n is the number of regions in the Slovenia map. The coefficients values are $\boldsymbol{\beta}^\top = (2, 1, -1)$. We selected 3 simulation scenarios:

1. (ICAR spatial-X) The spatial effect is generated using the ICAR structure (that is, $\boldsymbol{\theta} \sim N(0, \tau_\theta \boldsymbol{Q})$) with covariate X_{1i} generated as i.i.d. $N(0, 1)$ and covariate X_{2i} equal to s_{i1}, the first coordinate of i-th area centroid;

2. (RHZ) The spatial random effect is generated by the orthogonal RHZ proposal: $\boldsymbol{\theta} \sim N(0, \tau_\theta \boldsymbol{P}^\perp \boldsymbol{Q}(\boldsymbol{P}^\perp)^\top)$. The explanatory variables are the same as in scenario 1;

3. (ICAR non-spatial-X) Its is equal to scenario 1 except that the explanatory variable X_{2i} is also generated from independent standard normal distribution without any spatial information.

For each scenario, we generated 1000 datasets with the following combination of τ_θ and τ_e, the precision of the random effects and error term, respectively: (a) $\tau_e = 0.2$ and $\tau_\theta = 1$; (b) $\tau_e = 1$ and $\tau_\theta = 1$; (c) $\tau_e = 1$ and $\tau_\theta = 0.2$, to study their effects in model estimation and each running time. For each one of the nine scenarios, the posterior estimates of the following models were recorded: 1) SPOCK, 2) RHZ, 3) HH, 4) ICAR, 5) linear model (LM), as well as their execution time. The R software (R Development Core Team, 2011) was used to fit all proposed models. For the RHZ model, we used the R script freely available in http://www4.stat.ncsu.edu/~reich/Code/. The HH model was fitted using the R package ngspatial (Hughes and Cui, 2017, version 1.2) with half of the maximum Moran attractive eigenvectors. To fit the LM and ICAR models, we used the R-INLA package. Finally, SPOCK was fitted with one INLA version (SPOCK (R-INLA)) and one MCMC version (SPOCK (CARBayes)). Therefore, our time comparison is between the available implementation of the methods and not the computational complexity of the algorithms involved. In the simulation study, to make the computation time measuring fair between SPOCK (CARBayes) and the competitor models (RHZ and HH) that are MCMC based, all MCMC models were run with a single chain of size 15000, a burnin period of 5000 samples ending with a chain of 10000 samples.

We estimate the posterior mean of the fixed effects in each one of the 1000 replications. Table 1 in Section S1 in the supplementary material (Prates et al., 2018) presents the median and 2.5% and 97.5% percentiles of these estimates. We can see that, for all scenarios, the SPOCK model provides very similar inference compared to the LM, RHZ, and HH models. The ICAR model has a slightly different behavior. When the true generating model is the ICAR spatial-X, the ICAR model apparently outperforms the other models as τ_θ decreases. However, this is not so simple, as we explain soon. In the RHZ scenario, the ICAR model clearly has a large variance associated with β_2, specially when $\tau_\theta = 0.2$. This is expected because, as Reich et al. (2006) show, as the ratio τ_θ/τ_e decreases, the confounding problem becomes severe. Finally, in the ICAR non-spatial-X scenario, with no spatial information in the explanatory variables, all models seem to present a similar behavior, independently of precision differences. From Table 1 in Section S1 in the supplementary material, we can get two main conclusions. First, there is are major differences in the estimates provided by the SPOCK and the RHZ or HH models, making it a competitive model. The difference of the fitted parameters between the ICAR model and the other three models is large when the ratio between the spatial effect precision and the error precision is small ($\tau_\theta/\tau_e = 0.2$). The second one, is the apparent better behavior of the ICAR model in the ICAR spatial-X scenario. We discuss that this is not necessarily so in the following.

Hodges and Reich (2011) discuss the interpretation of confounding in spatial regression and link it with the multicollinearity problem in linear regression. By construction, the solution proposed by Reich et al. (2006) and more recently by Hughes and Haran (2013) assigns to $\boldsymbol{X}$ all the spatial variation that $\boldsymbol{\theta}$ and $\boldsymbol{X}$ are competing for. That is, RHZ, HH and our own method aims to estimate $\boldsymbol{\beta}^* = \boldsymbol{\beta} + (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top [\boldsymbol{1}, \boldsymbol{s}_1, \boldsymbol{s}_2] \boldsymbol{\gamma}$ in (7). There is not an universal agreement around this solution. Paciorek (2010) and Hanks et al. (2015) argue that making the spatial random effects orthogonal to the span of $\boldsymbol{X}$ is a very strong choice to alleviate the confounding problem.

To gain a better understanding of the effects of alleviating confounding we compare the $\hat{\boldsymbol{\beta}}$ Bayesian estimate (posterior mean) with $\boldsymbol{\beta}^*$, rather than with $\boldsymbol{\beta}$. To investigate this issue, we produce Figure 5. We calculate the ratio $\hat{\boldsymbol{\beta}}/\boldsymbol{\beta}^*$ in each one of the 1000 simulations and show them in the box plots for the 3 scenarios with $\tau_\theta = 0.2$ and $\tau_e = 1$ fixed. Similar figures appear with the other values for τ_θ and τ_e. Each scenario is shown in a different column. The first and second rows of plots present $\hat{\beta}_0/\beta_0^*$ and $\hat{\beta}_1/\beta_1^*$, respectively. Their coefficients β_0 and β_1 are not associated with spatial covariates. They have a similar behavior in all scenarios and the main conclusion is that the different methods produce very similar results, all of them centered around 1. The third row presents the distribution of the ratio $\hat{\beta}_2/\beta_2^*$. The first two scenarios, ICAR spatial-X and RHZ, have β_2 associated with a covariate with spatial structure. In this case, the traditional ICAR method has a much larger variance than the methods that alleviate spatial confounding. This shows that the ICAR method is not a good choice when the aim is to assign to $\boldsymbol{X}$ all the variation shared with $\boldsymbol{\theta}$. In the ICAR non-spatial-X scenario, all methods show a similar behavior, as expected.

After demonstrating that the SPOCK model is able to deal with the spatial confounding problem and provide very similar results to RHZ and HH models, we evaluate the average time to run each algorithm in their current available implementations.
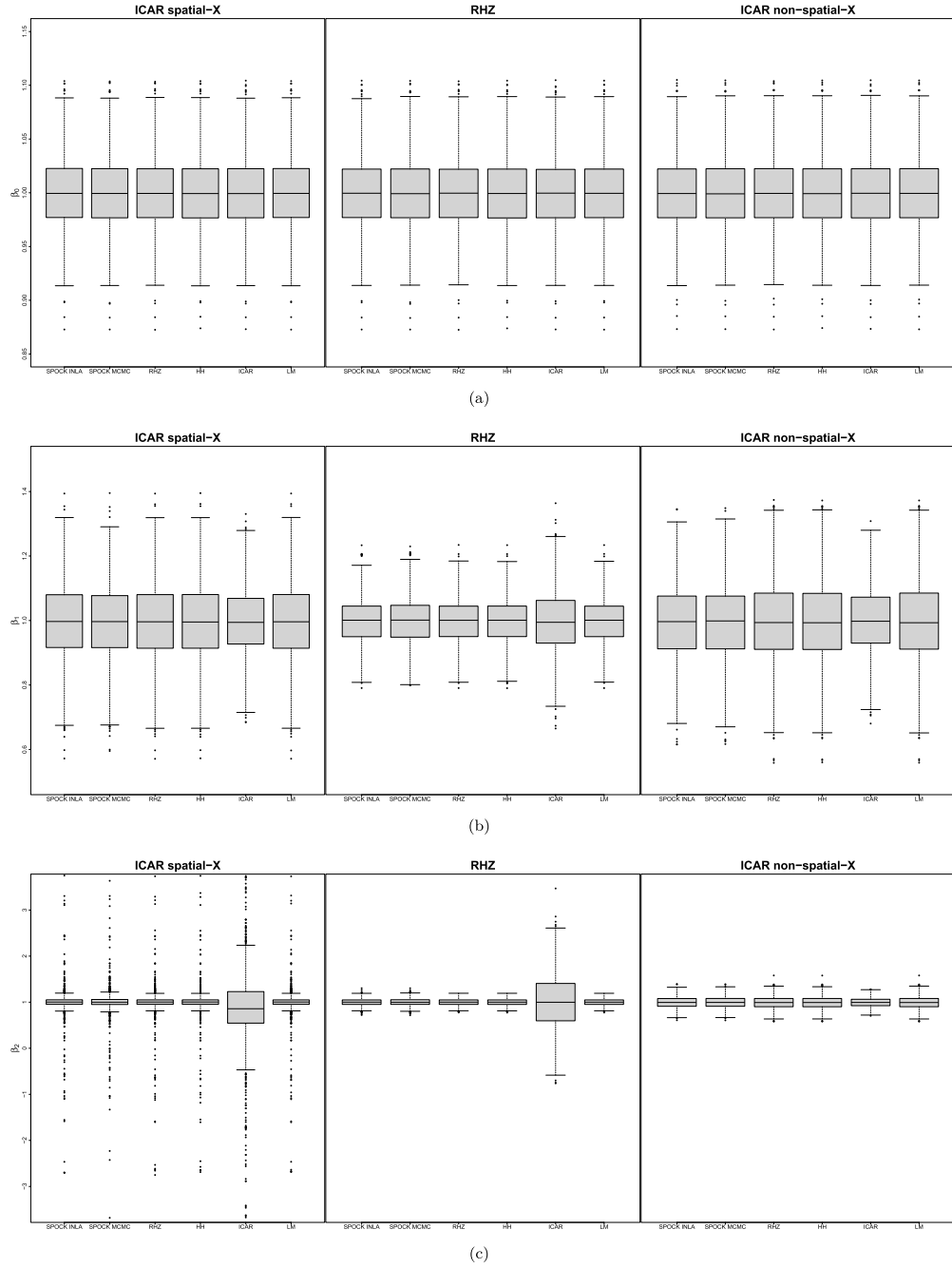
Figure 5: (a) Posterior estimates for β_0 under the different models. (b) Posterior estimates for β_1 under the different models. (c) Posterior estimates for β_2 under the different models.

Model	$\tau_e = 0.2, \tau_\theta = 1$		$\tau_e = 1, \tau_\theta = 1$		$\tau_e = 1, \tau_\theta = 0.2$	
	Time (sec)	sd	Time (sec)	sd	Time (sec)	sd
SPOCK (`R-INLA`)	0.466	0.025	0.471	0.033	0.487	0.047
SPOCK (`CARBayes`)	2.750	0.067	2.758	0.068	2.750	0.069
RHZ	38.911	0.141	38.995	0.153	39.193	0.281
HH	3.700	0.049	3.710	0.050	3.701	0.054
ICAR	0.429	0.033	0.385	0.041	0.353	0.076
LM	0.131	0.002	0.134	0.002	0.131	0.013

Table 2: Median execution time (in seconds) from 1000 replicates in scenario 2 (RHZ) and the execution standard deviation (sd) of the evaluation time of each method.

Table 2 presents the median execution time of 1000 replicates of each model for scenario RHZ. The LM method is substantially faster than the others. However, the LM method does not take into account the spatial variation that improves modeling. Among the spatial models, SPOCK combined with `R-INLA` clearly outperforms the RHZ and HH implementations in running time and it has comparable time with the ICAR fit by `R-INLA`. The RHZ has the largest median time, while the HH method seems to run around 10 times slower than SPOCK (`R-INLA`) and around 35% slower than the SPOCK (`CARBayes`).

5.2 Non-linear confounding

The second simulation study is focused on understanding how the methods alleviate spatial confounding when there is non-linear dependence between the spatial coordinates s and the covariates. To do so, we keep the Slovenia map and perform a simulation with the ICAR spatial-X scenario of Section 5.1 fixing $\tau_e = 1$ and $\tau_\theta = 0.2$ which is the most severe confounding scheme considered.

The data are generated by the following model $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 g(s_{i1}, s_{i2}) + \theta_i + e_i$ with $e_i \stackrel{ind}{\sim} N(0, \tau_e)$ and $i = 1, \ldots, n$ where the coefficients values $\boldsymbol{\beta}^\top$ are the same as in Section 5.1; X_{1i} and $\boldsymbol{\theta}$ are generated according to ICAR spatial-X scenario; and $X_{2i} = g(s_{i1}, s_{i2})$ is a non-linear function of the area centroids. To be more specific, a quadratic and a cubic scenario were created respectively as: (a) $X_{2i} = g(s_{i1}, s_{i2}) = s_{i1}^2 + s_{i2}^2$, and (b) $X_{2i} = g(s_{i1}, s_{i2}) = s_{i1}^3 + s_{i2}^3$. For each scenario 1000 datasets were generated.

As discussed in Section 5.1 we use the ratio of $\hat{\boldsymbol{\beta}}/\boldsymbol{\beta}^*$ as a more adequate measure of the spatial confounding. For each one of the 1000 simulations the ratio was calculated and Figure 1 in Section S2 in the supplementary material presents the box plot of the posterior bias estimates of $\hat{\boldsymbol{\beta}}/\boldsymbol{\beta}^*$. SPOCK provides very similar results in comparison to the RHZ and HH models. As expected, the estimates of fixed effects β_1 related with the covariate without spatial information present a similar behavior for all methods (Figure 1(b) in Section S2 in the supplementary material). All of them alleviate the spatial confounding for non-linear dependence, as can be seen in Figure 1(c) in Section S2 of the supplementary material.

5.3 Type-S error rates

Hanks et al. (2015) conducted a study to examine the Type-S error rate for restricted spatial regression (RSR) models and SGLMM in a geostatistical perspective. They consider that a Type-S error occurred when the regression parameters are equal to zero ($\beta = 0$ or $\beta^* = 0$) and the 95% equal-tailed posterior credible interval for the regression parameter does not contain zero. In their study, they concluded that the RSR model has higher Type-S error than the traditional SGLMM.

In this Section, we present a study to verify the Type-S error as in Hanks et al. (2015) and also the coverage of the regression coefficients for areal data. We used the Slovenia map and generate 1000 datasets from the following model: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \theta_i + e_i$ with $e_i \overset{ind}{\sim} N(0, \tau_e)$ and $i = 1, \ldots, n$. The number of regions in the map is $n = 192$ and $\boldsymbol{\theta}$ is the spatial effect following an ICAR model. The coefficients values are $\boldsymbol{\beta}^\top = (2, 1, 0)$ and both explanatory variables, $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, are generated from independent standard normal distribution without any spatial information. We fitted the same combinations of the precision parameters τ_θ and τ_e used in Section 5.1. In this scenario, $\boldsymbol{\beta} = \boldsymbol{\beta}^*$ since there is no spatial information in $\boldsymbol{X}$.

For each of the three scenarios, we recorded the coverage for β_1 and Type-S error for β_2 of the following models: SPOCK, RHZ, HH, ICAR, and LM. Table 3 shows that, as the precision of the spatial random effects decreases, the RHZ model drastically increases its Type-S error. This result agrees with the conclusion presented in Hanks et al. (2015). With a smaller spatial precision, SPOCK and HH slightly overestimates the Type-S error. The other methods seem to correctly control the nominal 5% Type-S error. Table 3 also presents the coverage rate of the 95% equal-tailed posterior credible interval for β_1. Only the RHZ model present smaller coverage rate. As previously observed with

		τ_e, τ_θ		
		$0.2, 1$	$1, 1$	$1, 0.2$
Type-S	SPOCK (R-INLA)	6.1%	7.3%	13.8%
	SPOCK (CARBayes)	5.9%	7.2%	14.2%
	RHZ	6.5%	15.7%	40.2%
	HH	5.6%	6.9%	11.5%
	ICAR	5.5%	6.1%	5.8%
	LM	5.6%	5.4%	6.4%
Coverage	SPOCK (R-INLA)	94.0%	94.0%	91.0%
	SPOCK (CARBayes)	94.0%	94.0%	91.0%
	RHZ	94.0%	88.0%	65.0%
	HH	94.0%	94.0%	92.0%
	ICAR	94.0%	94.0%	94.0%
	LM	95.0%	96.0%	96.0%

Table 3: Type-S error and coverage rates from 1000 replicates of the regression coefficients in the three scenarios for the fitted models.

the Type-S error, the coverage gets worse when the precision of the spatial effect is smaller. All other methods, including SPOCK, seem to control the coverage rate for the coefficient.

6 Applications

In this Section, different real data sets studied in the literature are presented to verify the presence of spatial confounding and the results presented by SPOCK in terms of estimation and computation time. bf In all real data examples, the HH model was fitted selecting half of the maximum Moran attractive eigenvectors (HH-Half) and the maximum Moran attractive eigenvectors (HH-Max). To make a fair comparison in terms of computation time between SPOCK and the competitor models (RHZ and HH) that are MCMC based, we also fitted SPOCK using the `CARBayes` package in addition to using INLA. To make them time comparable, all MCMC models were run with a single chain of size 50000, with a burnin period of 15000 samples ending with a chain of 35000 samples.

6.1 Slovenia dataset

The proposed model was adjusted to the same dataset used by Reich et al. (2006). The response variable Y_i is the number of cases of stomach cancer observed in the municipalities of Slovenia during the 1995–2001 period. The single covariate is the standardized value of a socioeconomic status measure for each area $i = 1, \ldots, 192$. Therefore, we have the following model $Y_i | \lambda_i, \sim \text{Poisson}(\lambda)$ with $\log(\lambda_i) = \boldsymbol{X}_i \boldsymbol{\beta} + \theta_i$ and $\boldsymbol{\beta}$ being the fixed effect and $\boldsymbol{\theta}$ the spatial effect.
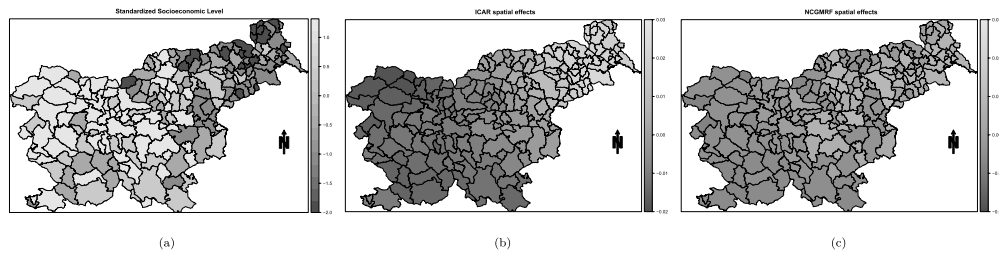

Figure 6: (a) Standardized socioeconomic level for the municipalities of Slovenia. (b) Posterior mean estimates for spatial random effects of the ICAR model. (c) Posterior mean estimates for spatial random effects of the SPOCK (`R-INLA`) model.

Using a simple exploratory analysis, the authors noticed that the data show a negative relationship between the response and the explanatory variable. This is expected based on the common knowledge of association between health risk and deprivation. Their first attempt was to fit the data with the traditional SGLMM to capture the spatial heterogeneity. However, from Figure 6(a), it is clear that the explanatory variable

has a diagonal spatial trend, presenting higher values in the southwest and smaller in the northeast. This is confirmed by the highly significant Moran index, with $I = 0.55$ with p-value approximately equal to zero. This is an indication of the presence of confounding between the random spatial effects and the socioeconomic level. The pattern in this dataset is similar to what we simulated in Section 3 when we took $X_i = s_{i1} + s_{i2}$ (see Figure 4(b)). After the SGLMM model was fitted by Reich et al. (2006), the coefficient associated with the socioeconomic level had a very wide credibility interval that covered even positive values. That is, the negative covariate effect disappeared.

Using our diagnostic test from Section 4, we can verify if there is evidence of spatial confounding. Calculating the first canonical correlation between the spatial centroids s and the covariate X, we find it equal to 0.670. This value is highly significant according to both, the asymptotic and the permutation test, obtained from Menzel (2012). The value is also high in absolute terms, indicating that the random spatial effects will be mixed up with the covariate effects.

To better understand the confounding effect between the explanatory variable and the random effects, we look at the spatial residuals from the ICAR model. From Figure 6(b), it is clear that both, the spatial random effect and the explanatory variable, share a southwest to northeast trend and therefore are competing for the spatial variability contained in the response Y_i. Figure 6(c) show the posterior mean estimates for the spatial effects under our SPOCK (R-INLA) model. After alleviating the spatial confounding, there is still some spatial trend left in the same southwest to northeast direction. However, the spatial dependence now is weaker and the spatial effects are smoothed toward zero. This is expected since, after alleviating the spatial confounding, we assume that the exploratory variable carries most of the spatial information in Y_i. Although weaker, it is important to notice that the spatial random effect structure from the SPOCK (R-INLA) model is still coherent with the space under analysis.

Table 4 shows the posterior mean estimates and credible intervals obtained applying all discussed models. The SPOCKs point estimate and credible interval are similar to those from the RHZ and HH-Max models. The HH-Half provide slightly different results than the other models. This could be due to a non adequate choice of the number of Moran attractive eigenvectors. As it can be seen in the last column of Table 4, SPOCK outperforms the other non-confounding methods algorithms in terms of running time.

Model	β	90% Credible Interval	Time (sec)
SPOCK (R-INLA)	-0.1214	$(-0.1674, -0.0752)$	2.64
SPOCK (CARBayes)	-0.1186	$(-0.1706, -0.0672)$	18.10
RHZ	-0.1216	$(-0.1665, -0.0759)$	249.49
HH-Half	-0.0798	$(-0.1257, -0.0342)$	24.45
HH-Max	-0.1157	$(-0.1556, -0.0749)$	25.05
ICAR	-0.0380	$(-0.0999, 0.0259)$	0.42
LM	-0.1358	$(-0.1682, -0.1032)$	0.22

Table 4: Posterior mean estimate and credible intervals of the coefficient associated with the socioeconomic variable (Se) for the five fitted models.

6.2   SIDS dataset

As a second example we fitted the models for the sudden infant death syndrome (SIDS) dataset, collected in the 100 counties of the North Carolina state during the year of 1974. These data were used by Cressie (1991). Rodrigues and Assunção (2012) extended these analyzes including covariates that explain the county mortality rate. We include in our models one of these covariates, the percentage of mothers who had prenatal care. The Moran index is equal to $I = 0.29$ with p-value approximately equal to zero, which indicates the presence of spatial correlation. From Figure 7, the centroids configuration is somewhat altered after the projection. The canonical correlation with the centroids is large, 0.394, and this value is significant according to the asymptotic test and the permutation test.

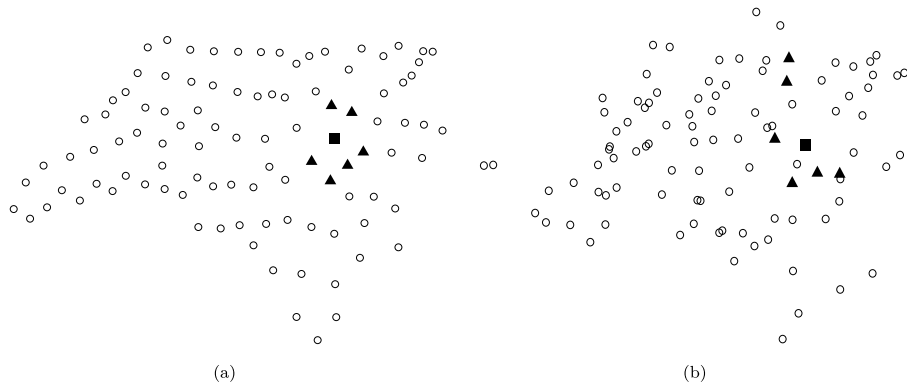(a)   (b)

Figure 7: (a) Original centroids of North Carolina. (b) New centroids after projecting the original centroids coordinates into the vector space orthogonal to the span of $\boldsymbol{X}$.

The first two columns of Table 5 show the parameter estimates for all models. The ICAR and HH-Half models are the ones for which the parameter seemed to be non-significant. The HH-Half was not able to alleviate the spatial confounding in this example, while HH-Max provide different estimates from the other methods but at least a significant one. The confidence intervals obtained for the other models were very close to each other. The last column of this table shows the required time to fit the models, and again, SPOCK has a much better performance in both cases than the RHZ and HH fits.

6.3   Scotland dataset

The models were also fitted to the famous lip cancer dataset recorded in Scotland districts from 1975 to 1986 (Breslow and Clayton, 1993). The covariate used in this case was the workforce percentage in each district employed in agriculture, fishing and forestry. According to the Moran index, this variable does not have a spatial structure ($I = 0.07$ and p-value $= 0.16$). The canonical correlation between the centroids and the

Model	β	90% Credible Interval	Time (sec)
SPOCK (R-INLA)	-1.2719	$(-2.3515, -0.1925)$	0.99
SPOCK (CARBayes)	-1.2479	$(-2.3577, -0.0988)$	11.51
RHZ	-1.5297	$(-2.7348, -0.2748)$	78.50
HH-Half	-0.7498	$(-1.6825, 0.1836)$	20.53
HH-Max	-0.9476	$(-1.8125, -0.0678)$	22.52
ICAR	-0.6244	$(-1.9563, 0.7514)$	1.04
LM	-1.1499	$(-1.9668, -0.3167)$	0.20

Table 5: Posterior mean estimate and credible intervals of the coefficient associated with the prenatal variable for the five fitted models.

covariate is small, 0.196, and not significant, suggesting that spatial confounding may not be present. Figure 8 shows that the centroids arrangement on the map is practically unchanged after their projection in the orthogonal space of this covariate.
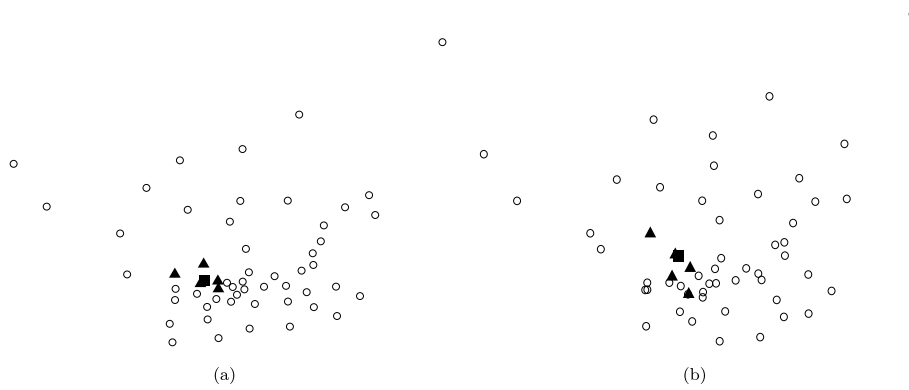

Figure 8: (a) Original centroids of Scotland. (b) New centroids after projecting the original centroids coordinates into the vector space orthogonal to the span of X.

Table 6 presents the results obtained after the estimation of the three models. The values estimated by all models are very similar to each other, as expected. And again, the proposed model outperforms its competitors implementations with both, using INLA (R-INLA) and MCMC (CARBayes).

7 Conclusion

In this paper, we introduced an alternative way to alleviate spatial confounding. The main idea is to construct a graph capable of capturing the spatial dependence orthogonal to the space generated by the span of X. By doing so, the introduced method maintains the original sparsity of the precision matrix Q and introduces no restriction in the spatial modeling setup.

Model	β	90% Credible Interval	Time (sec)
SPOCK (`R-INLA`)	0.0681	(0.0450, 0.0912)	0.86
SPOCK (`CARBayes`)	0.0574	(0.0326, 0.0823)	8.79
RHZ	0.0697	(0.0531, 0.0864)	37.66
HH-Half	0.0693	(0.0567, 0.0828)	18.82
HH-Max	0.0668	(0.0531, 0.0798)	19.38
ICAR	0.0625	(0.0431, 0.0815)	0.68
LM	0.0737	(0.0639, 0.0835)	0.23

Table 6: Posterior mean estimates and credible intervals of the coefficient associated with the workforce variable for the five fitted models.

As one reviewer pointed out, "the asymptotics are not clear when the size of the areal regions shrink toward zero and the density of spatial units increases, leading to continuous spatial support rather than discrete." In particular, the type of folding that Sampson and Guttorp (1992) caution against may also be an issue here. Therefore, it is not clear to us if our method can be extended to the continuous setting by an asymptotic argument. However, we think that for many applied researchers dealing with areal data this asymptotic approach is not of concern because it is artificial to think on a limiting process in this situation. For example, consider the case of Bayesian disease mapping. One needs risk population (and cases) in each small area and it does seem awkward to imagine a limiting process where the area is reduced to zero and hence no population is possible inside it. This is an interesting mathematical issue but we think that most areal data researchers will be concerned and limited to work with areas bounded from below.

From our simulation study and real data examples, we were able to show that the SPOCK approach provides similar results to the other methods that alleviate confounding. These alternative models project the spatial random effects on the orthogonal space spanned by X. Instead, SPOCK projects the original geographical structure on that same vector space. In all simulated scenarios and real data analyses, our method was faster than the available implementations of the other existing methodologies. This advantage can make it attractive to researchers in different areas, specially when working with very large spatial datasets. Another advantage is that our method can be used with any usual ICAR implementation such as `R-INLA`, `WinBUGS`, `OpenBUGS`, and `CARBayes`.

We showed that, when no spatial confounding is present, it does not matter which method one uses. However, when this is not the case, running the usual ICAR produces different coefficients and inference from those obtained from the three methods to alleviate spatial confounding. The reason for this behavior is that these two classes of models estimate different parameters. While ICAR estimates $\boldsymbol{\beta}$, the others estimate $\boldsymbol{\beta}^*$. Specific application considerations should guide which parameter is more meaningful. In the common situation where the spatial effects $\boldsymbol{\theta}$ represent geographically structured nuisance effects, it may be desirable to assign to X all spatial variation shared with the random spatial effect $\boldsymbol{\theta}$ and hence, to estimate $\boldsymbol{\beta}^*$. However, as mentioned by Paciorek (2010), the data generating system may have a non-observed spatial confounder

with the observed explanatory variables and setting all spatial variation to the observed covariates may not be appropriate.

As seen in the simulation study, the SPOCK methodology has coverage close to the nominal value and controls for the Type-S error, having a slight inflation when the spatial random effect has large variability. This is a nice property since the RHZ model can severely inflate the Type-S error and severely deflate the coverage rate. Another nice property is that SPOCK inspires a diagnostic test to decide when we should carry out the correction for spatial confounding. It is simple and can be calculated previous to any model fitting. When there is no spatial confounding, fitting ICAR will lead to the same inference as the spatial confounding alleviating methods as $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^*$ are the same. However, fitting ICAR in the other situation, when there is unobserved spatial confounders, leads to an estimate of $\boldsymbol{\beta}$, rather than of $\boldsymbol{\beta}^*$. The user should be aware of these differences so he does not use one believing to have the other.

Supplementary Material

Alleviating spatial confounding for areal data problems by displacing the geographical centroids: Supplementary Material (DOI: 10.1214/18-BA1123SUPP; .pdf). The supplementary material present some results for the simulation studies of the paper.

References

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice data systems (with discussion). *Journal of the Royal Statistical Society, Series B 36*, 192–225. MR0373208. 624, 627, 628

Besag, J., J. York, and A. Mollie (1991). Bayesian image restoration with two application in spatial statistics (with discussion). *Annals of the Institute Statistical Mathematics 43*, 1–59. MR1105822. doi: https://doi.org/10.1007/BF00116466. 623, 626, 628

Breslow, N. E. and D. G. Clayton (1993). Approximate inference in generalized linear mixed models. *Journal of the American statistical Association 88*(421), 9–25. MR1394064. doi: https://doi.org/10.2307/2291379. 625, 642

Clayton, D., L. Bernardinelli, and C. Montomoli (1993). Spatial correlation in ecological analysis. *International Journal of Epidemiology 6*, 1193–1202. 624, 626

Cressie, N. (1991). *Statistics for spatial data*. John Wiley & Sons. MR1127423. 625, 642

Gelman, A. and F. Tuerlinckx (2000). Type s error rates for classical and bayesian single and multiple comparison procedures. *Computational Statistics 15*(3), 373–390. 635

Hanks, E. M., E. M. Schliep, M. B. Hooten, and J. A. Hoeting (2015). Restricted spatial regression in practice: geostatistical models, confounding, and robustness under model misspecification. *Environmetrics 26*(4), 243–254. MR3340961. doi: https://doi.org/10.1002/env.2331. 624, 625, 634, 636, 639

Hefley, T. J., M. B. Hooten, E. M. Hanks, R. E. Russell, and D. P. Walsh (2017). The bayesian group lasso for confounded spatial data. *Journal of Agricultural, Biological and Environmental Statistics 22*(1), 42–59. MR3607653. doi: https://doi.org/10.1007/s13253-016-0274-1.   624, 625, 634

Hodges, J. S. and B. J. Reich (2011, January). Adding Spatially-Correlated Errors Can Mess Up the Fixed Effect You Love. *The American Statistician 64*(4), 325–334. MR2758564. doi: https://doi.org/10.1198/tast.2010.10052.   630, 636

Hughes, J. and X. Cui (2017). *ngspatial: Fitting the Centered Autologistic and Sparse Spatial Generalized Linear Mixed Models for Areal Data*. Denver, CO. R package version 1.2.   635

Hughes, J. and M. Haran (2013). Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *Journal of the Royal Statistical Society, Series B 75*, 139–159. MR3008275. doi: https://doi.org/10.1111/j.1467-9868.2012.01041.x.   624, 626, 627, 636

Lee, D. (2013). CARBayes: An R package for Bayesian spatial modeling with conditional autoregressive priors. *Journal of Statistical Software 55*(13), 1–24.   625

Leroux, B. G., X. Lei, and N. Breslow (1999). Estimation of disease rates in small areas: A new mixed model for spatial dependence. In M. E. Halloran and D. Berry (Eds.), *In Statistical Models in Epidemiology; the Environment and Clinical Trials*, pp. 179–192. New York: Springer–Verlag. MR1731684. doi: https://doi.org/10.1007/978-1-4612-1284-3_4.   624, 627, 628

Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter (2000). WinBUGS – a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing 10*, 325–337.   623

Menzel, U. (2012). *CCP: Significance Tests for Canonical Correlation Analysis (CCA)*. R package version 1.1.   641

Murakami, D. and D. A. Griffith (2015). Random effects specifications in eigenvector spatial filtering: a simulation study. *Journal of Geographical Systems 17*(4), 311–331. 624

Paciorek, C. J. (2010). The importance of scale for spatial-confounding bias and precision of spatial regression estimators. *Statistical Science 25*, 107–125. MR2741817. doi: https://doi.org/10.1214/10-STS326.   624, 625, 636, 644

Prates, M. O., Assunção, R. M., and Rodrigues, E. C. (2018). Alleviating spatial confounding for areal data problems by displacing the geographical centroids: Supplementary Material. *Bayesian Analysis*. doi: https://doi.org/10.1214/18-BA1123SUPP. 636

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.   635

Reich, B. J., J. S. Hodges, and V. Zadnik (2006). Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. *Biometrics 62*, 1197–1206. MR2307445. doi: https://doi.org/10.1111/j.1541-0420.2006.00617.x. 624, 625, 626, 627, 631, 636, 640, 641

Rodrigues, E. C. and R. Assunção (2012). Bayesian spatial models with a mixture neighborhood structure. *Journal of Multivariate Analysis 109*(0), 88–102. MR2922856. doi: https://doi.org/10.1016/j.jmva.2012.02.017. 624, 627, 628, 642

Rue, H. and L. Held (2005). *Gaussian Markov random fields: Theory and applications.* Chapman & Hall. MR2130347. doi: https://doi.org/10.1201/9780203492024. 623, 628

Rue, H., S. Martino, and N. Chopin (2009). Approximate bayesian inference for latent gaussian models using integrated nested laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B 71*, 319–392. MR2649602. doi: https://doi.org/10.1111/j.1467-9868.2008.00700.x. 625, 632

Sampson, P. D. and P. Guttorp (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association 87* (417), 108–119. 627, 644

Wilks, S. (1935). On the independence of k sets of normally distributed statistical variables. *Econometrica, Journal of the Econometric Society*, 309–326. 634

Zadnik, V. and B. J. Reich (2006). Analysis of the relationship between socioeconomic factors and stomach cancer incidence in Slovenia. *Neoplasma 53*, 103–110. 625

Acknowledgments