# RESTRICTED SPATIAL REGRESSION METHODS: IMPLICATIONS FOR INFERENCE

**Kori Khan**[1],
Department of Statistics, The Ohio State University

**Catherine A. Calder**
Department of Statistics and Data Sciences, University of Texas at Austin

## Abstract

The issue of spatial confounding between the spatial random effect and the fixed effects in regression analyses has been identified as a concern in the statistical literature. Multiple authors have offered perspectives and potential solutions. In this paper, for the areal spatial data setting, we show that many of the methods designed to alleviate spatial confounding can be viewed as special cases of a general class of models. We refer to this class as Restricted Spatial Regression (RSR) models, extending terminology currently in use. We offer a mathematically based exploration of the impact that RSR methods have on inference for regression coefficients for the linear model. We then explore whether these results hold in the generalized linear model setting for count data using simulations. We show that the use of these methods have counterintuitive consequences which defy the general expectations in the literature. In particular, our results and the accompanying simulations suggest that RSR methods will typically perform worse than non-spatial methods. These results have important implications for dimension reduction strategies in spatial regression modeling. Specifically, we demonstrate that the problems with RSR models cannot be fixed with a selection of "better" spatial basis vectors or dimension reduction techniques.

### Keywords

Confounding; Spatial Statistics; Bayesian; Dimension Reduction

## 1. INTRODUCTION

In our increasingly data rich world, large spatial data sets are becoming abundant. A booming field of research involves incorporating spatial dependence into models in a computationally efficient way. As is often the case with spatial statistics, much of this research has initially focused on developing and analyzing methods for spatial process models: models developed to make predictions at unobserved locations (e.g., Banerjee et al., 2008; Wikle, 2010; Fuentes, 2007; Stein, 2014). There have been few attempts to understand how these methods impact inference on regression coefficients.

---

[1] khan.746@osu.edu .

Recently, however, there has been a line of work designed to efficiently incorporate spatial dependence into regression models when the primary interest is on inference for the regression coefficients. In many ways, this set of work mirrors the work done in the spatial process model research. For some examples, in the context of areal data, Hughes and Haran (2013) presented a reduced-rank approach, Prates et al. (2019) developed a sparse approximation technique, and Burden et al. (2015) introduced an approximate likelihood method. Bradley et al. (2015) extended ideas of basis selection techniques to multivariate spatio-temporal mixed effects models, while Murakami and Griffith (2015) incorporate spatial dependence through eigen-vector spatial filtering. More recently, Thaden and Kneib (2018) introduced a structural equation approach for estimating regression coefficients when there is an unobserved spatially dependent covariate.

In contrast to the work in the spatial process model realm, these methods are either designed or inspired by methods designed to alleviate "spatial confounding." The first explicit reference to spatial confounding is often attributed to Clayton et al. (1993), who observed what he referred to as "confounding by location": the situation where estimates of a regression coefficient associated with a spatially-structured covariate are affected by the presence of a spatial random effect in the model. In recent work, the phenomenon is often explained as the presence of multicollinearity between the covariates, $X$, and the spatial random effect (Prates et al., 2019; Hanks et al., 2015; Thaden and Kneib, 2018; Hefley et al., 2017).

Reich et al. (2006) and Hodges and Reich (2010) brought recent attention to the issue (see also, Paciorek, 2010, for a study on the effects of spatial confounding on inference for regression coefficients). This work highlighted for the first time that in the presence of spatial confounding, not only can the estimates for regression coefficients change, but the uncertainty associated with these estimates can be "overinflated." To address these dual concerns, Reich et al. (2006) proposed a method employing synthetic predictors to smooth orthogonally to the fixed effects. Hughes and Haran (2013) then extended this work by suggesting the use of eigenvectors of the Moran operator as synthetic predictors. They argued these basis functions would allow for dimension reduction through the selection of only those synthetic predictors associated with "attractive" spatial dependence (as opposed to "repulsive" spatial dependence). The intuitive appeal of this set of work inspired a series of follow-up investigations (Prates et al., 2019; Burden et al., 2015; Bradley et al., 2015; Thaden and Kneib, 2018). However, it has been observed that these methods can lead to elevated levels of Type-S errors, the Bayesian analogue of Type I error (Prates et al., 2019; Hanks et al., 2015, the latter paper suggested a posterior predictive approach to address this concern). Recently, Hanks et al. (2015) and Prates et al. (2019) have noted that there is a need to better understand when it is appropriate to utilize such methods. Indeed, with the exception of Reich et al. (2006), there has been a lack of mathematical formalism to assess the impact of smoothing orthogonally to the fixed effects has on inference for regression coefficients.

In this work, we propose to fill that void. To do so, we consider a Bayesian analysis of Gaussian areal spatial data. We show that with respect to inference on regression coefficients, the current methods proposed to smooth orthogonally to the fixed effects can

be thought of as a subset of a larger class of models. Extending the terminology currently in use for these proposed methods, we will refer to this larger class as *Restricted Spatial Regression* (RSR) models. We find that RSR models transform a mixed-effects model into an overfit linear model. Specifically, any of these models will produce a posterior mean for the regression coefficients which is equivalent to the posterior mean obtained in the corresponding non-spatial model. The various approaches to smoothing orthogonally to fixed effects were designed to ensure that all the marginal posterior variances of the regression coefficients are greater than the corresponding posterior variances of the non-spatial model. However, we show that the exact opposite is often true. The reduction in the posterior variances is caused solely by the addition of the spatial basis functions. Furthermore, our analytic results and the included simulation studies indicate that sufficiently large credible intervals for the regression coefficients will generally be nested within the corresponding credible intervals from the non-spatial model. Importantly, these results are invariant to the spatial structure of the covariates and any spatial structure in the residuals. In short, our results indicate that with respect to coverage and Type-S error, one would be better off fitting a non-spatial model than any RSR model. Furthermore, simulations indicate that when there is spatial dependence unexplained by the covariates, RSR models for Gaussian data exasperate the decrease in coverage and increase in Type-S rates - even if the true unexplained spatial dependence in the response variable is generated orthogonally to the covariates. A simulation study and example application of RSR models suggest similar results might hold for count data.

The rest of the paper proceeds as follows: Section 2 provides an overview of deriving inference on regression coefficients in a spatial random effects model and a discussion of the methods developed to alleviate spatial confounding. Section 3 contains the results of this paper. Section 4 includes simulation studies, and Section 5 compares various approaches on the Slovenia stomach cancer data set. The proofs for all theorems are contained in Appendix C of the online supplement.

## 2. BACKGROUND

### 2.1. Inference on Regression Coefficients in Random Effects Models.

The spatial generalized linear mixed model (SGLMM), popularized by Diggle et al. (1998), assumes that $Y = \{y_i, ..., y_n\}^T$ is a realization from a random field where $y_i$ is observed at spatial location $s_i$, $i = 1 \ldots n$. The vector of transformed conditional means $Z = \{g(E(y_1|\eta(s_1))), ..., g(E(y_n|\eta(s_n)))\}^T$ for a given link function $g$ is then related to fixed effects and a spatial random effect:

$$Z = \mathbf{1}\beta_0 + X\beta^X + \eta. \tag{1}$$

Here, $\mathbf{1}$ is the $n \times 1$ column vector of 1's, $X = [X_1, ..., X_p]$ is the $n \times p$ design matrix whose $i$th row consists of the $p$ covariates associated with $Z_i$, $\beta^* = (\beta_0, \beta^X) \in \mathbb{R}^{p+1}$ is a vector of regression parameters, and $\eta = (\eta_1, ..., \eta_n)^T$ is a zero-mean random effect with a

spatial covariance matrix $\Sigma(\boldsymbol{\theta})$ parameterized by $\theta \in \mathbb{R}^m$. In this paper, the notation denoting dependence on location will be dropped when the meaning is unambiguous. We define $\boldsymbol{X}^* \equiv [1 \ \boldsymbol{X}]$. The distinction between $\boldsymbol{X}$ and $\boldsymbol{X}^*$ is important when we compare different models. However, unless otherwise stated, the points about one in this section are true for both.

Historically, much of the statistical community's intuition for the behavior of estimates of regression coefficients is based on work for the linear model when $g$ is the identity mapping. A non-spatial (NS) analysis in this case would correspond to an assumption that $\Sigma(\theta) \propto \boldsymbol{I}$. In this setting, the best unbiased linear estimator for the regression coefficients is the well-known ordinary least squares (OLS) estimator $\hat{\boldsymbol{\beta}}_{NS}^* = (\boldsymbol{X}^{*T}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*T}\boldsymbol{Y}$. Classical statistics textbooks remind us that if the assumption of i.i.d error structure is not met, these OLS estimates will be inefficient. Instead, if $\boldsymbol{\theta}$ were known, the most efficient estimator would be the generalized least squares estimator (GLS) $\hat{\boldsymbol{\beta}}_{GLS}^* = (\boldsymbol{X}^{*T}\Sigma^{-1}(\theta)\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*T}\Sigma(\theta)^{-1}\boldsymbol{Y}$.

The two estimates will generally not be the same. However, in spatial statistics there is often an assumption that the addition of a spatial random effect should not change the point estimates of the regression coefficients. This expectation seems to be driven by the fact that geostatistical software developed to facilitate spatial process modeling often only implemented ordinary kriging and not universal kriging, the former allowing for an unknown, but constant mean and the latter allowing the mean to be an unknown linear combination of covariates associated with points in space (Waller and Gotway, 2004, pg. 344). Conventional wisdom in the spatial statistics literature suggests that if there is spatial dependence unexplained by the covariates (i.e, processes in which things near in space are assumed to be more similar than things far away in space), the naive use of $\hat{\boldsymbol{\beta}}_{NS}^*$ will underestimate the variance of this estimator. This intuition is driven by well-understood examples which illustrate that ignoring spatial dependence will result in an underestimate of the variance of the mean of a spatial process (see, e.g., Cressie, 1993, pages 13-15). Paciorek (2010) pointed out that there is a lack of formal quantification of this belief for regression coefficients in general. Working in a setting where $\boldsymbol{X}$ is stochastic, Paciorek (2010) showed that generally the naive variance estimator $\text{Var}(\hat{\boldsymbol{\beta}}_{OLS}^X)$ will underestimate the uncertainty estimate of the correct variance estimator $\text{Var}(\hat{\boldsymbol{\beta}}_{GLS}^X)$, thereby lending support to the common expectation in spatial statistics. In any case, the expectation that the estimates of the variance associated with regression coefficient estimators will increase in models accounting for the presence of residual spatial dependence extends beyond the linear model. In practice this expectation is often true, but there are examples in the literature where it does not hold (for an example, see Banerjee et al., 2003).

## 2.2. Spatial Confounding.

The term spatial confounding is used to describe the effect that multicollinearity between the fixed covariates $\boldsymbol{X}$ and the spatial random effect $\boldsymbol{\eta}$ has on the point estimates and variance of the regression coefficient estimates. Most of the work in spatial confounding involves a fully Bayesian analysis where the point estimates for the regression coefficients are defined

to be the posterior mean $E(\beta^X|Y)$ and the variances of interest are the diagonal entries of $\text{Var}(\beta^X|Y)$ (Reich et al., 2006; Hughes and Haran, 2013; Hefley et al., 2017; Hanks et al., 2015; Prates et al., 2019). Efforts have been made to propose statistics designed to detect the presence of spatial confounding (Reich et al., 2006; Hefley et al., 2017; Prates et al., 2019). Thaden and Kneib (2018) provided a formalization of spatial confounding that assumes multicollinearity between $X$ and the spatial random effect occurs because of the absence of another unobserved spatially varying covariate. However, there is not currently a formal definition of spatial confounding in a more general setting, in part because it is a difficult concept to quantify. It is easier to define what spatial confounding is not, and so that is what we do in Definition 1. As will be seen shortly, all the methods designed to address spatial confounding are designed in the hopes of achieving the properties in Definition 1.

**Definition 1.**—A method which results in posterior mean $E(\beta^X|Y)$ and marginal posterior variances $\text{Var}(\beta_i^X|Y)$, $i = 1, \ldots, p$ alleviates spatial confounding if the following conditions are met:

1.  $E(\beta^X|Y) = E(\beta_{NS}^X|Y)$, and

2.  $\text{Var}(\beta_{NS,i}^X|Y) \leq \text{Var}(\beta_i^X|Y) \leq \text{Var}(\beta_{Spatial,i}^X|Y)$ for $i = 1, \ldots, p$,

where $\beta_{NS}^X$ are the regression coefficients of the corresponding non-spatial model and $\beta_{Spatial}^X$ are the regression coefficients from an unrestricted spatial random effect.

Hodges and Reich (2010) argued that spatial confounding can be a concern whenever a spatial random effect is included in a model. However, most of the proposed methods to address spatial confounding are developed in the context of Gaussian areal spatial data (Reich et al., 2006; Hodges and Reich, 2010; Hughes and Haran, 2013; Prates et al., 2019). In the areal data setting, spatial dependence is described by the introduction of an underlying, undirected graph $G = (V, E)$. Non-overlapping spatial regions that partition the study area are represented by vertices, $V = \{1, \ldots, n\}$, and edges $E$ defined so that each pair $(i, j)$ represents the proximity between region $i$ and region $j$. We represent $G$ by its $n \times n$ binary adjacency matrix $A$ with entries defined such that $\text{diag}(A) = 0$ and $A_{i,j} = \mathbb{1}_{(i,j) \in E, i \neq j}$. In models designed to address spatial confounding, the spatial random effect is typically assumed to follow the intrinsic conditional autoregressive (ICAR) model/ prior (Besag et al., 1991). We will refer to models utilizing this prior as ICAR models.

For Gaussian areal data, we observe that the ICAR model, the non-spatial model, and models designed to alleviate spatial confounding such as those proposed by Reich et al. (2006) (RHZ), Hughes and Haran (2013) (HH), and Prates et al. (2019) (PAR) are all special cases of the following more general form:

$$Y = X\beta^X + W\delta + \epsilon$$

$$p(\delta|\tau_s) \propto \tau_s^{\text{rank}(F)/2}\exp\{\frac{-\tau_s}{2}\delta^T F\delta\},$$

(2)

where $\epsilon \sim N(\mathbf{0}, \tau_\epsilon \mathbf{I})$, and $\tau_\epsilon$ and $\tau_s$ are precision parameters. $\mathbf{W}$ is a $n \times q$ set of basis vectors such that $q \leq n - p$, and $\mathbf{F}$ is a symmetric, non-negative definite matrix. Table 1 illustrates how the three proposed RSR methods can be considered special cases of the form (2).

The distinction between the use of $\mathbf{X}$ and $\mathbf{X}^*$ in Table 1 is a direct consequence of the impropriety of the ICAR prior. The ICAR model captures spatial dependence through a Gaussian Markov Random Field (GMRF). The precision matrix is the graph Laplacian $\mathbf{Q}$ $= \text{diag}(\mathbf{A1}) - \mathbf{A}$. Because the adjacency matrix $\mathbf{A}$ is defined to be the zero/one adjacency matrix on $G$, $\mathbf{Q}$ is the graph Laplacian of a simple graph. A well-known fact from spectral graph theory is that the kernel of $\mathbf{Q}$ will be constant on each connected component of $G$ (Von Luxburg, 2007). This fact means that the prior $p(\delta|\tau_s)$ will include an implicit intercept for each connected component of the graph in Bayesian analysis. For simplicity, this paper will assume that the $G$ is connected and therefore there is a single global intercept included in the ICAR prior.

Reich et al. (2006) investigated in depth the effect multicollinearity between $\mathbf{X}$ and the spatial random effect induced by the ICAR prior has on $E(\beta^X|\mathbf{Y})$ and $\text{Var}(\beta^X|\mathbf{Y})$ using a re-parameterization of the ICAR model:

$$\mathbf{Y} = \mathbf{X}\beta^X + \mathbf{V}\delta + \epsilon,$$

$$p(\delta|\tau_s) \propto \tau_s^\kappa \exp\left\{\frac{-\tau_s}{2}\delta^T \mathbf{\Lambda} \delta\right\}, \tag{3}$$

where $\mathbf{V\Lambda V}^T$ is its eigendecomposition of $\mathbf{Q}$, and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_n)$ where $\lambda_1 \geq \lambda_2, \ldots \geq \lambda_n$. Under the assumptions of this paper (i.e., that $G$ is connected), $\kappa = \frac{n-1}{2}$. Using a Bayesian approach with flat priors on $\boldsymbol{\beta}^X$ and independent gamma priors on $\tau_s$, $\tau_s$, Reich et al. (2006) and Hodges and Reich (2010) illustrated that $E[\beta^X|Y, \tau_s, \tau_\epsilon] \neq E[\beta^X_{NS}|\mathbf{Y}]$ in the presence of multicollinearity between $\mathbf{X}$ and $\mathbf{V}$. When regression coefficients are of primary interest, they argued that spatial random effects are added merely to account for spatial correlation in residuals when computing the posterior variance. Under this reasoning, the spatial random effect should not change the point estimates of the fixed effects. Furthermore, Reich et al. (2006) suggested that this multicollinearity caused "overinflation" of the posterior variance of the regression coefficients. This overinflation was noted to be exacerbated when $\mathbf{X}$ was correlated with low frequency eigenvectors of the graph Laplacian. In spatial regression, Reich et al. (2006) argued that two factors can lead to an increase in the posterior variance of regression coefficients as compared to the NS model: 1) collinearity with the spatial random effects and 2) a reduction in the effective number of observations because of spatial clustering. They proposed a model that they claimed would eliminate the former of these factors while preserving the latter. In other words, the RHZ model was designed to alleviate spatial confounding according to Definition 1:

1. $E(\beta^X_{RHZ}|\mathbf{Y}) = E(\beta^X_{NS}|\mathbf{Y})$

2. $\text{Var}(\beta^X_{NS,i}|\mathbf{Y}) \leq \text{Var}(\beta^X_{RHZ,i}|\mathbf{Y}) \leq \text{Var}(\beta^X_{ICAR,i}|\mathbf{Y})$ for $i = 1, \ldots, p$.

The authors provided theoretical support for first inequality of the second claim.[2]
To accomplish these goals, the authors proposed using synthetic predictors which are orthogonal to the column space of $X$. Define $P_X = X(X^T X)^{-1} X^T$ to be the projection matrix onto the column space of $X$, and $P_X^\perp = I - P_X$ to be its orthogonal complement. Let $L$ be the $n \times (n - p)$ matrix composed of the eigenvectors of $P_X^\perp$ associated with an eigenvalue of 1 and let $K$ be the $n \times p$ matrix composed of the eigenvectors associated with an eigenvalues of 0. Model (3) can be re-written as a function of $\delta_1 = L^T V \delta$ and $\delta_2 = K^T V \delta$. The RSR method proposed by Reich et al. (2006) involves setting $\delta_2 = 0$. Using this notation, the RHZ model can be expressed as:

$$Y = X\beta^X + L\delta_1 + \epsilon,$$

$$p(\delta_1 | \tau_s) \propto \tau_s^k \exp\{\frac{-\tau_s}{2} \delta_1^T Q_s \delta_1\},$$

$(4)$

where $Q_s = L^T Q L$. Hodges and Reich (2010) explained that this solution assigns all variability explained by both $V$ and $X$ to $X$ by restricting spatial smoothing to the orthogonal complement of the fixed effects.

Hughes and Haran (2013) also argued that variance inflation caused by spatial confounding is a concern in the SGLMM. However, they stated that the RHZ model (4) fails to account for the underlying graph in the construction of $L$, thereby permitting structure in the spatial random effects that corresponds to negative spatial dependence. Hughes and Haran (2013) suggested such negative spatial dependence should not be expected in the context in which spatial models are generally fit. They developed a model (hereafter the HH model) that they stated would eliminate negative dependence while alleviating the impact of spatial confounding on the marginal posterior variances. Utilizing a Bayesian framework with proper priors on the regression coefficients, they suggested that rather than using $L$, one should instead use $M_q$, where $M_q$ is the $n \times q$ matrix composed of the eigenvectors of what they referred to as the Moran operator $P_X^\perp A P_X^\perp$. In support of the argument that $M_q$ should be preferred to $L$, the authors pointed out that the column vectors of $M_q$ had more spatial structure than the column vectors of $L$.

Hughes and Haran (2013) also stated that this model naturally lent itself to dimension reduction. The Moran operator is the numerator of a generalization of the Moran's $I$ statistic, a popular non-parametric measure of spatial dependence introduced by Moran (1950). We will refer to this generalization of the Moran's $I$ as $I_X(A)$. Relying on work on the original Moran's $I$ statistic in the context of bounded regular tessellations by Boots and Tiefelsdorf (2000), Hughes and Haran (2013) noted the (standardized) spectrum of the Moran operator comprises all possible values of $I_X(A)$ and its eigenvectors comprise of all possible mutually distinct patterns of clustering after accounting for $X$ and the graph. Importantly,

---

[2]Section 3 of Reich et al. (2006) provides a mathematical argument that this should be true. However, the argument appears to rely on the stochastic ordering of a gamma distribution. It appears the authors assume they are using a gamma distribution with a scale parameter- which is stochastically increasing in the scale. Their work suggests they are instead working with a rate parameterization- which is stochastically decreasing in the rate parameter.

they argued that choosing only "attractive" eigenvectors (those associated with positive eigenvalues) of the Moran operator would improve the inference on regression coefficients by eliminating patterns of "repulsive" spatial dependence (eigenvectors associated with negative eigenvalues). Furthermore, they suggested that for most graphs only selecting the first $.1n$ eigenvectors should be sufficient to perform well for regression. The work in Hughes and Haran (2013) relied on simulations to illustrate the points made.

A downside to both the RHS and HH models is a loss of some of the computational efficiency inherent in the ICAR prior utilized in (3). ICAR models capture spatial dependence through a Gaussian Markov Random Field (GMRF). This is appealing because the precision matrix for the random field is very sparse, which allows the use of sparse matrix routines to facilitate efficient computations (Rue and Held, 2005; Paciorek, 2009). The matrix $Q_s$ need not be, and generally is not, sparse. Furthermore, the RHZ and HH methods do not extend naturally to spatial models that do not utilize the ICAR prior. Recent work by Prates et al. (2019) attempts to to overcome the computational limitations of RSR methods. Their work suggests projecting the vertices of the graph $G$ onto the orthogonal space of the design matrix $X$. Using the projected vertices, they then construct a new sparse precision matrix $Q^{\perp}$ for use in analysis. This model is referenced as the PAR model in Table 1. Guan and Haran (2018) also suggested a projection based approach to approximate covariance matrices of spatial random effects to improve the computationally efficiency of RSR methods. This method can also be extended to spatial models beyond the ICAR model.

RSR methods are not universally accepted; some find the assumption that the spatial random effect operates orthogonally to the fixed effects to be too strong (See e.g., Paciorek, 2010). Furthermore, recent work has highlighted that RSR methods can suffer from elevated rates of Type-S errors under model misspecification (Hanks et al., 2015; Prates et al., 2019). Type-S error is the Bayesian analogue to Type 1 error, and we define it to occur when the equal-tailed 95% credible interval for a regression coefficient that is truly 0 does not include 0. The model misspecification studied has involved a generating model which include spatial random effects that do not operate orthogonally to the fixed effects. Currently, the prevalent belief is that the elevated Type-S error is caused by such misspecification. However, in this paper we will show that the phenomena occurs even without it.

It remains an open question when RSR methods should be preferred over traditional spatial models. Understanding when RSR methods perform well is important because RSR methods can lead to a very different interpretations of the effect of regression coefficients (Hanks et al., 2015). Attempts to understand when RSR methods are appropriate have led to suggestions that when $X$ is correlated with low frequency eigenvectors of the graph Laplacian, RSR methods should be preferred (Reich et al. (2006); Prates et al. (2019), Hefley et al. (2017)).

## 3.  CONSEQUENCES OF ORTHOGONAL SMOOTHING

In this section, we investigate the impact that methods designed to alleviate spatial confounding have on inference for regression coefficients. Formally, we state the following definition of an RSR model:

**Definition 2.**

A *Restricted Spatial Regression Model* is any model of the form (2) with $W$ chosen such that $\mathscr{C}(W) \perp \mathscr{C}(X)$.

Here and throughout the paper $\mathscr{C}(\cdot)$ denotes the column space of a matrix. Both the RHZ and HH models are special cases of this class of models. Note that the ICAR model is not a special case of an RSR model. For RSR models, any results stated for $\beta^X$ (relying on $X$) are also true for $\beta^*$ (by replacing $X$ with $X^*$). Since the PAR model is an attempt to approximate these methods, we will not spend time directly investigating its performance. All of the following results assume a Bayesian analysis that uses flat priors on the regression coefficients and independent gamma priors on the precision parameters $\tau_s$ and $\tau_\epsilon$ with respective shape parameters $a_s$, $a_\epsilon > 0$ and scale parameters $b_s$, $b_\epsilon > 0$.

## 3.1. Properties of Marginal Posterior Distribution.

To begin to investigate the impact RSR models have on inference for regression coefficients, we consider the mean and variance of the marginal posterior distribution of $\beta^X$. As previously noted, RSR models are developed in the areal data setting with the expectation that the marginal posterior mean of $\beta^X$ will be $(X^T X)^{-1} X^T Y$. Working in a geostatistical setting and utilizing an empirical Bayes approach, Hanks et al. (2015) observed this would be the case for a RSR method in that context as well. However, all of the previous work has relied either on observations regarding distributions of $\beta^X$ conditional on the precision parameters (for example, Reich et al. (2006); Hodges and Reich (2010)) or on the full conditional distribution of $\beta^X$ (Hanks et al. (2015)). Theorem 1 offers the first rigorous proof of the existence of the mean and variance of the marginal posterior distributions of $\beta^X$ in RSR models, as well as the forms these moments take. It illustrates that the point estimates obtained from RSR models will be the same as the non-spatial point estimates. It also shows the variance of the distribution looks functionally similar to the variance we would have obtained in the non-spatial model. For the rest of the paper, we adopt the following notational shortcut: $\sigma_\epsilon = \frac{1}{\tau_\epsilon}$.

**Theorem 1.**—Under conditions A(1)- A(5) of the online supplement, a RSR model with $W$ a $n \times q$ matrix will give rise to the marginal posterior distributions of $\beta^X$ such that:

$$\mathrm{E}[\beta^X | Y] = (X^T X)^{-1} X^T Y$$

$$\mathrm{Var}[\beta^X | Y] = (X^T X)^{-1} E[\frac{1}{\tau_\epsilon^{RSR}} | Y] = (X^T X)^{-1} E[\sigma_\epsilon^{RSR} | Y].$$

**Proof:** See Appendix C.1 of the online supplement.

An immediate consequence of Theorem 1 is that the dimension of $W$ will not impact the point estimates for the regression coefficients. Thus, any choice of $q$ linearly independent

column vectors which are orthogonal to $\mathscr{C}(X)$ will give the same point estimates as a HH model using $M_q$. Of course, incorporating spatial dependence in a model is a computationally expensive and often theoretically complicated endeavor. If the concern were simply to obtain point estimates from a non-spatial model, there would be no need to use a spatial model. As noted previously, the focus in fitting a spatial model is often to account for spatial correlation in residuals when computing the posterior variance. Hence, our interest lies in $\text{Var}[\beta^X|Y]$, which Theorem 1 illustrates is a function of $E[\sigma_\epsilon^{R,SR}|Y]$.

The prevalent expectation is that the marginal posterior variances of the regression coefficients in RHZ and HH models will be greater than the marginal posterior variances obtained in the non-spatial model. In settings where RSR models are currently recommended for use, Theorem 2 illustrates that for a broader class of RSR models including the RHZ and HH model, the opposite is true. In the statement of Theorem 2, $r = \dfrac{\tau_s}{\tau_\epsilon}$ is the inverse of the signal-to-noise ratio.

**Theorem 2.—**Under conditions A(1)-A(4) of the online supplement, a RSR model with $W$ a $n \times q$ matrix with orthonormal columns and $F$ a positive definite and symmetric matrix will always result in a marginal posterior variance for $\beta^X_i$, $i = 1, \dots, p$ that is less than or equal to that of the posterior variance that would have been obtained in the non-spatial model whenever the following holds:

$$\frac{E[r|Y]}{E[\tau_s]} < E\left[\frac{1}{T_\epsilon^{NS}}|Y\right] = E[\sigma_\epsilon^{NS}|Y],$$

where $E[\tau_s]$ is the prior expectation of $\tau_s$ under the RSR model.

**<u>Proof:</u>** See Appendix C.2 of the online supplement.

As a practical matter, the conditions set forth in Theorem 2 will often be met. To see this, consider that spatial regression models, including RSR models, are typically used in settings in which the $X\beta^X$ in the NS model does not explain nearly all the variability in $Y$ (i.e., when $E[\sigma_\epsilon^{NS}|Y]$ is not nearly zero). Spatial confounding is thought to be problematic in settings where $r$ is "small" for traditional spatial regression models (see e.g., Reich et al. 2006 and Hodges and Reich 2010). "Small" is usually taken to mean substantially less than 1 (see e.g., Reich et al. 2006 and Prates et al. 2019 who consider values of $r = .1$ and $.2$ respectively). This corresponds to situations where there is residual spatial dependence unexplained by $X$ (and hence larger $E[\sigma_\epsilon^{NS}|Y]$). The better a RSR model is at explaining this residual spatial dependence, the smaller $E(r|Y)$ will be. Finally, in the literature, $E[\tau_s]$ ranges in value from 1 (see e.g., Hodges and Reich, 2010) to 1, 000 (see e.g., Hughes and Haran, 2013). Thus, the conditions in Theorem 2 will typically be satisfied for prominent examples of RSR models.

Theorem 2 offers the first fully rigorous proof detailing the bounds on the variance of the marginal posterior distribution of $\beta^X$ for RSR models. Working in a geostatistical setting with an empirical Bayesian analysis, Hanks et al. (2015) stated without proof that the

inequality related to posterior variances in Theorem 2 should hold for a particular RSR model. However, perhaps because of the difference in settings, this observation has not been perceived as contradicting the expectations for the more commonly studied RHZ and HH models, which are precisely the opposite. Interestingly, the work in Hanks et al. (2015) is often reported as illustrating that RSR methods can have elevated Type-S error rates *under model misspecification* (Prates et al., 2019; Page et al., 2017). Prates et al. (2019) further noted that these elevated Type-S error rates were exacerbated in settings in which *r* is small. Yet, Theorem 2 suggests that RSR models will have elevated Type-S error rates in the absence of model misspecification. In a frequentist setting, the OLS estimators would be normally distributed and an ordering of variance with the same mean would suggest that the $100(1 - a)$% confidence intervals would be nested (with the interval of the distribution with less variance completely contained in the other). In that setting, Theorem 2 would indicate that any RSR model would have higher rates of Type I error and lower rates of Type II error than the non-spatial model. Since the non-spatial model yields estimators with sampling distributions whose coverage probability matches the nominal coverage probability, this indicates that the RSR coverage probability would always be less than the nominal coverage probability under the conditions of Theorem 2. Hence, Theorem 2 would explain the elevated levels of Type-S error observed for RSR models, and it would mean these elevated levels could exist even without model misspecification.

In Bayesian analysis, when inference on regression coefficients is of primary interest, practitioners typically utilize an equal-tail credible interval. This is especially true in settings like those of this paper where improper priors preclude the use of Bayes factors. Because the marginal posterior distribution $f(\beta^X|\boldsymbol{Y})$ is not normal for any of the models discussed, the variance ordering in Theorem 2 need not necessarily indicate a relationship between the credible intervals.

Investigating the relationship between equal-tailed credible intervals for the non-spatial model and a RSR model for all possible choices of graph, data, and spatial basis vectors is difficult because these distributions will not generally be available in closed form. However, for the case that $\beta^X \in \mathbb{R}$, it is possible to make some observations that shed light on the relationship between the credible intervals of a RSR model and the NS model. This is formalized in Theorem 3.

**Theorem 3.**—Assume $\beta^X \in \mathbb{R}$, A.1[3] holds, and A.1[5] holds. Let $g(\beta^X|\boldsymbol{Y})$ be the marginal posterior probability distribution (pdf) from a RSR model with choice of $\boldsymbol{F}$ which is a symmetric and positive definite $q \times q$ matrix. Let $h(\beta^X|\boldsymbol{Y})$ be the marginal posterior pdf from the non-spatial model. Then, $g(\beta^X|\boldsymbol{Y}) = \mathrm{O}(h(\beta^X|\boldsymbol{Y}))$ as $\beta^X \to \infty$ and $\beta^X \to -\infty$.

**Proof:** See Appendix C.3 of the online supplement.

An immediate corollary to this result is the following.

**Corollary 1.**—Define $G$ and $H$ to be the respective cumulative distribution functions of $g(\beta^X|\boldsymbol{Y})$ and $h(\beta^X|\boldsymbol{Y})$. $\exists\ C > 0$ such that:

1.
$$\limsup_{\beta^X \to -\infty} \frac{G(\beta^X|\boldsymbol{Y})}{H(\beta^X|\boldsymbol{Y})} \le C,$$

2.
$$\limsup_{\beta^X \to \infty} \frac{1 - G(\beta^X|\boldsymbol{Y})}{1 - H(\beta^X|\boldsymbol{Y})} \le C$$

Furthermore, for $D_g^*$ and $K(C_q, D_h)$ as defined in the proof in Appendix C of the online supplement, if $D_g^* > K(C_q, D_h)$, then $\exists \, \beta^* > 0$, $\beta_* < 0$ such that:

1. $G(\beta|\boldsymbol{Y}) \le H(\beta|\boldsymbol{Y}) \forall \beta < \beta_*$

2. $1 - G(\beta|\boldsymbol{Y}) \le 1 - H(\beta|\boldsymbol{Y}) \forall \beta > \beta^*$

**Proof:** See Appendix C.3 of the online supplement.

Generally, what these results indicate is that the tails of the marginal posterior distribution for the non-spatial model and the tails of the marginal posterior distribution for RSR models decay roughly at the same rate. The combination of Theorem 1 and Theorem 3 suggests that that RSR models offer inference very similar to a non-spatial model with respect to regression coefficients. Importantly, these results are invariant to the choice of graph, data, and spatial basis vectors. To see how the inference will be similar, note the posterior distribution $f(\beta_{NS}^X|\boldsymbol{Y})$ is symmetric about its mean. While, $f(\beta_{RSR}^X|\boldsymbol{Y})$ is not symmetric about its mean, two applications of Jensen's inequality can bound the distance between a mean and median by the standard deviation. In practice, the standard deviation for an RSR model is always quite small. Thus, RSR and NS models produce marginal posterior distributions with a similar medians and similar tail decay. This result is important because RSR models are not offering much beyond what an NS model offers, yet they are much more computationally expensive. For example, Hughes and Haran (2013) reported that a NS model took under a minute, while their dimension reductions technique took over 6 hours.

If the condition put forth in Corollary (1) holds, it also means that for sufficiently large credible intervals, the credible interval for the RSR model will be completely contained in the corresponding credible interval for the non-spatial model. Importantly, these results require no assumptions about $E(t|\boldsymbol{Y})$. Just as in the frequentist setting, this means that RSR methods will only capture the regression coefficient if the non-spatial model does. Furthermore, the RSR methods will always have higher rates of Type-S error than the non-spatial model - even if there is no spatial confounding. However, in practice determining whether the condition in Corollary (1) is satisfied will be impractical. Therefore, we further the investigation with simulation studies in Section 4.

### 3.2. Implications for Inference.

The premise of RSR models is that incorporating spatial dependence will yield better inference on the regression coefficients. In particular, the HH model and those influenced by it have emphasized choosing spatial basis vectors which are spatially smooth. This emphasis is inherited from work on dimension reduction techniques developed for spatial process modelling where this technique is quite useful (See e.g., Cressie and Wikle, 2015, pages

353-354). Recall our definition of RSR models includes both models designed to capture spatial dependence as well as models with arbitrary choices of $W$ and $F$. Yet, the results of Theorem 1 – Theorem 3 hold whether the models are designed to incorporate spatial dependence or not. A natural question then is: For RSR models, do choices of $W$ (and perhaps $F$) which exhibit spatial patterns behave any differently than those which do not?

Theorem 4 begins to explore the answer to this question. Recall that one of the criticisms of the RHZ model (4) is that the basis vectors $L$ do not appear to be spatially smooth. Theorem 4 states that when $F$ is of the form $W^T Q W$, the inference on the regression coefficients will be invariant to $\mathscr{C}(W)$. This result suggests that the choice of basis vectors which exhibit spatial patterns (such as those from the Moran operator) does not affect inference on regression coefficients in the ways expected. As an example, for moderate $q$, given a selection of spatially smooth basis vectors, we could pick a different basis of the column space (which may not appear spatially smooth), and inference for $\beta^X$ would be equivalent.

**Theorem 4.—**A RSR model with $W_1$ a $n \times q$ matrix with orthonormal columns and $F_1 = W_1^T B W_1$ for arbitrary non-null symmetric $B$ will yield the same marginal posterior distribution $f(\boldsymbol{\beta}^X|\boldsymbol{Y})$ as any other choice $W_2$ with orthonormal columns such that $\mathscr{C}(W_1) = \mathscr{C}(W_2)$ and $F_2 = W_2^T B W_2$.

**Proof:** See Appendix C.4 of the online supplement.

Theorem 4 suggests that the assumptions underlying the development of RSR models may be worth re-evaluating. The addition of the spatial basis vectors in RSR models is assumed to make the regression coefficient estimates less biased and more precise. However, further investigation into RSR models indicates that by choosing these spatial basis vectors to be orthogonal to the column space of $X$, RSR models are effectively transforming a mixed-effects model into an overfit fixed effects model. Thus, RSR models may not be aiding in inference for regression coefficients as expected.

To understand the intuition behind this last claim, it is instructive to take a moment to review the non-spatial Gaussian model. The method of ordinary least squares makes the assumption that all the variability in the direction of $X$ can be described by the linear combination of $X\beta^X$. This assumption results in fitted values which are simply a projection of $Y$ onto the column space of $X$. In other words, the point estimates are $\hat{\boldsymbol{\beta}}^X = (X^T X)^{-1} X^T Y$. Importantly, this assumption is also reflected in the estimated variance of these point estimators. This can be seen in the familiar form for the estimated variance of the OLS estimators, where the $i^{th}$ regression coefficient $\hat{\beta}_i^X$ is simply the $(i, i)$ element of $(X^T X)^{-1} \hat{\sigma}_\epsilon^{NS}$:

$$\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}^X{}_i) = (X^T X)_{ii}^{-1} \hat{\sigma}_\epsilon^{NS} = (X^T X)_{ii}^{-1} \frac{\left\| P_X^\perp Y \right\|^2}{n - p}.$$

The first component reflects the relationship between the different covariates, which is not of interest in this paper. The second term, however, is a consequence of the assumption that the

variation in the direction of $X$ can be explained by $X\boldsymbol{\beta}^X$. Here, $\hat{\sigma}_\epsilon^{NS}$ is simply a function of the magnitude of the component of $Y$ unexplained by any linear combination of the column space of $X$.

In Bayesian analysis, the assumption that all the variability in the direction of $X$ can be explained by a linear combination of the columns of $X$ has analogous implications. For the non-spatial model, the point estimates for the regression coefficients will simply be those of the OLS model. The associated posterior variance for the $i^{th}$ regression coefficient $\boldsymbol{\beta}^X_i$ will be the $(i, i)$ element of $(X^T X)^{-1} \mathrm{E}(\sigma_\epsilon^{NS}|Y)$:

$$\mathrm{Var}[\boldsymbol{\beta}^X_{NS, i}|Y] = (X^T X)^{-1}_{ii} \mathrm{E}(\sigma_\epsilon^{NS}|Y) = (X^T X)^{-1}_{i,i} \frac{b_\epsilon^{-1} + .5\|P_X^{\perp}Y\|^2}{(a_\epsilon - 1) + .5(n - p)}. \tag{5}$$

Although a bit messier than the OLS setting, this relationship again shows that the uncertainty associated with the regression coefficients is a function of the magnitude of the component of $Y$ unexplained by the the column space of $X$. For $n \gg p$ and the typical choices of hyper-parameters ($a_\epsilon, b_\epsilon^{-1} < 1$), this will in fact be quite similar to the estimate obtained via OLS.

In a fixed effects model, researchers are well aware that overfitting the model with the addition of synthetic predictors can distort the true relationship between $Y$ and $X$. For example, consider the non-spatial model with new design matrix $T = [X\ W]$, with $\mathscr{C}(X) \perp \mathscr{C}(W)$. Because $W$ is constructed to belong to the orthogonal complement of the columns space of $X$, the point estimates remain unchanged. If $W$ was a $n \times (n - p)$ matrix of rank $n - p$, $T$ would be a basis of $\mathbb{R}^n$. Thus, by construction, $P_T^{\perp}Y = 0$. More generally, because $(T^T T)^{-1}$ is a block diagonal matrix with diagonals $(X^T X)^{-1}$ and $(W^T W)^{-1}$, the following will be true:

$$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}^X_i) = (X^T X)^{-1}_{ii} \hat{\sigma}_T = (X^T X)^{-1}_{ii} \frac{\|P_T^{\perp}Y\|^2}{n - p - q}.$$

For fixed $X$, $\|P_T^{\perp}Y\|^2$ is monotonically decreasing as a function of the number of columns of $W$. If we add enough (linearly independent) columns to $W$, we will eventually explain the variation in $Y$. Of course, the $n - p - q$ in the denominator ensures that $\hat{\sigma}_T$ is not monotonically decreasing as the number of columns of $W$ increases. However, a basis for the orthogonal complement to $\mathscr{C}(X)$ can still be added in such a way as to ensure that $\hat{\sigma}_T$ will be very, very close to zero (see Appendix D.1 of the online supplement for an example). Thus, even if there is truly no relationship between $Y$ and $X$, with enough synthetic covariates this approach can essentially ensure that the covariates are deemed "significant" (assuming that the columns $X$ are not very strongly correlated). Of course, as noted this problem is well-understood in fixed effect models and this approach is not considered acceptable in the practice of regression analysis.

Returning to RSR models, Theorem 1 illustrates that like the NS model, the point estimates for these models assume that all the variation in the direction of $X$ can be explained by the linear combination $X\boldsymbol{\beta}^X$. However, Theorem 2 suggests that when it comes to the posterior variance, RSR models are going beyond this assumption. In RSR models, $\boldsymbol{F}$ and $\tau_s$ act as regularization parameters that can smooth some of the elements of $\boldsymbol{\delta}$ to 0. These terms make it difficult to find a closed form expression for $E[\sigma_\epsilon^{RSR}|\boldsymbol{Y}]$. However, we can still gain intuition by looking at the behavior of $E[\sigma_\epsilon^{RSR}|\boldsymbol{Y}, r]$, where $r = \frac{\tau_s}{\tau_\epsilon}$ is again the inverse of the signal-to-noise ratio. Straightforward calculations show that $E[\sigma_\epsilon^{RSR}|\boldsymbol{Y}, r]$ can be expressed as:

$$\frac{\frac{1}{b_\epsilon} + \frac{r}{b_s} + .5\boldsymbol{Y}^T(\boldsymbol{P}_X^\perp - \boldsymbol{W}(\boldsymbol{I} + r\boldsymbol{F})^{-1}\boldsymbol{W}^T)\boldsymbol{Y}}{a_\epsilon + a_s + .5(n - p) - 1}.$$

Note that the $(n - p)$ term in the denominator will remain unchanged no matter how many columns are added to the spatial basis matrix $\boldsymbol{W}$. If we consider the behavior of $E[\sigma_\epsilon^{RSR}|\boldsymbol{Y}, r]$ as the signal-to-noise ratio increases (i.e., $r \to 0$) for $\boldsymbol{W}$ a $n \times (n - p)$ matrix of full rank, then

$$\lim_{r \to 0} E[\sigma_\epsilon^{RSR}|\boldsymbol{Y}, r] = \frac{1}{b_\epsilon(a_\epsilon + a_s + .5(n - p) - 1)}.$$

For moderate $p$ (relative to $n$) and the typical choices of $a_\epsilon$ and $b_\epsilon$, this value will be quite small. Note that this is slightly smaller than the value that would be obtained in a Bayesian NS model if the fixed effects $X$ alone completely explained the variation in $\boldsymbol{Y}$ (see Eq. (5) for the case when $\|\boldsymbol{P}_X^\perp\boldsymbol{Y}\|^2 = 0$). However, this reduction is a function of how well the additional columns of $\boldsymbol{W}$ explain the variability in $\boldsymbol{Y}$, and the signal-to-noise ratio will tend to increase as "useful" columns are added to $\boldsymbol{W}$. In other words, if $X$ did a poor job of explaining the variability of $\boldsymbol{Y}$, the addition of columns to $\boldsymbol{W}$ will lead to a more dramatic decrease in $E[\sigma_\epsilon^{RSR}|\boldsymbol{Y}]$ and hence $\text{Var}[\boldsymbol{\beta}^X|\boldsymbol{Y}]$. Because of the regularization induced by $\boldsymbol{F}$ and $\tau_s$, this decrease will not necessarily be monotonic.

To understand the implications of this discussion as well as Theorem 4 on inference, we analyze a real world dataset with a series of RSR models. We consider the SAT data originally analyzed by Wall (2004). We use the data set provided as an example in Bivand et al. (2008). The data include statewide averaged verbal scores on the SAT and the percent of students eligible to take the test in 1999 for the 48 contiguous states of the United States. Let $y_i$ denote the statewide averaged SAT verbal score for the $i$th state and $\boldsymbol{Y} = (y_i, \ldots, y_{48})$. If $X_1 =$ Percent of students eligible to take the exam, then we consider a model with covariates $X^* = [1\ \ X_1\ \ X_1^2]$ with associated regression coefficients $\beta_0, \beta_1, \beta_2$. We fit a series of HH models for various $M_q, q = 1, \ldots, 45$. For all these analyses, $a_s = .5$, $b_s = 2000$ as suggested by Hughes and Haran (2013).

In Fig. 1, we plot the marginal posterior variance for $\beta_0$, $\beta_1$, and $\beta_2$ for each of the various HH models. The green in the graphs represents vectors which are "attractive" and the red represents vectors which are "repulsive." All RSR models result in approximately the same point estimates as those obtained as in the non-spatial model (590.5, −2.84, .022) and have marginal posterior variances less than those obtained in the non-spatial model. Based on the discussion above, we expect that as the number of columns of $W$ increases, the posterior variances should tend to decrease. And in fact, that seems to be the general pattern. Since the "attractive" vectors are added to the model first, these tend to have a larger variance than models that include both "attractive" vectors and "repulsive" vectors. Note that this same pattern can be obtained by adding arbitrary synthetic covariates to a non-spatial model (see, Appendix D.1 of the online supplement).

However, note, that the distinction between "attractive" and "repulsive" eigenvectors is not directly linked to the posterior variances otherwise. For example, the posterior variance obtained for the set of "attractive" vectors proposed by Hughes and Haran (2013) is achieved again for choices that include up to 14 "repulsive" vectors, but is not obtained with the recommended dimension reduction to $.1n$. A restriction to "attractive" eigenvectors was originally proposed as a way to approximate the inference from the RHZ model (See, Section 7 of Hughes and Haran, 2013). However, there is no clear association between the variances that are obtained via the RHZ model and those obtained restricting the spatial basis vectors to "attractive" ones.

It has been claimed that RSR models assume that all the variability in the direction of $X$ can be described by the linear combination $X\beta^X$ (Hodges and Reich, 2010; Hanks et al., 2015). In this section, we point out that the NS already makes this assumption. In order to gain an intuitive understanding of why the posterior variances are RHZ models are always bounded above by the posterior variance of the corresponding NS model, we offer an alternative interpretation which indicates the the reduction in the posterior variance for RHZ models is akin to what would be observed an overfit fixed effects model. When this intuition holds, we would expect that RHZ models would always have a higher Type-S error rate than NS models even in the absence of model misspecification. Importantly, these problems cannot be resolved by selecting different spatial basis vectors within the class of RSR models, and dimension reduction techniques designed to mimic smoothing orthogonally to the fixed effects will likely suffer from the same problems. However, just as is the case for the implications of Theorem 3, exploring this rigorously is impractical. Therefore, in the next section we use simulations to further our investigation.

## 4. Simulation Studies

In this set of simulations, we investigate the relative performances of the ICAR model, the RHZ model, and the NS model for continuous areal data and count areal data. In particular, we explore how and when the RHZ model offers inference different than the inference which would have been obtained with a NS model. We also explore whether RHZ models improve the inference for regression coefficients relative to the ICAR model in the presence of spatial confounding.

Currently, spatial confounding is thought to be a concern when $X$ is spatially dependent and there is residual spatial dependence unexplained by the covariates. Therefore, all of the following simulations involve covariates which are spatially dependent and residual spatial dependence (either with a spatial random effect or with "unobserved" covariate). To simulate the covariates, we rely on previous work in spatial confounding. In the context of areal data, statistics developed by Reich et al. (2006) and Prates et al. (2019) essentially define spatial dependence in $X$ to be the existence of correlation between $X$ and low frequency eigenvectors of the graph Laplacian. These statistics reflect a common assumption in the literature that the eigenvectors of the graph Laplacian $Q$ associated with low eigenvalues exhibit spatially smooth patterns while those associated with high eigenvalues oscillate rapidly. In the continuous case, this phenomenon can be formalized. For the low frequency eigenfunctions, Bernstein estimates show smoothness (see e.g., Zelditch (2017) Theorem 5.17); whereas for high frequency eigenfunctions, rapid oscillation can be shown using upper bounds on the size of nodal domains (see e.g., Zelditch (2017) Theorem 13.1). We are not aware of any formalization of the phenomenon in the discrete case. However it does seem true in practice, and we will utilize this assumption in the following work. Therefore, in the following simulations we generate covariates that are correlated with low frequency eigenvectors of the graph Laplacian. To do so, we utilize the fact that on a connected graph, any column vector $l \in \mathbb{R}^n$ can be written as follows:

$$s_l\sqrt{n-1}\,V\rho_{lV} + \mathbf{1}\bar{l} \tag{6}$$

where $s_l$ and $\bar{l}$ are the sample standard deviation and sample mean of $l$, $\rho_{lV} = (\rho_{l,V_i}, ..., \rho_{l,V_{n-1}}, 0)^T$ is a column vector where $\rho_{l,V_i}$ is the correlation between $l$ and the $i^{th}$ column of $V$.

In the following simulation studies, all models are fit using Markov Chain Monte Carlo algorithms with Gibbs updates and/or Metropolis-Hastings random-walk updates. Gibbs updates are used for all parameters in the Gaussian models and for $\tau_s$ in the Poisson model. For the Poisson model, the $\beta^X$ are updated as in Hughes and Haran (2013) using a random walk with proposal $\beta^{X[j+1]} \sim N(\beta^{X[j]}, U^{-1})$ where $U$ is the estimated asymptotic covariance matrix from the non-spatial model, and $\delta$ is updated with multivariate random walks with a spherical normal proposal. For all models, where applicable, $a_\epsilon = a_s = .01$ and $b_\epsilon = b_s = 100$. For the Gaussian models, $\beta^X$ has a flat prior and for the Poisson model $\beta^X$ has a normal prior with standard deviation of 1000.

To ensure that the Monte Carlo standard errors were sufficiently small, trial simulations were run until a sample path length was found that ensured all models had Monte Carlo standard error $< .01$. For the Gaussian models, a sample path of 80,000 was sufficient; while for the Poisson models, a sample path of 1,000,000 was sufficient. Monte Carlo standard errors were calculated using batch means (Flegal et al., 2008, 2012).

### 4.1. Simulation 1: Gaussian Model.

For this set of simulations we consider Gaussian models of the form (2) for the graph of the 48 contiguous states. We generate data from three models. All generating models are

of the form $Y = 1\beta_0 + X_1\beta^X + v + \epsilon$ where $\epsilon \sim N(0, I)$, $\beta_0 = 1$, $\beta^X = 2$, and $X_1$ is randomly generated to be correlated with $.2n$ of the eigenvectors associated with the lowest non-zero eigenvalues. The vector $v$ is one of three options: $0$ (i.e., the model is a non-spatial linear model), a realization from $L\delta$ with $\delta \sim N(0, L^T QL)$ (the RHZ model (4)), or a realization from the ICAR prior with $\tau_s = 1$ (realizations from the ICAR prior are generated using Algorithm 2.6 in Rue and Held (2005)). For each of the choices of $v$ we simulated 1000 realizations of $Y$ and fit the NS model, RHZ model, and the ICAR model.

Table 2 lists these coverage probabilities based on 95% credible intervals with equal tail weights. 98% of the credible intervals from an analysis using the RHZ model were nested within the corresponding credible intervals derived from the NS model. To explore how the RHZ analysis model differs from the NS analysis model, we consider how the coverage differed for each data set in Table 3. We consider three outcomes:

1. The credible intervals for both models either both contained the true value of $\beta^X$ or both failed to contain it (Agree)

2. the credible interval of the RHZ model contained the true value of $\beta^X$ and the NS model did not (RHZ +)

3. The credible interval of the NS model contained the true value of $\beta^X$ and the RHZ model did not (NS +).

With respect to coverage, both the ICAR model and the NS model obtain higher coverage rates than the RHZ model, even when the data is generated from the RHZ model. When the data is generated from a NS or RHZ model, the differences between the three models are not substantial. These differences are driven primarily by the widths of the credible intervals. The ICAR model tends to have the largest width and the RHZ model tends to have the narrowest credible intervals.

Arguably, the setting of primary interest for this and the simulation studies that follow is when the data is generated with an ICAR prior. This setting is often thought to be most closely related to what would be seen in the real world. When the generating model is the ICAR model, both the RHZ model and the NS model achieve their lowest coverage rates. This suggests, as observed in Prates et al. (2019) and Hanks et al. (2015), that the RHZ model may suffer poor coverage in the presence of spatial random effects which do not operate orthogonally to the fixed effects. In investigating the cause of this poor performance, we found that it was a combination of bias and narrow credible intervals. Here, we define bias as $\beta^X - E[\beta^X|Y]$. In Appendix D.2.1 of the online supplement, we find that the bias for the point estimates for the RHZ and NS models is at its highest when the generating model is an ICAR model. Since by Theorem 1, the point estimates of the RHZ and NS models are the same, the further reduction in coverage rates for the RHZ are caused by narrower credible intervals. There is not a single case in which the credible interval of the RHZ model obtained coverage when the NS model did not.

### 4.2. Simulation 2: Type-S Error in the Gaussian Model.

For this set of simulations we consider Gaussian models of the form (2) for the graph of the 48 contiguous states. We again generate data from three models. All models are of the form $Y = 1\beta_0 + X_1\beta_1^X + X_2\beta_2^X + \nu + \epsilon$ where $\epsilon \sim N(\mathbf{0}, I)$, $\beta_0 = 1$, $\beta_1^X = 2$, and $\beta_2^X = 0$. $X_1$ is again randomly generated by the process described above to be correlated with $.2n$ of the eigenvectors associated with the lowest non-zero eigenvalues, and $X_2$ is independently generated to be correlated with the $.5n$ eigenvectors with the lowest non-zero eigenvalues. Once again, $\nu$ is one of the three options listed in Simulation 1.

With respect to inference on $\beta_1^X$, we again consider coverage rates. For inference on $\beta_2^X$ we now consider the Type-S error rate: the proportion of times a credible interval does not contain zero. As in Simulation 1, we use 95% equal-tailed credible intervals. We are interested in Type-S errors in the context of data which have spatial variation unexplained by the fixed effects. When the generating model is the RHZ or the ICAR model, $\nu$ itself provides such variation. In the context of a non-spatial model, the only way to achieve such dependence is by omitting a spatially varying covariate. Thus, when the generating model is the RHZ or ICAR, we use a design matrix of $X^* = [\mathbf{1}\ X_1\ X_2]$. When the generating model is the non-spatial model, we use $X^* = [\mathbf{1}\ X_2]$, omitting the spatially varying $X_1$. Recall that the associated $X$ will always be used when fitting the ICAR model.

The results are contained in Table 4- Table 7. Both the ICAR model and the NS model have higher coverage rates than the RHZ model in every scenario. As before, there is not a single synthetic data set in which the resulting credible interval obtained coverage for $\beta_1^X$ under the RHZ model and the NS model did not. The RHZ model once again suffers from poor coverage when the generating model is the ICAR model. The differences in the coverage rates are driven by the same factors as discussed in Section 4.1.

In Table 6, the RHZ model has extremely inflated Type-S error rates - even when the data is generated from a RHZ model. Notably, the RHZ model performs the worst with respect to Type-S errors when a spatially varying covariate is omitted from the design matrix. In Appendix D.2.2 of the online supplement, we find that the bias for the NS and RHZ point estimates is at its highest in this setting. This bias leads to the elevated Type-S error rates for the NS model. The extreme inflation from the RHZ model relative to the NS model is caused by narrow credible intervals. We also compare the performance of the RHZ and NS analysis models for each synthetic data set with respect to Type-S error in Table 7. Every data set for which the NS model resulted in a Type-S error, the RHZ model did as well. However, there were a a number of data sets for which the RHZ model resulted in a Type-S error, and the NS model did not.

### 4.3. Simulation Study 3: Investigating Type-S Error in Poisson Model.

Much of the research in the literature as well as the results of this paper focus on Gaussian models. However, in practice most of the models discussed are typically used for areal count data. It is not altogether clear whether work in the linear case can be translated into results for the non-linear case. In this set of simulations, we attempt to investigate whether the

results for the linear case are relevant for count data. To our knowledge, this is the first extensive simulation study of count data for RSR methods.

For this set of simulations, we use the graph of 194 Slovenia municipalities considered in Reich et al. (2006) and Prates et al. (2019). We consider count data generated from a Poisson model of the form $Z = \beta_0 + X_1\beta_1^X + X_2\beta_2^X + v$ using the log link function. The coefficients are defined as follows: $\beta_0 = 1$, $\beta_1^X = 1$, and $\beta_2^X = 0$. We generate 100 datasets for each scenario. The set-up of the simulation study is otherwise the same as in Section 4.2. Fitting the RHZ model for the Poisson model is different from that of the Gaussian model. The details of the RHZ model in the non-linear context are given in Appendix B of the online supplement.

It should be noted that the credible intervals from the RHZ model are nearly always the same as the credible intervals from the NS model. As an example, Fig. 2 shows the credible intervals associated with 25 of the 300 simulated datasets. In Table 8 and Table 9 we see that that the RHZ model performs essentially the same as the NS model. There are three cases in which the RHZ credible intervals lead to different conclusions than the NS credible intervals. In these cases, the respective bounds for the NS and RHZ credible intervals are within .05 of one another.

Unlike in the Gaussian case, the performance of the NS model suffers greatly in the presence of spatial variation unexplained by the covariates. This poor coverage seems to be driven by both bias and very narrow credible intervals. In Appendix D.2.3 of the online supplement, we find that the bias for point estimates obtained by the RHZ and NS models is always greater than that of the ICAR model, regardless of the generating model. This phenomena is distinct than the patterns observed in the Gaussian case. With respect to coverage rates and Type-S error, the bias is compounded with very narrow credible intervals for the RHZ and NS models, as illustrated by Table 9. Although the Poisson model has not been heavily studied in simulations, we note that Hughes and Haran (2013) observed a similar phenomenon with a single simulated dataset. In that setting, the data was simulated from a HH model. The widths of the 95% credible intervals obtained from a NS, HH and ICAR models were, respectively, .24, .24, and 1.84. It is not clear why the credible intervals of the NS model are so narrow.

## 5.  APPLICATION TO SLOVENIA DATA

In light of our findings, we now revisit the dataset which motivated the creation of RSR models. Reich et al. (2006) and Hodges and Reich (2010) used the Slovenia stomach cancer data as an example of how spatial confounding could distort inference on regression coefficients when they observed that the 95% credible interval for the NS model contained 0 and the respective credible interval for the ICAR model did not contain 0. They believed that this discrepancy occurred because $\mathrm{E}[\beta_{ICAR}^X|Z]$ is biased and $\mathrm{Var}[\beta_{ICAR}^X|Z]$ is overinflated when $X$ displays spatial dependence. This perspective and the Slovenia dataset has continued to influence subsequent research (See e.g., Prates et al. (2019)). However, the results of this paper offer an alternative interpretation of why there is a discrepancy between the NS and

ICAR models. To explore this, we now introduce the data and corresponding models in more detail.

The data was collected from 1995 to 2001 for each of the 194 municipalities in Slovenia. A full description of the data is available in Hodges (2016). The response variable $Y$ is defined so that $y_i$ is the observed count of stomach cancer cases for the *ith* municipality. We will let $E_i$ be the expected count of stomach cancer cases and $X = [x_1, \dots, x_{194}]$ be the vector of standardized socioeconomic scores for the municipalities (see, Hodges and Reich, 2010, for a description). We consider five Poisson models for this model: 1) the non-spatial, 2) the RHZ model, 3) the HH model with all the attractive eigenvectors chosen, 4) the HH model with $q = 19$, and 5) the ICAR model. In other words, we assume that conditional on $W_i$, $y_i \overset{ind}{\sim}$ Poisson($\lambda_i$). For the non-spatial model, $\log(\lambda_i) = \log(E_i) + \beta_0 + x_i\beta^X$. The RHZ model and the HH models are of the form: $\lambda_i = \log(E_i) + \beta_0 + x_i\beta^X + W_i\delta$. The details of the choices of $W$ and the prior for $\delta$ are given in Appendix B of the online supplement. The ICAR model takes the form $\lambda_i = \log(E_i) + x_i\beta^X + \delta_i$ where $\delta$ is assumed to have the ICAR prior. For all models, a normal prior with mean $\mathbf{0}$ and standard deviation 1,000 is given to the regression coefficients and $\tau_s$ is given a gamma prior with shape and scale, respectively, .01 and 100.

In Table 10, the ICAR model's credible interval is the only one to include zero. In applications, researchers would typically conclude that the socioeconomic scores were not significant for the ICAR model but were significant in the other models. The prevailing belief in spatial confounding literature is that the RSR and NS models offer better inference than the ICAR model for this dataset. However, in Section 4.3, we observed that if there is residual spatial dependence (either through a missing spatially varying covariate or through a spatial random effect), the NS and RSR models tend to have more biased point estimates than the ICAR models. Similarly, the NS and RSR models suffer from low coverage rates in this setting. Thus, if there is residual spatial dependence unexplained by $X$, the NS and RSR models will likely offer poor inference for $\beta^X$.

Unfortunately, the ICAR model also has drawbacks in this setting. Note unlike in Section 4.3, the point estimates in Table 10 are quite small. In this setting, the ICAR model will still have smaller bias and higher rates of coverage than the NS and RHZ models (see Appendix D.3.3 of the online supplement). However, the Bayesian analogue of power, which we define as percentage of time that the 95% credible interval does not contain 0, will be very low for the ICAR model (see again Appendix D.3.3 of the online supplement). Practically speaking, Theorem 1 – Theorem 3 suggest that RSR models give essentially the same inference as the NS model in the Gaussian case, and Section 4.3 seems to support this insight extends to count data. In particular, we note that for each of the RSR models, the 95% credible intervals are nested within (or equivalent to) the credible interval obtained with the NS model. While it may not be clear when the NS or ICAR model should be preferred, it does seem likely that RSR models will typically be a more computationally intensive method of arriving at the same inference as the NS model with respect to bias and coverage rates.

## 6. Discussion

In this paper we examine the impact that methods designed to alleviate spatial confounding have on inference for regression coefficients of areal SGLMMs. We emphasis that our inferential focus is limited to assessing whether there is a linear *association* between the response and the covariates in settings where an assumption of (spatial) independence for the errors is apparently invalid (e.g., evidence of residual spatial dependence in the non-spatial model). We note that the distinction between inference related to a linear association and causal inference is often blurred in the literature related to spatial confounding. For instance, the work in Paciorek (2010) and Thaden and Kneib (2018) is arguably more closely related to concerns in causal inference. Teasing apart whether common expectations for regression coefficients in spatial models come from an interest in inference for causal relationships or a linear association will likely be important in future related work. However, the implications of RSR methods for formal causal inference are beyond the scope of the current paper.

The results of Section 3 reexamine the intuition that prompted the development of RSR methods. Despite the expectation that RSR methods will result in marginal posterior variances that are greater than would have been obtained in a non-spatial model, we have shown the opposite is usually true. Recently, there has been a call for more research to understand when the RSR methods should be used. This interest is driven in part by the fact that some RSR methods have high rates of Type-S error, particularly when $\frac{\tau_s}{\tau_\epsilon}$ is small.

Currently, RSR methods are recommended when the covariates $X$ are correlated with low-frequency eigenvectors of the graph Laplacian. Section 3 and Section 4 illustrate how and why the elevated Type-S error rates are occurring for Gaussian RSR models. In particular, these results indicate that it is not the spatial structure of the covariates that drives the performance of RSR models. Rather, the posterior distribution of the regression coefficients for RSR models is primarily a function of how well the spatially varying $W$ (in RSR models) explains the residual spatial dependence. The question of when the RSR models should be preferred is typically asked with the implicit comparison is to that of a traditional spatial model. Surprisingly, our results indicate that in all cases studied, one would always be better off using a Gaussian non-spatial model than a Gaussian RSR model. For the count data, the RHZ model and non-spatial model result in almost equivalent inference for the regression coefficients and both perform poorly in the presence of spatial variation unexplained by the covariates. This work also shows that the ICAR model can provide poor inference as well, particularly in the context of small effect sizes.

In spatial statistics, expectations regarding dimension reduction strategies for areal data models have largely been shaped by work in the spatial process modeling realm. For example, in the process model setting, choosing "attractive" spatial vectors of the Moran operator is analogous to modeling low frequency components of spatial variation. This is the basic premise of approaches such as reduced rank approaches in spatial process modeling. It is quite likely that this method performs well for spatial process modeling. However, this paper shows that insights from spatial process modeling can lead to misleading and counterintuitive results when the interest is no longer on predicting the spatial process. Thus,

this work has important implications for dimension reduction when the primary interest is deriving inference about covariate effects.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

### REFERENCES

Banerjee S, Gelfand AE, Finley AO, and Sang H (2008). Gaussian predictive process models for large spatial data sets. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 70(4):825–848. [PubMed: 19750209]

Banerjee S, Wall MM, and Carlin BP (2003). Frailty modeling for spatially correlated survival data, with application to infant mortality in minnesota. Biostatistics, 4(1):123–142. [PubMed: 12925334]

Besag J, York J, and Mollié A (1991). Bayesian image restoration, with two applications in spatial statistics. Annals of the Institute of Statistical Mathematics, 43(1):1–20.

Bivand RS, Pebesma EJ, Gómez-Rubio V, and Pebesma EJ (2008). Applied Spatial Data Analysis with R, volume 747248717. Springer.

Boots B and Tiefelsdorf M (2000). Global and local spatial autocorrelation in bounded regular tessellations. Journal of Geographical Systems, 2(4):319–348.

Bradley JR, Holan SH, Wikle CK, et al. (2015). Multivariate spatio-temporal models for high-dimensional areal data with application to longitudinal employer-household dynamics. The Annals of Applied Statistics, 9(4):1761–1791.

Burden S, Cressie N, and Steel D (2015). The SAR model for very large datasets: a reduced rank approach. Econometrics, 3(2):317–338.

Clayton DG, Bernardinelli L, and Montomoli C (1993). Spatial correlation in ecological analysis. International Journal of Epidemiology, 22(6):1193–1202. [PubMed: 8144305]

Cressie N (1993). Statistics for Spatial Data. Wiley.

Cressie N and Wikle CK (2015). Statistics for Spatio-temporal Data. John Wiley & Sons.

Diggle PJ, Tawn J, and Moyeed R (1998). Model-based geostatistics. Journal of the Royal Statistical Society: Series C (Applied Statistics), 47(3):299–350.

Flegal JM, Haran M, and Jones GL (2008). Markov chain monte carlo: Can we trust the third significant figure? Statistical Science, pages 250–260.

Flegal JM, Hughes J, Vats D, and Dai N (2012). mcmcse: Monte carlo standard errors for MCMC. Riverside, CA and Minneapolis, MN. R package version, pages 1–0.

Fuentes M (2007). Approximate likelihood for large irregularly spaced spatial data. Journal of the American Statistical Association, 102(477):321–331. [PubMed: 19079638]

Guan Y and Haran M (2018). A computationally efficient projection-based approach for spatial generalized linear mixed models. Journal of Computational and Graphical Statistics, 27(4):701–714.

Hanks EM, Schliep EM, Hooten MB, and Hoeting JA (2015). Restricted spatial regression in practice: geostatistical models, confounding, and robustness under model misspecification. Environmetrics, 26(4):243–254.

Hefley TJ, Hooten MB, Hanks EM, Russell RE, and Walsh DP (2017). The bayesian group lasso for confounded spatial data. Journal of Agricultural, Biological and Environmental Statistics, 22(1):42–59.

Hodges JS (2016). Richly Parameterized Linear Models: Additive, Time Series, and Spatial Models using Random Effects. Chapman and Hall/CRC.

Hodges JS and Reich BJ (2010). Adding spatially-correlated errors can mess up the fixed effect you love. The American Statistician, 64(4):325–334.

Hughes J and 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.

Moran PA (1950). Notes on continuous stochastic phenomena. Biometrika, 37(1/2):17–23. [PubMed: 15420245]

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

Paciorek C (2009). Technical vignette 5: Understanding intrinsic gaussian markov random field spatial models, including intrinsic conditional autoregressive models. Technical report.

Paciorek CJ (2010). The importance of scale for spatial-confounding bias and precision of spatial regression estimators. Statistical Science: A Review Journal of the Institute of Mathematical Statistics, 25(1):107. [PubMed: 21528104]

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

Prates MO, Assunção RM, and Rodrigues EC (2019). Alleviating spatial confounding for areal data problems by displacing the geographical centroids. Bayesian Anal., 14(2):623–647.

Reich BJ, Hodges JS, and Zadnik V (2006). Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. Biometrics, 62(4):1197–1206. [PubMed: 17156295]

Rue H and Held L (2005). Gaussian Markov Random Fields: Theory and Applications. CRC press.

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

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

Von Luxburg U (2007). A tutorial on spectral clustering. Statistics and Computing, 17(4):395–416.

Wall MM (2004). A close look at the spatial structure implied by the CAR and SAR models. Journal of Statistical Planning and Inference, 121(2):311–324.

Waller LA and Gotway CA (2004). Applied Spatial Statistics for Public Health Data, volume 368. John Wiley & Sons.

Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wikle CK (2010). Low-rank representations for spatial processes. In Handbook of spatial statistics, pages 114–125. CRC Press.

Zelditch S (2017). Eigenfunctions of the Laplacian on a Riemannian Manifold. CBMS Regional Conference Series in Mathematics. Conference Board of the Mathematical Sciences.
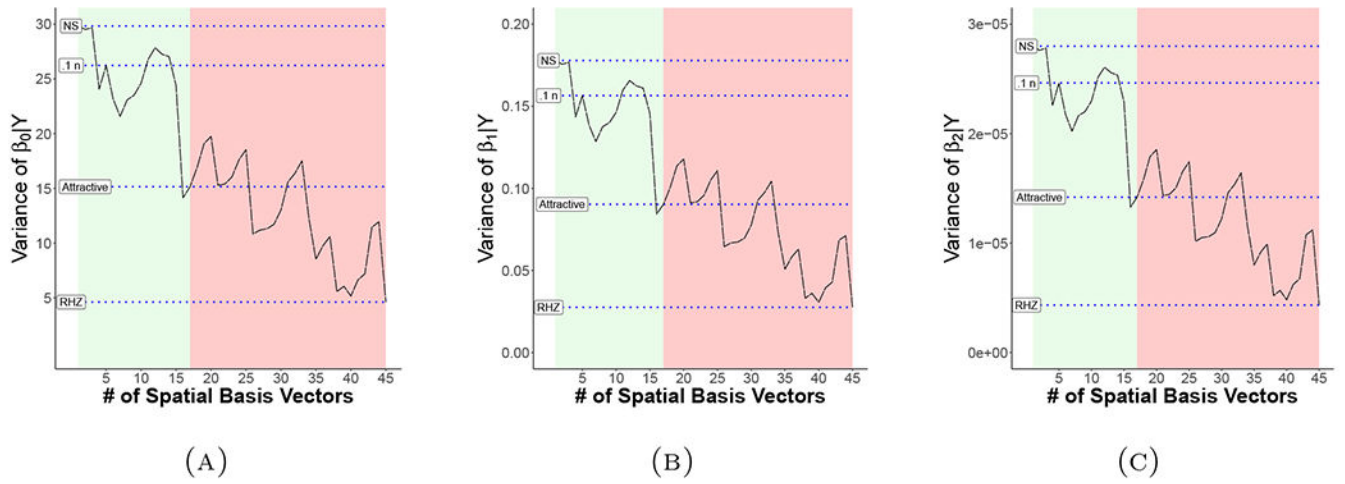
**Figure 1.**

These graphs depict the posterior variance of each of the regression coefficients. Green indicates spatial basis vectors which are "attractive." Red indicates spatial basis vectors which are "repulsive." These figures were constructed using the ggplot2 R package (Wickham, 2016).
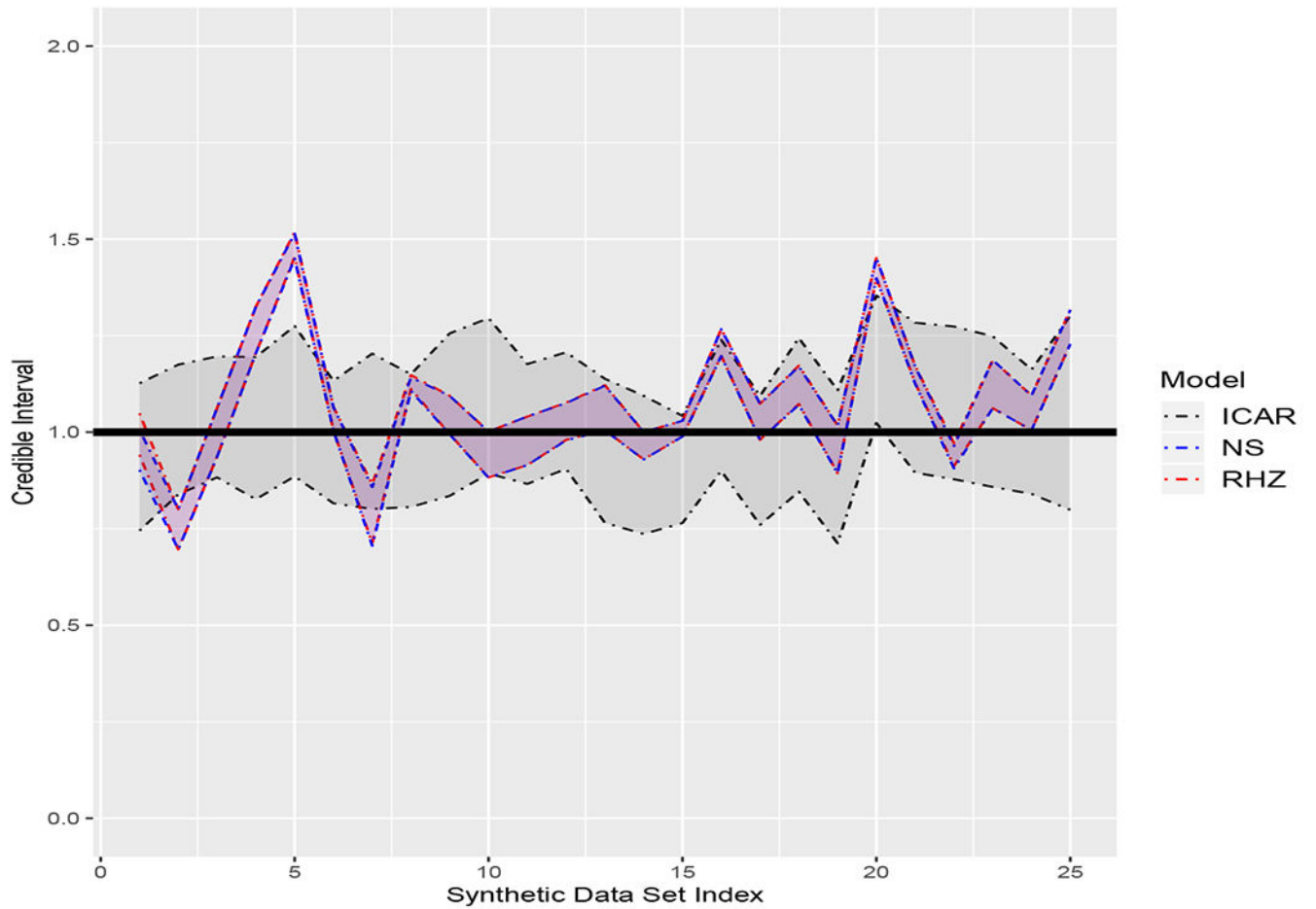
**Figure 2.**
Credible intervals for $\beta_1^X$. This figure was made using the ggplot2 R package (Wickham, 2016).

**Table 1.**

Special Cases of (2)

| Model | | Design Matrix | $W$ | $F$ |
|-------|---|---------------|-----|-----|
| NS | | $X*$ | $\mathbf{0}$ | $\mathbf{0}$ |
| ICAR | | $X$ | $I$ | $Q$ |
| RHZ | | $X*$ | $L$ | $L^T Q L$ |
| HH | | $X*$ | $M_q$ | $M_q^T Q M_q$ |
| PAR | | $X$ | $I$ | $Q^\perp$ |

**Table 2.**

Coverage of $\beta^X$

| Analysis Model | Generating Model | | |
|---|---|---|---|
| | NS | RHZ | ICAR |
| NS | 94.4% | 98.5% | 84.6% |
| RHZ | 93.1% | 93.8% | 72.5% |
| ICAR | 97.0% | 99.4% | 95.3% |

**Table 3.**

Comparison of Coverage of the RHZ and NS models

| Comparisons | Generating Model | | |
|---|---|---|---|
| | **NS** | **RHZ** | **ICAR** |
| Agree | 98.7% | 95.3% | 87.9% |
| RHZ + | 0.0% | 0.0% | 0.0% |
| NS + | 1.3% | 4.7% | 12.1% |

**Table 4.**

Coverage of $\beta_1^X$

| Analysis Model | Generating Model | |
|---|---|---|
| | **RHZ** | **ICAR** |
| NS | 97.9% | 85.3% |
| RHZ | 90.9% | 72.9% |
| ICAR | 99.0% | 95.8% |

**Table 5.**

Comparison of $\beta_1^X$ coverage of the RHZ and NS models

| Comparisons | Generating Model | |
|---|---|---|
| | **RHZ** | **ICAR** |
| Agree | 93.0% | 87.6% |
| RHZ + | 0.0% | 0.0% |
| NS + | 7.0% | 12.4% |

**Table 6.**

Type-S Error of $\beta_2^X$

| Analysis Model | Generating Model | | |
|---|---|---|---|
| | **NS** | **RHZ** | **ICAR** |
| NS | 13.9% | 8.5% | 10.3% |
| RHZ | 80.2% | 18.0% | 21.7% |
| ICAR | 3.7% | 4.8% | 6.5% |

**Table 7.**

Comparison of Type-S of the RHZ and NS models

| Comparisons | Generating Model | | |
|---|---|---|---|
| | NS | RHZ | ICAR |
| Agree | 33.7% | 90.5% | 88.6% |
| RHZ + | 0.0% | 0.0% | 0.0% |
| NS + | 66.3% | 9.5% | 11.4% |

**Table 8.**

Coverage of $\beta_1^X$

| Analysis Model | Generating Model | |
|---|---|---|
| | **RHZ** | **ICAR** |
| NS | 32% | 14% |
| RHZ | 33% | 14% |
| ICAR | 97% | 98% |

**Table 9.**

Type-S Error of $\beta_2^X$

| Analysis Model | Generating Model | | |
|---|---|---|---|
| | NS | RHZ | ICAR |
| NS | 68.0% | 79.0% | 67% |
| RHZ | 67.0% | 79.0% | 67% |
| ICAR | 7.0% | 7.0% | 3.0% |

**Table 10.**

Summaries of Posterior Distribution $\beta^X$

| Model | Posterior Mean | 95% Credible Interval |
|-------|----------------|------------------------|
| NS | −.137 | (−.175, −.098) |
| RHZ | −.137 | (−.175, −.098) |
| $HH_{19}$ | −.118 | (−.161, −.074) |
| $HH_{all\ pos.}$ | −.101 | (−.151, −.049) |
| ICAR | −.018 | (−.096, .065) |