

Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture

2015 - 27th Annual Conference Proceedings

On fixed effects estimation in spline-based semiparametric regression for spatial data

Guilherme Ludwig

University of Wisconsin - Madison, gvludwig@stat.wisc.edu

Jun Zhu

University of Wisconsin - Madison, jzhu@stat.wisc.edu

Chun-Shu Chen

National Changhua University of Education, cschen@cc.ncue.edu.tw

Follow this and additional works at: <https://newprairiepress.org/agstatconference>

 Part of the [Agriculture Commons](#), [Applied Statistics Commons](#), and the [Statistical Models Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

Recommended Citation

Ludwig, Guilherme; Zhu, Jun; and Chen, Chun-Shu (2015). "On fixed effects estimation in spline-based semiparametric regression for spatial data," *Conference on Applied Statistics in Agriculture*.

<https://doi.org/10.4148/2475-7772.1088>

This Event is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.

On fixed effects estimation in spline-based semiparametric regression for spatial data

Guilherme Ludwig^a, Jun Zhu^b, Chun-Shu Chen^c

^a*Department of Statistics, University of Wisconsin-Madison*

^b*Department of Statistics and Department of Entomology, University of Wisconsin-Madison*

^c*Institute of Statistics and Information Science, National Changhua University of Education, Changhua, Taiwan*

Abstract

Spline surfaces are often used to capture spatial variability sources in linear mixed-effects models, without imposing a parametric covariance structure on the random effects. However, including a spline component in a semiparametric model may change the estimated regression coefficients, a problem analogous to spatial confounding in spatially correlated random effects. Our research aims to investigate such effects in spline-based semiparametric regression for spatial data. We discuss estimators' behavior under the traditional spatial linear regression, how the estimates change in spatial confounding-like situations, and how selecting a proper tuning parameter for the spline can help reduce bias.

Key words: Semiparametric regression, spatial interaction, spatial statistics

1. Introduction

Spatial confounding refers to correlation between a spatial random error and covariates in a spatial regression model. The inclusion of a spatial random error can impact the estimates of the regression coefficients β in ways that are difficult to anticipate. This issue, as we outline in the following paragraphs, is found in many spatial statistics applications. The purpose of our research is to investigate an analogy to spatial confounding and its effect in semiparametric regression for spatial data.

Reich et al. (2006) derived bias and variance inflation terms for the regression coefficients in spatial lattice data models, but proposed no correction to the spatial confounding effect. Hodges and Reich (2010) further developed a Bayesian approach for spatial lattice data, and outlined common mechanisms of spatial confounding. They considered two situations, namely, (A) "Random effects are a formal device to implement a smoother", but also

Email addresses: gvludwig@stat.wisc.edu (Guilherme Ludwig), jzhu@stat.wisc.edu (Jun Zhu), cschen@cc.ncue.edu.tw (Chun-Shu Chen)

a device for when the analyst either knows there are missing covariates, or was unable to collect them, and therefore adds a random effect to the model to try to recover these effects. This strategy corresponds to approaches seen in Clayton et al. (1993), Ruppert et al. (2003), Reich et al. (2006) and Stroup (2012). Hodges and Reich (2010) concluded that for these scenarios, fitting a spatial random effects model can either remove or introduce bias, and is not conservative in the sense that if there's no actual spatial effect present, the model might produce spurious results. The second mechanism of spatial confounding is (B) "The random effect is a Scheffé-style random effect", which depends on how the covariates are observed. Fixed and known covariates result in unbiased estimates for the regression coefficients β , but when the regression covariates are actually randomly drawn from a probability distribution and the model is conditioned on their realization (as in Paciorek, 2010), then fitted models with or without a spatial random effect can be biased. Hodges and Reich (2010) suggested to attribute most of the variability of the response to the regression effect, by fitting the data to covariates alone first, and smoothing the residuals of that fit. Another strategy for spatial lattice data was employed by Hughes and Haran (2013), in which the spatial component is made orthogonal to the covariates, and a subset of the eigenvectors is used to model spatial variability, mitigating confounding.

In a different approach, Paciorek (2010) focused on the spatial regression where both covariates and the spatial random effects are realizations of a spatial process, and showed that bias is introduced in the estimators of β even when the variance of the spatial process is known. More recently, Hanks et al. (2015) considered spatial regression models and proposed to force a fitted spatial random effect to be orthogonal to the covariates. They emphasized that such approach works only to recover the ordinary least squares point estimates of β . Another important observation made in Hanks et al. (2015) is that even when fixed and random effects are generated independently, if both are spatially smooth, then they can still be collinear and unconditional inference on β through ordinary least squares or the orthogonal approach of their study could lead to spurious conclusions about the importance of the covariates in the model.

Most of the work on spatial confounding was done for Bayesian models and methods (Hodges and Reich, 2010; Hughes and Haran, 2013) or, more recently, kriging (Hanks et al., 2015). Here we consider a common semiparametric approach to spatial regression and in particular the role of spline-based random effects and spatial confounding. Research on semiparametric models with splines is available under specific conditions – the most important being that the unknown functional component of the data model is deterministic and smooth with independent errors. Rice (1986) derived the asymptotics for semiparametric univariate spline models, and argued that automatic smoothing (such as by cross-validation or generalized cross-validation) is invalid for the estimation of β , a conclusion shared by Green et al. (1985). Speckman (1988) introduced a different penalty such that the fixed components have negligible asymptotic bias, but in the context of kernel smoothing and for independent errors. Most of these results were summarized in Hastie and Tibshirani

(1990). More recently, Schimek (2000) considered computational implementation of semi-parametric spline models. The problem of spatial confounding is related to the *concurvity* issue observed in Hastie and Tibshirani (1990), in which linear combinations of smoother terms might approximate a covariate or another smoother term.

Here we focus on thin-plate splines (Wahba, 1990b), which are a natural extension of smoothing splines to spatial problems, and share a strong connection with kriging (Cressie, 1993). This connection was debated in a series of letters (Cressie, 1989; Wahba, 1990a) which concluded that splines are appropriate for surfaces that are deterministic, or deterministic with independent noise (Cressie, 1990). Later, Altman (2000) argued that splines often do not have the optimal prediction ability of kriging because of improper selection of the tuning parameter, since this selection does not reflect the dependency seen in spatial processes. On the other hand, Nychka (2000) argued that geostatistics literature does not emphasize enough that kriging, being a linear combination of basis functions prescribed by the covariance, can be seen as a smoother, and therefore, important parameters of the covariance structure can be found by cross-validation based methods or restricted maximum-likelihood.

The remainder of this document is organized as follows. In Section 2, we present a spatial data model where the spatial random effect is a stochastic spatial process, and consider thin-plate splines for model fitting. We then investigate the properties of the model estimators and develop a tuning parameter selection procedure. A simulation study is conducted in Section 3. In Section 4, we investigate the spatial confounding issue in the semiparametric thin-plate splines regression.

2. Semiparametric spline-based regression

2.1. Spatial data model

Let $\mathcal{D} \subset \mathbb{R}^2$ be a spatial domain of interest. Consider the spatial linear regression model for the response variable,

$$Y(\mathbf{s}) = \beta_0 + \beta_1 x_1(\mathbf{s}) + \cdots + \beta_p x_p(\mathbf{s}) + \eta(\mathbf{s}) + \varepsilon(\mathbf{s}), \quad i = 1, \dots, n, \quad (1)$$

where $\mathbf{s} = (s_1, s_2) \in \mathcal{D}$, $\eta(\mathbf{s})$ is a spatial process with spatial covariance function $\gamma(\mathbf{s}, \mathbf{s}^*) = \text{Cov}(Y(\mathbf{s}), Y(\mathbf{s}^*))$, $\varepsilon(\mathbf{s})$ are i.i.d. measurement errors with $\mathbb{E}(\varepsilon(\mathbf{s})) = 0$ and $\text{Var}(\varepsilon(\mathbf{s})) = \sigma^2$. Assume that the spatial process $\eta(\mathbf{s})$ has mean zero and is mean squared differentiable, in the sense that

$$\lim_{\|\delta\| \rightarrow 0} \mathbb{E}(\|\eta(\mathbf{s}) - \eta(\mathbf{s} + \delta)\|^2) = 0,$$

where $\|\cdot\|$ denotes the L_2 norm. Mean squared differentiability implies $\eta(\mathbf{s})$ has a spatial covariance function $\gamma(\mathbf{s}, \mathbf{s}^*)$ that is continuous. Define $\phi_k(\mathbf{s})$, $k = 1, 2, \dots$ to be continuous orthonormal functions such that

$$\int_{\mathcal{D}} \gamma(\mathbf{s}, \mathbf{s}^*) \phi_k(\mathbf{s}^*) d\mathbf{s}^* = \xi_k \phi_k(\mathbf{s}),$$

where $\xi_k > 0, k = 1, 2, \dots$. Then, the spatial covariance function γ has a spectral decomposition

$$\gamma(\mathbf{s}, \mathbf{s}^*) = \sum_{k=1}^{\infty} \xi_k \phi_k(\mathbf{s}) \phi_k(\mathbf{s}^*).$$

Furthermore, $\eta(\mathbf{s})$ admits a Karhunen-Loève decomposition

$$\eta(\mathbf{s}) = \sum_{k=1}^{\infty} \xi_k^{1/2} Z_k \phi_k(\mathbf{s}),$$

where $\{Z_k\}_{k=1}^{\infty}$ are i.i.d. $N(0, 1)$ (Loève, 1978).

With the response variable $Y(\mathbf{s})$ observed at sampling sites $\mathbf{s}_1, \dots, \mathbf{s}_n$, let $\mathbf{y} = (y(\mathbf{s}_1), \dots, y(\mathbf{s}_n))'$ denote the vector of observations. Let \mathbf{X} be an $n \times (p + 1)$ design matrix with a column of ones and covariates $\mathbf{x}_1, \dots, \mathbf{x}_p$ observed at $\mathbf{s}_1, \dots, \mathbf{s}_n$, and $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)'$ denote the regression coefficients. Let $\boldsymbol{\eta} = (\eta(\mathbf{s}_1), \dots, \eta(\mathbf{s}_n))'$ with $\mathbb{E}(\boldsymbol{\eta}) = \mathbf{0}$ and $\text{Var}(\boldsymbol{\eta}) = \sum_{k=1}^{\infty} \xi_k \boldsymbol{\phi}_k \boldsymbol{\phi}_k'$, where $\boldsymbol{\phi}_k = (\phi_k(\mathbf{s}_1), \dots, \phi_k(\mathbf{s}_n))'$. Suppose a standard linear regression model is fitted to spatial data, with ordinary least squares estimates of $\boldsymbol{\beta}$ given by

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}.$$

Since $\mathbb{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$, the estimates $\tilde{\boldsymbol{\beta}}$ are unbiased. However,

$$\text{Var}(\tilde{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} + \sum_{k=1}^{\infty} \xi_k \boldsymbol{\psi}_k \boldsymbol{\psi}_k', \quad (2)$$

where $\boldsymbol{\psi}_k = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\phi}_k$. The second term in (2) is due to the variance of the spatial process.

2.2. Thin-plate spline estimation

Often one wishes to fit the spatial data model (1) without pre-specifying a spatial covariance structure for the data, but somehow control for this spatial variability. In this case, a spline provides a viable approach (Stroup, 2012). In this section, we describe estimation based on thin-plate spline and its properties under the spatial data model (1).

Definition 2.1 (Thin-plate spline). A thin-plate spline (Wahba, 1990b) is the solution of the variational problem,

$$f_\lambda = \arg \min_{g \in \mathcal{W}_2^2} \sum_{i=1}^n \{Y(\mathbf{s}_i) - g(\mathbf{s}_i)\}^2 + \lambda J[g], \quad (3)$$

where \mathcal{W}_2^2 is the class of functions $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ that are differentiable and have bounded second derivatives, $J[g]$ is a roughness penalty on g , and $\lambda > 0$ is a tuning parameter.

Thin-plate splines are in general a combination of low-order polynomials and a linear combination of radial basis functions. For example, for a penalty based on the squared norm of the bending energy,

$$J[g] = \int_{\mathbb{R}^2} \left\{ \frac{\partial^2}{\partial s_1^2} g(\mathbf{s}) \right\}^2 + 2 \left\{ \frac{\partial^2}{\partial s_1 \partial s_2} g(\mathbf{s}) \right\}^2 + \left\{ \frac{\partial^2}{\partial s_2^2} g(\mathbf{s}) \right\}^2 ds, \quad (4)$$

where $\mathbf{s} = (s_1, s_2)'$, we have a thin-plate spline that has a closed form given by

$$f(\mathbf{s}) = \theta_0 + \theta_{s_1} s_1 + \theta_{s_2} s_2 + \sum_{i=1}^n \theta_i \varphi_i(\mathbf{s}),$$

where θ_0 is the intercept, θ_{s_1} and θ_{s_2} are the slopes for the (s_1, s_2) coordinates of \mathbf{s} , $\{\theta_i\}_{i=1}^n$ are the spline coefficients, subject to the constraints $\sum_{i=1}^n \theta_i = \sum_{i=1}^n \theta_i s_{1,i} = \sum_{i=1}^n \theta_i s_{2,i} = 0$, and φ_i are the collection of thin-plate spline basis functions, given by

$$\varphi_i(\mathbf{s}) = \|\mathbf{s} - \mathbf{s}_i\|^2 \log \|\mathbf{s} - \mathbf{s}_i\|.$$

The thin-plate spline can also be written as $f_\lambda(\mathbf{s}) = \theta_0 + \theta_{s_1} s_1 + \theta_{s_2} s_2 + \mathbf{\Phi}(\mathbf{s})\boldsymbol{\theta}$, where $\mathbf{\Phi}(\mathbf{s}) = (\varphi_1(\mathbf{s}), \dots, \varphi_n(\mathbf{s}))$, and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)'$. The next step is to fit a semiparametric model with covariates and a thin-plate spline to the spatial data. It requires adjustments for identifiability: θ_0 and β_0 are redundant. Furthermore, since $J[\theta_0 + \theta_{s_1} s_1 + \theta_{s_2} s_2] \equiv 0$, we treat s_1 and s_2 as part of the design matrix \mathbf{X} as well, so now each row of \mathbf{X} is given by $\mathbf{x}_i = (1, s_{1,i}, s_{2,i}, x_{1,i}, \dots, x_{p,i})$, $i = 1, \dots, n$. Let $\mathbf{\Phi}$ be the $n \times n$ matrix of basis functions $(\mathbf{\Phi}(\mathbf{s}_1)', \dots, \mathbf{\Phi}(\mathbf{s}_n)')$. The coefficients $\theta_0, \theta_{s_1}, \theta_{s_2}, \beta_1, \dots, \beta_p, \theta_1, \dots, \theta_n$ are estimated by minimizing

$$Q(\boldsymbol{\beta}, \boldsymbol{\theta}) = \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{\Phi}\boldsymbol{\theta}\|^2 + \lambda J[f] \quad (5)$$

subject to $\mathbf{X}'\boldsymbol{\theta} = \mathbf{0}$. The roughness penalty for thin-plate splines can be rewritten as

$$J[f] = \boldsymbol{\theta}'\mathbf{R}\boldsymbol{\theta} \quad (6)$$

where \mathbf{R} is an $n \times n$ matrix with entries

$$\mathbf{R}_{i,i'} = \varphi_i(\mathbf{s}_{i'}). \quad (7)$$

The result in (7) is particular to thin-plate splines (Wahba, 1990b, p. 32). The solutions for (5), for fixed λ are

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{f}} \end{pmatrix} = \begin{pmatrix} \{\mathbf{X}'(\mathbf{I} - \mathcal{S}_\lambda)\mathbf{X}\}^{-1}\mathbf{X}'(\mathbf{I} - \mathcal{S}_\lambda)\mathbf{y} \\ \mathcal{S}_\lambda(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}), \end{pmatrix} \quad (8)$$

where $\mathcal{S}_\lambda = \mathbf{\Phi}(\mathbf{\Phi}'\mathbf{\Phi} + \lambda\mathbf{R})^{-1}\mathbf{\Phi}'$ is the smoother matrix. We will refer to $\hat{\boldsymbol{\beta}}$ and \hat{f}_λ as the penalized least squares estimates. From (8), the regression coefficients estimates $\hat{\boldsymbol{\beta}}$ can be viewed as weighted least squares, with weights based on the complement of the smoother function. The estimator \hat{f}_λ is obtained by applying the smoother matrix to the detrended data.

A feature of thin-plate splines is that it shares a similar structure to the universal kriging, which is the best linear unbiased prediction (Cressie, 1993), in the sense that the optimization of problem (5) using variogram entries instead of thin-plate spline basis functions results in kriging estimates.

2.3. Bias and variance of estimators

Next, we show the expected value and the variance of the penalized least squares estimators.

Proposition 2.1. *Under the spatial data model (1), for any fixed choice of λ , the penalized least squares estimators in (8) are unbiased.*

Proposition 2.2. *Under the spatial data model (1), for any fixed choice of λ , the variance of $\hat{\boldsymbol{\beta}}$ can be written as*

$$\begin{aligned} \text{Var}(\hat{\boldsymbol{\beta}}) &= \sigma^2\{\mathbf{X}'(\mathbf{I} - \mathcal{S}_\lambda)\mathbf{X}\}^{-1}\mathbf{X}'(\mathbf{I} - \mathcal{S}_\lambda)^2\mathbf{X}\{\mathbf{X}'(\mathbf{I} - \mathcal{S}_\lambda)\mathbf{X}\}^{-1} + \sum_{k=1}^{\infty} \xi_k \boldsymbol{\psi}_{\lambda,k} \boldsymbol{\psi}'_{\lambda,k} \\ &= \mathbb{E} \left(\text{Var}(\hat{\boldsymbol{\beta}}|\boldsymbol{\eta}) \right) + \text{Var} \left(\mathbb{E}(\hat{\boldsymbol{\beta}}|\boldsymbol{\eta}) \right) \\ &\equiv \text{Var}_{\mathbf{X}}(\hat{\boldsymbol{\beta}}) + \text{Var}_{\boldsymbol{\eta}}(\hat{\boldsymbol{\beta}}) \end{aligned}$$

where $\boldsymbol{\psi}_{\lambda,k} = \{\mathbf{X}'(\mathbf{I} - \mathcal{S}_\lambda)\mathbf{X}\}^{-1} \mathbf{X}'(\mathbf{I} - \mathcal{S}_\lambda)\boldsymbol{\phi}_k$.

We focus now on $\text{Var}(\hat{\boldsymbol{\beta}})$, since our main interest is in the estimation of the regression coefficients. There is a trade-off in the two components of $\text{Var}(\hat{\boldsymbol{\beta}})$: $\text{Var}_{\mathbf{X}}(\hat{\boldsymbol{\beta}})$, which is the variability of $\hat{\boldsymbol{\beta}}$ attributed to the columns of \mathbf{X} , and $\text{Var}_{\boldsymbol{\eta}}(\hat{\boldsymbol{\beta}})$ which is due to the spatial component $\boldsymbol{\eta}(\mathbf{s})$. The component $\text{Var}_{\mathbf{X}}(\hat{\boldsymbol{\beta}})$ reduces to $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ when $\lambda = 0$. For $\lambda > 0$, there is a shrinkage effect similar to the one observed in ridge regression. Thus, the tuning parameter λ does not only control the smoothness of \hat{f}_λ , but also a variability trade-off in the estimation of $\boldsymbol{\beta}$. On the other hand, the component $\text{Var}_{\boldsymbol{\eta}}(\hat{\boldsymbol{\beta}})$ depends on the relation between \mathbf{X} and the eigenfunctions $\boldsymbol{\phi}_k(\mathbf{s})$, reflecting any non-smooth effects of $\boldsymbol{\eta}$ that are projected onto the columns of \mathbf{X} , due to the weight $\mathbf{I} - \mathcal{S}_\lambda$. An illustration of this trade-off, for a single $\hat{\beta}_1$, where $x_1(\mathbf{s})$ is uncorrelated with $\boldsymbol{\eta}$, is shown in Figure 1. We can see that there is a local minimum, as a function of λ , for $\text{Var}_{\mathbf{X}}(\hat{\beta}_1)$, but $\text{Var}_{\boldsymbol{\eta}}(\hat{\beta}_1)$ seems to increase monotonically, which is reflecting any non-smooth variation in $\boldsymbol{\eta}$ that is collinear with \mathbf{X} . The minimum of $\text{Var}(\hat{\boldsymbol{\beta}})$ would constitute an ideal smoothing parameter for the model, when the interest lies only in $\hat{\boldsymbol{\beta}}$, since it minimizes $\text{MSE}(\hat{\boldsymbol{\beta}})$.

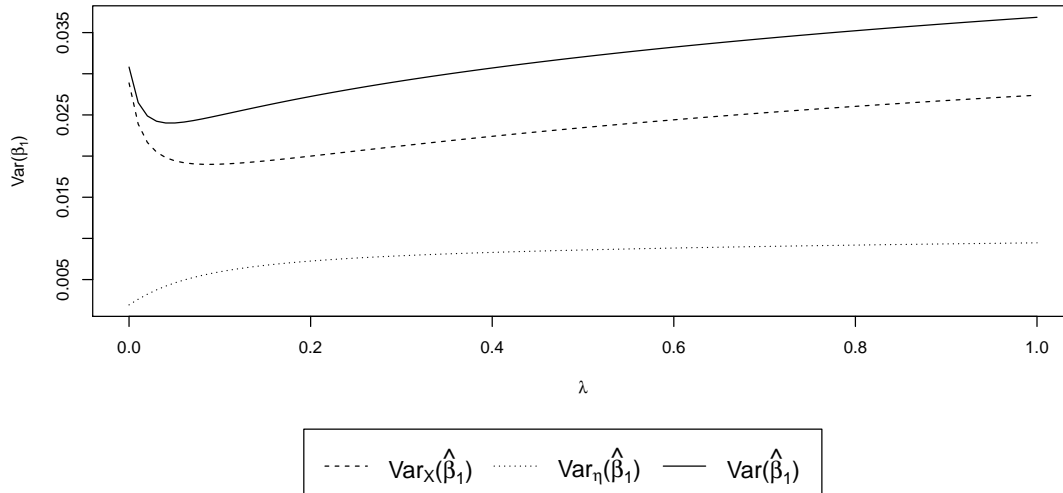


Figure 1: An illustration of the trade-off between the two components of $\text{Var}(\hat{\beta}_1)$ when $\eta(\mathbf{s})$ is independent of \mathbf{X} . $\text{Var}_{\mathbf{X}}(\hat{\beta}_1)$ is the variance attributed to the design matrix \mathbf{X} and $\text{Var}_{\eta}(\hat{\beta}_1)$ is the variance attributed to the spatial process $\eta(\mathbf{s})$.

2.4. Selection of the tuning parameter

Selection of the tuning parameter is usually based on cross-validation, which approximates the prediction squared error (PSE) (Hastie and Tibshirani, 1990). In other words, let

$$\text{PSE}(\lambda) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}\{y^*(\mathbf{s}_i) - \hat{y}_{\lambda}(\mathbf{s}_i)\}^2,$$

where $y^*(\mathbf{s}_i)$ is a new observation at sampling site \mathbf{s}_i , and $\hat{y}_{\lambda}(\mathbf{s}_i) = \mathbf{x}(\mathbf{s}_i)' \hat{\beta} - \hat{f}_{\lambda}(\mathbf{s}_i)$. If we let

$$\text{CV}(\lambda) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}\{y(\mathbf{s}_i) - \hat{y}_{\lambda}^{-i}(\mathbf{s}_i)\}^2,$$

where $\hat{y}_{\lambda}^{-i}(\mathbf{s}_i)$ is the prediction from a model fitted while excluding observation i , we have

$$\mathbb{E}\{\text{CV}(\lambda)\} \approx \text{PSE}(\lambda).$$

The λ^* that minimizes $\text{CV}(\lambda)$ is not necessarily optimal for $\text{MSE}(\hat{\beta})$. Since the estimators of β are unbiased under model (1), we can minimize $\text{tr}(\text{Var}(\hat{\beta}))$ instead. However, since $\eta(\mathbf{s})$ is not observable directly, this parameter is optimal only for minimizing $\text{Var}_{\mathbf{X}}(\hat{\beta})$, (i.e., in the ridge regression sense). Suppose the effect of $\text{Var}_{\eta}(\hat{\beta})$ is small. We now develop a new algorithm to find this optimal value:

Computational algorithm

- Find λ_{GCV}^* that minimizes

$$\text{GCV}(\lambda) = \frac{n \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\{n - \text{tr}(\mathbf{A}_\lambda)\}^2}$$

where $\hat{\mathbf{y}} = (\hat{y}(\mathbf{s}_1), \dots, \hat{y}(\mathbf{s}_n))' = \mathbf{A}_\lambda \mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{f}}_\lambda$.

- For a grid of positive weights $\{w\}$, evaluate $\text{tr}(\text{Var}_{\mathbf{x}}(\hat{\boldsymbol{\beta}}))$, where $\lambda = w\lambda_{\text{GCV}}^*$. Then, choose the λ value that minimizes the trace.

We use GCV instead of CV for faster computational speed.

3. Simulation study

3.1. Simulation setup

We conducted a simulation study to evaluate the effect of spatial dependence on regression coefficient estimates, as well as evaluating the performance of the selection procedure for the tuning parameter λ . We simulated from the spatial model (1)

$$Y(\mathbf{s}) = \boldsymbol{\beta}'\mathbf{x}(\mathbf{s}) + \eta(\mathbf{s}) + \varepsilon(\mathbf{s}),$$

where the spatial effect $\eta(\mathbf{s})$ is a Gaussian random field with a Matérn covariance function

$$\gamma(d; \rho, \kappa, \sigma_\eta^2) = \frac{\sigma_\eta^2}{2^{\kappa-1}\Gamma(\kappa)} \left(\sqrt{2\kappa}\frac{d}{\rho}\right)^\kappa K_\kappa\left(\sqrt{2\kappa}\frac{d}{\rho}\right),$$

where K_κ is the modified Bessel function of second kind. The parameter κ controls the smoothness of the process, with larger values of κ corresponding to smoother realizations. A process with Matérn covariance is κ -times differentiable. The parameter σ_η^2 is the variance of $\eta(\mathbf{s})$, the parameter ρ controls the scale of dependency, which means that larger values of ρ correspond to stronger spatial dependence. Figure 2 shows a realization of a Gaussian process simulated with the same random seed, but with different choices of ρ and κ for the Matérn covariance.

The simulation setup is as follows.

- The number of simulations was $S = 400$. A total of $n = 100$ locations $\mathbf{s}_1, \dots, \mathbf{s}_n$ were sampled randomly over the spatial domain $[0, 1] \times [0, 1]$, and were used as the sampling sites.
- Two sets of covariates, $x_1(\mathbf{s}_i), x_2(\mathbf{s}_i)$ were sampled from $N(0, 1)$, and then treated as fixed, for $i = 1, \dots, n$.

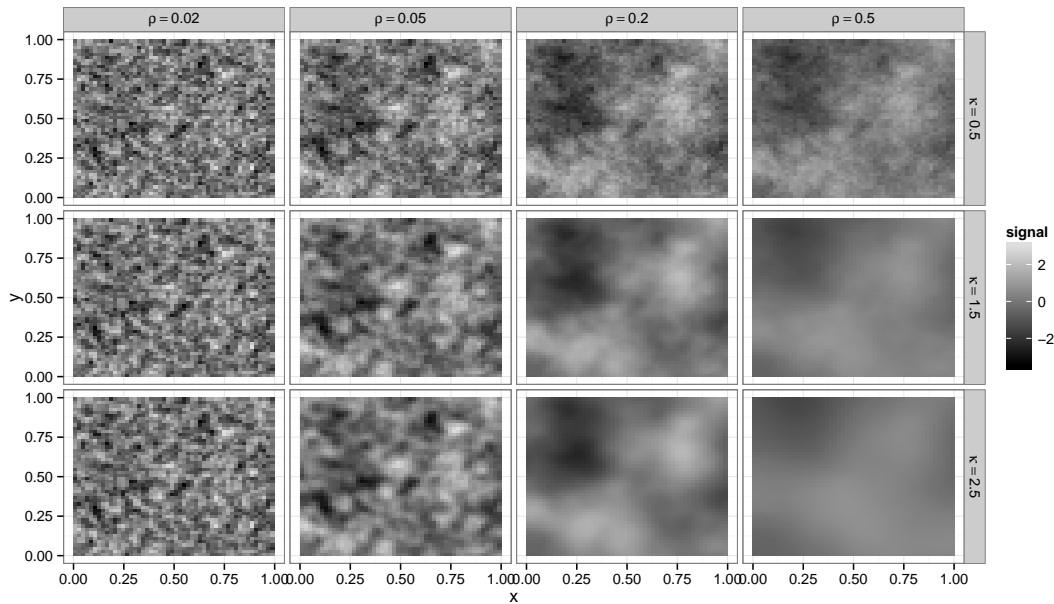


Figure 2: Examples of Matérn covariance processes for different parameters: each panel shows the realization of a single Gaussian process with choices of parameters $\kappa = 0.5, 1.5, 2.5$ and $\rho = 0.02, 0.05, 0.2, 0.5$.

- The intercept and slopes were set to $\beta_0 = \beta_1 = \beta_2 = 1$.
- The spatial process $\eta(\mathbf{s})$ has a Matérn covariance function with parameters $\kappa = 0.5, 1.5$ or 2.5 and $\rho = 0.02, 0.05, 0.2$ or 0.5 , while the variance is $\sigma_\eta^2 = 1$.
- The measurement errors $\varepsilon(\mathbf{s}_i)$ are i.i.d. $N(0, 1)$, for $i = 1, \dots, n$.

We compared the estimation of β using the standard linear model estimates (denoted hereafter as LM), the restricted maximum likelihood estimates (denoted as RM), and the semiparametric thin-plate spline fit, the latter using regular GCV (denoted as SP) and our correction to GCV to select the tuning parameter (denoted as SPc). For the RM method we estimated the Matérn covariance function parameters with the `geoR` package (Ribeiro Jr. and Diggle, 2015). For the correction in the SPc approach, we searched w in the interval of $[0.5, 1.5]$.

3.2. Simulation results

In Figure 3, we have the mean squared prediction error (MSPE) broken down by the spatial process parameters. The results suggest that including a thin-plate spline component improves the MSPE in general, and more so, when the scale parameter ρ is larger. This might be due to the relatively small number of sampled data points, which make it

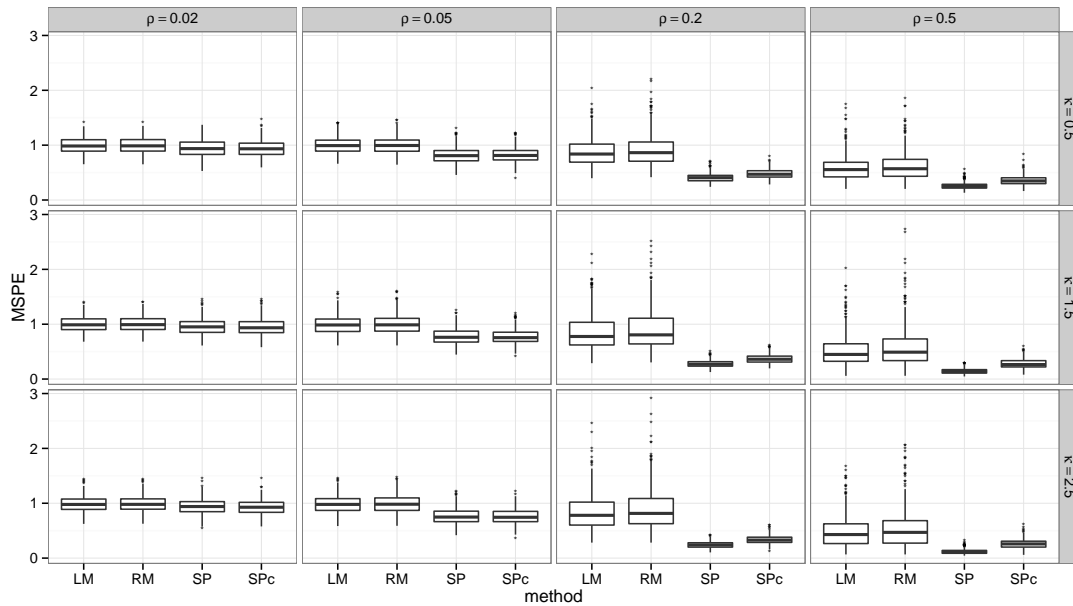


Figure 3: Mean squared prediction error (MSPE) for 400 simulations. LM refers to the standard linear model, RM refers to restricted maximum likelihood estimation, SP is the semiparametric thin-plate spline model and SPc is the SP model with λ estimated by the algorithm from Section 2.4. The spatial process has a Matérn covariance function with parameters $\kappa = 0.5, 1.5, 2.5$ and $\rho = 0.02, 0.05, 0.2, 0.5$.

harder for the thin-plate spline to fit smooth variability under a certain resolution. Overall, the semiparametric methods perform better as the scale parameter increases, yielding MSPEs that are smaller and varying less across simulated replications as ρ increases. The LM also yields smaller MSPE as ρ increases, but the MSPE is always larger than the semiparametric methods. Furthermore, the variability of the observed MSPE for LM increases with ρ . The performance of the RM approach is similar to LM, in terms of MSPE. Since the data are observed at sampling locations only, the smoothness parameter κ does not seem to affect the spline model substantially. The smoothing parameter obtained with our correction is in general 5% smaller than the smoothing parameter obtained by GCV. It does slightly increase the MSPE.

The estimates of β are shown in Figure 4, and we observe that the estimates under LM, RM and the SP models are not very different. The ratio of the sum of squared deviances from the true parameters ($\beta_1 = 1, \beta_2 = 1$), which is the relative efficiency of the estimators, suggests that the SP model compared to the LM model is roughly 5% less efficient when $\rho = 0.02$, but with about the same efficiency when $\rho = 0.05$, and roughly 10% more efficient when $\rho = 0.2$ or $\rho = 0.5$. For this case, the correction of the tuning parameter did not improve the results from regular GCV-based tuning.

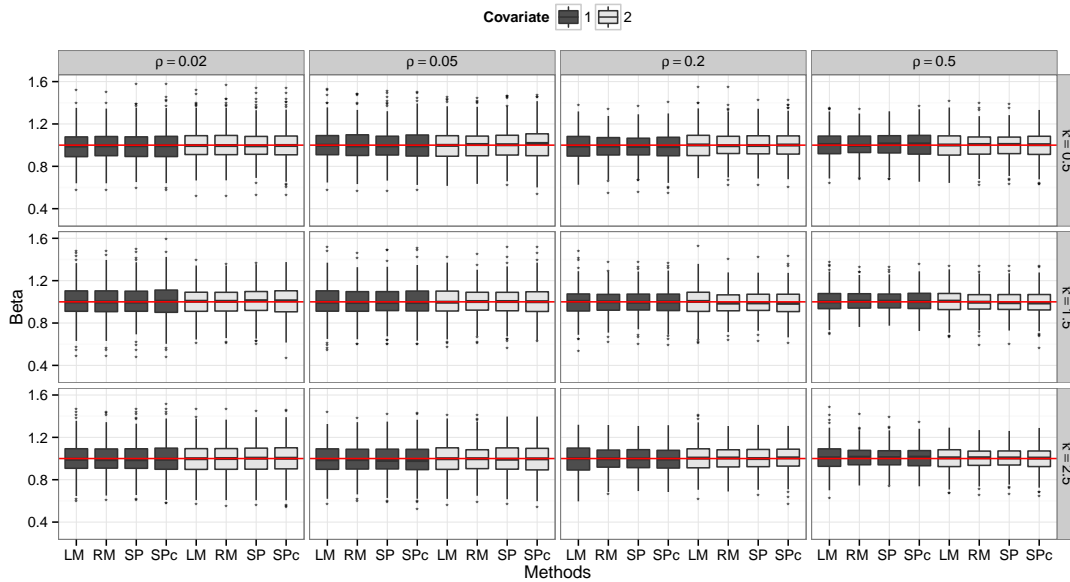


Figure 4: Boxplots of $\hat{\beta}$ using the standard linear model (LM), the restricted maximum likelihood model (RM), the semiparametric thin-plate spline model (SP) and the SP model with λ estimated by the correction algorithm from Section 2.4 (SPc); the color coding indicates $\hat{\beta}_1$ and $\hat{\beta}_2$. The spatial process has a Matérn covariance function with parameters $\kappa = 0.5, 1.5, 2.5$ and $\rho = 0.02, 0.05, 0.2, 0.5$. The horizontal solid line indicates the true value of β .

4. Spatial confounding

The simulation results in the Section 3 were derived under the assumption that the covariates $\mathbf{x}(\mathbf{s})$ and the spatial random process $\eta(\mathbf{s})$ are independent. Furthermore, sampling the covariates from independent Gaussian processes makes it unlikely that the realizations are going to exhibit any sort of collinearity. In this section, we will expand upon the simulation by considering the possibility of relations between the covariates and the spatial processes.

4.1. Simulation setup

We expanded the simulation in Section 3 by considering the following two scenarios:

- (A) Generate $x_2(\mathbf{s})$ as a Gaussian random field with a Matérn covariance function and parameters $\kappa = 2.5, \rho = 0.2$. This experiment considers the case that $x_j(\mathbf{s}), j = 1, 2$ and $\eta(\mathbf{s})$ are independent, but varying smoothly on the same scale (see scenario (A) of Hodges and Reich, 2010 and Hanks et al., 2015).

- (B) Generate $x_2(\mathbf{s})$ such that $\text{Corr}(\eta(\mathbf{s}), x_2(\mathbf{s})) \approx 0.4$. This experiments with spatial confounding such that $x_j(\mathbf{s}), j = 1, 2$ and $\eta(\mathbf{s})$ are correlated (see scenario (B) of Hodges and Reich, 2010 and Paciorek, 2010).

We illustrate the behavior of estimates (8) under spatial confounding in a similar way to Figure 1. Scenario (A) is illustrated in Figure 5(A). There are two sets of curves now, one corresponding to $\hat{\beta}_1$ for reference, and the other corresponding to $\hat{\beta}_2$, which is the coefficient of the covariate confounded with the spatial process. In this case, the effect of the spatial process is strong on the variability of $\hat{\beta}_2$. Increasing the tuning parameter λ does seem to mitigate this issue, forcing most of the smooth variability to be attributed to \hat{f}_λ .

Scenario (B) is shown in Figure 5(B). Since by assumption $x_2(\mathbf{s})$ is known and fixed, then the conditional distribution of $[\eta(\mathbf{s})|x_2(\mathbf{s})]$ can be calculated. Let $\boldsymbol{\eta} = (\eta(\mathbf{s}_1), \dots, \eta(\mathbf{s}_n))'$ and $\mathbf{x}_2 = (x_2(\mathbf{s}_1), \dots, x_2(\mathbf{s}_n))'$. We have

$$\boldsymbol{\eta}|\mathbf{x}_2 \sim N(0.4\mathbf{x}_2, \Sigma_\eta - 0.4^2\mathbf{I}),$$

which shows that the increasing effect of the random component on $\text{Var}(\hat{\beta}_2)$ (Figure 5(A)) is a bias term. While in the model fitting most of the smooth variability is attributed to \hat{f}_λ , in theory $\eta(\mathbf{s})$ itself given $x_2(\mathbf{s})$ is not very smooth. We will see how this will be reflected back into $\text{Var}(\hat{\beta}_2)$ in the next section.

4.2. Simulation results

We will focus on the estimation of $\boldsymbol{\beta}$, setting the comparison of MSPE aside for future work. For Scenario (A), Figure 6 suggests improvement in estimation of $\boldsymbol{\beta}$ when the spatial process itself is smooth ($\rho \geq 0.2$), but the β_2 parameter estimation is worse when the spatial process is less smooth than x_2 . The ratio of the sum of squared deviations to the true value is roughly 0.75 when $\rho < 0.2$, but about 1.5 when $\rho \geq 0.2$. Values for $\hat{\beta}_1$ are also reported as a reference. It appears that as long as the covariates are not varying in a fashion smoother than the actual spatial process, it is worth including a spline component in the model. Moreover, the correction of the tuning parameter does improve the estimation of $\boldsymbol{\beta}$ in comparison to the GCV, at a rate of about 10%. While this improvement is not enough to justify including a spline in the case when $\rho < 0.2$, it does mitigate the spatial confounding.

For scenario (B), the bias effect of the dependence between $\eta(\mathbf{s})$ and $x_2(\mathbf{s})$ is shown in Figure 7. It is also clear that including a spline in such models is beneficial when the spatial process is smooth, and at least does not worsen the regression coefficients estimation otherwise.

In both cases, the performance of the spline based fitting is similar to the restricted maximum likelihood approach. This is remarkable because using the spline method does not require the knowledge of the covariance structure of the data.

The choice of spatial confounding-like scenarios presented here is arbitrary. We investigate spatial confounding further by looking at cases for which spatial confounding is present

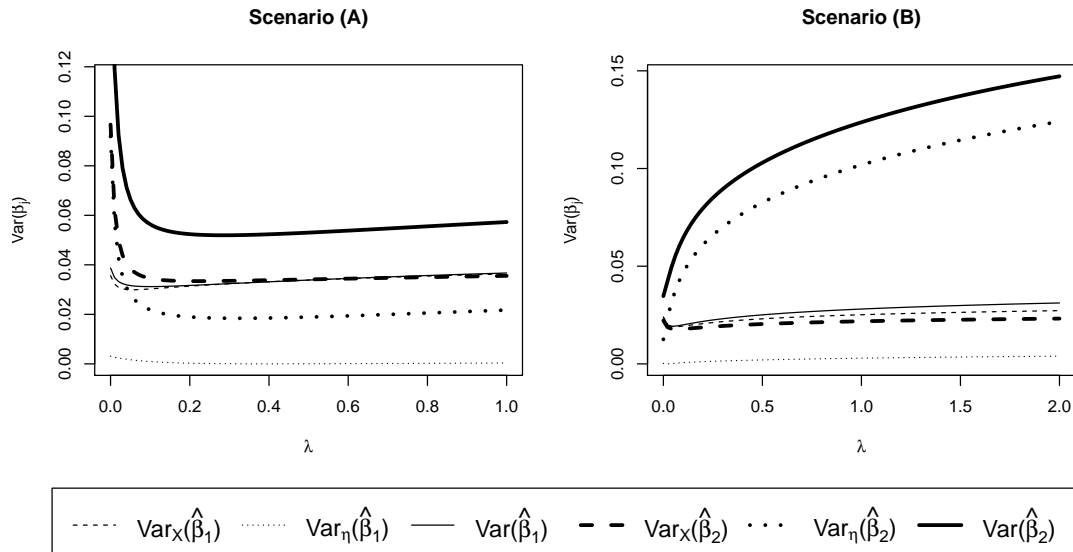


Figure 5: Illustration of the trade-off between the two components of $\text{Var}(\hat{\beta}_j)$, $j = 1, 2$, given by $\text{Var}_{\mathbf{X}}(\hat{\beta}_j)$ (variance attributed to the design matrix \mathbf{X}) and $\text{Var}_{\eta}(\hat{\beta}_j)$ (variance attributed to the spatial process $\eta(\mathbf{s})$). Scenario (A): In this case, $\eta(\mathbf{s})$ is independent of \mathbf{X} but $x_2(\mathbf{s})$ is a Gaussian random field with Matérn covariance and parameters $\kappa = 2.5$, $\rho = 0.2$. Scenario (B): $\eta(\mathbf{s})$ is correlated with $x_2(\mathbf{s})$ (with a correlation of 0.4).

at varying degrees of complexity. For example, in Scenario (A), $x_2(\mathbf{s})$ is a Gaussian random field with Matérn covariance function and parameters $\kappa = 2.5$, and $\zeta = 0.2, 0.4, 0.6$ and 0.8 for the range parameter. Similarly, for Scenario (B), we have $x_2(\mathbf{s})$ and $\eta(\mathbf{s})$ be correlated, with correlations $\zeta = 0.2, 0.4, 0.6$ and 0.8 .

Since the potential number of parameter combinations in the simulation becomes too large, we focus on two cases, in which $\eta(\mathbf{s})$ is a Gaussian random field with Matérn covariance function of parameters $\kappa = 2.5$, $\rho = 0.02$ and $\rho = 0.5$. Figure 8 shows the estimates of β_2 when $\eta(\mathbf{s})$ has $\rho = 0.02$, and Figure 9 shows the estimates of β_2 when $\eta(\mathbf{s})$ has $\rho = 0.5$. We observe in Figure 8 that when $\eta(\mathbf{s})$ has a small ρ (weak spatial dependence), using the spline is detrimental for Scenario (A) and only moderately helpful for Scenario (B). This suggests that when there is no spatially dependent error, using a thin-plate spline is just overfitting the data, with no improvement on estimation of β . On the other hand, when $\eta(\mathbf{s})$ has a large ρ (strong spatial dependence) we observe in Figure 9 the spline is very helpful in improving the estimates of β and controlling the spatial confounding effect for Scenario (B). For Scenario (B) in general, the spline is effective at controlling the spatial confounding effect. For Scenario (A) our conjecture of the spline only being helpful as long as $x_2(\mathbf{s})$ have a weaker spatial dependence than $\eta(\mathbf{s})$; the larger the ζ value is, the worse is the impact on estimation of β_2 for having a spline smoother.

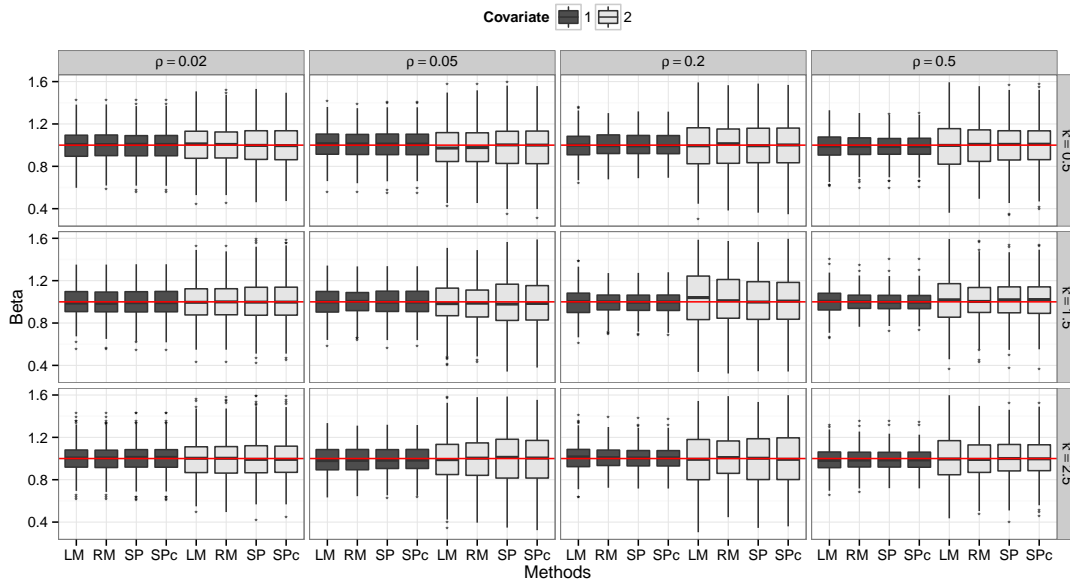


Figure 6: Scenario (A), in which $x_2(\mathbf{s})$ is a Gaussian random field with Matérn covariance function with parameters $\kappa = 2.5$ and $\rho = 0.2$. Boxplots of $\hat{\beta}$ using the standard linear model (LM), the restricted maximum likelihood model (RM), the semiparametric thin-plate spline model (SP) and the SP model with λ estimated by the correction algorithm from Section 2.4 (SPc); the color coding indicates $\hat{\beta}_1$ and $\hat{\beta}_2$. The spatial process has a Matérn covariance function with parameters $\kappa = 0.5, 1.5, 2.5$ and $\rho = 0.02, 0.05, 0.2, 0.5$. The horizontal solid line indicates the true value of β .

Finally, we also investigated changes in σ_η^2 , which lead to different signal-to-noise ratios for the spatial process $\eta(\mathbf{s})$. Figure 10 shows the case in which there is spatial confounding according to Scenarios (A) or (B), and σ_η^2 is set to $0.1\sigma^2$, $0.5\sigma^2$, $2\sigma^2$, or $10\sigma^2$. We can observe that for different signal-to-noise ratios in Scenario (A), the semiparametric method has a comparable performance with the restricted maximum likelihood estimates using the Matérn covariance. Aside from the increased variability of the $\hat{\beta}$ estimates, the results are consistent with what was previously observed. For Scenario (B), however, we observe that a higher the signal-to-noise ratio makes the restricted maximum likelihood estimates of β closer to the semiparameteric estimates.

4.3. Nonstationarity

The simulation study conducted considered only stationary spatial processes $\eta(\mathbf{s})$. The semiparametric model, however, does not assume stationarity. We will briefly look at two cases of nonstationarity, which are based on transformations of the Matérn covariance function:

- In the case labelled “Anisotropic”, let η^* be a process with Matérn covariance as

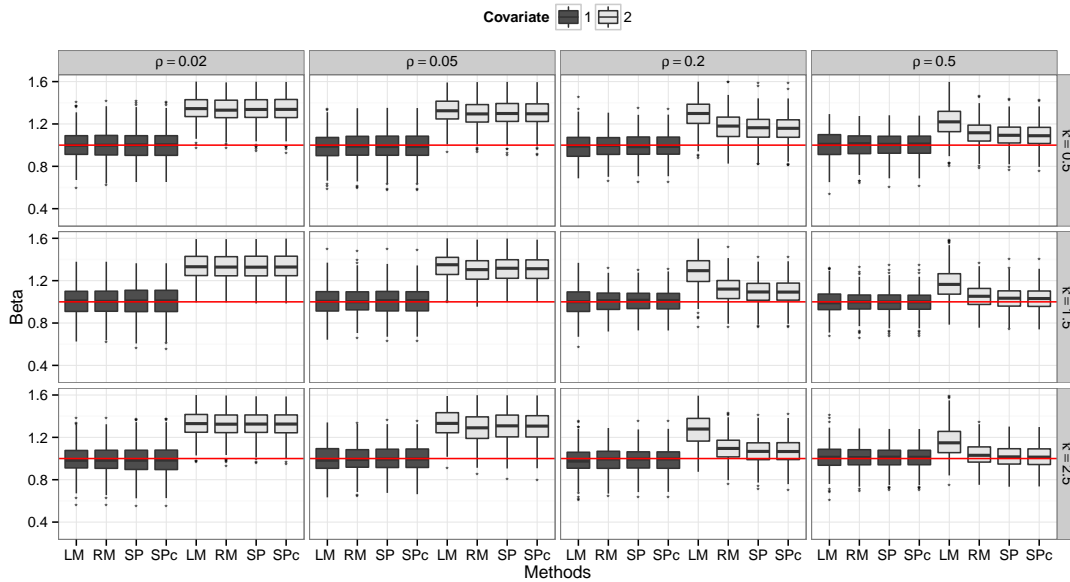


Figure 7: Scenario (B), in which $\text{Corr}(\eta(\mathbf{s}), x_2(\mathbf{s})) \approx 0.4$. Boxplots of $\hat{\beta}$ using the standard linear model (LM), the restricted maximum likelihood model (RM), the semiparametric thin-plate spline model (SP) and the SP model with λ estimated by the correction algorithm from Section 2.4 (SPc); the color coding indicates $\hat{\beta}_1$ and $\hat{\beta}_2$. The spatial process has a Matérn covariance function with parameters $\kappa = 0.5, 1.5, 2.5$ and $\rho = 0.02, 0.05, 0.2, 0.5$. The horizontal solid line indicates the true value of β .

described previously, generated on a set of coordinates $\mathbf{s}_1^*, \dots, \mathbf{s}_n^*$. Then η is obtained by rotation and scale of the coordinates, i.e. setting $\mathbf{s}_i^* = \mathbf{P}\mathbf{s}_i$ for $i = 1, \dots, n$, while the values of η^* are retained (see, e.g. Wackernagel, 2003). In particular, we used $\theta = \pi/4, a = 1, b = 8$ in

$$\mathbf{P} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} \sqrt{a} & 0 \\ 0 & \sqrt{b} \end{pmatrix}$$

- In the case labelled “Nonstationary”, we use the Matérn covariance model, but let the spatial dependence parameter ρ vary in the domain by employing a weighting function on the coordinates. Thus $\text{Cov}_{\text{NS}}(\eta(\mathbf{s}), \eta(\mathbf{s}')) = \gamma(w(\mathbf{s})\mathbf{s}, w(\mathbf{s}')\mathbf{s}')$, and we chose $w(\mathbf{s}) = 0.3 + (1 - 0.3)/(1 + \exp\{(s_1 - 0.4)/0.05\})$, so that the resulting processes show a transition of spatial dependence on the s_1 coordinate.

To illustrate the nonstationary processes, we consider in Figure 11 a diagram similar to Figure 2, showing the realization of a Gaussian process with the corresponding nonstationary covariance functions, for different values of spatial dependence ρ .

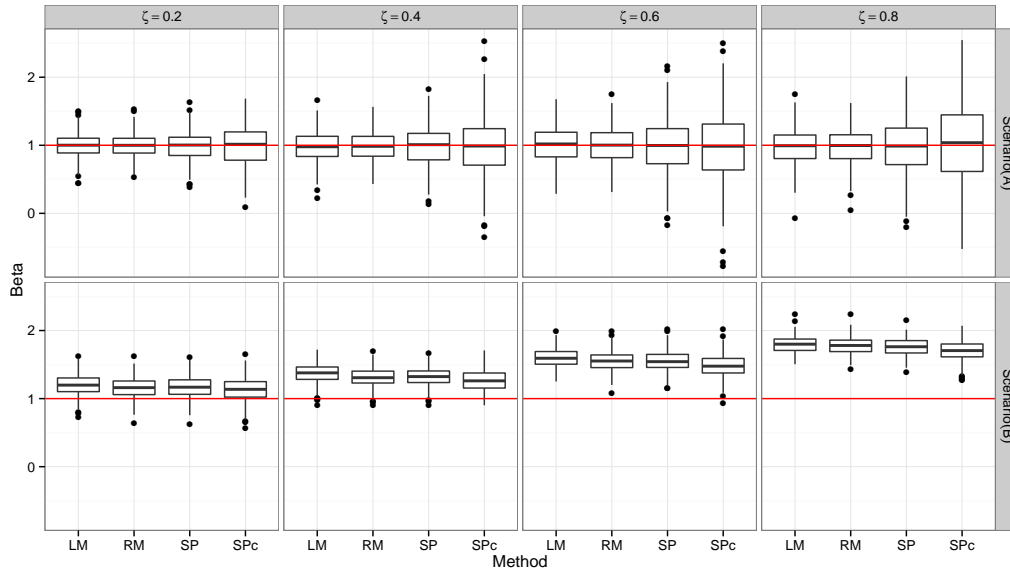


Figure 8: Comparison of $\hat{\beta}_2$ for different spatial confounding degrees of complexity when $\eta(\mathbf{s})$ is a Gaussian random field with Matérn covariance of parameters $\kappa = 2.5$, $\rho = 0.02$. For Scenario (A), ζ is the parameter ρ for generating $x_2(\mathbf{s})$; for Scenario (B), ζ is the correlation coefficient between $\eta(\mathbf{s})$ and $x_2(\mathbf{s})$.

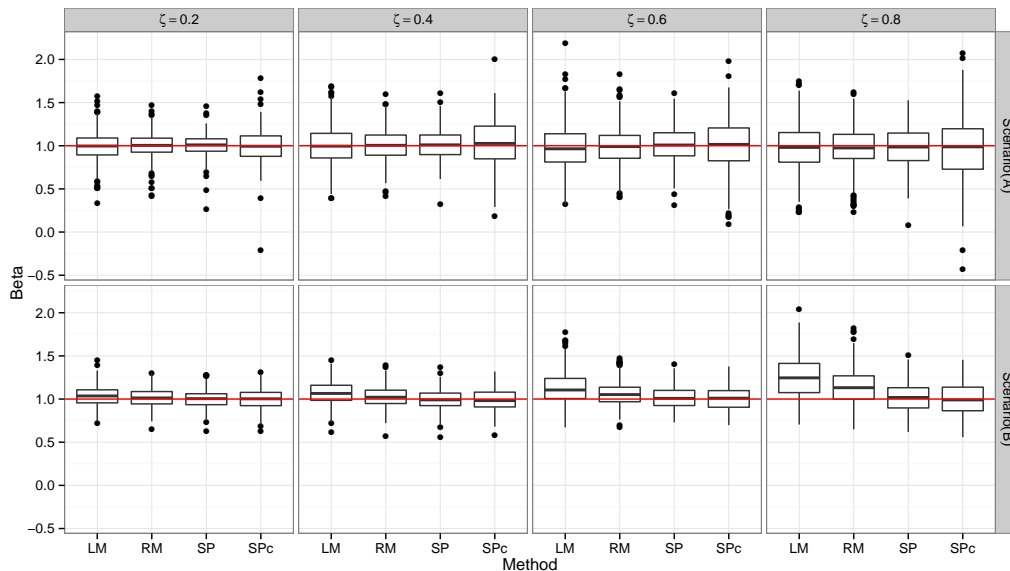


Figure 9: Comparison of $\hat{\beta}_2$ for different spatial confounding degrees of complexity when $\eta(\mathbf{s})$ is a Gaussian random field with Matérn covariance of parameters $\kappa = 2.5$, $\rho = 0.5$. For Scenario (A), ζ is the parameter ρ for generating $x_2(\mathbf{s})$; for Scenario (B), ζ is the correlation coefficient between $\eta(\mathbf{s})$ and $x_2(\mathbf{s})$.

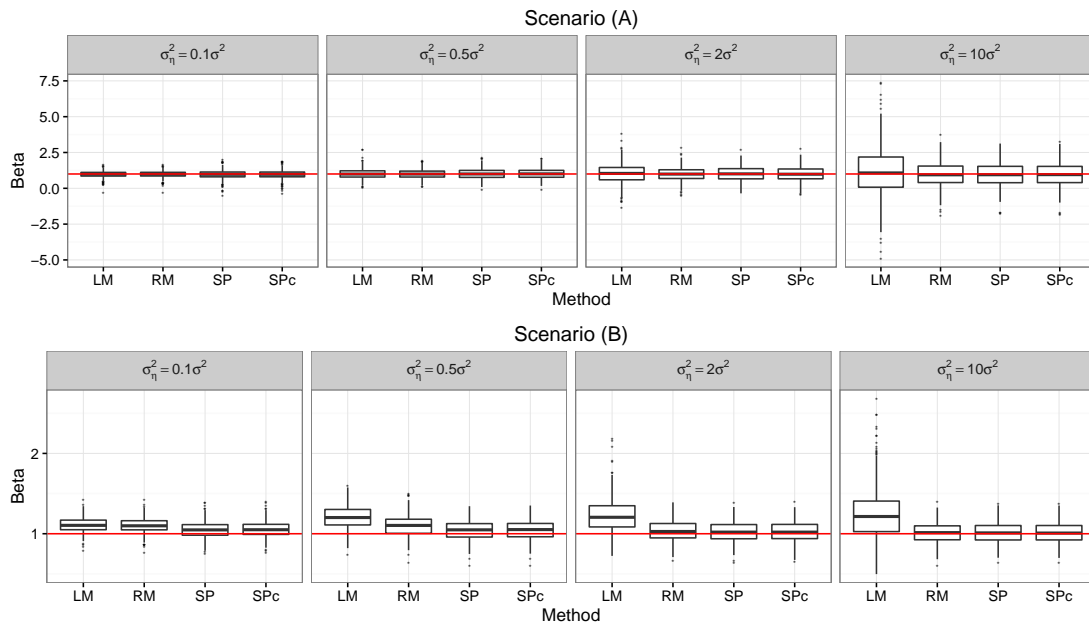


Figure 10: Comparison of $\hat{\beta}$ when σ_η^2 is set to $0.1\sigma^2$, $0.5\sigma^2$, $2\sigma^2$, or $10\sigma^2$. In this case, $\eta(\mathbf{s})$ is a Matérn covariance Gaussian process with parameters $\nu = 2.5$, $\rho = 0.2$.

In Figure 12, we show the distribution of $\hat{\beta}$ after simulating $\eta(\mathbf{s})$ with the anisotropic and nonstationary transformed covariance functions, based on the Matérn covariance function with $\kappa = 2.5$ and $\rho = 0.02, 0.05, 0.2$, and 0.5 . We observe that the results are similar to the stationary case.

5. Discussion

We investigated semiparametric spline-based regression when the data comes from spatial processes confounded with fixed covariates. Our conclusion is that including a thin-plate spline in a regression model for spatially confounded data is helpful, as long as the resolution of the sampling sites is fine enough to capture spatial variability. On the other hand, if the spatial process and the covariates are uncorrelated but have similar degrees of spatial dependence, then using a spline needs careful consideration, since the spline term might compete with the regression covariate. We also proposed a new strategy for the selection of the spline tuning parameter in semiparametric regression. We anticipate that a good selection of the tuning parameter is key to mitigate the spatial confounding effects on the estimation of the regression coefficients, which we leave for future investigation.

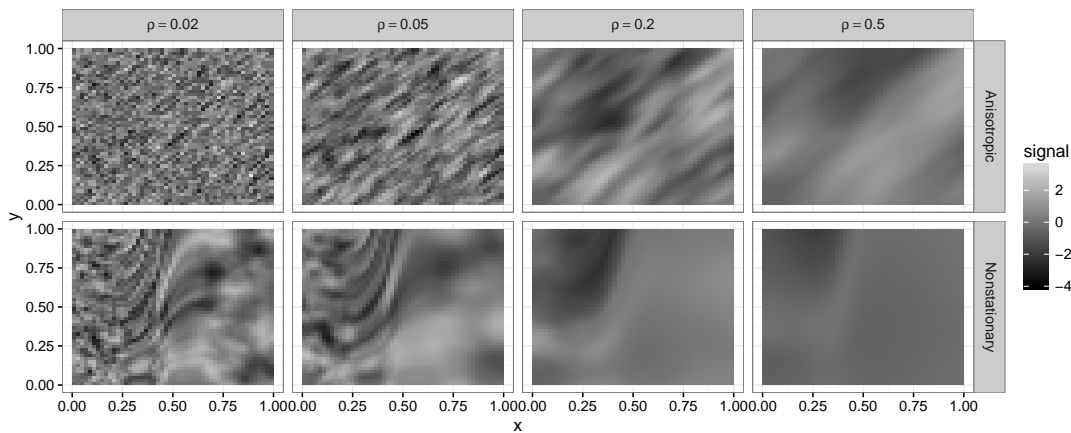


Figure 11: Examples of nonstationary covariance processes for different covariance functions: each facet shows the realization of a single gaussian process. The anisotropic case is a deformation of a Matérn process with choices of parameters $\kappa = 2.5$ and $\rho = 0.02, 0.05, 0.2, 0.5$. The nonstationary case has a weighting function applied to the ρ coefficient, so that the dependence range between $Y(\mathbf{s})$ and $Y(\mathbf{s}^*)$ changes depending on whether $s_1, s_1^* > 0.4$.

References

- N. Altman. Theory & methods: kriging, smooth, both or neither? *Australian & New Zealand Journal of Statistics*, 42(4):441–461, 2000.
- D. G. Clayton, L. Bernardinelli, and C. Montomoli. Spatial correlation in ecological analysis. *International Journal of Epidemiology*, 22(6):1193–1202, 1993.
- N. Cressie. Geostatistics. *The American Statistician*, 43(4):197–202, 1989.
- N. Cressie. Reply to letter by G. Wahba. *The American Statistician*, 44(3):256–258, 1990.
- N. Cressie. *Statistics for Spatial Data*. Wiley, New York, 2nd edition, 1993.
- P. J. Green, C. Jennison, and A. Seheult. Analysis of field experiments by least squares smoothing. *Journal of the Royal Statistical Society. Series B*, 47(2):299–315, 1985.
- E. M. Hanks, E. M. Schliep, M. B. Hooten, and J. A. Hoeting. Restricted spatial regression in practice: geostatistical models, confounding, and robustness under model misspecification. *Environmetrics*, 26(4):243–254, 2015.
- T. J. Hastie and R. J. Tibshirani. *Generalized Additive Models*, volume 43. Chapman & Hall, Baton Rouge, 1990.
- J. S. Hodges and B. J. Reich. Adding spatially-correlated errors can mess up the fixed effect you love. *The American Statistician*, 64(4):325–334, 2010.

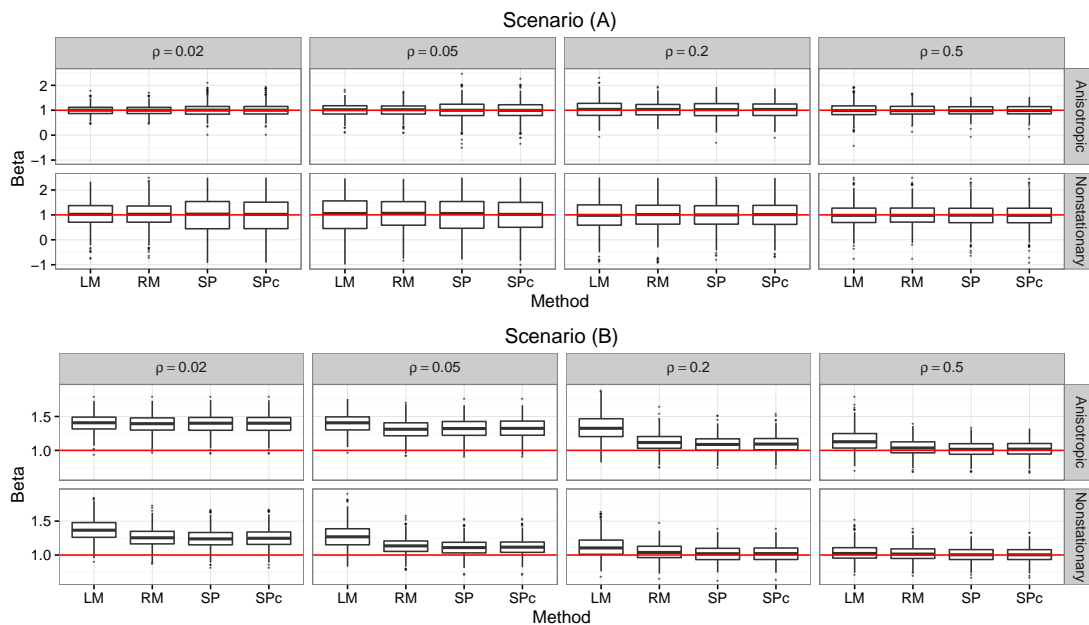


Figure 12: Comparison of $\hat{\beta}$ for the Anisotropic and Nonstationary cases, for Scenarios (A) and (B). This figure is analogous to Figure 6 and Figure 7.

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

M. Loève. *Probability Theory*, volume II. Springer, New York, 4th edition, 1978.

D. W. Nychka. Spatial-process estimates as smoothers. In M. G. Schimek, editor, *Smoothing and regression: approaches, computation, and application*, pages 393–424. Wiley, New York, 2000.

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

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

P. J. Ribeiro Jr. and P. J. Diggle. *geoR: Analysis of Geostatistical Data*, 2015. URL <http://CRAN.R-project.org/package=geoR>. R package version 1.7-5.1.

J. Rice. Convergence rates for partially splined models. *Statistics & Probability Letters*, 4(4):203–208, 1986.

- D. Ruppert, M. P. Wand, and R. J. Carroll. *Semiparametric Regression*. Cambridge University Press, Cambridge, 2003.
- M. G. Schimek. Estimation and inference in partially linear models with smoothing splines. *Journal of Statistical Planning and Inference*, 91(2):525–540, 2000.
- P. Speckman. Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society. Series B*, 50(3):413–436, 1988.
- W. W. Stroup. *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*. Chapman & Hall, Baton Rouge, 2012.
- H. Wackernagel. *Multivariate geostatistics*. Springer Science & Business Media, 2003.
- G. Wahba. Comment on Cressie. *The American Statistician*, 44:255–256, 1990a.
- G. Wahba. *Spline Models for Observational Data*, volume 59. SIAM, Philadelphia, 1990b.