Spectral adjustment for spatial confounding

Biometrika (2023), **00**, 0, *pp.* 1–21 https://doi.org/10.1093/biomet/asac069
Advance Access Publication 21 December 2022

Spectral adjustment for spatial confounding

By YAWEN GUAN

Department of Statistics, University of Nebraska,
343C Hardin Hall, Lincoln, Nebraska 68583, U.S.A.
yawen.guan@unl.edu

GARRITT L. PAGE

Department of Statistics, Brigham Young University,
238 TMCB, Provo, Utah 84602, U.S.A.
page@stat.byu.edu

BRIAN J. REICH

Department of Statistics, North Carolina State University,
2311 Stinson Drive, Raleigh, North Carolina 27695, U.S.A.
bjreich@ncsu.edu

MASSIMO VENTRUCCI

Department of Statistical Sciences, University of Bologna,
Via Zamboni 33, Bologna 40126, Italy
massimo.ventrucci@unibo.it

and SHU YANG

Department of Statistics, North Carolina State University,
2311 Stinson Drive, Raleigh, North Carolina 27695, U.S.A.
syang24@ncsu.edu

Summary

Adjusting for an unmeasured confounder is generally an intractable problem, but in the spatial setting it may be possible under certain conditions. We derive necessary conditions on the coherence between the exposure and the unmeasured confounder that ensure the effect of exposure is estimable. We specify our model and assumptions in the spectral domain to allow for different degrees of confounding at different spatial resolutions. One assumption that ensures identifiability is that confounding present at global scales dissipates at local scales. We show that this assumption in the spectral domain is equivalent to adjusting for global-scale confounding in the spatial domain by adding a spatially smoothed version of the exposure to the mean of the response variable. Within this general framework, we propose a sequence of confounder adjustment methods that range from parametric adjustments based on the Matérn coherence function to more robust semiparametric methods that use smoothing splines. These ideas are applied to areal and geostatistical data for both simulated and real datasets.

Some key words: Coherence; Conditional autoregressive prior; COVID-19; Matérn covariance; Spatial confounding.

1. Introduction

A fundamental task in environmental and epidemiological applications is to use spatially correlated observational data to estimate the effect of exposures. A key assumption needed to identify the effect of the exposures is that all relevant confounding variables have been included in the statistical model. This no-missing-confounder assumption is generally impossible to verify, but in the spatial setting it may be possible to remove the effects of unmeasured confounding variables if they have strong spatial dependence. An unmeasured spatial confounder exists when spatially varying factors that influence both the exposures and response are not observed. When the unmeasured spatial confounder is not taken into account, the effect estimate of an exposure can be biased, and the bias depends on the spatial scales of the exposure and the unmeasured spatial confounder (Paciorek, 2010; Page et al., 2017).

A slightly different but related confounding issue in modelling spatial data was discussed in Clayton et al. (1993). They proposed to model the geographical patterns in the response using a spatial random effect term independent of exposure to account for unmeasured spatially structured covariates that influence the response. Then, confounding, i.e., confounding due to location, may arise if exposure also varies smoothly with location, and the location may act as a confounder. In this case, regressions with and without spatial random effects can give different inference results on regression coefficients (Reich et al., 2006; Hodges & Reich, 2010), a phenomena also known as spatial confounding. For confounding due to location, the unmeasured spatial covariates do not directly influence the exposure, but rather are multicollinear with the exposure, and the multicollinearity is more likely to occur when the spatial random effects and exposure are both spatially smooth. In this paper, we focus on spatial confounding due to an unmeasured spatial confounder, in which the confounding due to location can be considered as a special case when the association between exposure and missing spatial covariates is zero.

To account for spatial confounding due to location, Reich et al. (2006), Hughes & Haran (2013) and Prates et al. (2019) restricted the residual spatial process to be orthogonal to the exposure, an approach referred to as restricted spatial regression. However, the approach makes a strong orthogonality assumption and can perform poorly in coefficient inference when the model is misspecified (Hanks et al., 2015; Khan & Calder, 2022; Zimmerman & Hoef, 2022). Alternative approaches with a focus on decomposing spatial scales following Paciorek (2010) have appeared in the literature. The main idea is that the location confounding can be eliminated by first removing the smooth components from both the exposure and response (Thaden & Kneib, 2018) or just the exposure (Keller & Szpiro, 2020; Dupont et al., 2022), then the exposure effect can be estimated by assessing the local variations in the covariate and response. More recently, Marques et al. (2022) proposed a joint Gaussian Markov random field model for exposure and response. Their work was developed independently around the same time as ours and is related to a special case of our parametric model. The listed works attempt to alleviate spatial confounding, but, in general, adjusting for a missing confounding variable is impossible without further information or assumptions. It remains unclear how to specify assumptions and methods that lead to consistent estimation of the exposure effect in the presence of an unmeasured spatial confounding variable, which may posit a more complex confounding structure than location confounding.

Connections with causal inference have been made. For areal data, Thaden & Kneib (2018) and Schnell & Papadogeorgou (2020) proposed jointly modelling the spatial structure in exposure and the unmeasured confounder. The former uses a structural equation model

while the latter uses a Gaussian Markov random field construction. Both these methods have connections with causal inference; see Reich et al. (2021) for a recent review of spatial causal inference. In the spatial causal effect setting, Osama et al. (2019) permitted the spatial causal effect to vary across space. Different assumptions on the confounding relationship may lead to a variety of approaches.

We propose new methods to couch spatial regression with missing spatial confounding variables using spectral methods. Similarly to Paciorek (2010), Page et al. (2017) and Keller & Szpiro (2020), the spatial scales of exposure and missing confounder are the focus, but we explicitly specify a joint model for these variables in the spectral domain and study their coherence, i.e., their correlation at different spatial scales. As an aside, from a temporal perspective, Stokes & Purdon (2017) and Faes et al. (2019) considered a frequency-domain measure of causality, although their estimators are quite different than those proposed here. The resulting effect estimate from our approach reveals that the optimal confounder adjustment is a function of the coherence function, providing fundamental insights on spatial confounding. We show that the optimal confounder adjustment is not estimable without further assumptions, and provide a set of conditions that allow us to identify the exposure effect. Parametric and nonparametric methods are developed to approximate the optimal confounding adjustment and identify the exposure effect, while accounting for uncertainty in this approximation. We consider both areal and point-referenced data for Gaussian and non-Gaussian responses. Proofs are given in the Supplementary Material.

2. Continuous-space modelling framework

2.1. *Preliminaries*

Let $X(s)$ and $Z(s)$ be the observed exposure and confounder processes, respectively, at the spatial location $s \in D \subset \mathcal{R}^2$, and let vectors $X = (X_1, \ldots, X_n)^{\mathrm{T}}$ and $Z = (Z_1, \ldots, Z_n)^{\mathrm{T}}$ be the process evaluated at the set of locations $\mathcal{S} = \{s_1, \ldots, s_n\} \in D$, where $X_i = X(s_i)$ and $Z_i = Z(s_i)$. For simplicity, we consider only a single exposure and confounding variable, but results extend to a multivariate setting; see the Supplementary Material. We do not assume that $\mathcal{S}$ is on a complete grid nor that the n observation locations are distinct, i.e., we allow for multiple observations at the same location with the inclusion of a nugget effect. Both X and Z are assumed to be spatial processes, potentially with some nonspatial nugget variability.

Following the commonly used spatial regression model, we assume a linear additive relationship for the response $Y = (Y_1, \ldots, Y_n)^{\mathrm{T}}$. Here, we present the method for continuous response, but the extension to non-Gaussian cases is straightforward, and therefore the details for the generalized model are presented in the Supplementary Material.

We have

$$Y = \beta_0 + \beta_x X + \beta_z Z + \varepsilon, \tag{1}$$

where $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^{\mathrm{T}}$ and $\varepsilon_i \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$, a normal distribution with mean 0 and variance σ^2. The regression coefficient β_x has a causal interpretation under the potential outcomes framework and the stable-unit-treatment-value, consistency and conditional-treatment-ignorability assumptions. If we observe the confounder Z, identification and estimation of β_x is straightforward using multiple linear regression. However, we assume that Z is an unmeasured confounder, making β_x not identifiable in general. We propose

to exploit the spatial structure of Z to mitigate the effects of the unobserved confounder and specify assumptions in the spectral domain for identifying β_x. Our inference procedure involves introducing a confounder adjustment variable $\hat{Z}$ to the linear model. The derivation of $\hat{Z}$ is based on the association between X and Z in the spectral domain, while in the spatial domain, $\hat{Z}$ can be viewed as a smoothed version of X under one of our assumptions for identification. Our work aims to identify and estimate β_x, which has a causal interpretation under the potential outcomes framework and assumptions described in the Supplementary Material. Here instead of restating all elements of the causal framework, in the remaining sections, we assume that the confounder process $Z(s)$ contains all unmeasured confounders and mainly focus on mitigating the effects of $Z(s)$. Therefore, we interpret β_x as the effect of exposure under model (1) rather than the causal effect, as the latter requires additional assumptions that are not essential for β_x inference in this work.

2.2. *Spectral representation of confounding and identification*

We model the dependence between $X(s)$ and $Z(s)$ using their spectral representations. This allows for different dependencies at different spatial scales as each frequency corresponds to a spatial scale, with low frequency corresponding to large spatial scale. We assume that both $X(s)$ and $Z(s)$ are mean-zero stationary Gaussian processes, and thus have spectral representations $X(s) = \int \exp(i\omega^{\mathrm{T}}s)\mathcal{X}(\omega)\, d\omega$ and $Z(s) = \int \exp(i\omega^{\mathrm{T}}s)\mathcal{Z}(\omega)\, \mathrm{d}\omega$, where $\omega \in \mathcal{R}^2$ is a frequency. The spectral processes $\mathcal{X}(\omega)$ and $\mathcal{Z}(\omega)$ are Gaussian with $\mathrm{E}\{\mathcal{X}(\omega)\} = \mathrm{E}\{\mathcal{Z}(\omega)\} = 0$ and are independent across frequencies, so that, for any $\omega \neq \omega'$, $\mathrm{cov}\{\mathcal{X}(\omega), \mathcal{X}(\omega')\} = \mathrm{cov}\{\mathcal{Z}(\omega), \mathcal{Z}(\omega')\} = \mathrm{cov}\{\mathcal{X}(\omega), \mathcal{Z}(\omega')\} = 0$. At the same frequency, the covariance of the joint spectral process has the form

$$\mathrm{cov}\begin{Bmatrix}\mathcal{X}(\omega)\\ \mathcal{Z}(\omega)\end{Bmatrix} = \begin{Bmatrix}\sigma_x^2 f_x(\omega) & \rho\sigma_x\sigma_z f_{xz}(\omega)\\ \rho\sigma_x\sigma_z f_{xz}(\omega) & \sigma_z^2 f_z(\omega)\end{Bmatrix},$$

where σ_x^2 and σ_z^2 are variance parameters, $f_x(\omega) > 0$ and $f_z(\omega) > 0$ are spectral densities that determine the marginal spatial correlation of $X(s)$ and $Z(s)$, respectively, and the cross-spectral density $f_{xz}(\omega)$ determines the dependence between the spectral processes.

Normalizing the cross-spectral density by each marginal standard deviation, we can derive the coherence function that determines the correlations between the two spectral processes (Kleiber, 2017),

$$\gamma(\omega) = \rho\frac{f_{xz}(\omega)}{\sqrt{f_x(\omega)f_z(\omega)}} \in [-1, 1]. \tag{2}$$

The scalar parameter ρ controls the overall strength of cross-correlation.

Returning to the response model (1), we let $Y(s) = \int \exp(i\omega^{\mathrm{T}}s)\mathcal{Y}(\omega)\, \mathrm{d}\omega$ be the spectral representation of the response. The conditional distribution of $\mathcal{Y}(\omega)$ given $\mathcal{X}(\omega)$, marginalizing over $\mathcal{Z}(\omega)$, is

$$\mathcal{Y}(\omega) \mid \mathcal{X}(\omega) \overset{\text{indep}}{\sim} N\{\beta_x\mathcal{X}(\omega) + \beta_z\alpha(\omega)\mathcal{X}(\omega), \tau^2(\omega) + \sigma^2\}, \tag{3}$$

$$\alpha(\omega) = \rho\frac{\sigma_z f_{xz}(\omega)}{\sigma_x f_x(\omega)} = \frac{\sigma_z\sqrt{f_z(\omega)}}{\sigma_x\sqrt{f_x(\omega)}}\gamma(\omega),$$

$$\tau^2(\omega) = \beta_z^2\sigma_z^2 f_z(\omega)\left\{1 - \rho^2\frac{f_{xz}(\omega)^2}{f_x(\omega)f_z(\omega)}\right\}.$$

The regression coefficient for $\mathcal{X}(\omega)$ is $\beta(\omega) = \beta_x + \beta_z\alpha(\omega) \neq \beta_x$. The additional term $\hat{\mathcal{Z}}(\omega) = \mathrm{E}\{\mathcal{Z}(\omega) \mid \mathcal{X}(\omega)\} = \alpha(\omega)\mathcal{X}(\omega)$ is a result of attributing the effect of the unmeasured confounder on the response to the exposure, potentially inducing bias in estimating β_x.

Therefore, β_x is identified only if the projection operator $\alpha(\omega)$ can be assumed to be known or estimated for some prespecified ω. Of course, $\alpha(\omega)$ is generally not known and cannot be estimated without further assumptions because $Z(s)$, and therefore $\mathcal{Z}(\omega)$, is not observed. We consider two approaches for identification: assume unconfoundedness at high frequencies, i.e., $\alpha(\omega) \approx 0$ for large $\|\omega\|$, so that high-frequency terms identify β_x; or specify a parsimonious coherence function with constraints on the parameters to ensure identification of β_x through estimation of $\alpha(\omega)$. We detail both approaches next.

For the case of unconfoundedness at high frequencies, if we assume that $\alpha(\omega) \to 0$ for large $\|\omega\|$ then $\mathrm{E}\{\mathcal{Y}(\omega) \mid \mathcal{X}(\omega)\} \approx \beta_x\mathcal{X}(\omega)$ and thus β_x is identified. The assumption that $\alpha(\omega) \to 0$ for large $\|\omega\|$ implies that the cross-spectral density decreases to zero faster than the spectral density of X, which means that confounding dissipates as the frequency increases, that is, as the scale of the spatial variation becomes smaller. High-frequency terms provide the most reliable information about β_x because they correlate local changes in the exposure with local changes in the response. An extreme case of local information about the exposure effect is the difference in the response for two nearby sites with different levels of exposure. This local difference eliminates problems caused by omitted variables that vary smoothly over space. Of course, this cannot completely rule out missing confounding variables that covary with both the exposure and the response at high frequencies, but it does lessen the likelihood of spurious confounding effects.

For the case of parsimonious coherence, if we assume that $f_{xz}(\omega) = C\sqrt{f_x(\omega)f_z(\omega)}$ for a constant C, then the coherence in (2) simplifies to the constant function $\gamma(\omega) = \rho C \in (-1, 1)$. Generalizing the use of a term from Gneiting et al. (2010), we refer to this as the parsimonious coherence model. This imposes the assumption that the correlation between the exposure and missing confounder is frequency invariant, and this greatly simplifies estimation because the model involves only two spectral densities that can be estimated using the marginal spatial covariances of the response and exposure, as described below. Moreover, if the marginal spectral densities differ for some frequencies then these frequencies can be used to identify β_x. The expression in (3) simplifies under the parsimonious model to

$$\mathcal{Y}(\omega) \mid \mathcal{X}(\omega) \overset{\text{ind.}}{\sim} N\left[\left\{\beta_x + \rho C\beta_z\frac{\sigma_z\sqrt{f_z(\omega)}}{\sigma_x\sqrt{f_x(\omega)}}\right\}\mathcal{X}(\omega), \{1-(\rho C)^2\}\beta_z^2\sigma_z^2 f_z(\omega) + \sigma^2\right]. \quad (4)$$

It is important to establish the identifiability of parameters in (4). The ρ and C are not uniquely identified, nor are β_z and σ_z, as they both appear in the model only through products. However, $\rho^* = \rho C$ is identified, as is σ_z if $\beta_z = 1$. An alternative parameterization is to let $\sigma_z^* = \beta_z\sigma_z$ and σ_z^* can be identified. The two expressions are equivalent, which does not affect the result of Theorem 1 below. The identifiability of ρ^*, σ_z under these conditions and that of the remaining parameters in (4) is established in the following theorem.

THEOREM 1. *Assume that $f_x(\omega) \neq f_z(\omega)$ for some ω, and set $\beta_z = 1$. Then parameters β_x, ρ^*, σ_x and σ_z, and functions f_x and f_z in model* (4) *are all identified.*

2.3. *Spatial representation of confounding and identification*

Returning to the spatial domain, the response process can be written as

$$Y(s) \mid X(s) = \beta_0 + \beta_x X(s) + \beta_z\hat{Z}(s) + \delta(s),$$

$$\hat{Z}(s) = \int \exp(i\omega^{\mathrm{T}} s)\hat{\mathcal{Z}}(\omega)\,\mathrm{d}\omega = \int \exp(i\omega^{\mathrm{T}} s)\alpha(\omega)\mathcal{X}(\omega)\,\mathrm{d}\omega,$$

and $\delta(s)$ is a mean-zero Gaussian process with spectral density $\tau^2(\omega) + \sigma^2$ independent of $X(s)$ and $\hat{Z}(s)$. The function $\alpha(\omega)$ acts as a smoothing operator and, if it were known, then $\hat{Z}(s)$ would be an appropriate adjustment to the mean to account for the unmeasured confounder. The form of the oracle confounder adjustment, i.e., $\hat{Z}(s)$ if $\alpha(\omega)$ is known, is established in the following lemma.

LEMMA 1. *If $\alpha(\omega)$ is known then $\hat{Z}(s) = \int K(s - s')X(s')\,\mathrm{d}s'$, where the kernel function $K(s - s')$ is the inverse Fourier transform of $\alpha(\omega)$.*

The appealing consequence of Lemma 1 is that the oracle confounder adjustment is conveniently expressed as a kernel-smoothed function of the covariate of interest. It is also straightforward to show that, for any n locations $\mathcal{S}$,

$$Y \mid X = \beta_0 + \beta_x X + \beta_z \hat{Z} + \delta, \quad \text{where} \quad \hat{Z} = \Sigma_{zx}\Sigma_x^{-1} X, \tag{5}$$

$\mathrm{cov}(\delta) = \beta_z^2(\Sigma_z - \Sigma_{zx}\Sigma_x^{-1}\Sigma_{zx}^T) + \sigma^2 I_n$ and $\Sigma_{zx} = \mathrm{cov}(Z, X)$, $\Sigma_x = \mathrm{cov}(X)$ and $\Sigma_z = \mathrm{cov}(Z)$. The product $\Sigma_{zx}\Sigma_x^{-1}$ serves as a smoothing operator on X. This representation is convenient for estimation and to determine the strength of confounding dependence between X and Z.

Since $\hat{Z}$ is a smoothed version of X, including it as a covariate in (5) effectively removes effects of the large-scale spatial trends in X, so that the estimate of β_x is largely determined by high-frequency terms. This expression also lays bare the importance of assuming that $\alpha(\omega)$ converges to zero for large frequencies or that X and Z have different spectral densities. If restrictions are not placed on $\alpha(\omega)$ then it may be that $\Sigma_{zx} = \Sigma_x$ and thus $\hat{Z} = X$, giving the nonidentifiable model $Y = \beta_0 + \beta_x X + \beta_z X + \delta$.

3. CONTINUOUS-SPACE ESTIMATION STRATEGIES

3.1. *The bivariate Matérn parametric model*

The bivariate Matérn model (Gneiting et al., 2010; Apanasovich et al., 2012) is a flexible parametric model for the spectral densities f_x, f_z and f_{xz}. The Matérn spectral density function for a process in two dimensions is $m(\omega; \nu, \phi) = \nu\phi^{-2\nu}(\phi^{-2} + \|\omega\|^2)^{-(\nu+1)}$, with smoothness $\nu > 0$ and spatial range $\phi > 0$. The bivariate Matérn may have different parameters for each process, $f_j(\omega) = m(\omega; \nu_j, \phi_j)$ for $j \in \{x, z, xz\}$, but constraints on the range and smoothness parameters are needed to ensure that the coherence is positive definite for all ω (Gneiting et al., 2010; Apanasovich et al., 2012). Another advantage of the Matérn parametric model is the closed-form expressions for both the spectral density and covariance functions. Under a common range assumption, as described in the next paragraph, the projection operator $\alpha(\omega)$ also has a closed-form Fourier transformation, allowing the estimation procedure to be performed completely in the spatial domain using (5).

With the bivariate Matérn modelling assumption, the projection operator has the form $\alpha(\omega) = \rho\sigma_z m(\omega; \nu_{xz}, \phi_{xz})/\{\sigma_x m(\omega; \nu_x, \phi_x)\}$. Therefore, if the cross-spectral density decays faster than the covariate spectral density, the confounding adjustment will be smaller for higher frequencies, i.e., $\alpha(\omega) \to 0$. Comparison of the ratio of spectral densities such that $\alpha(\omega) \to 0$ is complicated in general, and so we explore the special cases of common range,

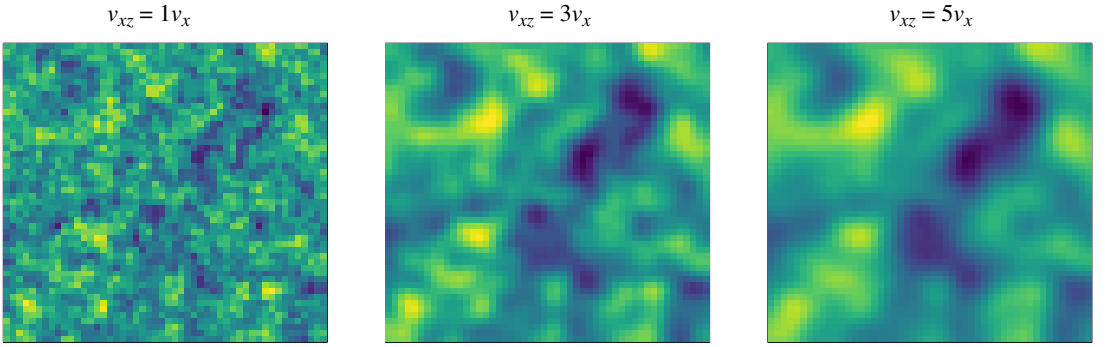


Fig. 1. Example confounder adjustment for the bivariate Matérn: $X(s)$ is generated from Matérn with $\phi_x = 1$ and $\nu_x = 1$ on a 50×50 grid with grid spacing one. The panels show the confounder adjustment $\hat{Z}(s)$ for $\phi_{xz} = \phi_x$ and $\nu_{xz} = c\nu_x$ for $c \in \{1, 3, 5\}$. For $c = 1$, $\hat{Z}(s) = X(s)$.

common smoothness and parsimonious models below. For each special case, we discuss the parameter settings that ensure unconfoundedness at high frequencies.

The common range model takes $\phi_j = \phi$ for $j \in \{x, z, xz\}$,

$$\alpha(\omega) = \rho \frac{\sigma_z}{\sigma_x} (\phi^{-2} + \|\omega\|^2)^{-(\nu_{xz} - \nu_x)}. \tag{6}$$

In this case, we have unconfoundedness at high frequencies, $\alpha(\omega) \to 0$, if and only if $\nu_{xz} > \nu_x$, i.e., the cross-covariance is smoother than the covariate covariance. On the other hand, if we assume a common smoothness $\nu_j = \nu$ for $j \in \{x, z, xz\}$ then $\alpha(\omega) \to (\phi_x/\phi_{xz})^{2\nu}$ and confounding persists at high frequencies, regardless of the range parameters. Therefore, a common range parameterization allows us to identify the exposure effect by reducing high-resolution confounding while a common smoothness parameterization will not. For simplicity, we assume a common range in the remainder of this section.

Figure 1 illustrates the confounder adjustment $\hat{Z}$ in (5) for the bivariate Matérn model with a common range and different smoothness $\nu_{zx} = c\nu_x$ for increasing values of c. Increasing c implies increasing decay rates of $\alpha(\omega)$. The original simulated X is plotted in the left panel of Fig. 1 with $c = 1$, in which case we have $\hat{Z} = X$ and thus a completely confounded model. In the cases with $c > 1$ the confounder adjustment $\hat{Z}$ is a smoothed version of X. Therefore, including $\hat{Z}$ as a covariate in the model removes large-scale trends in X to adjust for confounding at low frequencies.

The unmeasured confounder cannot be observed, making it difficult to estimate all the parameters in the bivariate Matérn model. Therefore, additional constraints are required for identifiability of the remaining parameters in addition to β_x. These are provided in the next theorem

THEOREM 2. *Assuming that $\beta_z = 1$ and a common range parameter, sufficient conditions for identifiability of the remaining parameters $\{\beta_x, \rho, \nu_x, \nu_z, \nu_{xz}, \sigma_x, \sigma_z\}$ are a large cross-smoothness parameter $\nu_{xz} > \max\{\nu_x, (\nu_x + \nu_z)/2\}$ and $\rho^2 < \nu_x \nu_z / \nu_{xz}^2$.*

The common range model simplifies further under the parsimonious model in (4) with $C = (\nu_x + \nu_z)/(2\sqrt{\nu_x \nu_z})$, $|\rho| < 1/C$ and $f_{xz}(\omega) = m\{\omega; (\nu_x + \nu_z)/2, \phi\}$. This is the parsimonious Matérn model of Gneiting et al. (2010), i.e., the cross-smoothness equals the average of the marginal smoothness parameters, $\nu_{xz} = (\nu_x + \nu_z)/2$. Under this model, the

confounder adjustment becomes

$$\alpha(\omega) = \rho \frac{\sigma_z}{\sigma_x} (\phi^{-2} + \|\omega\|^2)^{-(\nu_z - \nu_x)/2}, \tag{7}$$

and thus $\alpha(\omega) \to 0$ if and only if $\nu_z > \nu_x$, i.e., the missing confounder is smoother than the exposure. On the other hand, if $\nu_z < \nu_x$ then $\alpha(\omega) \to \infty$. However, $\alpha(\omega) \to 0$ is not needed here as the identification strategy for the parsimonious coherence is established in Theorem 1.

3.2. *Semiparametric model*

Rather than indirectly modelling the projection operator $\alpha(\omega)$ via a model for the cross-covariance function, in this section we directly model $\alpha(\omega)$ using a flexible mixture model. We use a linear combination of cubic B-splines of order 4,

$$\alpha(\omega) = \sum_{l=1}^{L} B_l(\|\omega\|) b_l, \qquad 0 < \|\omega\| < \pi/\triangle_s,$$

where the $B_l(\cdot)$ are B-spline basis functions, the b_l are the associated coefficients and $\triangle_s = \max\{ds_1, ds_2\}$ for grid spacing (ds_1, ds_2). A uniform sequence of knots $\{\omega_1^*, \ldots, \omega_L^*\}$ is placed to cover the interval $[0, \pi/\triangle_s]$, such that $0 \in (\omega_1^*, \omega_2^*)$ and $\pi/\triangle_s \in (\omega_{L-1}^*, \omega_L^*)$. The interval upper bound $\pi/\triangle_s$ is the largest spectrum that can be observed from uniformly spaced data due to aliasing (Fuentes & Reich, 2010). We have restricted the projection operator to be isotropic by taking the Euclidean norm $\|\cdot\|$ of the two-dimensional frequency ω, but this can be relaxed by using bivariate spline functions. Other mixture priors (Reich & Fuentes, 2012; Jang et al., 2017; Chen et al., 2021) can also be used for modelling the projection operator.

The B-spline mixture model for $\alpha(\omega)$ does not have a closed-form inverse Fourier transformation. We approximate the kernel-smoothed function $K(s-s')$ with a finite sum at a set of equally spaced frequencies $\mathcal{F} = \{\omega_1^f, \ldots, \omega_m^f\}$ with spacing $\triangle_{\mathcal{F}}$ and $\omega_m^f = \pi/\triangle_s$ following Qadir & Sun (2020):

$$K(s-s') = \sum_{\omega^f \in \mathcal{F}} h \left(\frac{2\pi\omega^f}{h} \right)^{\kappa+1} \mathcal{J}_\kappa(\omega^f h) \alpha(\omega^f) \triangle_{\mathcal{F}}.$$

Here $h = \|s - s'\|$, $\kappa = d/2 - 1$ and $\mathcal{J}_\kappa(\cdot)$ is a Bessel function of the first kind of order κ (Watson, 1995). This approximation allows us to directly compute confounder adjustment in the spatial domain, which would otherwise require the Fourier transform of data to perform analysis in the spectral domain. The confounder adjustment is then given by $\hat{Z}(s) = \sum_{l=1}^{L} b_l \hat{Z}_l(s)$, where

$$\hat{Z}_l(s) = \sum_{\omega^f \in \mathcal{F}} (2\pi\omega^f)^{\kappa+1} B_l(\omega^f) \triangle_{\mathcal{F}} \int \frac{\mathcal{J}_\kappa(\omega^f h)}{h^\kappa} X(s') \, \mathrm{d}s'. \tag{8}$$

When $X(s)$ is observed on a grid, the integral can be approximated as $\int \mathcal{J}_\kappa(\omega^f h) h^{-\kappa} X(s') \, \mathrm{d}s' = (1/n) \sum_{i=1}^{n} \mathcal{J}_\kappa(\omega^f h_i) h_i^{-\kappa} X(s_i)$ with $h_i = \|s - s_i\|$. For nongridded data, the

covariate can be interpolated to a grid and this discrete approximation to the grid can be applied. Other numerical approximations to integrals can be applied to nongridded data such as the finite element method (Johnson, 2012, Ch. 12). The confounder adjustment covariates $\hat{Z}_l(s)$ are precomputed to reduce computation during model fitting. We then fit the spatial model

$$Y(s) = \beta_0 + \beta_x X(s) + \sum_{l=1}^{L} b_l \hat{Z}_l(s) + \delta(s), \tag{9}$$

where $\beta_z = 1$ for identification and $\delta(s)$ is modelled as a Gaussian process with nugget. The coefficients $(b_1, \ldots, b_L)^{\mathrm{T}}$ are given intrinsic autoregressive priors with full conditional distributions $b_k \mid b_{(-k)} \sim N(\bar{b}_k, \sigma_b^2/N_k)$, where $\bar{b}_k$ is the mean of the N_k coefficients b_l with $|l-k| = 1$, so $N_1 = N_L = 1$ and $N_2 = \cdots = N_{L-1} = 2$.

4. Discrete-space methodology

4.1. *A spectral model for confounding*

We extend our methodology to the discrete case for a spatial domain comprised of n regions. For region i, let Y_i, X_i and Z_i be the response, exposure and confounding variables, respectively. Let $Y = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ and define X, Z similarly. We model Z using the conditional autoregressive model (Gelfand et al., 2010) with the Leroux parameterization (Leroux et al., 2000),

$$Z \sim N[\mu, \sigma_z^2\{(1-\lambda_z)I_n + \lambda_z R\}^{-1}],$$

where $\mu = (\mu_1, \ldots, \mu_n)^{\mathrm{T}}$ is the mean vector, σ_z^2 determines the overall variance, $\lambda_z \in [0, 1]$ controls the strength of spatial dependence and R is an $n \times n$ matrix specifying the spatial dependence. For the discrete case, the spatial dependence between the regions is often described by an adjacency structure. Let $a_{ij} = 1$ if regions i and j are adjacent and 0 otherwise, and let m_i be the number of regions adjacent to region i. Then R has (i, j) off-diagonal element $-a_{ij}$ and ith diagonal element m_i. We denote this model as $Z \sim \mathrm{CAR}(\mu, \sigma_z^2, \lambda_z)$.

An advantage of the Leroux parameterization is that the spatial covariance can be written as

$$\sigma_z^2\{(1-\lambda_z)I_n + \lambda_z R\}^{-1} = \sigma_z^2 \Gamma\{(1-\lambda_z)I_n + \lambda_z W\}^{-1}\Gamma^{\mathrm{T}},$$

where the spectral decomposition of R is $R = \Gamma W \Gamma^{\mathrm{T}}$ for orthonormal eigenvector matrix Γ and diagonal eigenvalue matrix W with kth diagonal element $\omega_k \geqslant 0$, ordered so that $\omega_1 \leqslant \cdots \leqslant \omega_n$. Assuming that all variables have the same adjacency structure R, this allows us to project the model into the spectral domain using the graph Fourier transform (Sandryhaila & Moura, 2013), $Y^* = \Gamma^{\mathrm{T}} Y = (Y_1^*, \ldots, Y_n^*)^{\mathrm{T}}$, $X^* = \Gamma^{\mathrm{T}} X = (X_1^*, \ldots, X_n^*)^{\mathrm{T}}$ and $Z^* = \Gamma^{\mathrm{T}} Z = (Z_1^*, \ldots, Z_n^*)^{\mathrm{T}}$. This transformation decorrelates the model and gives

$$Y_k^* \mid X_k^*, Z_k^* \overset{\text{indep}}{\sim} N(\beta_0 M_k + \beta_x X_k^* + \beta_z Z_k^*, \sigma^2),$$

where M_k is the sum of the kth column of Γ and (X_k^*, Z_k^*) are independent across k. To exploit this decorrelation property of the graph Fourier transform, we conduct all analyses of Gaussian data for the discrete spatial domain in the spectral scale.

Comparing the discrete to continuous cases, the eigenvalue ω_k is analogous to frequency ω. Terms with small ω_k have large variance and measure large-scale trends in the data. For example, it can be shown that if the n locations form a connected graph then $\omega_1 = 0$ and Y_1^* is proportional to the mean of Y. In contrast, terms with large ω_k have small variance and represent small-scale features. Using this analogy, in the remainder of this section we extend two of the continuous-domain methods of § 3 to the discrete case.

4.2. *Bivariate conditional autoregressive model*

As in § 2, we assume a joint model for X^* and Z^*. We assume that the pairs (X_k^*, Z_k^*) are independent across k, and Gaussian with mean zero and covariance

$$\operatorname{cov}\begin{Bmatrix} X_k^* \\ Z_k^* \end{Bmatrix} = \begin{Bmatrix} \sigma_x^2 f_x(\omega_k) & \rho\sigma_x\sigma_z f_{xz}(\omega_k) \\ \rho\sigma_x\sigma_z f_{xz}(\omega_k) & \sigma_z^2 f_z(\omega_k) \end{Bmatrix}, \tag{10}$$

where σ_x^2 and σ_z^2 are variance parameters, $f_x(\omega_k) > 0$ and $f_z(\omega_k) > 0$ are variance functions that determine the covariance of X and Z, respectively, and scalar ρ and function $f_{xz}(\omega_k)$ determine the dependence between X and Z. For the Leroux conditional autoregressive model, we have $f_j(\omega_k) = 1/(1 - \lambda_j + \lambda_j\omega_k)$ for $j \in \{x, z\}$ so that the marginal distributions are $X \sim \text{CAR}(0, \sigma_x^2, \lambda_x)$ and $Z \sim \text{CAR}(0, \sigma_z^2, \lambda_z)$.

One possible parametric cross-covariance model is $f_{xz}(\omega) = 1/(1 - \lambda_{xz} + \lambda_{xz}\omega)$. As with the bivariate Matérn, f_{xz} has the same functional form as f_x and f_z. Constraints are required to ensure that the covariance in (10) is positive definite, i.e., that

$$\rho^2(1 - \lambda_x + \lambda_x w)(1 - \lambda_z + \lambda_z w) < (1 - \lambda_{xz} + \lambda_{xz} w)^2 \tag{11}$$

for all $w \in \{\omega_1, \ldots, \omega_n\}$. Necessary conditions for (11) to hold for all $w \geqslant 0$ are $\rho^2(1 - \lambda_x)(1 - \lambda_z) < (1 - \lambda_{xz})^2$ and $\rho^2\lambda_x\lambda_z < \lambda_{xz}^2$, but these conditions are not sufficient and not even necessary when considering only $w \in \{\omega_1, \ldots, \omega_n\}$.

Assuming that the covariance parameters give a valid covariance, then marginalizing over Z_k^* and setting $\beta_z = 1$, as in § 2, for identification gives

$$Y_k^* \mid X_k^* \overset{\text{indep}}{\sim} N\{\beta_0 M_k + \beta_x X_k^* + \alpha(\omega_k) X_k^*, \tau^2(\omega_k) + \sigma^2\}, \tag{12}$$

where $\alpha(\omega_k) = \rho\sigma_z\sigma_x^{-1}(1 - \lambda_x + \lambda_x\omega_k)(1 - \lambda_{xz} + \lambda_{xz}\omega_k)^{-1}$ and $\tau^2(\omega_k) = \sigma_z^2(1 - \lambda_z + \lambda_z\omega_k)^{-1} - \rho^2\sigma_z^2(1 - \lambda_x + \lambda_x\omega_k)(1 - \lambda_{xz} + \lambda_{xz}\omega_k)^{-2}$. Therefore, $\alpha(\omega) \to \rho\sigma_z\sigma_x^{-1}\lambda_x\lambda_{xz}^{-1}$ as $\omega \to \infty$, and thus the high-resolution confounding effect is smallest when λ_x is smaller than λ_{xz}.

The parsimonious cross-covariance model is $f_{xz}(\omega_k) = \sqrt{f_x(\omega_k) f_z(\omega_k)}$, giving $\operatorname{cor}(X_k^*, Z_k^*) = \rho$ for all k. With this simplification, any $\rho \in (-1, 1)$ and $\lambda_x, \lambda_z \in (0, 1)$ give a valid covariance, and the terms in (12) reduce to $\alpha(\omega_k) = \rho\sigma_z\sigma_x^{-1}(1 - \lambda_x + \lambda_x\omega_k)^{1/2}(1 - \lambda_z + \lambda_z\omega_k)^{-1/2}$ and $\tau^2(\omega_k) = \sigma_z^2(1 - \rho^2)/(1 - \lambda_z + \lambda_z\omega_k)$. Here, the missing confounder need not be smoother than the exposure, i.e., $\lambda_z > \lambda_x$, for identifying β_x as the assumption for identification is parsimonious coherence. As long as $\lambda_z \neq \lambda_x$, the remaining parameters can also be identified, as established in the next theorem.

THEOREM 3. *Assuming that $\lambda_x \neq \lambda_z$ and $\beta_z = 1$, then the parameters β_x, σ^2, ρ, λ_x, λ_z, σ_x^2, σ_z^2 in the parsimonious model* (13) *below are all identifiable.*

We have fixed $\beta_z = 1$ in Theorem 3 so we can estimate σ_z, but this is unnecessary. The alternative parameterization $\sigma^* = \beta_z \sigma_z$ can be used, which does not affect the results, as discussed in § 2.2.

In the spatial domain, the parsimonious model is

$$Y \mid X, V \sim N(\beta_0 + \beta_x X + \Gamma A \Gamma^{\mathrm{T}} X + V, \sigma^2 I_n), \tag{13}$$

where $V \sim \textsc{car}\{0, \sigma_z^2(1-\rho^2), \lambda_z\}$ and A is diagonal with kth diagonal element $\alpha(\omega_k)$. The term $\Gamma A \Gamma^{\mathrm{T}} X$ adjusts for missing spatial confounders and the term V captures spatial variation that is independent of X. In this case with $\lambda_z > \lambda_x$, the confounder adjustment $\Gamma A \Gamma^{\mathrm{T}} X$ smooths X by first projecting into the spectral domain by multiplying by Γ^{T}, then dampening the high-frequency terms with large ω and thus small $\alpha(\omega)$ by multiplying by A, and finally projecting back in the spatial domain by multiplying by Γ. Marques et al. (2022) developed a Bayesian method to mitigate spatial confounding. Their model is related to our parsimonious Leroux conditional autoregressive model, except they used a Gaussian Markov random field model using the stochastic partial differential equation approach (Lindgren et al., 2011) for X and Z, and a penalized complexity prior for ρ.

4.3. *Semiparametric conditional autoregressive model*

Mirroring § 3.2, rather than specifying a parametric joint model for (X_k^*, Z_k^*), we directly specify a flexible model for the confounder adjustment, $\alpha(\omega)$. The joint model is specified first with the conditional model

$$Z_k^* \mid X_k^* \overset{\text{indep}}{\sim} N\left\{\alpha(\omega_k) X_k^*, \frac{\sigma_z^2}{1 - \lambda_z + \lambda_z \omega_k}\right\}.$$

In the spatial domain, this implies that $Z \mid X \sim \textsc{car}(\Gamma A \Gamma^{\mathrm{T}} X, \sigma_z, \lambda_z)$, where A is diagonal with diagonal elements $\{\alpha(\omega_1), \ldots, \alpha(\omega_n)\}$. Therefore, with any valid marginal distribution of X, the joint model of X and Z is well defined. Since X is observed, we do not need a model for its marginal distribution.

Marginalizing over the unknown Z^* gives

$$Y_k^* \mid X_k^* \overset{\text{indep}}{\sim} N\left\{\beta_0 M_k + \beta(\omega_k) X_k^*, \frac{\sigma_z^2}{1 - \lambda_z + \lambda_z \omega_k} + \sigma^2\right\}, \tag{14}$$

where $\beta(\omega_k) = \beta_x + \alpha(\omega_k)$. Following § 3.2, we assume that $\alpha(\omega_n) = 0$ so that X_n^* and Z_n^* are uncorrelated for the highest-frequency term. This implies that $\beta(\omega_n) = \beta_x$ and $\mathrm{E}(Y_n^*) = \beta_0 M_n + \beta_x X_n$, and thus the final term supplies unbiased information about the true exposure effect β_x. Of course, a single unbiased term is insufficient for estimation, and so we further assume that $\alpha(\omega)$ varies smoothly over ω to permit semiparametric estimation of β_x.

We fit model (14) with a covariate effect that is allowed to vary with k to separate associations at different spatial resolutions. Although other smoothing techniques are possible, the frequency-specific coefficients are smoothed using the basis expansion $\beta(\omega) = \sum_{l=1}^{L} B_l(\omega) b_l$, where the $B_l(\omega)$ are cubic B-spline basis functions and the b_l are the associated coefficients. Analogously to the semiparametric continuous-space model in § 3.2, we employ equally spaced B-splines (Eilers & Marx, 1996) with an intrinsic autoregressive prior on $(b_1, \ldots, b_L)^{\mathrm{T}}$. Under the assumption that $\alpha(\omega_n) = 0$, we use the posterior distribution of $\beta(\omega_n)$ to summarize the effect β_x.

In the spatial domain, the semiparametric conditional autoregressive model can be written as (9), i.e.,

$$Y|X, V \sim N\left(\beta_0 + \sum_{l=1}^{L} \hat{Z}_l b_l + V, \sigma^2 I_n\right),$$

where $V \sim \text{CAR}(0, \sigma_z^2, \lambda_z)$, $\hat{Z}_l = \Gamma B_l \Gamma^{\mathrm{T}} X$, B_l is the diagonal matrix with spline basis functions, $\{B_l(\omega_1), \ldots, B_l(\omega_n)\}$, on the diagonal and the regression coefficients are modelled as described below (8). The constructed covariates $\hat{Z}_l$ can be precomputed prior to estimation, and thus computation resembles a standard spatial analysis with L known covariates. As above, under the assumption of no confounding for large ω, we use the posterior of $\beta_x = \sum_{l=1}^{L} B_l(\omega_n) b_l$ to summarize the exposure effect. As our estimation of β_x relies on a B-spline estimate at an endpoint, it may be associated with relatively large uncertainty. Therefore, other smoothing techniques may also be considered to test the sensitivity of the estimate to the chosen method of estimation.

5. Simulation study

5.1. *Discrete space*

Data are generated at n locations $s_1, \ldots, s_n$ on a 40×40 square grid with grid spacing one. The conditional autoregressive model uses rook neighbourhood structure so that $a_{ij} = 1$ if and only if $\|s_i - s_j\| = 1$. Data are generated from $X \sim \text{CAR}(0, \sigma_x^2, \lambda)$, $Z \mid X \sim \text{CAR}(\beta_{xz} WX, \sigma_z^2, \lambda)$ and $Y \mid X, Z \sim N(\beta_x X + \beta_z Z, \sigma^2 I_n)$, where W is the kernel smoothing matrix with bandwidth ϕ, i.e., $W_{ij} = w_{ij}/(\sum_{l=1}^{n} w_{il})$ and $\log(w_{ij}) = -(\|s_i - s_j\|/\phi)^2$. Including the kernel smoothed X in the mean of Z induces low-resolution dependence between X and Z. In all cases we take $\sigma_x^2 = 1.7$, $\sigma_z^2 = 1$, $\lambda = 0.95$, $\beta_x = \beta_z = 0.5$ and $\sigma^2 = 0.25^2$, and we vary the strength of dependence via $\beta_{xz} \in \{0, 1, 2\}$, and the kernel bandwidth $\phi \in \{1, 2\}$. The value of β_z can be chosen without loss of generality, as only the product of β_z and σ_z can be uniquely identified. For each parameter combination, we generate 500 datasets; see the data examples in the Supplementary Material.

Figure 2 plots the induced correlations in the spectral domain for each scenario with $\beta_{xz} > 0$. The correlation is nonzero for only low-frequency terms when $\phi = 2$, but correlation spills over to high-frequency terms when $\phi = 1$, especially when $\beta_{xz} = 2$. Therefore, the assumption of no confounding at high frequencies is questionable when $\phi = 1$, and these scenarios are used to examine sensitivity to this key assumption. Also, these scenarios violate the parsimonious assumption of constant correlation across frequency, and so they illustrate the effects of misspecifying the parametric model.

For each simulated dataset, we fit the standard Leroux conditional autoregressive model that has $\beta_k = \beta_x$ for all k, the parametric parsimonious bivariate conditional autoregressive model and the semiparametric model with β_k varying across k using a cubic B-spline basis expansion. We compare two priors for σ_b^2, the variance of the coefficient process β_k, for the semiparametric model. The penalized complexity prior shrinks the process towards the constant function $\beta_k = \beta_x$ to avoid overfitting (Franco-Villoria et al., 2019); the second prior for the variance induces a $\text{Un}(0, 1)$ prior on the proportion of overall model variance explained by variation in β_k to balance all levels of spatial confounding. The prior distributions for all models are given in the Supplementary Material. We fit the semiparametric

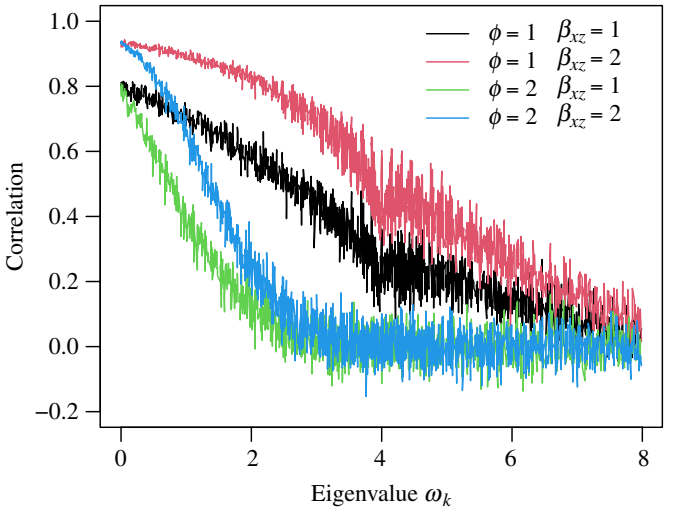


Fig. 2. Correlations in the spectral domain for the simulation study: $\mathrm{cor}(X_k^*, Z_k^*)$ by ω_k for different kernel bandwidths (ϕ) and strengths of exposure/confounder dependence (β_{xz}); the correlations in the spatial domain (over locations) $\mathrm{cor}(X_i, Z_i)$ are 0.62 when $\phi = 1$ and $\beta_{xz} = 1$, 0.80 when $\phi = 1$ and $\beta_{xz} = 2$, 0.44 when $\phi = 2$ and $\beta_{xz} = 1$, and 0.62 when $\phi = 2$ and $\beta_{xz} = 2$.

models for all $L \in \{1, 5, 10, 20, 30, 40\}$ and select the number of basis functions using the deviance information criterion (Spiegelhalter et al., 2002). All methods are fit using Markov chain Monte Carlo with 25 000 iterations and the first 5000 discarded as burn-in.

Table 1 compares methods in terms of the root mean squared error, bias, posterior standard deviation averaged over datasets and empirical coverage of 95% intervals for β_x, and Fig. 3 summarizes the sampling distribution of β_k against ω_k. The standard method performs well in the first scenario with no unmeasured confounder, $\beta_{xz} = 0$, but in all other scenarios the standard method is biased and has coverage at or near zero. The standard method allows for spatially dependent residuals, but this does not eliminate spatial confounding bias. Since the standard model assumes that X and Z are independent, when X and Z are highly correlated, all spatial variability is attributed to the exposure effect, leading to bias and small posterior standard deviation.

The parametric model performs well in scenario 1, and respectably in scenarios 4 and 5 where there is no confounding at high frequencies and parametric model assumptions do not hold. In fact, the parametric model is nearly identical to the standard model in the first case with no spatial confounding, suggesting that little is lost by allowing for a parametric confounding adjustment when it is not needed. However, the parametric model gives bias and low coverage in the cases with $\phi = 1$, and thus the form of spatial confounding does not match the parametric model. The estimated β_k curves in Fig. 3 show that the parametric form of the β_k model cannot match the slow decline in the true correlation of Fig. 2 when $\phi = 1$. However, the parametric model is still able to recover reasonably well the truth for $\phi = 2$; it appears that there is some robustness to the parsimonious coherence assumption for the parametric model if the unconfoundedness at high frequencies assumption holds.

The semiparametric methods have low bias and coverage near the nominal level for all five scenarios. However, the posterior standard deviation is always larger for the semiparametric models than the standard or parametric models. Therefore, in these cases, the semiparametric methods are robust, but conservative for estimating a casual effect in the presence of

Table 1. *Discrete-space simulation study comparing four methods: the standard Leroux model, the parametric parsimonious model, the semiparametric model with a penalized complexity prior and the semiparametric model with a uniform prior on the proportion of variance (semi-R^2). Data are generated with dependence between exposure and confounder controlled by β_{xz} and kernel bandwidth ϕ. Standard errors are given in parentheses and all results are multiplied by 100*

Scenario	Method	ϕ	β_{xz}	RMSE	Bias	SD	Cov
1	Standard	–	0	1.2 (0.0)	0.0 (0.1)	1.3 (0.0)	95.2 (1.0)
	Parametric			1.3 (0.0)	0.0 (0.1)	1.4 (0.0)	95.4 (0.9)
	Semiparametric			4.4 (0.3)	0.0 (0.2)	2.5 (0.1)	95.0 (1.0)
	Semi-R^2			4.5 (0.4)	0.0 (0.2)	2.5 (0.1)	94.8 (1.0)
2	Standard	1	1	19.0 (0.1)	18.9 (0.1)	1.4 (0.0)	0.0 (0.0)
	Parametric			14.0 (0.1)	13.8 (0.1)	1.7 (0.0)	0.0 (0.0)
	Semiparametric			7.7 (0.3)	−0.9 (0.3)	8.3 (0.1)	96.6 (0.8)
	Semi-R^2			8.0 (0.3)	−0.7 (0.4)	8.9 (0.1)	97.2 (0.7)
3	Standard	1	2	34.3 (0.1)	34.3 (0.1)	1.7 (0.0)	0.0 (0.0)
	Parametric			20.6 (0.2)	20.4 (0.1)	2.0 (0.0)	0.0 (0.0)
	Semiparametric			9.5 (0.3)	0.6 (0.4)	9.3 (0.1)	94.4 (1.0)
	Semi-R^2			9.5 (0.3)	0.8 (0.4)	9.4 (0.1)	96.0 (0.9)
4	Standard	2	1	5.6 (0.1)	5.4 (0.1)	1.4 (0.0)	3.8 (0.9)
	Parametric			2.1 (0.1)	1.2 (0.1)	1.6 (0.0)	85.4 (1.6)
	Semiparametric			8.2 (0.2)	−0.9 (0.4)	9.3 (0.1)	96.6 (0.8)
	Semi-R^2			8.2 (0.3)	−0.7 (0.4)	9.2 (0.1)	95.8 (0.9)
5	Standard	2	2	8.8 (0.1)	8.7 (0.1)	1.5 (0.0)	0.0 (0.0)
	Parametric			2.5 (0.1)	−1.1 (0.1)	1.8 (0.0)	84.0 (1.6)
	Semiparametric			10.2 (0.3)	−0.6 (0.5)	10.3 (0.1)	94.8 (1.0)
	Semi-R^2			10.4 (0.3)	−0.6 (0.5)	10.3 (0.1)	94.0 (1.1)

RMSE, root-mean-squared error; SD, average posterior standard deviation; Cov, coverage of 95% posterior intervals.

spatial confounding. Surprisingly, the semiparametric methods are insensitive to the choice of prior. Despite the two prior specifications having very different motivations, the results are similar, likely because the deviance information criterion often selects a small number of basis functions that negates the influence of the prior for σ_b^2.

5.2. *Continuous space*

Data in the continuous space are generated similarly to the discrete case. The data are simulated on a 23×23 unit square grid. We simulated $X \sim N(0, \sigma_x^2 \Sigma_x)$, $Z \mid X \sim N(\beta_{xz} WX, \sigma_z^2 \Sigma_z)$ and $Y \mid X, Z \sim N(\beta_x X + \beta_z Z, \sigma^2 I_n)$, where Σ_j is the $n \times n$ Matérn correlation matrix defined by parameters ϕ_j and ν_j, and W is the kernel smoothing matrix with bandwidth ϕ. In all cases we take $\sigma_x^2 = \sigma_z^2 = 1$, spatial range parameters $\phi_x = \phi_z = 0.1$, $\nu_x = \nu_z = 0.5$, $\beta_x = \beta_z = 1$ and $\sigma^2 = 0.25^2$, and we vary $\beta_{xz} \in \{0, 1, 2\}$, and the kernel bandwidth $\phi \in \{1/15, 2/15\}$. For each combination of these parameters, we generate 100 datasets. For each simulated dataset, we fit four models: the standard Matérn model in § 3.1 with $\rho = 0$, and thus no confounding adjustment, the bivariate Matérn model with common range in (6), the parsimonious Matérn model in (7) and the semiparametric model in § 3.2. Prior distributions and computing details are given in the Supplementary Material.

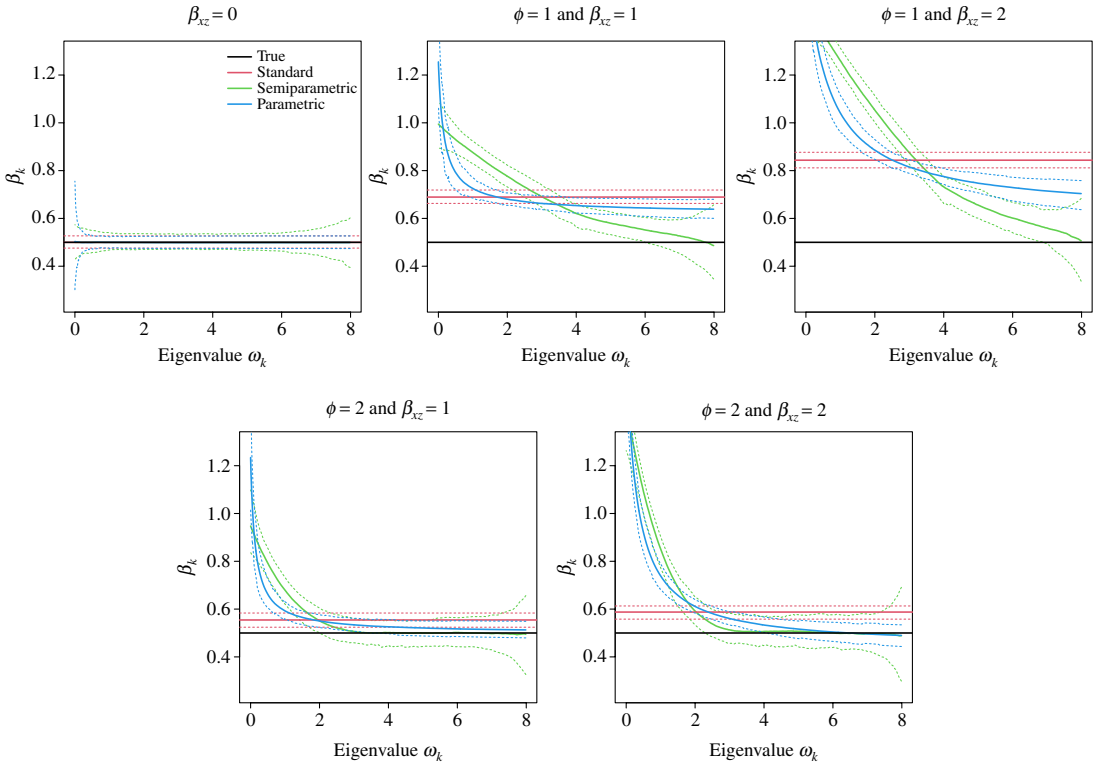


Fig. 3. Performance for the discrete simulation study: median (solid) and 95% confidence interval (dashed) for β_k for the standard model (red), semiparametric model with the penalized complexity prior (green), and parametric model (blue) for data generated with dependence between exposure, and confounder controlled by β_{xz} and kernel bandwidth ϕ. The black lines are the true $\beta_x = 0.5$.

The results mirror those in the discrete case and therefore the result table is deferred to the Supplementary Material. The semiparametric method maintains nearly the nominal coverage and low bias across all scenarios. The parametric Matérn models have bias and low coverage for the simulation settings where the data are not simulated with a Matérn covariance. The common range Matérn model dramatically reduces root-mean-square error and improves coverage compared to the parsimonious Matérn model, but neither is sufficiently flexible for these cases.

6. Real data examples

6.1. *Analysis of the Scottish lip cancer dataset*

All model fits in this section were carried out using the `eCAR` package found in `R` (R Development Core Team, 2023) that was created to fit the discrete space methods. We first consider the well-known lip cancer dataset, see Fig. 4, available in the `R` package `CARBayesdata`. The data cover $n = 56$ districts in Scotland. Three variables are recorded for each district: the recorded number of lip cancer cases, Y_i, the expected number of lip cancer cases computed using indirect standardization based on Scotland-wide disease rates, E_i, and the percentage of the district's workforce employed in agriculture, fishing and forestry, X_i. Since we have non-Gaussian responses, the generalized models presented in the Supplementary Material are used.

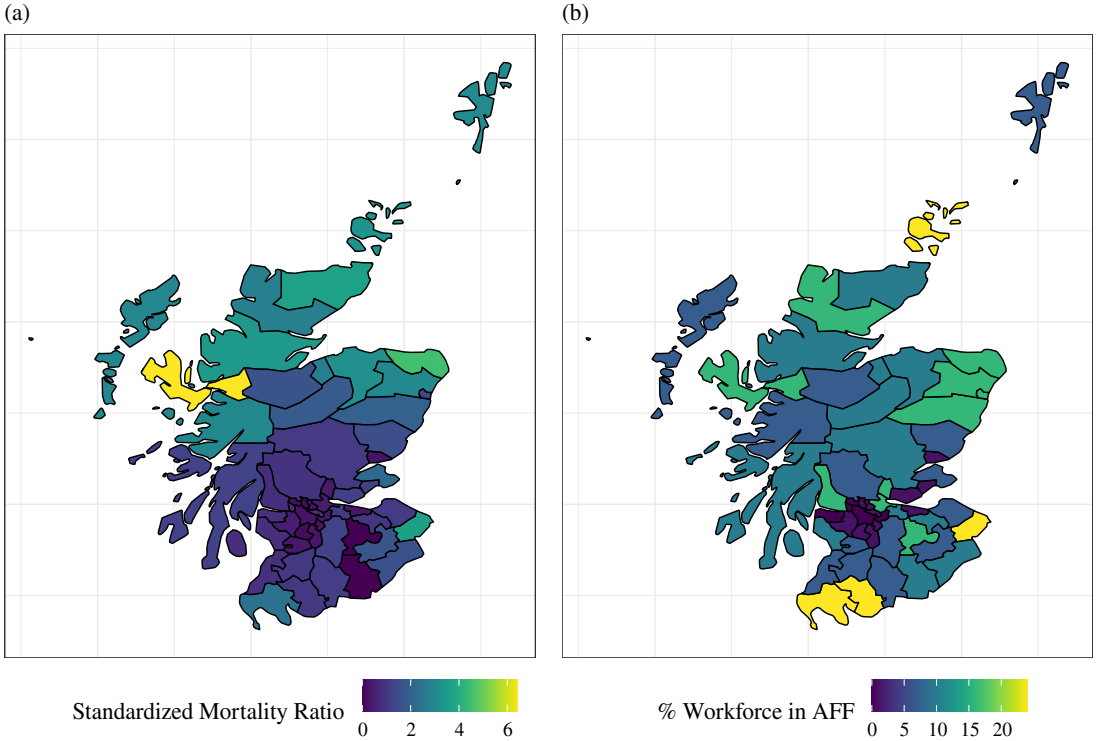


Fig. 4. Maps for the Scotland lip cancer dataset: (a) standard mortality ratio and (b) mortality rate in the agriculture, fishing and forestry workforce.

We fit the spatial Poisson regression model $Y_i \mid \theta_i \overset{\text{indep}}{\sim} \text{Po}\{E_i \exp(\theta_i)\}$, where θ_i is the log relative risk in district i. We model $\theta = (\theta_1, \ldots, \theta_n)^{\mathrm{T}}$ as $\theta \mid X \sim \text{CAR}(\beta_0 + \beta_x X + \Gamma A \Gamma^{\mathrm{T}} X, \sigma_z^2, \lambda_z)$. For the parametric approach, we use the same priors as in § 5.1 and, for the semiparametric model, we employ INLA (Rue et al., 2009) and use the penalized complexity prior with $L = 10$ basis functions, chosen via the deviance information criterion; results are stable for $L = \{10, 20, 30, 40, 50\}$. We also fit two standard nonspectral methods where β_k is constant over ω_k: a Poisson regression with the percentage of the workforce employed in agriculture, fishing and forestry as the covariate and a Poisson regression that includes Leroux conditional autoregressive random effects.

Figure 5 plots the posterior of $\exp(\beta_k)$ by the eigenvalue ω_k for each model. The standard methods have positive posterior mean and their 95% intervals exclude 1, indicating a significant increase in risk for lip cancer for a unit difference in the percentage of the workforce employed in agriculture, fishing and forestry. The spectral methods, which attempt to account for spatial confounding, do not agree with the standard methods: the estimated $\exp(\beta_k)$ trends toward 1 for large ω_k, meaning that the results of the standard models should be interpreted with caution because the strength of the relationship between these variables is weak at the local spatial scale. These results are consistent with a missing confounding variable with the same large-scale spatial pattern as lip cancer disease and the percentage of the workforce employed in agriculture, fishing and forestry.

6.2. *Analysis of COVID-19 mortality and* $\text{PM}_{2.5}$ *exposure*

Wu et al. (2020) noted that many of the pre-existing conditions that increase the mortality risk of COVID-19 are connected with long-term exposure to air pollution. They conducted

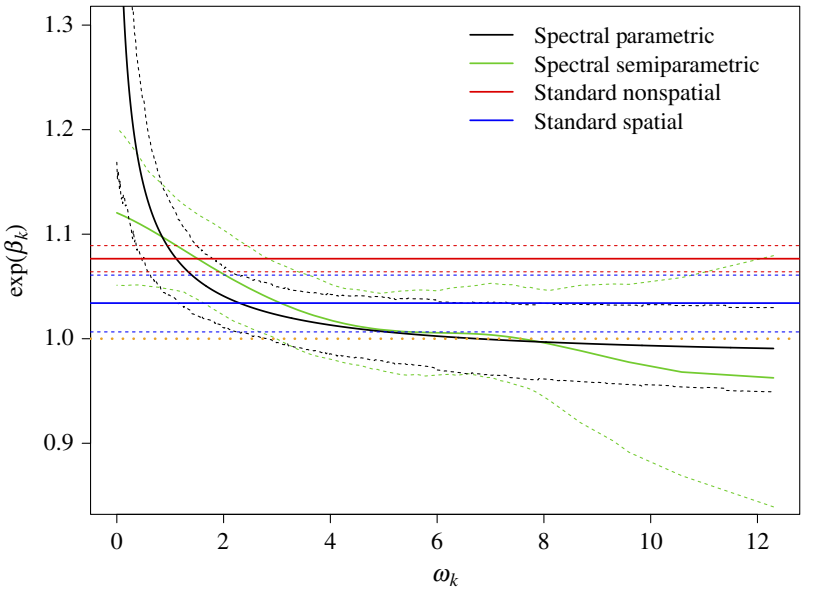


Fig. 5. Effect of the percentage of the workforce employed in agriculture, fishing and forestry on lip cancer in Scotland: posterior mean (solid lines) and 95% credible interval (dashed lines) of $\exp(\beta_k)$ for the spectral parametric model (black), the spectral semiparametric model with $L = 10$ (green), a Poisson regression on the percentage of the workforce employed in agriculture, fishing and forestry (red) and a Poisson regression with residuals modelled as Leroux (blue).

a study and found that a difference of 1 $\mu g/m^3$ in ambient fine particulate matter, $PM_{2.5}$, is positively associated with a 15% difference in the COVID-19 mortality rate. To further illustrate our proposed methods, we analyse the data collected by Wu et al. (2020) in an attempt to estimate the effect of $PM_{2.5}$ on COVID-19 mortality using spatial methods.

The response is the cumulative COVID-19 mortality counts through May 12, 2020 for US counties. County-level exposure to $PM_{2.5}$ was calculated by averaging results from an established exposure prediction model for the years 2000–16; see Wu et al. (2020) for more details. Eight counties and 12 Virginia cities were missing from the database, so we imputed their values using neighbourhood means with neighbours defined by counties that share a boundary. This resulted in mortality counts and $PM_{2.5}$ measures for $n = 3109$ counties; see Fig. 6. In addition to $PM_{2.5}$ exposure, 20 potential confounding variables, e.g., the percentage of the population at least 65 years old, are included in our modelling; see Wu et al. (2020) for the complete set of potential confounding covariates. For county i, denote Y_i as the number of deaths attributed to COVID-19, E_i as the population, X_i as the average $PM_{2.5}$ and C_i as the vector of 20 known confounding variables. Similar to Wu et al. (2020), we fit a negative-binomial regression model $Y_i \mid X_i, Z_i, C_i \overset{\text{indep}}{\sim} \text{NegBin}(r_i, p_i)$, where r_i is the size parameter and p_i the probability of success. Under this model, the mean is $\text{E}(Y_i \mid X_i, Z_i, C_i) = \lambda_i = r_i(1 - p_i)/p_i$. We parameterize the model in terms of λ_i and r_i. The prior is $\log(r_i) \sim N(0, 10)$ and the mean is linked to the linear predictor as $\log(\lambda_i) = \log(E_i) + \theta_i$, where $\theta_i = \beta_x X_i + Z_i + C_i'\beta_c$, the offset term E_i is the county population and β_c is a vector of regression coefficients associated with the confounding variables. Following the non-Gaussian models in the Supplementary Material, the linear predictor becomes $\theta \mid X \sim \text{CAR}(\beta_0 + \beta_x X + \Gamma A \Gamma^{\text{T}} X + \beta_c C, \sigma_z^2, \lambda_z)$, where C is a design matrix that includes an intercept term. We fit this model using the parametric and semiparametric

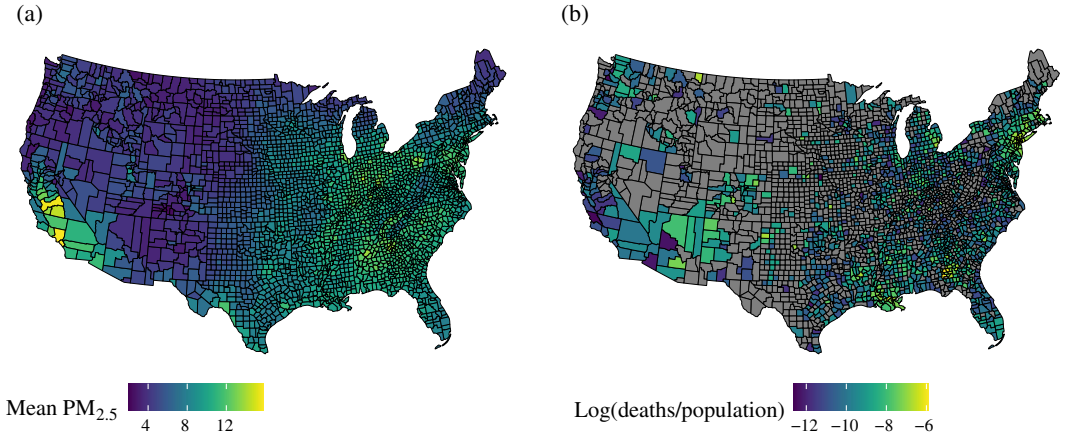


Fig. 6. (a) Average $PM_{2.5}$ ($\mu g/m^3$) over 2000–16 and (b) the log COVID-19 mortality rate through May 12, 2020. Counties with no deaths are shaded grey.

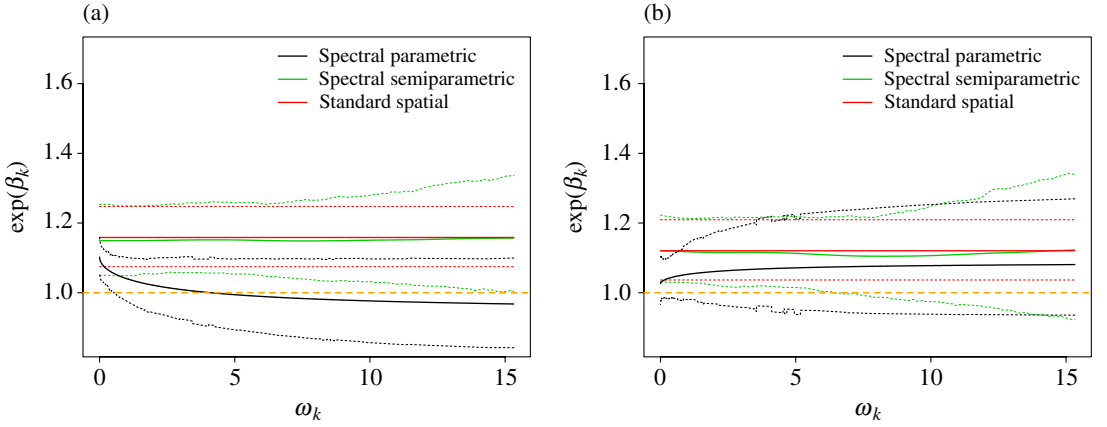


Fig. 7. Results for the COVID-19 example: the posterior mean (solid) and 95% credible interval (dashed) of the mortality rate ratio associated with a difference of $1\mu g/m^3$ of $PM_{2.5}$, $\exp(\beta_k)$. Results are for (a) all $n = 3109$ counties and (b) $n = 1977$ counties that reported at least 10 confirmed COVID-19 deaths. The standard spatial approach refers to a regression model including all confounders and a spatial Leroux model for county random effects.

approaches detailed in § 4.2 and § 4.3. The negative-binomial model is chosen to mimic the analysis in Wu et al. (2020). We also considered a binomial and Poisson model and inferences were relatively unchanged.

We compare our method to a variant of the model employed in Wu et al. (2020), which we refer to as the standard spatial model, i.e., a negative-binomial regression with all control variables and county random effects modelled using the Leroux model. Following Wu et al. (2020), two separate analyses using all $n = 3109$ counties and $n = 1977$ counties that reported at least 10 confirmed COVID-19 deaths were conducted; this was done to account for the fact that the size of an outbreak in a given county may be positively associated with both the COVID-19 mortality rate and PM2.5, thus introducing confounding bias.

Figure 7 displays model fits using the full and reduced data. The estimated difference in the COVID-19 mortality rate, associated with a difference of $1\mu g/m^3$ of $PM_{2.5}$, under the standard spatial model is 16% (95% confidence interval: 1.08, 1.25), and 12% (95% confidence interval: 1.04, 1.21) in the full and reduced analyses, respectively. The posterior mean

estimates from the parametric and semiparametric spectral models generally agree with the standard spatial approach, but the posterior standard deviation is higher for the spectral methods. In this analysis, the spectral methods support the standard spatial model and serve as a check of sensitivity to adjustments for missing confounders.

7. Discussion

Our work is the first to study the problem of spatial confounding in the spectral domain and model the coherence between the exposure and the unmeasured spatial confounder. We provide theoretical results that help understand the study in Paciorek (2010), where extensive simulations illustrate the bias obtained under different combinations of spatial scales for X and Z. New spectral methods are also proposed to adjust for unmeasured spatial confounding variables, and sufficient conditions are provided to ensure that the exposure effect is identifiable, including the important case without a nugget effect in the treatment and/or response variable. These ideas are developed for continuous and discrete spatial domains, and Gaussian and non-Gaussian data.

We have assumed that X and Z are stationary Gaussian processes in developing our methodology. Such modelling assumptions often work well with weak nonstationarity. With strong nonstationarity under model misspecification, we believe that the semiparametric model is more robust, and that β_x estimation will be primarily driven by observations from the regions that have variations at smaller spatial scales. We recommend gravitating towards the semiparametric approach. The parametric parsimonious model depends on scale-invariant coherence between the exposure and the unmeasured confounder; it also relies on parsimonious parameterization for estimation of $\alpha(\omega)$. The semiparametric model depends on the assumption that their coherence tends to zero for large frequencies. While neither of these assumptions are empirically verifiable, we believe that the latter assumption is more reasonable in practice. The semiparametric methods are also easier to implement computationally as the confounding adjustment takes the form of smoothed covariates; it is straightforward to pass these constructed variables into standard spatial computing packages. However, the continuous-space semiparametric model requires numerical approximations to integrals. When the exposure observations are highly spatially irregular, our implementation in § 3.2 can be problematic. In this case, we recommend seeking other numerical approximations to integrals.

Acknowledgement

This work was partially supported by the National Institutes of Health and the King Abdullah University of Science and Technology.

Supplementary material

The Supplementary Material includes an extension of the proposed methods to multiple predictors and non-Gaussian observations, details of the causal assumptions for the spatial framework, proofs of the lemmas and theorems, prior distributions and computing details for both discrete and continuous cases, and a results table for the continuous-space simulation study. The R code is available at https://github.com/yawenguan/spatial_confounding.

References

Apanasovich, T. V., Genton, M. G. & Sun, Y. (2012). A valid Matérn class of cross-covariance functions for multivariate random fields with any number of components. *J. Am. Statist. Assoc.* **107**, 180–93.

Chen, K., Dai, F., Marchiori, E. & Theodoridis, S. (2021). Novel compressible adaptive spectral mixture kernels for Gaussian processes with sparse time and phase delay structures. *arXiv:* 1808.00560v8.

Clayton, D. G., Bernardinellli, L. & Montomoli, C. (1993). Spatial correlation in ecological analysis. *Int. J. Epidemiol.* **22**, 1193–202.

Dupont, E., Wood, S. N. & Augustin, N. (2022). Spatial+: a novel approach to spatial confounding. *Biometrics* **78**, 1279–90.

Eilers, P. H. C. & Marx, B. D. (1996). Flexible smoothing with *b*-splines and penalties. *Statist. Sci.* **11**, 89–102.

Faes, L., Krohova, J., Pernice, R., Busacca, A. & Javorka, M. (2019). A new frequency domain measure of causality based on partial spectral decomposition of autoregressive processes and its application to cardiovascular interactions. In *2019 41st Ann. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, pp. 4258–61. Piscataway, NJ: IEEE Press.

Franco-Villoria, M., Ventrucci, M. & Rue, H. (2019). A unified view on Bayesian varying coefficient models. *Electron. J. Statist.* **13**, 5334–59.

Fuentes, M. & Reich, B. (2010). Spectral domain. In *Handbook of Spatial Statistics*, A. E. Gelfand, P. Diggle, M. Fuentes & P. Guttorp, eds. Boca Raton, FL: CRC Press, pp. 57–77.

Gelfand, A. E., Diggle, P., Fuentes, M. & Guttorp, P., Ed. (2010). *Handbook of Spatial Statistics*. Boca Raton, FL: CRC Press.

Gneiting, T., Kleiber, W. & Schlather, M. (2010). Matérn cross-covariance functions for multivariate random fields. *J. Am. Statist. Assoc.* **105**, 1167–77.

Hanks, E. M., Schliep, E. M., Hooten, M. B. & Hoeting, J. A. (2015). Restricted spatial regression in practice: geostatistical models, confounding, and robustness under model misspecification. *Environmetrics* **26**, 243–54.

Hodges, J. S. & Reich, B. J. (2010). Adding spatially-correlated errors can mess up the fixed effect you love. *Am. Statistician* **64**, 325–34.

Hughes, J. & Haran, M. (2013). Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *J. R. Statist. Soc.* B **75**, 139–59.

Jang, P. A., Loeb, A., Davidow, M. & Wilson, A. G. (2017). Scalable Levy process priors for spectral kernel learning. In *Advances in Neural Information Processing Systems*, pp. 3943–52. New York: Curran Associates.

Johnson, C. (2012). *Numerical Solution of Partial Differential Equations by the Finite Element Method*. New York: Dover Publications.

Keller, J. P. & Szpiro, A. A. (2020). Selecting a scale for spatial confounding adjustment. *J. R. Statist. Soc.* A **183**, 1121–43.

Khan, K. & Calder, C. A. (2022). Restricted spatial regression methods: implications for inference. *J. Am. Statist. Assoc.* **117**, 482–94.

Kleiber, W. (2017). Coherence for multivariate random fields. *Statist. Sinica* **27**, 1675–97.

Leroux, B. G., Lei, X. & Breslow, N. (2000). Estimation of disease rates in small areas: a new mixed model for spatial dependence. In *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, M. E. Halloran & D. Berry, eds. New York: Springer, pp. 179–91.

Lindgren, F., Rue, H. & Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *J. R. Statist. Soc.* B **73**, 423–98.

Marques, I., Kneib, T. & Klein, N. (2022). Mitigating spatial confounding by explicitly correlating Gaussian random fields. *Environmetrics* **33**, e2727.

Osama, M., Zachariah, D. & Schön, T. B. (2019). Inferring heterogeneous causal effects in presence of spatial confounding. In *Proc. 36th Int. Conf. Mach. Learn., Long Beach, California*, vol. 97, K. Chaudhuri & R. Salakhutdinov, eds. PMLR, pp. 4942–50.

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

Page, G. L., Liu, Y., He, Z. & Sun, D. (2017). Estimation and prediction in the presence of spatial confounding for spatial linear models. *Scand. J. Statist.* **44**, 780–97.

Prates, M. O., Assunção, R. M. & Rodrigues, E. C. (2019). Alleviating spatial confounding for areal data problems by displacing the geographical centroids. *Bayesian Anal.* **14**, 623–47.

Qadir, G. A. & Sun, Y. (2020). Semiparametric estimation of cross-covariance functions for multivariate random fields. *Biometrics* **77**, 547–60.

R Development Core Team (2023). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, http://www.R-project.org.

Reich, B. J. & Fuentes, M. (2012). Nonparametric Bayesian models for a spatial covariance. *Statist. Methodol.* **9**, 265–74.

Reich, B. J., Hodges, J. S. & Zadnik, V. (2006). Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. *Biometrics* **62**, 1197–206.

Reich, B. J., Yang, S., Guan, Y., Giffin, A. B., Miller, M. J. & Rappold, A. (2021). A review of spatial causal inference methods for environmental and epidemiological applications. *Int. Statist. Rev.* **89**, 605–34.

Rue, H., Martino, S. & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Statist. Soc.* B **71**, 319–92.

Sandryhaila, A. & Moura, J. M. F. (2013). Discrete signal processing on graphs: graph Fourier transform. In *2013 IEEE Int. Conf. Acoust. Speech Sig. Proces.*, pp. 6167–70. Piscataway, NJ: IEEE Press.

Schnell, P. & Papadogeorgou, G. (2020). Mitigating unobserved spatial confounding when estimating the effect of supermarket access on cardiovascular disease deaths. *Ann. Appl. Statist.* **14**, 2069–95.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *J. R. Statist. Soc.* B **64**, 583–639.

Stokes, P. A. & Purdon, P. L. (2017). A study of problems encountered in Granger causality analysis from a neuroscience perspective. *Proc. Nat. Acad. Sci.* **114**, E7063–72.

Thaden, H. & Kneib, T. (2018). Structural equation models for dealing with spatial confounding. *Am. Statistician* **72**, 239–52.

Watson, G. N. (1995). *A Treatise on the Theory of Bessel Functions*. Cambridge: Cambridge University Press.

Wu, X., Nethery, R. C., Sabath, B. M., Braun, D. & Dominici, F. (2020). Air pollution and COVID-19 mortality in the United States: strengths and limitations of an ecological regression analysis. *Sci. Adv.* **6**, 1–6.

Zimmerman, D. L. & Hoef, J. M. V. (2022). On deconfounding spatial confounding in linear models. *Am. Statistician* **76**, 159–67.

[*Received on* 12 *January* 2022. *Editorial decision on* 5 *December* 2022]

Biometrika (2022), **103**, 1, *pp*. 1–11
Printed in Great Britain

Supplementary material for "A spectral adjustment for spatial confounding"

BY YAWEN GUAN
Deparment of Statistics, University of Nebraska, Nebraska 68583, U.S.A
yawen.guan@unl.edu

GARRITT L. PAGE
Department of Statistics, Brigham Young University, Utah 84602, U.S.A
page@stat.byu.edu

BRIAN J REICH
Department of Statistics, North Carolina State University, North Carolina 27695, U.S.A
bjreich@ncsu.edu

MASSIMO VENTRUCCI
Department of Statistics, University of Bologna, 40126 Bologna, Italy
massimo.ventrucci@unibo.it

AND SHU YANG
Department of Statistics, North Carolina State University, North Carolina 27695, U.S.A
syang24@ncsu.edu

1. SPATIAL CAUSAL FRAMEWORK

Extending the notation in § 2.1 of the main document and following the potential outcomes framework, we define $Y_i(x_i)$ to be the potential outcome at location s_i if the treatment at s_i is x_i. Similarly, define $Y(x) = \{Y_1(x_1), \ldots, Y_n(x_n)\}^T$ as the vector of potential outcomes at $\mathcal{S}$ if treatment $x = (x_1, ..., x_n)^T$ were received. For the potential outcomes to be well defined, we make the stable-unit-treatment-value assumption (Rubin, 1980), which states that there is no interference between units and there is a single version of each treatment level. In the spatial context, the no interference assumption implies that treatment applied at one location does not affect the outcome at other locations.

Following the commonly used spatial regression model in the spatial statistics literature, we assume a linear additive structural model for the potential outcomes,

$$Y(x) = \beta_0 + \beta_x x + \beta_z Z + \varepsilon, \tag{1}$$

where $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)^T$ and $\varepsilon_i \overset{iid}{\sim} \text{Normal}(0, \sigma^2)$. The structural coefficient β_x determines the causal relationship of the potential outcome and the treatment, and it is assumed to be spatially invariant. In this model, we do not specify the interaction of treatment and covariates, and therefore β_x is both the marginal and conditional causal effect of treatment. Our objective is to estimate β_x. Model (1) is also used in Schnell & Papadogeorgou (2020) in a discrete-space case.

To connect the observed and potential outcomes, we make the causal consistency assumption that the observed outcomes are exactly the potential outcomes for the observed level of treatment X_i, $Y_i = Y_i(X_i)$. For the identifiability of β_x, we make the conditional treatment ignorability assumption that $X \perp Y(x) \mid Z$ for all x. The latter assumption holds if all factors that are associated with both the treatment and outcome variables are accounted for in Z. With these assumptions, the parameter β_x in the regression model

$$Y = \beta_0 + \beta_x X + \beta_z Z + \varepsilon. \tag{2}$$

is the same as β_x in (1) and thus has a causal interpretation.

If we observe the confounder Z, then identification and estimation of β_x is straightforward using multiple linear regression. However, we assume that Z is an unmeasured confounder. In the presence of an unmeasured confounder, β_x is not identifiable in general. Therefore, we propose to exploit the spatial structure of Z to mitigate the effects of this unknown confounder. Although Z confounds the causal relationship of X and Y in the spatial domain, in § 2.2 of the main text we propose assumptions in the spectral domain that identify the causal effect in the presence of an unmeasured spatial confounder.

2. Extension to multiple predictors

2·1. *Parsimonious continuous-space model*

In the spatial domain, let $X_0(s) = Z(s)$ be the unmeasured confounder and $X_1(s), ..., X_p(s)$ be the p observed covariates. Given the confounder, the model is $Y(s) = \beta_0 + \sum_{j=0}^{p} \beta_j X_j(s) + \varepsilon(s)$. For frequency ω, let $\mathcal{X}_0(\omega) = \mathcal{Z}(\omega)$ be the Fourier transform of the unmeasured confounder and $\mathcal{X}_1(\omega), ..., \mathcal{X}_p(\omega)$ be the the Fourier transforms of the measured covariates. In the spectral domain the joint model in (2) is extended to $\text{cov}\{\mathcal{X}_j(\omega), \mathcal{X}_k(\omega)\} = \Omega_{jk}\sigma_j\sigma_k f_{jk}(\omega)$ where $\Omega = \{\Omega_{jk}\}$ is a $(p+1) \times (p+1)$ correlation matrix, $\sigma_j > 0$ are standard deviation parameters and $f_{jk}(\omega) = f_{kj}(\omega)$ are spectral densities. The multivaritate extension of the parsimonious model in § 2.2 of the main text sets $f_{jk}(\omega) = \sqrt{f_j(\omega) f_k(\omega)}$. This is clearly a valid model because the $(p+1) \times (p+1)$ covariance of $\mathcal{X}_0(\omega), ..., \mathcal{X}_p(\omega)$ is $F(\omega)\Omega F(\omega)$, where $F(\omega)$ is diagonal with j^{th} diagonal element equal $\sigma_j \sqrt{f_j(\omega)}$, which is positive definite.

Blocking Ω to have first row column $(1, \Omega_{zx}^T)$ and bottom right $p \times p$ matrix Ω_x and marginalizing over $\mathcal{X}_0(\omega) = \mathcal{Z}(\omega)$ gives

$$\mathcal{Y}(\omega) | \mathcal{X}_1(\omega), ..., \mathcal{X}_p(\omega) \overset{indep}{\sim} \text{Normal}\left[\sum_{j=1}^{p} \{\beta_j + \alpha_j(\omega)\}\mathcal{X}_j(\omega), q_0 \sigma_z^2 f_z(\omega)\right], \tag{3}$$

where $\alpha_j(\omega) = \{\sigma_z \sqrt{f_z(\omega)}\}\{\sigma_j \sqrt{f_j(\omega)}\}^{-1} q_j \mathcal{X}_j(\omega)$, $(q_1, ..., q_p) = \Omega_{zx}^T \Omega_x^{-1}$ and $q_0 = \Omega_{zz} - \Omega_{zx}^T \Omega_x^{-1} \Omega_{zx}$. In the spatial domain this the causal adjustment for covariate j becomes $\hat{Z}_j = w_j C_{0j} C_{jj}^{-1} X_j$ where $C_{jk} = \text{cov}(X_j, X_k)$ is defined by the Matérn covariance parameters. The

spatial model is

$$Y(s_i) = \beta_0 + \sum_{j=1}^{p} \{X_j(s_i)\beta_j + \hat{Z}_j(s_i)\} + V(s_i) + \varepsilon(s_i)$$

where V is a mean-zero Gaussian process with variance $\sigma_z^2 q_o$ and Matérn correlation with smoothness η_z and range ϕ and $\varepsilon(s) \overset{iid}{\sim} \text{Normal}(0, \sigma^2)$.

2·2. *Semi-parametric continuous-space model*

Extending the semi-parametric model in § 3.2 to have multiple covariates is straightforward because the method is defined via the conditional distribution of the missing confounder given the covariates, rather than the joint distribution of the confounder and the covariates. We regress the confounder onto the p covariates using the additive model

$$\mathrm{E}\{\mathcal{Z}(\omega)\} = \sum_{j=1}^{p} \alpha_j(\omega)\mathcal{X}_j(\omega) = \frac{1}{2}\sum_{j=1}^{p}\sum_{l=1}^{L} b_{jl}B_l(\omega)\mathcal{X}_j(\omega).$$

In this spatial domain this gives

$$Y(s) = \beta_0 + \sum_{j=1}^{p} \beta_j X_j(s) + \sum_{j=1}^{p}\sum_{l=1}^{L} \beta_{jl}\hat{Z}_{jl}(s) + \delta(s),$$

where $\delta(s)$ is modeled as a Gaussian process and

$$\hat{Z}_{jl}(s) = \sum_{\omega^f \in \mathcal{F}} (2\pi\omega^f)^{\kappa+1} B_l(\omega^f) \triangle_{\mathcal{F}} \int \frac{\mathcal{J}_\kappa(\omega^f ||s - s'||)}{||s - s'||^\kappa} X(s')ds'.$$

As in the univariate case, β_j can be interpreted as the causal effect under the assumption that $\alpha_j(\omega) \to 0$.

2·3. *Parsimonious conditional autoregressive model*

In the discrete domain with multiple predictors, the oracle model is $Y \sim \text{Normal}(\beta_0 + \sum_{j=1}^{p} \beta_j X_j + Z, \sigma^2 I_{n\times n})$. Assuming marginal distribution $X_j \sim \text{CAR}(0, \sigma_j^2, \lambda_j)$ and $Z \sim \text{CAR}(0, \sigma_z^2, \lambda_z)$, parsimonious joint distributions in § 4.2 of the main text and following the ideas above in the extension of the parsimonious Matérn to the case of multiple predictors, the marginal distribution of Y over Z is

$$Y|X, V \sim \text{Normal}\left(\beta_0 + \sum_{j=1}^{p} \beta_j X_j + \sum_{j=1}^{p} \Gamma A_j \Gamma^T X_j + V, \sigma^2 I_{n\times n}\right),$$

where A_j is diagonal with k^{th} diagonal element $\alpha_j(\omega_k) = w_j \sigma_z {\sigma_j}^{-1}(1 - \lambda_j + \lambda_j\omega_k)^{1/2}(1 - \lambda_z + \lambda_z\omega_k)^{-1/2}$ and $V \sim \text{CAR}(0, q_0\sigma_z^2, \lambda_z)$.

2·4. *Semi-parametric conditional autoregressive model*

Extending the semi-parametric model in § 4.3 of the main text to the multivariate case follows the same steps as the semi-parametric model for continuous space above. In the spectral domain with $Z^* = \Gamma Z$ and $X_j^* = \Gamma X_j$, a natural extension to the semi-parametric conditional

autoregressive model for the missing confounder is

$$Z_k^*|X_{1k}^*,...X_{pk}^* \overset{indep}{\sim} \text{Normal}\left(\sum_{j=1}^{p}\alpha_j(\omega_k)X_j^*, \frac{\sigma_Z^2}{1-\lambda_z+\lambda_z\omega_k}\right).$$

Defining $\beta_j(\omega) = \beta_j + \alpha_j(\omega) = \sum_{l=1}^{L} B_l(\omega)b_{jl}$, gives the spatial model

$$Y|X,V \sim \text{Normal}\left(\beta_0 + \sum_{j=1}^{p}\sum_{l=1}^{L}\hat{Z}_{jl}b_{jl} + V, \sigma^2 I_n\right). \tag{4}$$

where $V \sim \text{CAR}(0, \sigma_z^2, \lambda_z)$. Under the assumption that $\alpha_j(\omega_n) = 0$, we summarize the causal effect of the jth covariate using the posterior of $\beta_j = \sum_{l=1}^{L} B_l(\omega_n)b_{jl}$.

3. Generalized linear mixed spatial models

Many of the methods proposed for continuous and discrete domains can be applied to non-Gaussian data using the generalized linear mixed modeling framework. Let $g\{\text{E}(Y_i \mid X_i, Z_i)\} = \theta_i = \beta_x X_i + Z_i$ for link function g. Our general approach is to build a joint model for (X_i, Z_i), and then compute the conditional distribution of $\theta = (\theta_1, ..., \theta_n)^T$ given X. For the continuous spatial domain, the conditional distribution is $\theta \mid X \sim \text{Normal}\left(\beta_0 + \beta_x X + \hat{Z}, \Sigma_z - \Sigma_{zx}\Sigma_x^{-1}\Sigma_{zx}^T\right)$. Then for the continuous spatial domain, a bivariate Matérn model or the semi-parametric model in § 3 of the main text can be used. For the discrete spatial domain, the parsimonious bivariate conditional autoregressive model gives $\theta \mid X \sim \text{CAR}(\beta_0 + \beta_x X + \Gamma A \Gamma^T X, (1-\rho^2)\sigma_z^2, \lambda_z)$. Or we can use the semi-parametric model in § 4 of the main text which gives $\theta \mid X \sim \text{CAR}\left(\beta_0 + \sum_{l=1}^{L}\hat{Z}_l b_l, \sigma_z^2, \lambda_z\right)$.

Unlike for Gaussian data, the MCMC cannot proceed in the spectral domain and matrix multiplication $\Gamma A \Gamma^T X = \Gamma A X^*$ is required when updating the parameters in $\alpha(\omega_k)$. For the semi-parametric model much of the computational burden can be shouldered outside the MCMC loop by precomputing matrices $\Gamma A \Gamma^T X = \sum_{l=1}^{L}\hat{Z}_l b_l$. After this computation, the method can be fit using standard software for spatial generalized linear models such as `INLA` or `OpenBUGS`, and the posterior of $\beta_x = \sum_{l=1}^{L} B_l(\omega_n)b_l$ summarizes the exposure effect.

4. Proofs

4·1. *Proof of Theorem 1: Identification under the parsimonious model*

The joint model

$$\mathcal{Y}(\omega)|\mathcal{X}(\omega) \overset{indep}{\sim} \text{Normal}\left\{\left(\beta_x + \rho\frac{\sigma_z\sqrt{f_z(\omega)}}{\sigma_x\sqrt{f_x(\omega)}}\right)\mathcal{X}(\omega), (1-\rho^2)\sigma_z^2 f_z(\omega) + \sigma^2\right\}$$

$$\mathcal{X}(\omega) \overset{indep}{\sim} \text{Normal}\left\{0, \sigma_x^2 f_x(\omega)\right\}$$

is defined by parameters $\theta = \{\beta_x, \rho, \sigma_x, \sigma_z, \sigma, F_x, F_z\}$ where $F_j = \{f_j(\omega); \omega \in \mathcal{R}^2\}$ and $\int f_j(\omega)d\omega = 1$ for $j \in \{x, z\}$. Define $\text{E}\{\mathcal{Y}(\omega)|\mathcal{X}(\omega)\} = \mu_y(\omega;\theta)$, $\text{var}\{\mathcal{Y}(\omega)|\mathcal{X}(\omega)\} = \tau_y^2(\omega;\theta)$ and $\text{var}\{\mathcal{X}(\omega)\} = \tau_x^2(\omega;\theta)$.

Denote $\theta^{(j)} = \{\beta_x^{(j)}, \rho^{(j)}, \sigma_x^{(j)}, \sigma_z^{(j)}, F_x^{(j)}, F_z^{(j)}\}$. The parameters θ are identified if and only if $\mu_y(\omega;\theta^{(1)}) = \mu_y(\omega;\theta^{(2)})$, $\tau_y^2(\omega;\theta^{(1)}) = \tau_y^2(\omega;\theta^{(2)})$ and $\tau_x^2(\omega;\theta^{(1)}) = \tau_x^2(\omega;\theta^{(2)})$ for all ω im-

plies that $\theta^{(1)} = \theta^{(2)}$. First assume that $\tau_x^2(\omega;\theta^{(1)}) = \tau_x^2(\omega;\theta^{(2)})$ for all ω. Since $\int \tau_x^2(\omega;\theta)d\omega = \sigma_x^2$, we must have $\sigma_x^{(1)} = \sigma_x^{(2)}$. This clearly implies that $F_x^{(1)} = F_x^{(2)}$, and thus the parameters in the marginal distribution of $\mathcal{X}$ are identified. Applying similar arguments to the $\tau_y^2(\omega;\theta)$ shows that $\tau_y^2(\omega;\theta^{(1)}) = \tau_y^2(\omega;\theta^{(2)})$ for all ω implies that $F_z^{(1)} = F_z^{(2)}$, $\{1-(\rho^{(1)})^2\}(\sigma_z^{(1)})^2 = \{1-(\rho^{(2)})^2\}(\sigma_z^{(2)})^2$ and $\sigma^{(1)} = \sigma^{(2)}$.

The remaining parameters are identified by $\mu_y(\omega;\theta)$. If we assume $\mu_y(\omega;\theta^{(1)}) = \mu_y(\omega;\theta^{(2)})$ for all ω and apply previous identifiability results that show $F_x^{(1)} = F_x^{(2)}$, $\sigma_x^{(1)} = \sigma_x^{(2)}$ and $F_z^{(1)} = F_z^{(2)}$, then the assumption that $f_x(\omega) \neq f_z(\omega)$ for some ω implies that $\beta_x^{(1)} = \beta_x^{(2)}$ and $\rho^{(1)}\sigma_z^{(1)} = \rho^{(2)}\sigma_z^{(2)}$. Combined with the result that $\{1-(\rho^{(1)})^2\}(\sigma_z^{(1)})^2 = \{1-(\rho^{(2)})^2\}(\sigma_z^{(2)})^2$, we have $\rho^{(1)} = \rho^{(2)}$ and $\sigma_z^{(1)} = \sigma_z^{(2)}$.

4·2. *Proof of Lemma 1: Oracle confounder adjustment*

We derive the oracle confounder adjustment for a given projection operator $\alpha(\omega)$. If the inverse Fourier transform of $\alpha(\omega)$ has a closed-form denoted by $K(s-s')$, then

$$\begin{aligned}
\hat{Z}(s) &= \int_{\mathcal{R}^2} \exp(i\omega^T s)\hat{\mathcal{Z}}(\omega)d\omega \\
&= \int_{\mathcal{R}^2} \exp(i\omega^T s)\alpha(\omega)\mathcal{X}(\omega)d\omega \\
&= \int_{\mathcal{R}^2} \exp(i\omega^T s)\alpha(\omega)\left[\frac{1}{(2\pi)^d}\int_{\mathcal{R}^2}\exp(-i\omega^T s')X(s')ds'\right]d\omega \\
&= \frac{1}{(2\pi)^d}\int_{\mathcal{R}^2}\left[\int \exp(i\omega^T(s-s'))\alpha(\omega)d\omega\right]X(s')ds' \\
&= \frac{1}{(2\pi)^d}\int_{\mathcal{R}^2} K(s-s')X(s')ds'.
\end{aligned}$$

4·3. *Proof of Theorem 2: Parameter identification for the bivariate Matérn model*

The parameters that define the marginal distribution of X, σ_x, ν_x and ϕ, are identified following the arguments in supplementary material § 4·1. Therefore, we assume they are fixed and known in this section. For the bivariate Matérn model, define $\mathrm{E}\{\mathcal{Y}(\omega)|\mathcal{X}(\omega)\} = \mu(\omega;\theta)$ and $\mathrm{V}\{\mathcal{Y}(\omega)|\mathcal{X}(\omega)\} = \tau^2(\omega;\theta)$, where $\theta = \{\beta_x, \rho, \nu_z, \nu_{xz}, \sigma_z\}$ are the remaining unknown parameters. Defining $\delta = (1+\phi^2||\omega||^2)^{-1}$, the mean and variance are

$$\begin{aligned}
\mu(\omega;\theta) &= \left(\beta_x + \rho\frac{\sigma_z\nu_{xz}}{\sigma_x\nu_x}\delta^{\nu_{xz}-\nu_x}\right)\mathcal{X}(\omega) \\
\tau^2(\omega;\theta) &= \sigma_z^2\nu_z\phi^2\delta^{\nu_z+1}\left[1-\rho^2\frac{\nu_{xz}^2}{\nu_x\nu_z}\delta^{2\nu_{xz}-\nu_x-\nu_z}\right].
\end{aligned}$$

Denote $\theta^{(j)} = \{\beta_x^{(j)}, \rho^{(j)}, \nu_z^{(j)}, \nu_{xz}^{(j)}, \sigma_z^{(j)}\}$. The parameters θ are identified if and only if $\mu(\omega;\theta^{(1)}) = \mu(\omega;\theta^{(2)})$ and $\tau^2(\omega;\theta^{(1)}) = \tau^2(\omega;\theta^{(1)})$ for all ω implies that $\theta^{(1)} = \theta^{(2)}$.

Now assume that $\mu(\omega;\theta^{(1)}) = \mu(\omega;\theta^{(2)})$ and $\tau^2(\omega;\theta^{(1)}) = \tau^2(\omega;\theta^{(2)})$ for all ω. By assumption, $\nu_{xz}-\nu_x > 0$, $2\nu_{xz}-\nu_x-\nu_z > 0$ and $\rho^2\nu_{xz}^2(\nu_x\nu_z)^{-1} < 1$. Therefore, $\mu(\omega;\theta)/\mathcal{X}(\omega) \to \beta_x$, and therefore $\beta_x^{(1)} = \beta_x^{(2)} = \beta_x^*$. Considering $\mu(\omega;\theta_j)/\mathcal{X}(\omega) - \beta_x^*$ over ω we have $\nu_{xz}^{(1)} = \nu_{xz}^{(2)} = \nu_{xz}^*$ and $\rho^{(1)}\sigma_z^{(1)} = \rho^{(2)}\sigma_z^{(2)}$. Turning to the variance terms, for large ω we have

$\tau^2(\omega;\theta) \approx \sigma_z^2 \nu_z \phi^2 \delta^{\nu_z+1}$. Therefore $\nu_z^{(1)} = \nu_z^{(2)}$ and $[\sigma_z^{(1)}]^2 \nu_z^{(1)} = [\sigma_z^{(2)}]^2 \nu_z^{(2)}$, and thus and $\sigma_Z^{(1)} = \sigma_Z^{(2)}$. Finally, since $\rho^{(1)}\sigma_Z^{(1)} = \rho^{(2)}\sigma_Z^{(2)}$ from above we have $\rho^{(1)} = \rho^{(2)}$, and thus $\theta_1 = \theta_2$.

4·4. *Proof of Theorem 3: Parameter identification for the parsimonious CAR model*

We aim to establish that the parameters $\theta = \{\beta_x, \sigma^2, \rho, \lambda_x, \lambda_z, \sigma_x^2, \sigma_z^2\}$ found in the joint model for (Y^*, X^*) that is partially described in equation (13) of the main text are identifiable. Here we will denote the joint model using $p(Y^*, X^*|\theta)$. As before, to establish identifiability we need to show that $p(Y^*, X^*|\theta^{(1)}) = p(Y^*, X^*|\theta^{(2)}) \Leftrightarrow \theta^{(1)} = \theta^{(2)}$. Note that we can express the joint model after setting $\beta_0 = 0$ w.l.o.g. in the following way

$$\begin{aligned} p(Y^*, X^*|\theta) &= p(Y^*|X^*, \theta)p(X^*|\theta) \\ &= N_n(Y^*; \beta_x X^* + \rho\frac{\sigma_z}{\sigma_x}AX^*, \sigma_z^2(1-\rho^2)\{(1-\lambda_z)I_{n\times n} + \lambda_z W\}^{-1} + \sigma^2 I_{n\times n})\times \\ &\quad N_n(X^*; 0, \sigma_x^2\{(1-\lambda_x)I_{n\times n} + \lambda_x W\}^{-1}). \end{aligned}$$

Here $N_n(\cdot;, m, V)$ denotes a n dimensional multivariate density function with n-dimensional mean vector m and $n \times n$-dimensional covariance matrix V, and, as before, A is a diagonal matrix whose ith diagonal entry is $(1-\lambda_x+\lambda_x\omega_i)^{1/2}(1-\lambda_z+\lambda_z\omega_i)^{-1/2}$. Note that we can establish identifiability by way of the mean structure, covariance structure, or both from the joint model. Since $p(X^*|\theta)$ is simply a Leroux model and its parameters are known to be identifiable we have that σ_x^2 and λ_x are both identifiable. Thus, we focus on $(\beta_x, \sigma^2, \rho, \lambda_z, \sigma_z^2)$. It is straightforward to see that

$$p(Y^*|X^*, \beta_x, \sigma^{2(1)}, \rho, \lambda_x, \lambda_z, \sigma_x^2, \sigma_z^2) = p(Y^*|X^*, \beta_x, \sigma^{2(2)}, \rho, \lambda_x, \lambda_z, \sigma_x^2, \sigma_z^2) \Leftrightarrow \sigma^{2(1)} = \sigma^{2(2)}$$

which implies that σ^2 identifiable.

Next since matrix inverse is a bijection we can work with $\sigma_z^{-2}(1-\rho^2)^{-1}\{(1-\lambda_z)I_{n\times n} + \lambda_z W\}$ which can be expressed as $\sigma_z^{-2}(1-\rho^2)^{-1}I_{n\times n} + \sigma_z^{-2}(1-\rho^2)^{-1}\lambda_z(W - I_{n\times n})$. Letting $\tau = \sigma_z^{-2}(1-\rho^2)^{-1}$ and we have

$$\begin{aligned} &\tau^{(1)}I_{n\times n} + \tau^{(1)}\lambda_z^{(1)}(W - I_{n\times n}) + \sigma^2 I_{n\times n} - \{\tau^{(2)}I_{n\times n} + \tau^{(2)}\lambda_z^{(2)}(W - I_{n\times n}) + \sigma^2 I_{n\times n}\} = 0 \\ &\Leftrightarrow \{\sigma^2 + \tau^{(1)} - (\sigma^2 + \tau^{(2)})\}I_{n\times n} + (\tau^{(1)}\lambda_z^{(1)} - \tau^{(2)}\lambda_z^{(2)})(W - I_{n\times n}) = 0 \\ &\Leftrightarrow (\tau^{(1)} - \tau^{(2)}) + (\tau^{(1)}\lambda_z^{(1)} - \tau^{(2)}\lambda_z^{(2)})(\omega_i - 1) = 0 \;\forall\, \omega_i \\ &\Leftrightarrow \tau^{(1)} - \tau^{(2)} = 0 \text{ and } \tau^{(1)}\lambda_z^{(1)} - \tau^{(2)}\lambda_z^{(2)} = 0. \end{aligned}$$

Now $(\tau^{(1)} - \tau^{(2)}) = 0$ establishes $\tau^{(1)} = \tau^{(2)}$ and $\tau^{(1)} = \tau^{(2)}$ along with $\tau^{(1)}\lambda_z^{(1)} - \tau^{(2)}\lambda_z^{(2)} = 0$ establishes $\lambda_z^{(1)} = \lambda_z^{(2)}$. Thus, τ and λ_z are identifiable. In summary, to this point, we have shown that λ_x, λ_z, σ_x^2, σ^2 and τ are all identifiable. We now show that β_x and $\eta = \rho\sigma_z$ are identifiable. Note that

$$\beta_x X^* + \rho\frac{\sigma_z}{\sigma_x}AX^* = \left(\beta_x I_{n\times n} + \frac{\eta}{\sigma_x}A\right)X^*.$$

Thus, we can focus on $\beta_x I_{n\times n} + \frac{\eta}{\sigma_x} A$ such that

$$
\begin{aligned}
&(\beta_x^{(1)} I_{n\times n} + \frac{\eta^{(1)}}{\sigma_x} A) - (\beta_x^{(2)} I_{n\times n} + \frac{\eta^{(2)}}{\sigma_x} A) = 0 \\
\Leftrightarrow\ &(\beta_x^{(1)} - \beta_x^{(2)}) I_{n\times n} + (\eta^{(1)} - \eta^{(2)}) \frac{1}{\sigma_x} A = 0 \\
\Leftrightarrow\ &(\beta_x^{(1)} - \beta_x^{(2)}) + (\eta^{(1)} - \eta^{(2)}) \frac{1}{\sigma_x} \sqrt{\frac{1-\lambda_x+\lambda_x\omega_i}{1-\lambda_z+\lambda_z\omega_i}} = 0 \,\forall\, \omega_i \\
\Leftrightarrow\ &(\beta_x^{(1)} - \beta_x^{(2)}) = 0 \text{ and } (\eta^{(1)} - \eta^{(2)}) = 0,
\end{aligned}
$$

establishing the identifiability of β_x and $\eta = \rho\sigma_z$. Now since $\tau = \sigma_z^{-2}(1-\rho^2)^{-1}$ and $\eta = \rho\sigma_z$, are identifiable, then so to are ρ and σ_z^2 where $\rho = \sqrt{\eta^2(\tau^{-1}+\eta^2)^{-1}}$ and $\sigma_z^2 = \tau^{-1} + \eta^2$.

5. Sample datasets for the simulation study

Supplementary material Figure 1 plots a dataset from each simulated scenario in § 6.1 of the main text to illustrate the spatial dependence of each variable and the relationships between variables.

6. Prior distributions and computational details for discrete cases

For the Gaussian responses in the simulation study, the variances in the standard and parametric models are parameterized as $\sigma_z^2 = \tau^2 r$ and $\sigma^2 = \tau^2(1-r)$ and the priors are $\beta_0, \beta_x \sim$ Normal$(0, 100)$, $\tau^2 \sim$ InvGamma$(0.1, 0.1)$ and $r, \lambda_z \sim$ Uniform$(0, 1)$. For the parsimonious bivariate conditional autoregressive model we fix $\sigma_x = 1$ and select priors $\lambda_x \sim$ Uniform$(0, \lambda_z)$ and $\rho \sim$ Uniform$(-1, 1)$.

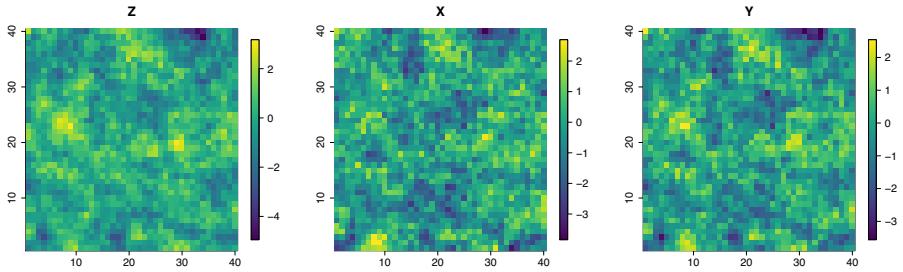
In the parsimonious bivariate conditional autoregressive model we used the following priors for the variance components $\sigma^2 \sim$ Gamma$(1, 1)$, $\sigma_x^2 \sim$ InvGamma$(1, 1)$, and $\tau = \sigma_z^2(1-\rho^2) \sim$ Gamma$(1, 1)$ where both the Gamma and Inverse Gamma distributions are parameterized using shape and rate. For spatial dependence parameters we used $\lambda_z \sim$ Uniform$(0, 1)$ and $\lambda_x \sim$ Uniform$(0, \lambda_z)$ (although we also ran simulation with $\lambda_x \sim$ Uniform$(0, 1)$ and results where similar). Finally for the regression type parameters we use $\beta_x \sim N(0, 100)$ and $\psi = \rho\sigma_z(\sigma_x)^{-1} \sim N(0, 100)$. To improve mixing (β, ψ) were updated in a blocked Gibbs step. Apart from σ_x^2, all other parameters were updated with a random walk Metropolis step using a Gaussian distribution to generate candidate values.

For the semi-parametric models with $L > 1$ we use the CAR prior for b given in § 3.2 of the main text with either a penalized complexity prior (PCP; Simpson et al., 2017) or R^2 prior (Zhang et al., 2020) for σ_b^2, as described below. The semi-parametric CAR model in § 4.3 of the main text is $Y|V \sim$ Normal$(\beta_0 + \beta_x X + \tilde{Z}b + V, \sigma^2 I_n)$. To set priors, we parameterize the covariance parameters as $\sigma_z^2 = \tau^2 c(\lambda_z) r$ and $\sigma^2 = \tau^2(1-r)$ for $c(\lambda_z) = n/\sum_{k=1}^n f_z(\omega_k)$ and $r \in (0, 1)$. Under this parameterization, the total over space error variance is

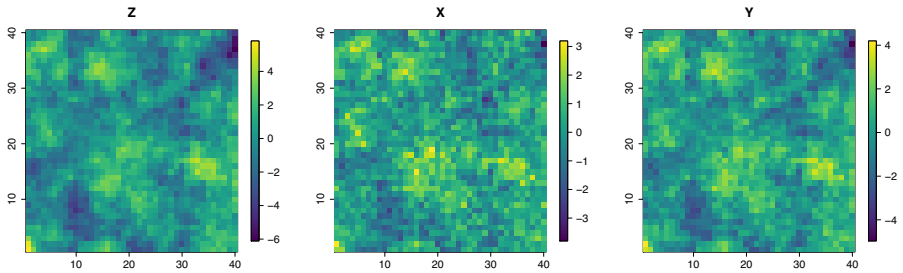
$$\text{Trace}\{\text{cov}(V) + \sigma^2 I_n\} = \tau^2 nr + \tau^2 n(1-r) = \tau^2 n,$$

and thus τ^2 controls the overall variance and r is the proportion of variance attributed to the spatial component of the error. Also, assume the prior for the adjustment coefficients b is normal with mean zero and precision $\sigma_b^{-2}\Omega$, where Ω has diagonal elements N_l and (j, l) off-diagonal element $-I(|j-l| = 1)$. The first is the penalized complexity prior (PCP; Simpson et al., 2017) on the

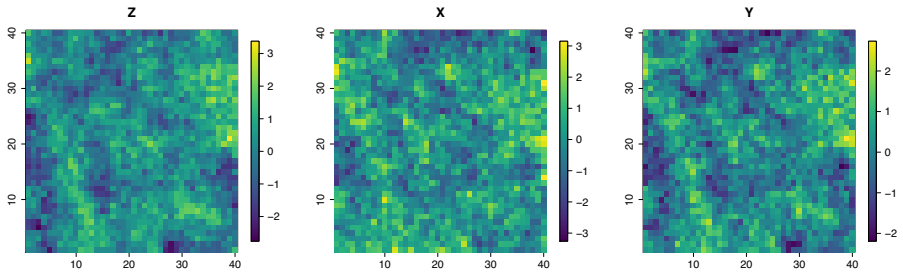
(a) $\phi = 1$ and $\beta_{xz} = 1$



(b) $\phi = 1$ and $\beta_{xz} = 2$



(c) $\phi = 2$ and $\beta_{xz} = 1$



(d) $\phi = 2$ and $\beta_{xz} = 2$

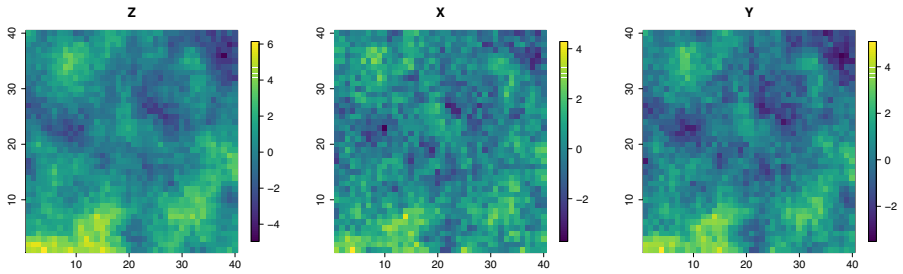


Fig. 1. **Simulated data in discrete space**: Realizations from the discrete-space simulation's data-generating process for different kernel bandwidth (ϕ) and strength of exposure/confounder dependence (β_{xz}).

standard deviation σ_b, i.e. $\sigma_b \sim \text{Exponential}(\xi)$ with scale parameter $\xi = -\log(0.01)0.31/U$, with $U = 0.5$; this choice sets the marginal standard deviation for b approximately equal to 0.5, following the rule of thumb proposed by Simpson et al. (2017). The PCP shrinks the model to the simpler base model without a confounder adjustment. The second is based on the R2D2 prior of Zhang et al. (2020). The variance is written $\sigma_b^2 = \tau^2\sigma_R^2$ where the prior density for σ_R^2 is $f(\sigma_R^2) \propto (\sigma_R^2 + 1)^{-2}$, which induces a Uniform(0,1) prior on the proportion of variance explained by the confounding adjustment, $\text{Trace}\{\text{cov}(\tilde{Z}b)\}/\text{Trace}\{\text{cov}(\tilde{Z}b) + \text{cov}(V) + \sigma^2 I\} = \sigma_R^2/(\sigma_R^2 + 1)$, and in this sense balanced the prior evenly over negligible to complete confounding adjustment. The priors for the remaining parameters are the same for the standard model.

Finally, when fitting the semi-parametric CAR model to the lip cancer and covid data in § 7.1 and § 7.2 of the main text, we exploit the default implementation available in `INLA` which runs very fast; this is based on the original parameterization of the semi-parametric, i.e. $\theta|X \sim \text{CAR}\left(\beta_0 + \sum_{l=1}^{L} \hat{Z}_l b_l, \sigma_z^2, \lambda_z\right)$. Priors are as follows: PCP on σ_z^2, i.e. $\sigma_z \sim \text{Exponential}(\xi)$, with scale parameter $\xi = -\log(0.01)0.31$; PCP on the variance of the adjustment coefficients b, i.e. $\sigma_b \sim \text{Exponential}(\xi)$, with scale parameter $\xi = -\log(0.01)0.31/0.1$, and $\text{logit}(\lambda_z) \sim \text{Normal}(0, 10)$.

7. Prior distributions and computational details for continuous cases

7·1. *Common range Matérn.*

To ensure that all parameters are identifiable, we fit the Matérn model of Gneiting et al. (2010) with common range $\phi_x = \phi_z = \phi_{xz} = \phi$, and additional constraints of $\nu_{xz} > \nu_x$ and $2\nu_{xz} \geq \nu_x + \nu_z$. With these constraints, the maximum for the term $\rho^2 f_{xz}(\omega)^2/\{f_x(\omega)f_z(\omega)\}$ in (4) is obtained at $\omega = 0$ with value $\rho^2\nu_{xz}^2/(\nu_x\nu_z)$, hence the Matérn model with the identifiability parameter constraints is valid if $|\rho| < \sqrt{\nu_x\nu_z}/\nu_{xz}$. We use the R package `GpGp` (Guinness & Katzfuss, 2020) to estimate ϕ_x, ν_x and σ_x^2 by fitting a spatial regression model with Matérn covariance function to the exposure. The estimated values are plugged into the conditional bivariate Matérn model

$$Y = X\beta_x + \rho\sigma_z R_{zx}\hat{R}_x^{-1}X + \delta \tag{5}$$

where $\hat{R}_x$ is the correlation matrix for X computed from the Matérn correlation function with plug-in values of $\hat{\phi}_x$, $\hat{\nu}_x$, and R_{zx} is the cross-correlation matrix between Z and X determined by parameters ν_{xz} and the common range $\phi = \hat{\phi}_x$. To reduce computation cost $\hat{\Sigma}_x^{-1}X$ can be pre-computed to avoid repetitive evaluation of matrix inversion inside the MCMC algorithm. We specify a joint prior for the spatial parameters ρ, ν_z and ν_{xz}, with $\pi(\rho, \nu_z, \nu_{xz}) = \pi_1(\rho|\nu_z, \nu_{xz})\pi_2(\nu_z|\nu_{xz})\pi_3(\nu_{xz})$ where $\pi_1 \sim$ Unif(-$\sqrt{\nu_x\nu_z}/\nu_{xz}$,$\sqrt{\nu_x\nu_z}/\nu_{xz}$), $\pi_2 \sim$ Unif(0,$2\nu_{xz} - \nu_x$) and π_3 is a half Cauchy distribution with mean ν_x and variance 1000. This specification will ensure that all parameter constrains are met. To complete specification of the hierarchical model, we use an exponential distribution with rate parameter 1 for the sum of variances $\sigma_z^2 + \sigma^2$, a uniform prior for the signal to noise ratio $\sigma_z^2/(\sigma_z^2 + \sigma^2)$, and a normal prior with mean 0 and variance 100 for the causal effect β_x. The model is fit using Metropolis-Hastings within Gibbs algorithm with a block update for the spatial parameters.

7·2. *Parsimonious Matérn.*

The estimation of the conditional model in supplementary material (5) simplifies to the estimation of parameters $\{\beta_x, \sigma_z, \sigma, \rho, \nu_z\}$. Priors similar to the common range Matérn model are assigned for β_x and ρ. We use an exponential prior with rate parameter 1 for the sum

$\sigma_z^2(1-\rho^2)+\sigma^2$, and a uniform prior for the signal ratio $\sigma_z^2(1-\rho^2)\{\sigma_z^2(1-\rho^2)+\sigma^2\}^{-1}$. A joint prior is used for ρ, ν_z to ensure $|\rho| \leq 2\sqrt{\nu_x\nu_z}/(\nu_x+\nu_z)$ is $\pi(\rho,\nu_z) = \pi_1(\rho|\nu_z)\pi_2(\nu_z)$ where $\pi_1 \sim Unif\left(-2\sqrt{\nu_x\nu_z}/(\nu_x+\nu_z), 2\sqrt{\nu_x\nu_z}/(\nu_x+\nu_z)\right)$ and π_2 is a half Cauchy distribution with mean ν_x and variance 1000.

8. TABLES

Table 1. *Continuous-space simulation study comparing four methods: The standard Matérn model with $\rho = 0$ and thus no confounding adjustment (Standard), the bivariate Matérn model with common range (Common range), the parsimonious Matérn model (Parsimonious) and the semi-parametric model (Semiparametric). Data are generated with dependence between exposure and confounder controlled by β_{xz} and kernel bandwidth ϕ. All results are multiplied by 100.*

Scenario	Method	ϕ	β_{xz}	RMSE	Bias	SD	Coverage
1	Standard	-	0	4.6 (0.4)	0.1 (0.5)	4.6 (0.0)	96.0 (2.0)
	Common range			14.3 (2.0)	-3.9 (1.4)	13.0 (0.7)	94.9 (2.2)
	Parsimonious			33.2 (3.6)	-6.5 (3.3)	34.7 (1.2)	90.9 (2.9)
	Semiparametric			6.9 (0.5)	0.5 (0.7)	6.9 (0.1)	94.9 (2.2)
2	Standard	$\frac{1}{15}$	1	13.0 (0.5)	11.8 (0.5)	4.9 (0.0)	35.0 (4.8)
	Common range			26.2 (1.9)	-20.8 (1.6)	15.1 (0.6)	68.0 (4.7)
	Parsimonious			82.6 (3.1)	-77.1 (3.0)	31.9 (1.0)	7.0 (2.6)
	Semiparametric			7.4 (0.6)	0.3 (0.7)	6.9 (0.1)	93.0 (2.6)
3	Standard	$\frac{1}{15}$	2	17.9 (0.5)	17.2 (0.5)	5.2 (0.0)	8.0 (2.7)
	Common range			50.6 (2.6)	-45.2 (2.3)	21.4 (0.8)	27.0 (4.5)
	Parsimonious			116.4 (3.5)	-110.8 (3.6)	35.3 (1.2)	0.0 (0.0)
	Semiparametric			6.7 (0.4)	0.1 (0.7)	6.9 (0.1)	95.0 (2.2)
4	Standard	$\frac{2}{15}$	1	5.9 (0.5)	3.2 (0.5)	4.7 (0.0)	90.0 (3.0)
	Common range			15.4 (1.5)	-10.6 (1.1)	10.9 (0.6)	82.0 (3.9)
	Parsimonious			49.2 (4.4)	-34.8 (3.5)	33.3 (1.0)	76.0 (4.3)
	Semiparametric			6.8 (0.5)	0.4 (0.7)	6.9 (0.1)	95.0 (2.2)
5	Standard	$\frac{2}{15}$	2	7.3 (0.5)	4.7 (0.6)	4.8 (0.0)	77.0 (4.2)
	Common range			16.3 (1.2)	-12.2 (1.1)	9.4 (0.4)	70.0 (4.6)
	Parsimonious			63.0 (3.9)	-54.7 (3.1)	28.8 (1.2)	30.0 (4.6)
	Semiparametric			7.9 (0.5)	-0.8 (0.8)	6.9 (0.1)	90.0 (3.0)

RMSE, root-mean-squared error; SD, average posterior standard deviation; Cov, coverage of 95% posterior intervals.

REFERENCES

GNEITING, T., KLEIBER, W. & SCHLATHER, M. (2010). Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association* **105**, 1167–1177.

GUINNESS, J. & KATZFUSS, M. (2020). *GpGp: Fast Gaussian process computation Using Vecchia's approximation*. R package version 0.2.1.

RUBIN, D. B. (1980). Comment on "Randomization analysis of experimental data in the Fisher randomization test" by D. Basu. *Journal of the American statistical association* **75**, 591–593.

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

SIMPSON, D., RUE, H., RIEBLER, A., MARTINS, T. G. & SØRBYE, S. H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science* **32**, 1–28.

ZHANG, Y. D., NAUGHTON, B. P., BONDELL, H. D. & REICH, B. J. (2020). Bayesian regression using a prior on the model fit: The R2-D2 shrinkage prior. *Journal of the American Statistical Association* , 1–37.