Regularized Principal Spline Functions to Mitigate Spatial Confounding

Zaccardi, C., Valentini, P., Ippoliti, L., and Schmidt, A. M. (2025). To appear in **Biometrics**.

ABS25 Applied Bayesian Statistics School

03-06, June 2025

Research funded by European Union - NextGenerationEU.

Research project PRIN2022 *CoEnv - Complex Environmental Data and Modeling*.

Research project PRIN PNRR 2022 (Bando 2023) *SLIDE - Stochastic Modeling of Compound Events*.

Finanziato dall'Unione europea NextGenerationEU

Ministero dell'Università e della Ricerca

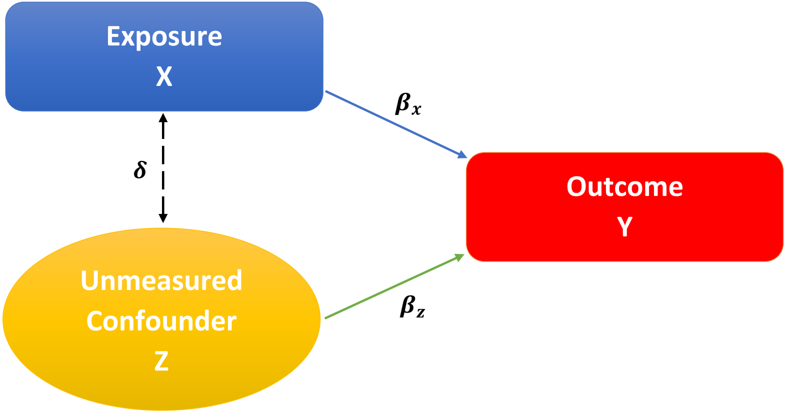
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

Outline

Introduction

Confounding

Goal: correctly recover the **direct** effect of $\boldsymbol{X}$ on $\boldsymbol{Y}$. Not interested in making predictions.



Confounding

- In epidemiological or environmental studies, the relationship between **exposure (X)** and **health outcome (Y)** is of main interest, but it is often **confounded**: one or more other variables (i.e. *confounders*) are associated with exposure and health outcome.
- For example, the *association* between $\mathbf{PM_{2.5}}$ and **mortality** could be confounded by air temperature, infectious disease events, power plant production levels, human habits, etc. (Peng and Dominici 2008).
- Ideally, all confounders must be included into the statistical model in order to obtain unbiased estimators.
- What happens when confounders are **unmeasured**?

Confounding (Independent Data)

- For any sample of size $n > 0$, let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$, $\boldsymbol{X} = (X_1, \ldots, X_n)'$ and $\boldsymbol{Z} = (Z_1, \ldots, Z_n)'$ be normally distributed random vectors.
- Assume the following linear model:

$$\boldsymbol{Y} = \beta_0 + \beta_x \boldsymbol{X} + \beta_z \boldsymbol{Z} + \boldsymbol{u}, \quad \boldsymbol{u} \sim \mathcal{N}(\boldsymbol{0}, \sigma_u^2 \boldsymbol{I}_n), \tag{1}$$

  where $\boldsymbol{u}$ is a vector of pure errors.
- Goal: recover the effect of $\boldsymbol{X}$ on $\boldsymbol{Y}$, represented by β_x, given that there is no information about $\boldsymbol{Z}$.
- We can ignore $\boldsymbol{Z}$ if it is independent of $\boldsymbol{X}$: the **OLS estimator** remains unbiased.

Confounding (Independent Data)

- If $\boldsymbol{Z}$ correlated with $\boldsymbol{X}$, i.e., $\delta = Cor(X_i, Z_i) \neq 0$, it is known as *unmeasured confounder*.
- If we decide to ignore $\boldsymbol{Z}$ anyway, the **confounding problem** arises.
- The **confounding bias** of the OLS estimator of β_x is:

$$\beta_z \delta \sqrt{\frac{Var[\boldsymbol{Z}]}{Var[\boldsymbol{X}]}}.$$

- How can we address this issue?
- In Zaccardi et al. (2025) we discuss a possible solution when the data are spatially structured.

Spatial Confounding

Motivating Example 1

- If the variables vary in space, the unmeasured confounding problem is known as spatial confounding.
- Reich et al. (2006) explore the relationship between socioeconomic status and stomach cancer incidence in Slovenia.
- They compare results from two models:
  - *Without spatial random effect*: the model suggests a negative association, and 95% credible interval (CI) does not include zero
  - *With spatial random effect (SRE)*: the model has a smaller DIC, but there is "**a dramatic effect**" on the posterior mean and variance of the parameter of interest, and now **CI includes zero**

Motivating Example 2

- Paciorek (2010) analyzes spatial data on the association between birth weight and black carbon (BC) concentrations (at the geocoded address of the mother) in eastern Massachusetts.
- To account for potential confounding in the regression model, the author includes census tract income, smooth terms for mother's age, gestational age, and mother's cigarette use, and several categorical variables (sex of baby, maternal education, ...)
- Next he considers what might have happened if most of the covariates were not measured: **much more substantial estimated effect** than the fully adjusted model.
- If a spatial random effect is added to the reduced model, the new estimate **approaches the fully adjusted estimate**.

Spatial random effects: beneficial or harmful?

- As seen, the consequences of adding a spatial random effect (SRE) are not always clear, and *spatial confounding* needs further investigation:
  - ME1: the spatial model results in counterintuitive inference
  - ME2: the spatial model restores the desired association
- Recently, Khan and Berrett (2023) noted that there are two perspectives on spatial confounding ("data analysis" and "data generation") with different goals.
- In this work, we focus on the **data generation** perspective, where the covariates are stochastic.

Analytic Framework

The Regression Model

- Following Paciorek (2010) and Page et al. (2017), consider a finite set of locations, $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ observed within a spatial domain $\mathcal{S} \subseteq \mathbb{R}^2$, and assume the following model:

$$Y(\mathbf{s}_i) = \beta_0 + \beta_x X(\mathbf{s}_i) + \beta_Z Z(\mathbf{s}_i) + \epsilon(\mathbf{s}_i), \quad \epsilon(\mathbf{s}_i) \overset{iid}{\sim} N(0, \sigma_\epsilon^2) \tag{2}$$

  where:
  - $Y(\mathbf{s}_i)$ and $X(\mathbf{s}_i)$ indicate health outcome and exposure, respectively
  - $Z(\mathbf{s}_i)$ is a spatial process representing the unmeasured confounder, which is correlated to the exposure
  - $\beta_z = 1$ without loss of generality
- **Main interest**: correctly recovering β_x
- In the rhs, we only know the exposure, so we will consider the **conditional distribution**, $Z(\mathbf{s}_i)|X(\mathbf{s}_i)$, instead of simply $Z(\mathbf{s}_i)$.

Joint and Conditional Distributions

- Let $\mathbf{X} = (X(\mathbf{s}_1), \ldots, X(\mathbf{s}_n))'$ and $\mathbf{Z} = (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))'$ be two jointly normally-distributed random variables:

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Z} \end{pmatrix} \sim N\left[\begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_z \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_x & \boldsymbol{\Sigma}_{xz} \\ \boldsymbol{\Sigma}'_{xz} & \boldsymbol{\Sigma}_z \end{pmatrix}\right], \tag{3}$$

- If $\mathbf{A}^{1/2}\mathbf{A}^{1/2\prime} = \mathbf{A}$ for any p.d. matrix $\mathbf{A}$, we write:

$$\boldsymbol{\Sigma}_{xz} = \delta \boldsymbol{\Sigma}_x^{1/2} \boldsymbol{\Sigma}_z^{1/2\prime}$$

  where $\delta \in (-1, 1)$ is the correlation between $\mathbf{X}$ and $\mathbf{Z}$.

Joint and Conditional Distributions

- Since $\mathbf{X}$ and $\mathbf{Z}$ are correlated, it would be appropriate to marginalize $\mathbf{Y} = (Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n))'$ over $(\mathbf{Z}|\mathbf{X})$ to obtain:

$$(\mathbf{Y}|\mathbf{X}) = \beta_0 \mathbf{1}_n + \beta_x \mathbf{X} + (\mathbf{Z}|\mathbf{X}) + \boldsymbol{\epsilon}, \;\; \boldsymbol{\epsilon} \sim N(0, \sigma_\epsilon^2 \mathbf{I}_n) \tag{4}$$

- Conditional distribution: $\mathbf{Z}|\mathbf{X} \sim N(\boldsymbol{\mu}_{z|x}, \boldsymbol{\Sigma}_{z|x})$ where

$$\begin{aligned}
\boldsymbol{\mu}_{z|x} &= \boldsymbol{\mu}_z + \boldsymbol{\Sigma}_{xz}' \boldsymbol{\Sigma}_x^{-1} (\mathbf{X} - \boldsymbol{\mu}_x) = \boldsymbol{\mu}_z + \delta \boldsymbol{\Sigma}_z^{1/2} \boldsymbol{\Sigma}_x^{-1/2} (\mathbf{X} - \boldsymbol{\mu}_x)\,, \\
\boldsymbol{\Sigma}_{z|x} &= \boldsymbol{\Sigma}_z - \boldsymbol{\Sigma}_{xz}' \boldsymbol{\Sigma}_x^{-1} \boldsymbol{\Sigma}_{xz} = (1 - \delta^2) \boldsymbol{\Sigma}_z
\end{aligned}$$

- $\mathbf{Y}|\mathbf{X} \sim N(\boldsymbol{\mu}_{y|x}, \boldsymbol{\Sigma}_{y|x})$ where

$$\boldsymbol{\mu}_{y|x} = \beta_0 \mathbf{1}_n + \beta_x \mathbf{X} + \boldsymbol{\mu}_{z|x}\,, \qquad \boldsymbol{\Sigma}_{y|x} = \sigma_\epsilon^2 \mathbf{I}_n + \boldsymbol{\Sigma}_{z|x}$$

Joint and Conditional Distributions

- With geo-referenced data we have:

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Z} \end{pmatrix} \sim N\left[\begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_z \end{pmatrix}, \begin{pmatrix} \sigma_x^2 \mathbf{R}_{\phi_x} & \delta\sigma_x\sigma_z \mathbf{R}_{\phi_x}^{1/2}\mathbf{R}_{\phi_z}^{1/2'} \\ \delta\sigma_x\sigma_z \mathbf{R}_{\phi_z}^{1/2}\mathbf{R}_{\phi_x}^{1/2'} & \sigma_z^2 \mathbf{R}_{\phi_z} \end{pmatrix}\right], \tag{5}$$

  where $\mathbf{R}_{\phi_x}$ and $\mathbf{R}_{\phi_z}$ are defined by parametric correlation functions, $\rho(|\mathbf{s}_i - \mathbf{s}_j|; \phi)$.
- Conditional distribution: $\mathbf{Z}|\mathbf{X} \sim N(\boldsymbol{\mu}_{z|x}, \boldsymbol{\Sigma}_{z|x})$ where

$$\boldsymbol{\mu}_{z|x} = \boldsymbol{\mu}_z + \delta\frac{\sigma_z}{\sigma_x}\mathbf{R}_{\phi_z}^{1/2}\mathbf{R}_{\phi_x}^{-1/2}(\mathbf{X} - \boldsymbol{\mu}_x)\,, \qquad \boldsymbol{\Sigma}_{z|x} = \sigma_z^2(1-\delta^2)\mathbf{R}_{\phi_z}$$

OLS vs GLS

- Paciorek (2010) and Page et al. (2017) fit two models and compare their results:
  - **Non-spatial (OLS)**: $\mathbf{Y} = \beta_0 \mathbf{1}_n + \beta_x \mathbf{X} + \boldsymbol{\epsilon}, \ \boldsymbol{\epsilon} \sim N(0, \sigma_\epsilon^2 \mathbf{I}_n)$
  - **Structured residuals (GLS)**: $\mathbf{Y} = \beta_0 \mathbf{1}_n + \beta_x \mathbf{X} + \boldsymbol{\epsilon}^*, \ \boldsymbol{\epsilon}^* \sim N(0, \boldsymbol{\Sigma}_{y|x})$

- Although both estimators are **biased** there are some differences.

Biased Estimators

- If variance components were known and the true model is Eq. 4, it can be shown that (Paciorek 2010; Page et al. 2017):

$$\Delta_{OLS} = E[\hat{\boldsymbol{\beta}}_{OLS}] - \boldsymbol{\beta} = \mathbf{H}_{OLS}\boldsymbol{\mu}_z + \delta\frac{\sigma_z}{\sigma_x}\mathbf{H}_{OLS}\mathbf{R}_{\phi_z}^{1/2}\mathbf{R}_{\phi_x}^{-1/2}(\mathbf{X} - \boldsymbol{\mu}_x)\,, \tag{6}$$

$$\Delta_{GLS} = E[\hat{\boldsymbol{\beta}}_{GLS}] - \boldsymbol{\beta} = \mathbf{H}_{GLS}\boldsymbol{\mu}_z + \delta\frac{\sigma_z}{\sigma_x}\mathbf{H}_{GLS}\mathbf{R}_{\phi_z}^{1/2}\mathbf{R}_{\phi_x}^{-1/2}(\mathbf{X} - \boldsymbol{\mu}_x)\,, \tag{7}$$

where $\hat{\boldsymbol{\beta}}_{OLS}, \hat{\boldsymbol{\beta}}_{GLS}$ are estimators for $\boldsymbol{\beta} = (\beta_0, \beta_x)'$, and

$$\mathbf{H}_{OLS} = \left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)\tilde{\mathbf{X}}', \quad \tilde{\mathbf{X}} = [\mathbf{1}_n, \mathbf{X}]\,,$$

$$\mathbf{H}_{GLS} = \left(\tilde{\mathbf{X}}'\boldsymbol{\Sigma}_{y|x}^{-1}\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}'\boldsymbol{\Sigma}_{y|x}^{-1}\,.$$

- **Note**: we are interested only in the bias of the estimator for β_x

The Importance of Scales

- When residuals are assumed spatially structured, **confounding bias**:
  - can be **reduced** (i.e., $\Delta_{GLS} < \Delta_{OLS}$) only if exposure varies at a scale smaller than that of confounder, $\phi_x < \phi_z$
  - is **amplified** when exposure varies at a scale greater than that of confounder, $\phi_x > \phi_z$
  - remains the same when $\phi_x = \phi_z$ or $\phi_z \to 0$

- If we have data, how can one know which situation represents our case?

Confounding Adjustment through Basis Functions

- Many models proposed in the literature to account for confounding consider the use of **basis functions**, e.g. Dupont et al. (2022), Guan et al. (2023), and Keller and Szpiro (2020).
- Is the exposure's effect recovered *in any case* when some bases are included into the regression model?
- We set up a simulation study to answer this question: **it depends**...
  - on the scales of variation of exposure and unmeasured confounder
  - on the type of basis expansion considered
  - on the number of bases chosen and in which order
- Do we really need more complex models (than a simple non-spatial fit) to tackle confounding problems?

Connection Between Smoothing and Kriging

- Consider the model, conditionally on the exposure $X(\mathbf{s})$:

$$Y(\mathbf{s}_i) = f_x(\mathbf{s}_i) + \widetilde{\epsilon}_y(\mathbf{s}_i), \qquad \widetilde{\epsilon}_y(\mathbf{s}_i) \sim N(0, \widetilde{\sigma}^2_\epsilon)\,, \tag{8}$$

  where the effects of the exposure and the spatial process are represented by an unknown smooth function $f_x(\mathbf{s}_i)$ defined over the spatial domain of the data.
- Two different but related strategies can be used for estimating the function $f_x(\mathbf{s})$:
  - One is based on the theory of *reproducing kernel Hilbert space* (RKHS), and involves minimizing a penalized function
  - The other is based on optimal prediction in the stochastic process setting (*kriging*).
- The connection between optimal smoothing in a separating RKHS framework and optimal prediction (kriging) for a stochastic process is well-established (see Wahba 1990; Cressie 1993; Kent and Mardia 1994).

Smoothing and interpolation for stochastic processes

- Let $\mathcal{H}$ be an RKHS with reproducing kernel $k(\mathbf{s}_i, \mathbf{s}_j)$ s.t. $\mathcal{H} = \mathcal{H}_0 \bigoplus \mathcal{H}_1$ (Wahba 1990), where $\mathcal{H}_0$ is a null space with basis $\{u_l(\mathbf{s}_i), l = 1, \ldots, q\}$.
- Let $Y(\mathbf{s}_i)$ be a stochastic process with mean and covariance structure

$$E\left[Y(\mathbf{s}_i)\right] = \sum_{l=1}^{q} g_l u_l(\mathbf{s}_i), \qquad Cov\big(Y(\mathbf{s}_i), Y(\mathbf{s}_j)\big) = k(\mathbf{s}_i, \mathbf{s}_j)\,.$$

- Then for any $\mathbf{s}$, the best linear unbiased (or kriging) predictor of $Y(\mathbf{s})$ is identical to the value of optimal smoothing function $f_x(\mathbf{s})$ in an RKHS (Kent and Mardia 1994), and both can be expressed in the form

$$f_x(\mathbf{s}) = \Big(\mathbf{u}(\mathbf{s})'\mathbf{G}_\theta + \mathbf{k}(\mathbf{s})'\mathbf{M}_\theta\Big)\mathbf{y}\,,$$

where $\mathbf{u}(\mathbf{s}) = (u_1(\mathbf{s}), \ldots, u_q(\mathbf{s}))'$, $\mathbf{k}(\mathbf{s}) = \big(k(\mathbf{s}, \mathbf{s}_1), \ldots, k(\mathbf{s}, \mathbf{s}_n)\big)'$

Smoothing and interpolation for stochastic processes

$$f_x(\mathbf{s}) = \Big(\mathbf{u}(\mathbf{s})'\mathbf{G}_\theta + \mathbf{k}(\mathbf{s})'\mathbf{M}_\theta\Big)\mathbf{y}\,,$$

- The matrices $\mathbf{G}_\theta$ and $\mathbf{M}_\theta$ can be found as blocks in the inverse matrix,

$$\begin{bmatrix} \mathbf{K}_\theta & \mathbf{U} \\ \mathbf{U}' & \mathbf{0} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{M}_\theta & \mathbf{G}_\theta' \\ \mathbf{G}_\theta & \mathbf{L} \end{bmatrix}\,,$$

  where $\mathbf{K}_\theta = \mathbf{K} + \theta \mathbf{I}_n$, $\theta \geq 0$ is a smoothing parameter
- **Remark**: The columns of $\mathbf{M}_\theta$ and of $\mathbf{U}$ are orthogonal, i.e. $\mathbf{M}_\theta \mathbf{U} = \mathbf{0}$. Also, this implies that the first q eigenvalues of $\mathbf{M}_\theta$ are 0 and the corresponding eigenvectors are given by the q columns of $\mathbf{U}$.

The reduced-rank random effects model

- Let $\theta = 0$ and consider the equation $f_x(\mathbf{s}) = \Big(\mathbf{u}(\mathbf{s})'\mathbf{G} + \mathbf{k}(\mathbf{s})'\mathbf{M}\Big)\mathbf{y}$.
- Consider the spectral decomposition of $\mathbf{M}$, $\mathbf{M} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}'$, such that $\mathbf{M}\mathbf{v}_l = \lambda_l \mathbf{v}_l$ and $\mathbf{y} = \mathbf{V}\boldsymbol{\xi}$. Then, the Kriging predictor can be rewritten as

$$f_x(\mathbf{s}) = \Big(\mathbf{u}(\mathbf{s})'\mathbf{G} + \mathbf{k}(\mathbf{s})'\mathbf{M}\Big)\mathbf{V}\boldsymbol{\xi} = \sum_{l=1}^{n}\Big\{\big(\mathbf{u}(\mathbf{s})'\mathbf{G} + \mathbf{k}(\mathbf{s})'\mathbf{M}\big)\mathbf{v}_l\Big\}\xi_l\,. \tag{9}$$

- Let $\psi_l(\mathbf{s}) = \big(\mathbf{u}(\mathbf{s})'\mathbf{G} + \mathbf{k}(\mathbf{s})'\mathbf{M}\big)\mathbf{v}_l$ define the l-th **principal kriging function** (PKF; Kent, Mardia, et al. 2001; Fontanella et al. 2019).
- For any $\mathbf{s} \in \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$, this function acts as an interpolator, $\psi_l(\mathbf{s}) = \mathbf{v}_l(\mathbf{s})$ for $l = 1, \ldots, n$.
- If the linear combination above is restricted to a limited number of eigenvectors, say $p < n$, Equation (8) represents a **reduced-rank random effects model**, which can be expressed as:

$$Y(\mathbf{s}_i) \;=\; \sum_{l=1}^{q} u_l(\mathbf{s}_i)\xi_l^a + \sum_{l=q+1}^{p} \psi_l(\mathbf{s}_i)\xi_l^b + \widetilde{\epsilon}_y(\mathbf{s}_i) \tag{10}$$

Proposed Regularization of the Regression Model

Proposed Model

- We reformulate Equation (8) as:

$$\begin{aligned} Y(\mathbf{s}_i) &= \beta_0 + \beta_x X(\mathbf{s}_i) + \widetilde{f}(\mathbf{s}_i) + \widetilde{\epsilon}_y(\mathbf{s}_i) \\ &= \beta_0 + \beta_x X(\mathbf{s}_i) + \sum_{l=2}^{q} u_l(\mathbf{s}_i)\xi_l^a + \sum_{l=q+1}^{p} \psi_l(\mathbf{s}_i)\xi_l^b + \widetilde{\epsilon}_y(\mathbf{s}_i) \\ &= \beta_0 + \beta_x X(\mathbf{s}_i) + \mathbf{b}(\mathbf{s}_i)'\widetilde{\boldsymbol{\xi}} + \widetilde{\epsilon}_y(\mathbf{s}_i)\,, \end{aligned} \tag{11}$$

where:

- the link between thin plate splines and kriging (Kent and Mardia 1994) shows that $\widetilde{f}(\mathbf{s}_i)$ defines a *thin plate spline*.
- the basis functions $u_j(\mathbf{s}_i)$ are monomials in the spatial coordinates of $\mathbf{s}$, such that $q = 3$ and $\mathbf{u}(\mathbf{s}_i) = \left(1, \mathbf{s}_i[1], \mathbf{s}_i[2]\right)'$
- $\mathbf{b}(\mathbf{s}_i) = \left(u_2(\mathbf{s}_i), u_3(\mathbf{s}_i), \psi_{q+1}(\mathbf{s}_i), \ldots, \psi_p(\mathbf{s}_i)\right)'$
- $\widetilde{\boldsymbol{\xi}} = \left(\xi_2^a, \ldots, \xi_q^a, \xi_{q+1}^b, \ldots, \xi_p^b\right)'$ is a $(p-1)$-dimensional vector of expansion coefficients
- the kernel function is $k(\mathbf{s}_i, \mathbf{s}_j) = \frac{1}{8\pi}\|\mathbf{s}_i - \mathbf{s}_j\|^2 \log \|\mathbf{s}_i - \mathbf{s}_j\|$

Prior Specification

- By setting the $\theta = 0$, model flexibility is controlled by the basis dimension, so that model selection becomes a matter of choosing p rather than estimating a smoothing parameter.
- We propose to set $p = n - 1$ and use spike-and-slab priors (George and McCulloch 1997) to identify the most important bases in a **non-sequential** fashion.
- We consider a **non-local** spike-and-slab prior structure known as the **first-order product moment** *(pMOM)* proposed by Johnson and Rossell (2012) and Rossell and Telesca (2017).
- The l-th element of $\widetilde{\boldsymbol{\xi}}$ is considered as a mixture between a point mass at zero and a bimodal density, namely

$$\xi_l | \widetilde{\sigma}_\epsilon^2 \overset{ind}{\sim} \xi_l^2 (\nu \widetilde{\sigma}_\epsilon^2)^{-1} N\left(0, \nu \widetilde{\sigma}_\epsilon^2\right)$$

  where $\nu = 0.348$ is a hyperparameter controlling the prior variance.
- We assign independent priors on the remaining parameters, namely $\beta_0, \beta_x \overset{iid}{\sim} N(0, 10^6)$, and $\widetilde{\sigma}_\epsilon^2 \sim IG(0.01, 0.01)$

Simulations

Setup

- **Purpose**: compare our proposal with existing approaches. Benchmark: OLS (no confounding adjustment).
- The data generating mechanism is inspired to that proposed by Marques and Kneib (2022).
- The data are sampled from a 64×64 grid over the unit square. We consider $n = 500$ randomly-sampled locations as fixed.
- We set the **relative OLS bias**, $\frac{\Delta_{OLS}}{\beta_x} = 0.15$. Therefore, we fix $\delta = Cor(\mathbf{X}, \mathbf{Z}) = 0.5$ and we allow σ_z to vary:

$$\frac{\Delta_{OLS}}{\beta_x} = \frac{1}{\beta_x} \left[\delta \frac{\sigma_z}{\sigma_x} \mathbf{H}_{OLS} \mathbf{R}_{\phi_z}^{1/2} \mathbf{R}_{\phi_x}^{-1/2} \mathbf{X} \right]_2$$

$$\mathbf{H}_{OLS} = \left(\tilde{\mathbf{X}}' \tilde{\mathbf{X}} \right) \tilde{\mathbf{X}}'$$

Data Generating Mechanism

- Recall Eq. 5:

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Z} \end{pmatrix} \sim N\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sigma_x^2 \mathbf{R}_{\phi_x} & \delta\sigma_x\sigma_z \mathbf{R}_{\phi_x}^{1/2}\mathbf{R}_{\phi_z}^{1/2'} \\ \delta\sigma_x\sigma_z \mathbf{R}_{\phi_z}^{1/2}\mathbf{R}_{\phi_x}^{1/2'} & \sigma_z^2 \mathbf{R}_{\phi_z} \end{pmatrix}\right],$$

- The exposure is sampled from its marginal distribution, then:

$$\mathbf{g} = (\mathbf{Z}|\mathbf{X}) \sim N(\boldsymbol{\mu}_{z|x}, \boldsymbol{\Sigma}_{z|x})$$
$$\boldsymbol{\mu}_{z|x} = \delta\frac{\sigma_z}{\sigma_x}\mathbf{R}_{\phi_z}^{1/2}\mathbf{R}_{\phi_x}^{-1/2}\mathbf{X}\,, \qquad \boldsymbol{\Sigma}_{z|x} = \sigma_z^2(1-\delta^2)\mathbf{R}_{\phi_z}$$

with $\sigma_x^2 = \sigma_z^2 = 1$, and $\phi_x, \phi_z \in [0.05, 0.5] \times [0.05, 0.5]$.

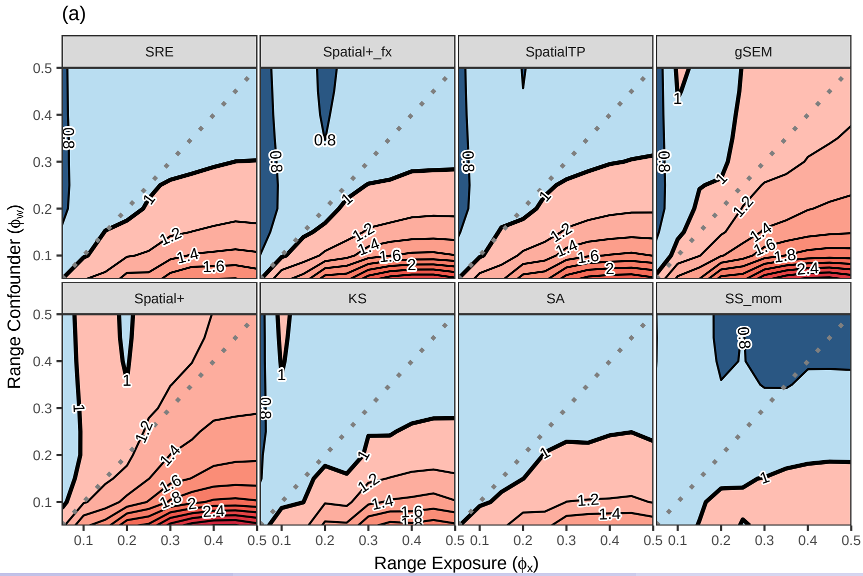
Data Generating Mechanism

- The outcome is simulated from the following:

$$\mathbf{Y} = \beta_0 \mathbf{1}_n + \beta_x \mathbf{X} + \mathbf{g} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(0, \sigma_\epsilon^2 \mathbf{I}_n)$$

  with $\beta_0 = 1, \beta_x = 2, \sigma_\epsilon^2 = 0.25$.
- Results are based on 100 replicates.
- The following figure shows contour plots of the ratio of MAE of competing methods over that of the OLS (non-spatial) model, for all ϕ_x, ϕ_z combinations.

Simulation Results



Reducing Confounding Bias in Ozone–NO_x Association

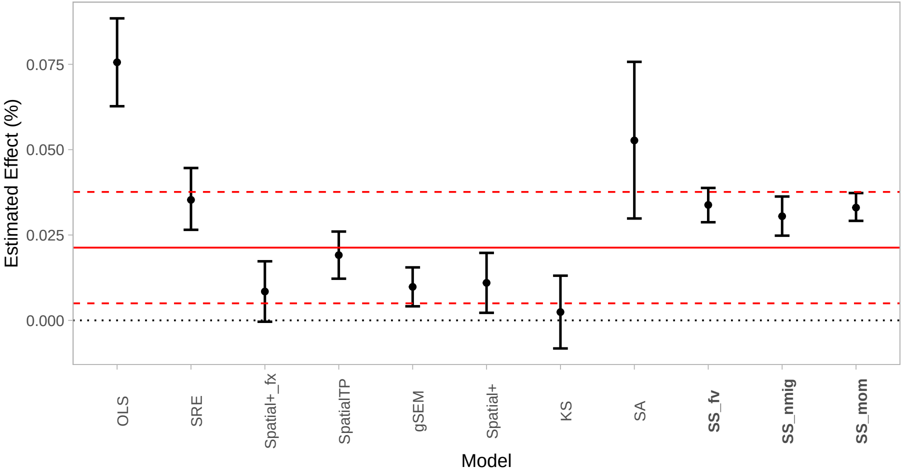
Real Data Application

- The relationship between $\mathbf{NO_x}$ and $\mathbf{O_3}$ is inferred from June to August 2019 in three Italian regions, namely Lazio, Abruzzo and Molise.
- Data for all the variables are available as measurements on a 0.1° latitude-longitude grid, from the Copernicus Atmosphere Monitoring Service (CAMS) Atmosphere Data Store and from the Copernicus Climate Change Service (C3S) Climate Data Store.
- To control for potential confounders, we employ the model, for $i = 1, \ldots, n = 353$:

$$\begin{aligned} log(O_3(\mathbf{s}_i)) &= \beta_1 log(NO_x(\mathbf{s}_i)) + \beta_2 u_{10}(\mathbf{s}_i) + \beta_3 v_{10}(\mathbf{s}_i) + \beta_4 Temp(\mathbf{s}_i) + \beta_5 SSR(\mathbf{s}_i) \\ &+ \beta_6 VOC(\mathbf{s}_i) + \beta_7 RH(\mathbf{s}_i) + \epsilon(\mathbf{s}_i)\,, \qquad \epsilon(\mathbf{s}_i) \overset{iid}{\sim} N(0, \sigma^2_\epsilon)\,, \end{aligned} \tag{12}$$

Real Data Application

Full Model: $\boxed{log(O_3(\mathbf{s}_i)) = \beta_1 log(NO_x(\mathbf{s}_i)) + \beta_2 u_{10}(\mathbf{s}_i) + \beta_3 v_{10}(\mathbf{s}_i) + \beta_4 Temp(\mathbf{s}_i) + \beta_5 SSR(\mathbf{s}_i) + \beta_6 VOC(\mathbf{s}_i) + \beta_7 RH(\mathbf{s}_i) + \epsilon(\mathbf{s}_i)}$



References I

Cressie, Noel (1993). *Statistics for spatial data*. Rev. ed. Wiley series in probability and mathematical statistics. New York: Wiley. ISBN: 9780471002550.

Dupont, Emiko, Wood, Simon N., and Augustin, Nicole H (2022). "Spatial+: a novel approach to spatial confounding". In: *Biometrics* 78.4, pp. 1279–1290.

Fontanella, Lara, Ippoliti, Luigi, and Kume, Alfred (2019). "The offset normal shape distribution for dynamic shape analysis". In: *Journal of Computational and Graphical Statistics* 28.2, pp. 374–385.

George, Edward I and McCulloch, Robert E (1997). "Approaches for Bayesian variable selection". In: *Statistica Sinica*, pp. 339–373.

Guan, Yawen et al. (2023). "Spectral adjustment for spatial confounding". In: *Biometrika* 110.3, pp. 699–719.

Johnson, Valen E and Rossell, David (2012). "Bayesian model selection in high-dimensional settings". In: *Journal of the American Statistical Association* 107.498, pp. 649–660.

Keller, Joshua P and Szpiro, Adam A (2020). "Selecting a scale for spatial confounding adjustment". In: *Journal of the Royal Statistical Society Series A: Statistics in Society* 183.3, pp. 1121–1143.

Kent, John T and Mardia, Kanti V (1994). "The Link Between Kriging and Thin-Plate Splines". In: *Probability, Statistics and Optimization: A Tribute to Peter Whittle*. Ed. by F.P. Kelly. Wiley, pp. 325–339.

Kent, John T, Mardia, Kanti V, et al. (2001). "Functional models of growth for landmark data". In: *Proceedings in Functional and Spatial Data Analysis* 109115.

References II

Khan, Kori and Berrett, Candace (2023). "Re-thinking Spatial Confounding in Spatial Linear Mixed Models". In: *arXiv preprint arXiv:2301.05743.*

Marques, Isa and Kneib, Thomas (2022). "Discussion on "Spatial+: A novel approach to spatial confounding" by Emiko Dupont, Simon N. Wood, and Nicole H. Augustin". In: *Biometrics* 78.4, pp. 1295–1299. DOI: 10.1111/biom.13650.

Paciorek, Christopher J (2010). "The importance of scale for spatial-confounding bias and precision of spatial regression estimators". In: *Statistical science: a review journal of the Institute of Mathematical Statistics* 25.1, p. 107.

Page, Garritt L et al. (2017). "Estimation and prediction in the presence of spatial confounding for spatial linear models". In: *Scandinavian Journal of Statistics* 44.3, pp. 780–797.

Peng, Roger D and Dominici, Francesca (2008). *Statistical methods for environmental epidemiology with R: a case study in air pollution and health*. Springer.

Reich, Brian J, Hodges, James S, and Zadnik, Vesna (2006). "Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models". In: *Biometrics* 62.4, pp. 1197–1206.

Rossell, David and Telesca, Donatello (2017). "Nonlocal priors for high-dimensional estimation". In: *Journal of the American Statistical Association* 112.517, pp. 254–265.

Wahba, G. (1990). *Spline Models for Observational Data*. CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 59, Philadelphia, PA: Society for Industrial and Applied Mathematics.

Zaccardi, Carlo et al. (2025). "Regularized Principal Spline Functions to Mitigate Spatial Confounding". In: *Biometrics (to appear).*

Thank you!