Environmetrics

WILEY

RESEARCH ARTICLE OPEN ACCESS

Causal Inference for Geostatistical Data Using an INLA-based Spatial Propensity Score

Chiara Di Maria | Alessandro Albano | Mariangela Sciandra | Antonella Plaia

Department of Economics, Business and Statistics, University of Palermo, Palermo, Italy

Correspondence: Chiara Di Maria (chiara.dimaria@unipa.it)

Received: 7 May 2025 | **Revised:** 10 February 2026 | **Accepted:** 1 April 2026

Keywords: counterfactuals | spillover effects | SUTVA

ABSTRACT

In this paper, we propose a Bayesian approach for spatial causal inference based on combining spatial propensity scoring with Integrated Nested Laplace Approximation. The method models both local and spillover exposure effects via multiple likelihoods and treats counterfactuals as missing data, allowing inference also for non-Gaussian outcomes. We validated the proposed method through simulations and an application to U.S. county-level cancer data, demonstrating the critical importance of properly accounting for spatial dependence when drawing causal conclusions from geostatistical data. Our results show that the proposed method achieves MCMC-comparable accuracy with substantially reduced computational time.

1 | Introduction

Understanding causal relationships from observational data is a central goal in many scientific disciplines. Recently, some scholars started to analyze spatial data within a causal framework, a relevant task in several real-world applications, such as studying the effect of a public policy on the duration of unemployment (Huber and Steinmayr 2021) in different areas, how opening new gyms at different locations influences the prevalence of colorectal cancer (Smith et al. 2023) or how armed conflicts impact forest loss (Christiansen et al. 2022). Recent reviews on spatial causal inference (Reich et al. 2021; Gao et al. 2022; Akbari et al. 2023) highlighted the complexities connected to drawing causal conclusions in spatial settings. First of all, the classical assumption of units' independence, on which most causal approaches rely, does not hold. When analyzing spatially indexed data, such as county-level health outcomes or pollution exposure, nearby units often influence each other, violating the independence assumption and potentially biasing causal estimates if not properly accounted for. This phenomenon, commonly referred to as *spillover* in spatial contexts, falls under the broader concept of *interference* in the causal inference literature. Interference has been studied in various settings beyond spatial data, such as in clustered randomized trials (Hudgens and Halloran 2008; VanderWeele and Tchetgen 2011; Sinclair et al. 2012), and in network-based analyses where dependencies between units are explicitly modeled (Aronow and Samii 2017; Bhattacharya et al. 2020; Zigler and Papadogeorgou 2020).

Another problem arising in spatial causal inference is confounding, which poses challenges to the estimation of causal effects in general. However, this issue is exacerbated in the presence of spatial data. Indeed, spatial confounding arises when unobserved or unmeasured variables exhibit spatial patterns that correlate with both the treatment (or exposure) assignment and the outcome of interest (Paciorek 2010). This spatial structure in confounding variables creates complex dependencies that standard statistical approaches often fail to address adequately.

Abbreviations: DIC, deviance information criterion; INLA, integrated nested Laplace approximation; JPS, joint propensity score; MCMC, Markov chain Monte Carlo; PS, propensity score; RCT, randomized controlled trial; SAR, spatial autoregressive model; SDAG, spatial directed acyclic graph; SPDE, stochastic partial differential equation.

Chiara Di Maria and Alessandro Albano are co-first authors.

In spatial settings, confounding typically manifests in several distinct ways. First, many environmental, socioeconomic, and demographic factors cluster geographically, creating spatially structured confounding that can bias causal effect estimates if not properly accounted for (Reich et al. 2021). For instance, when evaluating the health effects of environmental exposures, socioeconomic status often exhibits spatial clustering and simultaneously influences both exposure levels and health outcomes. Second, spatial proximity can be a confounder when it affects both treatment assignment and outcomes through mechanisms beyond measured covariates (Schnell and Papadogeorgou 2020). For example, in studies of neighborhood effects, geographic location may determine access to resources that influence both the likelihood of receiving an intervention and the potential outcomes. Evidence of such confounding can often be found through spatial autocorrelation in model residuals, which may signal the presence of unmeasured spatial confounders (Hodges and Reich 2010). This spatial dependence violates the independence assumptions underlying many statistical models and can lead to inflated precision in effect estimates when not properly modeled.

Several methodological approaches have been developed to address spatial confounding. Spatial random effects models incorporate spatial correlation structures to account for unmeasured spatial confounders (Reich et al. 2021). However, these models must be carefully specified, as research has shown that naively including spatial random effects can sometimes increase rather than reduce bias in treatment effect estimates (Hodges and Reich 2010). Restricted spatial regression (RSR) represents another approach that constrains spatial random effects to be orthogonal to the treatment variable, helping to prevent adjustment away from the true treatment effect (Thaden and Kneib 2018). More recently, researchers have explored causal graphical models extended to spatial contexts to clarify the structural relationships between spatial processes, treatments, and outcomes (Datta et al. 2019). These spatial directed acyclic graphs (SDAGs) help identify appropriate adjustment strategies for spatial confounding while avoiding adjustment for colliders or mediators. Sensitivity analyses specifically designed for spatial settings have also been developed to quantify the potential impact of unmeasured spatial confounding on causal estimates (VanderWeele and Ding 2017; Schnell and Papadogeorgou 2020). These approaches allow researchers to determine how strong an unmeasured spatial confounder would need to be to meaningfully alter study conclusions.

One of the most widely spread approaches to address confounders in causal inference is the propensity score approach (Thoemmes and Kim 2011; Rosenbaum 2023). The key idea is to estimate the probability of treatment assignment given covariates and then use this score to balance treated and untreated groups. However, spatial dependencies complicate the application of propensity scores in spatial contexts. This can lead to unreliable causal estimates when these spatial structures are ignored, as is often the case with traditional propensity score estimation methods.

Recent work has begun to address these challenges. Davis and colleagues (Davis et al. 2017, 2021) analyzed racial disparities in health by proposing a spatial propensity score for areal data obtained by including region-specific random effects in the propensity score model. Papadogeorgou et al. (2019) focus on point-referenced data and discuss a novel propensity score that takes into account the distance between pairs of locations. Giffin et al. (2023) proposed a generalized spatial propensity score that jointly models the local and spillover exposure effect, offering improved adjustment for confounding. However, most of these approaches rely on Markov Chain Monte Carlo (MCMC) methods, which can be computationally intensive and difficult to scale, especially when dealing with large spatial datasets or complex models that include both local and spillover effects.

An alternative for spatial causal inference may be represented by the Integrated Nested Laplace Approximation (INLA, Rue et al. 2009), which provides computationally efficient Bayesian inference for latent Gaussian models, making it well-suited for handling complex spatial dependence structures. By incorporating spatially structured random effects and allowing for flexible prior specifications, INLA can help account for unmeasured spatial confounders while maintaining computational tractability.

In this work, we introduce a novel approach to spatial causal inference that combines spatial propensity score modeling with the computational efficiency of the INLA framework. While conceptually related to the method proposed by Giffin et al. (2023), in which local exposure and spillover effects are modeled jointly, our approach offers several key improvements. Crucially, INLA's built-in support for multiple likelihoods is particularly well-suited for modeling both direct (local) and indirect (spillover) effects simultaneously. An additional crucial aspect lies in the Bayesian nature of INLA. Indeed, we treat counterfactual outcomes as missing data and use posterior predictive distributions to estimate causal effects. This formulation allows for the estimation of causal effects as differences in posterior means of counterfactual distributions, without requiring the response variable to follow a Normal distribution, as assumed in Giffin et al. (2023). Finally, our approach offers substantial computational advantages, providing a faster and more scalable alternative to traditional MCMC-based methods. The implementation code used to reproduce the simulation study and the real-data application is publicly available at the GitHub repository INLA-causal-inference.

The remainder of the paper is organized as follows. In Section 2, we introduce basic notation for causal inference with interference. Section 3 presents the proposed method and details its INLA-based estimation strategy. In Section 4, we validate the approach via simulation experiments, while in Section 5, we illustrate its application to U.S. county-level cancer data. Finally, in Section 6, we conclude with a discussion of findings and future work.

2 | Background

Causal inference in observational studies traditionally relies on the potential outcomes framework (Rubin 1974), where each unit $i = 1, \ldots, n$ is associated with a set of potential outcomes $Y_i(x)$, one for each level of a binary or continuous exposure X. The causal effect of the exposure on the outcome can be defined as a contrast of the counterfactuals $Y(x)$ and $Y(x')$, for $x \neq x'$, i.e., one

compares the value of the outcome if all subjects were exposed to X at level x and its value if all were exposed to level x'.

The fundamental problem of causal inference is that only one potential outcome is observed for each unit, while all others remain counterfactual. To identify causal effects and express them in terms of observed data, some assumptions need to hold (Pearl 2009), among which is the so-called *Stable Unit Treatment Value Assumption* (SUTVA; Rubin 1974), which consists of two components: (i) no interference between units, and (ii) no hidden versions of the treatment.

In the context of spatial data, however, SUTVA is typically violated due to interference (or spillover), that is, the treatment assigned to one unit may influence the outcome of another. In such cases, the potential outcomes must be indexed not only by the unit's own treatment, but also by the treatments assigned to other units in the system. For example, suppose that a new power plant is opened in a city: The number of asthma cases could increase not only in the city, but also in neighboring cities, because of the power plant's emissions.

Formally, let us assume that data are collected at n spatial locations $s \in \{s_1, \ldots, s_n\} \in \mathcal{D} \subset \mathbb{R}^2$, and let a variable with the s subscript denote the value of that variable at location s. Let $\mathbf{X} = (X_1, \ldots, X_n)$ denote the vector of treatment assignments across all locations. Then, the potential outcome for unit s is written as $Y_s(\mathbf{X})$. This creates a rapidly growing set of potential outcomes, complicating both estimation and interpretation. To manage this complexity, we adopt the *exposure mapping* framework (Aronow and Samii 2017), which assumes that each unit's outcome depends only on its own treatment and an exposure function summarizing the influence of others. Specifically, we define the potential outcome as

$$Y_s(x_s, \tilde{x}_{-s}) := Y_s\big(X_s = x_s, \tilde{X}_{-s} = \tilde{x}_{-s}\big),$$

where $\tilde{X}_{-s}$ is a real-valued summary of the treatment assignments in all locations but s, and which is often referred to as the spillover exposure.

The potential outcome framework is then extended to account for both the *direct effect* of the treatment X_s and the *indirect effect* arising from $\tilde{X}_{-s}$. This requires redefining common causal estimands. For instance, under a binary treatment setting $X_s \in \{0, 1\}$, we can define:

- The local (direct) effect: $Y_s(1, \tilde{x}_{-s}) - Y_s(0, \tilde{x}_{-s})$, holding the spillover level fixed;
- The spillover (indirect) effect: $Y_s(x_s, \tilde{x}_{-s}) - Y_s(x_s, \tilde{x}'_{-s})$, holding own treatment fixed;

This decomposition of effects requires stronger assumptions for identification, such as exposure-conditional ignorability:

$$(Y_s(x_s, \tilde{x}_{-s}) \perp (X_s, \tilde{X}_{-s})) | \boldsymbol{Z},$$

which states that, conditional on observed covariates $\boldsymbol{Z}$, the treatment and spillover assignments are as good as random. This assumption is particularly delicate in spatial settings due to latent spatial confounding: Unobserved spatially varying factors may affect both treatment assignment and outcomes. To address this, modern approaches—such as the one proposed in this paper—explicitly model the spatial structure of the data using random fields.

3 | Causal Inference With Joint Propensity Score

This section presents the proposed approach for spatial causal inference in three main steps. First, in Section 3.1, we recall the fundamentals and computational advantages of INLA. Next, in Section 3.2, we describe how to model the exposure local and spillover effects through a joint propensity score. Finally, in Section 3.3, we show how causal effects are estimated by treating counterfactual outcomes as missing data and drawing on posterior predictive distributions.

3.1 | Basics of INLA

To deal with geostatistical data and model spatial effects, we adopted INLA (Rue et al. 2009) combined with a stochastic partial differential equation approach (SPDE; Lindgren et al. 2011), which allows us to explicitly include spatial effects in our models. Such effects are modeled as a Gaussian Random Field (GRF), which is a generalization of a Gaussian process to multiple dimensions. One of its most important properties is its covariance function, which describes how values of the random field at different locations are related.

The covariance can be modeled in different ways (Cressie 1993), a typical choice being a Matérn function

$$C(h) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\kappa h)^\nu K_\nu(\kappa h) \tag{1}$$

where $h = \|\mathbf{s}_1 - \mathbf{s}_2\|$ is the Euclidean distance between locations, σ^2 is the marginal variance, $\kappa > 0$ is a scaling parameter controlling the range of correlation, $\nu > 0$ is a smoothness parameter, and K_ν is the modified Bessel function of the second kind.

Lindgren et al. (2011) proved that a Matérn covariance model can be expressed as the stationary solution to an SPDE:

$$(\kappa^2 - \Delta)^{\alpha/2}\omega u(\mathbf{s}) = \mathcal{W}(\mathbf{s}), \tag{2}$$

where ω controls the variance, α determines the smoothness, Δ is the Laplacian operator, and $\mathcal{W}(\mathbf{s})$ is spatial Gaussian white noise.

The SPDE is solved numerically using the finite element method (FEM), where the spatial domain is discretized into a triangular mesh. The solution $u(\mathbf{s})$ is approximated as a linear combination of basis functions:

$$u(\mathbf{s}) \approx \sum_{k=1}^{n} \psi_k(\mathbf{s}) w_k, \tag{3}$$

where $\{\psi_k(\mathbf{s})\}$ are piecewise linear basis functions defined on the mesh vertices and $\mathbf{w} = (w_1, \ldots, w_n)^T$ are Gaussian weights.

This representation transforms the continuous spatial model into a Gaussian Markov Random Field (GMRF; Rue and Held 2005) with sparse precision matrix $\mathbf{Q}$, enabling efficient computation $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}^{-1})$. The precision matrix $\mathbf{Q}$ is sparse because the basis functions have local support, with $Q_{ij} \neq 0$ only when vertices i and j share an edge in the triangulation. This sparsity is critical for the computational efficiency of INLA.

The aim of INLA is to obtain the posterior distribution of the Gaussian field $\boldsymbol{\phi}$, that is,

$$p(\phi_i|\mathbf{y}) = \int p(\phi_i|\boldsymbol{\theta}, \mathbf{y}) p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}, \tag{4}$$

where $\mathbf{y}$ represents the observed data and $\boldsymbol{\theta}$ the vector of hyperparameters. INLA computes the hyperparameters' posteriors through a series of nested Laplace approximations, while $p(\phi_i|\boldsymbol{\theta}, \mathbf{y})$ becomes particularly easy to compute if ϕ is a GMRF, given the sparsity of the precision matrix.

In addition to being particularly suited for analyzing spatial data, INLA offers an advantage in the context of causal inference. Indeed, since it is Bayesian in nature, we can estimate counterfactuals using a missing-value approach. This approach treats unobserved potential outcomes as missing data, allowing us to generate posterior distributions for these counterfactuals and obtain posterior estimates of causal effects. The computational efficiency of INLA makes it especially valuable for causal inference problems with complex spatial or temporal structures. By specifying appropriate priors and likelihood functions, we can incorporate domain knowledge and account for spatial autocorrelation when estimating treatment effects. This is particularly valuable in observational studies where confounding may vary spatially.

This leads us to another advantage in using INLA, i.e., how confounders are addressed. Dealing with confounders is a ubiquitous issue in causal inference within an observational framework. In a spatial setting, this is complicated by the fact that the confounders can be spatial in nature, and they may have a spillover effect as well; thus, as noted by Giffin et al. (2023), adjusting only for the value of the confounder at each location is not sufficient. Like them, we propose to use a propensity score approach to adjust for confounding, which jointly models the probability of X_s and $\tilde{X}_{-s}$. They proved that conditioning on a joint propensity score is sufficient to estimate the causal effect of X on Y unbiasedly. However, throughout their paper, they assume that X_s and $\tilde{X}_{-s}$ are conditionally independent given the covariates, and, as a consequence, they can be modeled separately. We extend their framework by explicitly modeling the local and spillover exposure jointly using the INLA ability to fit joint likelihood models. We provide more details in the following.

3.2 | Joint Modeling of Treatment and Spillover

To address confounding in the presence of interference, one should notice that the treatment and the outcome at each location s do not depend only on covariates measured at that location, but can be influenced by the values of the covariates at other locations. This easily makes models high-dimensional and computationally difficult to handle. Thus, a convenient approach may be the use of a propensity score (Rosenbaum 2023), that, however, unlike the traditional one in settings without interference, has to model both the local and spillover exposure as functions of covariates. We propose a joint model for the bivariate exposure vector $(X_s, \tilde{X}_{-s})$, conditional on covariates $\boldsymbol{Z}$, but, unlike Giffin et al. (2023), who assume conditional independence between X_s and X_{-s} given covariates, we explicitly model their dependence structure using INLA's joint likelihood specification.

Specifically, INLA handles multivariate responses, modeling the local treatment X_s and the spillover effect $\tilde{X}_{-s}$ simultaneously. Their joint density serves as the foundation for our *Joint Propensity Score* (JPS),

$$e_s(X_s, \tilde{X}_{-s}|\boldsymbol{Z}) := f(X_s, \tilde{X}_{-s}|\boldsymbol{Z}, \boldsymbol{\phi}),$$

representing the joint conditional density of observing a certain treatment–spillover configuration given the covariates, as implied by the INLA model, and generalizing the classical propensity score to a setting with structured interference. Here, $\boldsymbol{\phi} = \{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\phi}_{12}\}$ denotes the vector of spatial effects of the model. We model the JPS components as:

$$\begin{aligned} X_s|\boldsymbol{Z}, \boldsymbol{\phi}_1 &\sim F_1\left(g_1^{-1}(\eta_1)\right), \\ \eta_1 &= \beta_0 + \boldsymbol{Z}\,\boldsymbol{\beta}_1 + \boldsymbol{\phi}_1(s), \\ \tilde{X}_{-s}|\boldsymbol{Z}, \boldsymbol{\phi}_2, \boldsymbol{\phi}_{12} &\sim F_2\left(g_2^{-1}(\eta_2)\right), \\ \eta_2 &= \gamma_0 + \boldsymbol{Z}\,\boldsymbol{\gamma}_1 + \boldsymbol{\phi}_2(s) + \boldsymbol{\phi}_{12}(s). \end{aligned} \tag{5}$$

Here, F_1 and F_2 represent appropriate distributions for the treatment and spillover effects, respectively, which can be chosen based on the nature of the data (e.g., Binomial for binary treatments). The functions g_1 and g_2 are corresponding link functions (e.g., logit), and the linear predictors η_1 and η_2 contain the same set of confounders $\boldsymbol{Z}$ and different spatial effects. Specifically, the spatial fields are defined as:

$$\begin{aligned} \boldsymbol{\phi}_1 &\sim \text{GMRF}(\mathbf{0}, \mathbf{Q}_{\boldsymbol{\phi}_1}^{-1}) \\ \boldsymbol{\phi}_2 &\sim \text{GMRF}(\mathbf{0}, \mathbf{Q}_{\boldsymbol{\phi}_2}^{-1}) \\ \boldsymbol{\phi}_{12} &= \lambda \cdot \boldsymbol{\phi}_1, \end{aligned}$$

where $\boldsymbol{\phi}_1$ represents the spatial random effect for the treatment assignment, $\boldsymbol{\phi}_2$ is an independent spatial random effect for the spillover component, and $\boldsymbol{\phi}_{12}$ is a scaled copy of the treatment spatial effect, with scaling parameter λ estimated from the data. This specification induces a structured dependence between the linear predictors for treatment and spillover: The spillover predictor η_2 includes both an independent spatial effect ϕ_2 and a shared spatial component $\phi_{12} = \lambda \cdot \phi_1$ The shared component ϕ_{12} creates a direct link between the spatial pattern in treatment assignment and the spillover effects, with the scaling parameter λ controlling the strength of this relationship. Meanwhile, ϕ_2 captures additional spatial variation in spillover that is independent of the treatment spatial pattern.

This structure explicitly incorporates the dependence between the treatment and spillover processes through the shared spatial component $\boldsymbol{\phi}_{12}$. The precision matrices $\mathbf{Q}_{\boldsymbol{\phi}_1}$ and $\mathbf{Q}_{\boldsymbol{\phi}_2}$ are derived from the SPDE approach with Matérn covariance using the finite element method on a triangular mesh, as described earlier.

This JPS approach offers several advantages over separate models. First, it explicitly accounts for the correlation between local treatment assignment and spillover effects. Second, it provides a unified propensity score estimation framework that incorporates both direct and indirect exposure mechanisms. Finally, it leverages the computational efficiency of INLA to handle complex spatial structures that would be computationally prohibitive with traditional MCMC methods. Indeed, INLA's deterministic approximations, based on sparse matrix algorithms and numerical integration, offer fast and accurate inference even for large spatial datasets.

3.3 | Estimation of Causal Effects

Once the JPS has been estimated, $\hat{e}_s(X_s, \tilde{X}_{-s}|\boldsymbol{Z}_s)$, it can be included in the outcome model to adjust for confounding. The adjusted outcome model takes the form:

$$g(E[Y_s|X_s, \tilde{X}_{-s}, \hat{e}_s]) = \delta_0 + \delta_1 X_s + \delta_2 \tilde{X}_{-s} + \kappa\ \hat{e}_s + \boldsymbol{\phi}_Y(s) \quad (6)$$

where $g(\cdot)$ is an appropriate link function, and $\boldsymbol{\phi}_Y(s)$ is a spatial random effect that captures residual spatial correlation in the outcome.

When the outcome model is linear with additive Gaussian errors,

$$Y_s = \delta_0 + \delta_1 X_s + \delta_2 \tilde{X}_{-s} + \kappa\ \hat{e}_s + \boldsymbol{\phi}_Y(s) + \epsilon_s, \quad \epsilon_s \sim \mathcal{N}(0, \sigma^2),$$

the regression coefficients δ_1 and δ_2 can be directly interpreted as average causal effects, under unconfoundedness. This is because the expected value of the counterfactual can be expressed as the conditional expectation in Equation (6). Therefore, the causal effects are obtained by taking the differences between counterfactual scenarios. For instance, the direct effect of changing the treatment from 0 to 1 (while keeping spillover fixed) is:

$$\begin{aligned} &\mathbb{E}\left[Y_s(1, \tilde{x}_{-s})\right] - \mathbb{E}\left[Y_s(0, \tilde{x}_{-s})\right] \\ &= \left[\delta_0 + \delta_1 \cdot 1 + \delta_2\ \tilde{x}_{-s} + \hat{e}_s + \boldsymbol{\phi}_Y(s)\right] \\ &\quad - \left[\delta_0 + \delta_1 \cdot 0 + \delta_2\ \tilde{x}_{-s} + \hat{e}_s + \boldsymbol{\phi}_Y(s)\right] \\ &= \delta_1 \end{aligned} \quad (7)$$

Similarly, the spillover effect is given by δ_2. The linearity of the model ensures that these coefficients directly represent the causal effects, without requiring further calculations or transformations.

However, when the outcome is modeled through a nonlinear link function, the regression coefficients do not directly correspond to causal effects. In this case, we must recover the counterfactuals explicitly. To address this, we adopt a Bayesian predictive approach. For each location s, we define a set of hypothetical exposure configurations, assuming a binary exposure for simplicity, and construct a data-augmented matrix that includes all possible counterfactual scenarios. Let $\mathcal{Y}$ denote the observed data and model estimates. The counterfactual potential outcome under exposure $(x_s, \tilde{x}_{-s})$ is then imputed via the posterior predictive distribution:

$$f(Y_s(x_s, \tilde{x}_{-s})|\mathcal{Y}) = \int f(Y_s|X_s = x_s, \tilde{X}_{-s} = \tilde{x}_{-s}, \boldsymbol{\theta})\ f(\boldsymbol{\theta}|\mathcal{Y})\ d\boldsymbol{\theta} \quad (8)$$

where $\boldsymbol{\theta}$ represents all model parameters, including fixed effects, spatial random effects, and error variance components.

In practice, to estimate the causal estimands of interest, we extract the posterior means of the counterfactual values for each configuration and compare them using appropriate contrasts (difference or ratio). For example, on the difference scale:

$$\text{Local Effect:} \quad \mathbb{E}[Y_s(1, \tilde{x}_{-s})] - \mathbb{E}[Y_s(0, \tilde{x}_{-s})] \quad (9)$$

$$\text{Spillover Effect:} \quad \mathbb{E}[Y_s(0, \tilde{x}_{-s})] - \mathbb{E}[Y_s(0, \tilde{x}'_{-s})] \quad (10)$$

$$\text{Total Effect:} \quad \mathbb{E}[Y_s(1, \tilde{x}_{-s})] - \mathbb{E}[Y_s(0, \tilde{x}'_{-s})] \quad (11)$$

where $\tilde{x}_{-s}$ and $\tilde{x}'_{-s}$ denote two fixed spillover configurations.

This strategy aligns with the Rubin causal model, treating unobserved potential outcomes as missing data and recovering them through coherent posterior predictive inference. In addition, it accommodates both homogeneous and heterogeneous treatment effects. By conditioning on location-specific covariates and spatial random effects when computing the counterfactuals, we can estimate spatially varying causal effects, providing insights into how treatment impacts may differ across the spatial domain.

4 | Simulation Study

We consider two simulation settings to evaluate the performance of our proposed JPS method. Scenario I follows the data generation scheme of Giffin et al. (2023) with Gaussian outcomes, while Scenario II extends this framework to binary responses.

In both scenarios, data are generated on a regular grid defined over the unit square $[0, 1] \times [0, 1]$. A grid of size $n^{1/2} \times n^{1/2}$ is constructed with $n = 100^2 = 10,000$ locations, and for each complete dataset $N = 100$ independent replications are generated, yielding a total of 10000×100 observations. While this spatial structure follows Giffin et al. (2023), we extend their single-covariate setup by simulating three spatially-structured confounders Z_1, Z_2, Z_3. Each confounder is generated independently from a mean-zero, unit-variance Gaussian process with an isotropic exponential covariance function and spatial range of 0.6.

The binary treatment $X_s \in \{0, 1\}$ is drawn independently across units, conditional on their covariates Z_s, as a Bernoulli random variable with success probability

$$P(X_s = 1|\mathbf{Z}_s) = \text{expit}(Z_{1s} + Z_{2s} + Z_{3s} - 3).$$

To account for spatial interference, the spillover exposure is defined as a weighted sum (Simulation I) or a weighted mean (Simulation II) of the treatment indicators in neighboring locations using a Gaussian kernel with bandwidth parameter τ.

Specifically, for the Gaussian outcome setting (Simulation I), the spillover exposure is defined as:

$$\tilde{X}_{\tau,-s} = \sum_{s' \neq s} \exp\left(-\frac{d^2_{s,s'}}{\tau^2}\right) X_{s'},$$

while for binary outcomes (Simulation II), it is constructed as a weighted average to constrain its values between 0 and 1:

$$\tilde{X}_{\tau,-s} = \frac{\sum_{s' \neq s} \exp\left(-\frac{d_{s,s'}^2}{\tau^2}\right) X_{s'}}{\sum_{s' \neq s} \exp\left(-\frac{d_{s,s'}^2}{\tau^2}\right)}.$$

For Simulation I (Gaussian outcomes), the response is generated as:

$$Y_s = \delta_1 X_s + \delta_2 \tilde{X}_{\tau,-s} + h(\mathbf{Z}_s) + \epsilon_s, \quad \epsilon_s \sim \mathcal{N}(0,1),$$

For Simulation II (binary outcomes), we generate:

$$Y_s \sim \text{Bernoulli}\left(\text{expit}(\delta_1 X_s + \delta_2 \cdot \lambda \tilde{X}_{\tau,-s} + h(\mathbf{Z}_s))\right),$$

where the true causal parameters are $\delta_1 = \delta_2 = 1$. Here, $\lambda = 2$ is included only in Scenario II to amplify the spillover effect, which would otherwise be compressed on the probability scale. The spatial confounding, denoted by $h(\mathbf{Z}_s)$, is constructed in three alternative forms described in Table 1.

Each confounding setting uses a spatially smoothed term, W_{ks}, computed as a Gaussian-kernel weighted average of neighboring Z_{ks} values with bandwidth 0.5, thereby introducing spatial structure in the confounding. While the Linear form is identical in both scenarios, the Exponential and Cubic forms are multiplied by a constant (-1 and -2, respectively) in Scenario II to produce comparable effect sizes on the logit scale of the binary outcome.

In the Gaussian setting, causal effects are directly estimated through regression coefficients. Therefore, the local effect δ_1 corresponds to the coefficient of X_s, while the spillover effect δ_2 is captured by the coefficient of $\tilde{X}_{\tau,-s}$ (as shown in Equation 7). In contrast, for binary outcomes, causal effects are estimated via contrasts of predicted probabilities derived from the fitted linear predictor, denoted by

$$\hat{\eta}_i = \hat{\delta}_0 + \hat{\delta}_1 X_s + \hat{\delta}_2 \tilde{X}_{\tau,-s} + \hat{\kappa} \hat{e}_s,$$

where $\hat{e}_s$ is the estimated Joint Propensity Score. Specifically, the spillover effect is defined as the difference in predicted probabilities among treated units ($X_s = 1$) in the groups of those exposed and unexposed to spillover:

$$\mathbb{E}\left[\text{logit}^{-1}(\hat{\eta}_i) | X_s = 1, \tilde{X}_{\tau,s} = 1, \hat{e}_s\right] - \mathbb{E}\left[\text{logit}^{-1}(\hat{\eta}_i) | X_s = 1, \tilde{X}_{\tau,s} = 0, \hat{e}_s\right].$$

Similarly, the direct effect compares units with and without direct exposure, holding spillover exposure fixed at zero:

$$\mathbb{E}\left[\text{logit}^{-1}(\hat{\eta}_i) | X_s = 1, \tilde{X}_{\tau,s} = 0, \hat{e}_s\right] - \mathbb{E}\left[\text{logit}^{-1}(\hat{\eta}_i) | X_s = 0, \tilde{X}_{\tau,s} = 0, \hat{e}_s\right].$$

TABLE 1 | Forms of spatial confounding $h(\mathbf{Z}_s)$.

Type	$h(\mathbf{Z}_s)$
Linear	$W_{1s} + W_{2s} + W_{3s}$
Exponential	$\exp(W_{1s}) + \exp(W_{2s}) + \exp(W_{3s})$
Cubic	$-0.5(W_{1s}^3 + W_{2s}^3 + W_{3s}^3)$

The corresponding true effects are obtained by applying the logistic transformation to the linear predictor components. Specifically, the true spillover effect is:

$$\mathbb{E}\left[\text{logit}^{-1}(h(Z_s) + \delta_1 + \lambda \cdot \delta_2)\right] - \mathbb{E}\left[\text{logit}^{-1}(h(Z_s) + \delta_1)\right],$$

and the true direct effect is given by:

$$\mathbb{E}\left[\text{logit}^{-1}(h(Z_s) + \delta_1)\right] - \mathbb{E}\left[\text{logit}^{-1}(h(Z_s))\right].$$

In the estimation phase, the smoothing parameter τ used to construct $\tilde{X}_{\tau,-s}$ is selected via grid search. Specifically, a sequence of plausible candidate values is evaluated, and the optimal τ is chosen based on the Deviance Information Criterion (DIC) obtained from an INLA model for the outcome that includes all covariates.

We compare our JPS method against a naive model that omits the propensity score adjustment, thereby ignoring spatial confounding. In other words, the naive model is a spatial model for the outcome that includes just the local exposure and the spillover as covariates. Tables 2 and 3 present simulation results for Scenario I and Scenario II, respectively, across three bandwidth values ($\tau \in \{0.3, 0.7, 1.5\}$), three confounding specifications (Linear, Exponential, Cubic), and each simulation setting is repeated 500 times to assess estimator performance robustly.

As regards Scenario I (Table 2), the JPS method demonstrates high stability across all simulation scenarios. For the direct effect δ_1, the bias remains consistently low, ranging from 26.78 to 31.20 units (per 1000) across all combinations of τ values and confounding specifications. The spillover effect δ_2 shows a good performance as well, with biases between 2.39 and 20.40 units. Most notably, the JPS method achieves perfect recovery of the bandwidth parameter τ in all configurations, with zero bias. In contrast, the naive model exhibits substantially higher and more variable biases. Overall, as τ increases from 0.3 to 1.5 (representing more diffuse spatial spillover), the method's bias for δ_1 remains stable, while spillover effect bias slightly decreases, as expected. This pattern holds across all confounding specifications, indicating that the method is robust to varying degrees of spatial interaction.

Moving to Scenario II, the JPS method continues to perform well in the more challenging binary outcome setting, though with some notable differences from the Gaussian case. Direct effect estimates remain highly accurate, with biases ranging from 10.19 to 15.16 across all scenarios. Spillover effect biases are still small, ranging from 13.95 to 16.65 units. The primary challenge in the binary setting concerns bandwidth estimation. When the true τ is small (0.3), the method achieves good recovery. However, as the true τ increases to 1.5, estimation becomes more difficult, with biases rising to 385.80–421.80. This difficulty arises from the combination of the nonlinear link function and the weighted average construction of spillover, which compresses information about the spatial scale. Importantly, this bandwidth estimation challenge does not substantially compromise causal effect estimates: Even with $\tau = 1.5$, the direct effect bias remains below 15 and spillover bias below 47 across linear and cubic confounding. Once again, the naive model performs much worse, particularly for spillover effects under exponential confounding.

TABLE 2 | Simulation bias (multiplied by 1000) for Gaussian outcomes (Scenario I) across different τ values and confounding specifications.

		Linear			Exponential			Cubic		
τ	Model	δ_1	δ_2	τ	δ_1	δ_2	τ	δ_1	δ_2	τ
0.3	JPS	28.31	9.21	0.00	30.91	16.66	0.00	31.20	20.40	0.00
	Naive	394.55	26.16	100.00	427.89	115.61	102.20	253.65	258.83	44.60
0.7	JPS	26.78	3.70	0.00	27.76	8.74	0.00	27.45	5.08	0.00
	Naive	112.59	185.68	0.00	229.41	259.16	4.00	108.03	140.04	1.80
1.5	JPS	27.01	2.39	0.00	28.86	5.04	0.00	27.79	3.16	0.00
	Naive	194.18	147.09	99.40	349.14	205.84	119.20	172.52	110.13	101.00

TABLE 3 | Simulation bias (multiplied by 1000) for binary outcomes (Scenario II) across different τ values and confounding specifications.

		Linear			Exponential			Cubic		
τ	Model	δ_1	δ_2	τ	δ_1	δ_2	τ	δ_1	δ_2	τ
0.3	JPS	15.16	16.23	27.20	11.76	47.01	33.00	12.25	15.16	19.40
	Naive	162.60	170.12	162.40	167.00	774.53	506.60	159.49	197.76	163.60
0.7	JPS	15.02	16.65	101.20	11.46	55.18	125.60	13.00	15.31	94.80
	Naive	156.39	179.30	110.20	148.97	755.38	198.80	153.61	206.67	109.20
1.5	JPS	14.89	13.95	398.20	10.19	46.88	421.80	13.39	14.80	385.80
	Naive	195.00	145.53	500.00	85.56	702.81	500.00	192.47	173.50	500.00

TABLE 4 | Comparison of computation times.

Method	Average time per run	Standard Deviation
MCMC (Giffin et al. 2023)	2735.8 s (~45.6 min)	620.4 s (~ 10.3 min)
JPS	36.4 s	2.5 s

Overall, in both Scenario I and II (a total of 18 different simulation settings), the JPS method consistently outperforms the naive approach. The JPS demonstrates particular strength in Gaussian settings and maintains robust causal effect estimation even when bandwidth estimation becomes challenging in binary outcome scenarios.

In addition, we carried out other simulation studies using the same data-generating mechanism described in Giffin et al. (2023), including just a single covariate and considering only the Gaussian case. To quantify the computational gain with respect to the MCMC approach, we run 10 replicates with this configuration, whose results are reported in the Appendix. The bias levels observed in our simulations are comparable to those reported by Giffin et al. (2023), thereby confirming the validity of our approach in recovering both direct and spillover effects under spatial interference and confounding. However, a notable advantage of JPS lies in its computational efficiency. Specifically, computations were performed on a Dell Precision 3680 Tower workstation with an Intel Core i9-14,900 processor, 32GB RAM, 1TB SSD, and an NVIDIA GeForce RTX 4080 GPU (16GB). As shown in Table 4, the MCMC-based approach with $S = 30,000$ iterations—matching the setup of the original study—required approximately 45 min per run, whereas our INLA-based JPS method completed in just 36 s on average.

5 | Application to Real Data

In this section, we apply the proposed JPS to real-world data, focusing on United States county-level cancer incidence rates and associated health determinants. Specifically, we investigate the impact of $PM_{2.5}$ concentration, along with its spillover effects in neighboring counties, on cancer incidence rates, while controlling for various potential confounding factors.

The analysis integrates data from multiple sources. The primary outcome of interest, cancer incidence rates, is obtained from the State Cancer Profiles database (https://www.statecancerprofiles.cancer.gov/). These data report the number of cancer cases per 100,000 residents in each U.S. county, averaged over the period from 2017 to 2021. The explanatory variables used in the analysis are derived from the County Health Rankings & Roadmaps dataset (https://www.countyhealthrankings.org/health-data/methodology-and-sources/data-documentation), which includes various health and socioeconomic indicators averaged over the 2013 to 2016 period. The dataset includes: The average daily concentration of particulate matter ($PM_{2.5}$), the percentage of adult smokers, the percentage of individuals classified as obese, the percentage of physically inactive individuals, the percentage of individuals reporting excessive alcohol consumption, the percentage of uninsured residents, high school graduation rates, unemployment rates, and the percentage of children living in poverty.

A key aspect of this analysis is the temporal shift between the exposure variables and the cancer incidence data. Specifically, cancer incidence rates reflect diagnoses between 2017 and 2021, while the covariates are averaged over the 2013–2016 period. Using a lag of four years between PM2.5 exposure and the recording of lung cancer cases is a reasonable approach based on the available literature. Several studies have highlighted the

long-term effects of PM2.5 exposure on lung cancer development, indicating that prolonged exposure can lead to significant health outcomes, including cancer progression and increased mortality rates (Acharjee 2018; Liao et al. 2020; Ciabattini et al. 2021; Wang et al. 2023).

These covariates, along with the cancer incidence data and geographic coordinates (latitude and longitude), were merged using the Federal Information Processing Standards (FIPS) code. Alaska and Hawaii were excluded due to their geographic separation from the continental US. Additionally, within the included states, counties with missing data belong to the states of California, part of Montana, Colorado, part of Nevada, part of North Dakota, South Dakota, Nebraska, Kansas, Arkansas, and Alabama. The final dataset used in the analysis includes data from 2790 counties with complete information.

Table 5 reports the main descriptive statistics for the ten variables used in the analysis. The variables display considerable variation, particularly in health-related indicators such as incidence rate and child poverty, which suggests potential heterogeneity across the dataset.

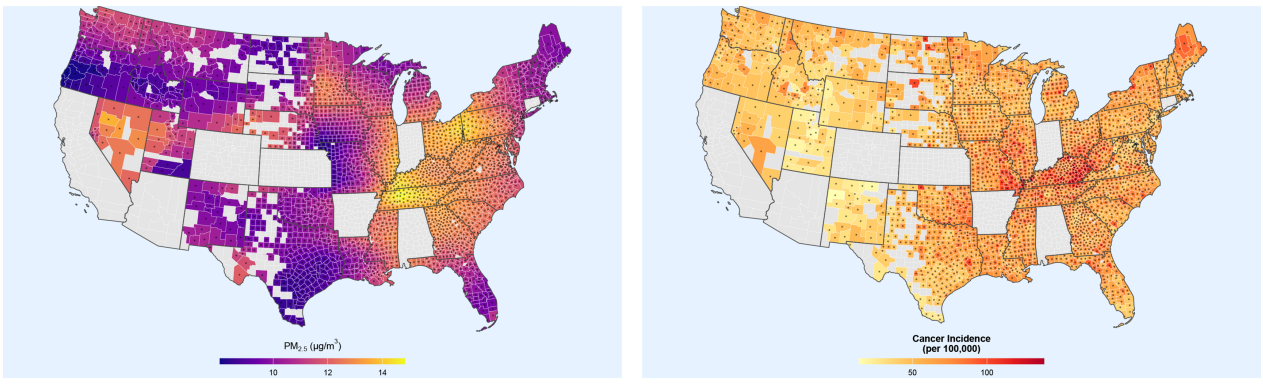
Figure 1 displays the spatial distribution of $PM_{2.5}$ concentrations (panel a) and cancer incidence rates (panel b) across US counties. Both variables exhibit substantial spatial heterogeneity and clustering patterns that warrant careful attention in causal inference.

Particulate matter concentrations show marked regional variation, with notably higher levels concentrated in the Midwest and parts of the South. Lower exposure levels are observed in western mountain states and parts of the Northeast. This spatial clustering likely reflects regional variations in emission sources, environmental conditions, and socioeconomic factors affecting air quality. Cancer rates (per 100,000 population) also demonstrate considerable spatial structure, with higher incidence concentrated in specific regional clusters, particularly visible in parts of the South and Midwest, while lower rates are more prevalent in western states.

Interesting spatial co-occurrence patterns between high $PM_{2.5}$ concentrations and elevated cancer rates can be noted, particularly in certain regions of the Midwest and South. However, the relationship is not uniform across the country: Some areas with high pollution show moderate cancer rates (see Nevada), while others with lower exposure exhibit elevated incidence (e.g., the counties of Missouri). Moreover, the presence of spatial spillover effects is also evident: Counties neighboring high-pollution areas may experience indirect exposure through atmospheric transport.

TABLE 5 | Descriptive statistics of variables used in the analysis.

Variable Name	Role	Min	Mean	Median	Max	SD
Incidence Rate (per 100k)	Outcome	14.10	63.60	62.40	138.30	17.30
$PM_{2.5}$ Concentration	Treatment	8.05	11.57	11.75	14.82	1.46
Smoking Prevalence (%)	Covariate	5.75	20.38	20.00	43.00	5.08
Obesity Prevalence (%)	Covariate	13.25	30.90	31.00	45.25	3.75
Physical Inactivity (%)	Covariate	11.00	27.68	27.75	41.75	4.89
Excessive Drinking (%)	Covariate	5.25	15.72	15.50	34.00	4.48
Uninsured Rate (%)	Covariate	3.25	17.48	17.50	38.50	5.38
High School Graduation Rate (%)	Covariate	24.00	82.98	84.50	100.00	8.90
Unemployment Rate (%)	Covariate	1.00	7.58	7.40	17.38	2.40
Child Poverty Rate (%)	Covariate	3.00	24.62	24.25	58.00	9.00



(a) (b)

FIGURE 1 | Spatial patterns of exposure and outcome variables across the continental United States. (a) shows $PM_{2.5}$ concentrations (in $\mu g/m^3$) measured at county centroids. (b) displays age-adjusted cancer incidence rates (per 100,000 population). Grey-shaded counties represent locations with missing data.

TABLE 6 | Comparison of causal effect estimates across modeling approaches: Posterior means with 95% credible intervals and computational time.

Parameter	JPS	Giffin et al. (MCMC)	Naive
Local Effect (δ_1)	4.26 [2.93, 5.57]	5.10 [3.10, 7.14]	4.21 [3.82, 4.60]
Spillover Effect (δ_2)	0.004 [0.003, 0.005]	0.004 [0.003, 0.006]	0.003 [0.003, 0.004]
Computation time	36 s	1,661 s (~28 min)	0.18 s

To account for potential spatial spillover effects, we construct an extended exposure variable incorporating information on ambient air pollution from neighboring areas, defined as a weighted sum of $PM_{2.5}$ concentrations from adjacent counties. The weights are computed using a Gaussian kernel based on the Haversine distances between county centroids. Formally, for county s, the spillover exposure is given by

$$\tilde{X}_{\tau,-s} = \sum_{s' \neq s} \exp\left(-\frac{d_{s,s'}^2}{\tau^2}\right) \mathrm{PM}_{2.5,s'}.$$

We explore a range of candidate values for τ through a grid search procedure, selecting the optimal value based on the DIC from the INLA outcome model, which includes relevant health-related covariates. This process yields an optimal estimate of $\hat{\tau} = 770$, and the corresponding spillover exposure variable is constructed accordingly.

To ensure that the estimated spillover exposure captures only the influence of neighboring areas and not local pollution, we residualize the spillover variable with respect to the local $PM_{2.5}$ concentration. This step is necessary because the local measurement of $PM_{2.5}$ inherently includes pollution from adjacent regions due to atmospheric diffusion and measurement proximity. As a result, the constructed spillover variable may be partially collinear with the local exposure.

To disaggregate these two sources of exposure, we fit a simple linear regression of the spillover term on local $PM_{2.5}$:

$$\tilde{X}_{\tau,-s} = \alpha + \delta\ \mathrm{PM}_{2.5,s} + \zeta_s,$$

and use the residuals $\hat{\zeta}_s$ (i.e., the portion of $\tilde{X}_{\tau,-s}$ orthogonal to $PM_{2.5,s}$) from this regression as the adjusted spillover component. This residualized variable isolates the spatial externalities attributable exclusively to neighboring areas, allowing for a more precise estimation of local and spillover effects in subsequent analyses.

Next, we estimate the bivariate JPS for the treatment–spillover variables as specified in Equation (5), where both likelihoods are assumed to be Gaussian. We fit this joint model using INLA, and then extract the posterior mean

$$\hat{e}_s = \mathbb{E}\left[(\mathrm{PM}_{2.5,s}, \zeta_s) | \boldsymbol{Z}_s\right],$$

as our estimated propensity score for each county.

Finally, we fit the spatial outcome INLA model for the cancer incidence rate Y_s, incorporating the estimated propensity score $\hat{e}_s$ to adjust for residual confounding. The model takes the form:

$$Y_s = \delta_0 + \delta_1\ \mathrm{PM}_{2.5,s} + \delta_2\ \zeta_s + \kappa\ \hat{e}_s + \boldsymbol{\phi}_Y(s) + \varepsilon_s,$$

where $\boldsymbol{\phi}_Y(s)$ is a spatial latent effect modeled via an SPDE approach and $\varepsilon_s \sim N(0, \sigma^2)$ is independent noise. Here, δ_1 represents the direct effect of local $PM_{2.5}$, δ_2 the spillover effect, and κ the coefficient on the propensity adjustment, which does not have a causal interpretation.

Table 6 presents a comparison of causal effect estimates across three modeling approaches: Our proposed JPS method implemented via INLA, the MCMC-based approach of Giffin et al. (2023), and a naive model that ignores spatial confounding. For each method, we report posterior mean estimates with 95% credible intervals for the local effect (δ_1) and the spillover effect (δ_2), along with computational time.

The three methods yield similar conclusions regarding the presence and magnitude of causal effects, while differing in uncertainty quantification. All three approaches detect a statistically significant local effect, with point estimates ranging from 4.21 to 5.10. The JPS estimate of 4.26 is nearly identical to the naive estimate of 4.21 but slightly lower than Giffin et al.'s 5.10. Critically, the naive model produces the narrowest credible interval [3.82, 4.60], which likely reflects underestimation of uncertainty due to not accounting for spatial confounding. In contrast, both propensity score-based methods (JPS and Giffin et al.) yield wider intervals that more appropriately reflect uncertainty in the presence of spatial confounding.

For the spillover effect, remarkable agreement emerges across all three methods, with estimates of 0.004 (JPS and Giffin et al.), 0.003 (Naive), and substantially overlapping credible intervals. This estimate suggests that the spillover effect is robustly identified regardless of whether spatial confounding is explicitly modeled. However, we emphasize that this apparent robustness should not be interpreted as justification for the naive approach: In settings with stronger spatial confounding, the bias could be far more severe.

From a model fit perspective, the JPS approach achieves a DIC of 18, 565.09, substantially lower than the naive model's DIC of 19, 800.42. We note that DIC comparison with Giffin et al.'s MCMC approach is not feasible, as INLA's deterministic approximations and MCMC's sampling-based inference calculate DIC through fundamentally different mechanisms that preclude direct numerical comparison. Beyond model fit considerations, the computational advantage of the JPS approach is substantial: Our method completes in 36 s compared to approximately 28 min (1661 s) for the MCMC approach.

From a public health perspective, the estimated local effects indicate a strong impact of local $PM_{2.5}$ exposure and cancer incidence: A one-unit increase in $PM_{2.5}$ concentration ($\mu g/m^3$) is associated with statistically significant 4.26 additional cancer cases per 100,000 population (JPS). The spillover effect is statistically

significant as well, as its credible interval excludes zero. This effect quantifies the additional cancer risk attributable to $PM_{2.5}$ exposure in neighboring counties, beyond what is captured by local exposure. The spillover mechanism likely operates through atmospheric transport of pollutants across county boundaries, population mobility patterns (e.g., commuting to nearby counties), and potentially shared environmental or socioeconomic determinants that span geographic areas. This finding aligns with the extensive epidemiological literature documenting the carcinogenic effects of fine particulate matter.

6 | Conclusion

This paper introduces a novel JPS approach to spatial causal inference that leverages the computational efficiency of INLA to address the challenges of confounding and interference in spatial settings. Our methodology builds upon existing spatial propensity score approaches by explicitly modeling the joint distribution of local treatment and spillover effects, while incorporating spatial random effects to account for unobserved spatial confounding.

Several key contributions distinguish our approach. First, we developed a joint model for treatment and spillover effects that explicitly accounts for their spatial dependence structure, providing a more realistic representation of exposure mechanisms in spatial settings. Second, our Bayesian predictive approach treats counterfactual outcomes as missing data, enabling coherent inference for causal effects through posterior predictive distributions. This formulation accommodates nonlinear response models, allowing for flexible specification of outcome distributions beyond the Gaussian case. Third, by employing INLA's deterministic approximations rather than MCMC methods, we achieve substantial computational gains without sacrificing accuracy, making our approach feasible for large spatial datasets and complex models.

Our simulation studies were conducted using two distinct settings. The first setting directly adopted the Gaussian data generation framework proposed by (Giffin et al. 2023), while the second setting represents our novel extension of this framework to accommodate non-Gaussian (binary) responses. The proposed method successfully recovered both local and spillover effects across various confounding scenarios, maintaining low bias even under complex nonlinear confounding functions. Notably, our approach exhibited comparable accuracy to Giffin's MCMC-based methods while reducing computation time from approximately 45 min to just 36 s per run, representing a significant advance in computational efficiency for spatial causal inference.

The application to U.S. county-level cancer incidence data further validated our methodology in a real-world context. Our analysis revealed that $PM_{2.5}$ concentrations have both direct local and meaningful spillover effects on cancer incidence rates, even after controlling for potential confounders and accounting for spatial dependence. These findings highlight the importance of considering spatial interference when evaluating environmental health impacts and carry important implications for environmental health policy and regulatory decision-making. First, the magnitude of the local effect suggests that even modest reductions in $PM_{2.5}$ concentrations could yield substantial public health benefits at the population level. This underscores the importance of continued efforts to meet current National Ambient Air Quality Standards (U.S. Environmental Protection Agency 2024).

Second, the presence of significant spillover effects highlights the need for coordinated regional approaches to air quality management rather than isolated county-level interventions. Pollution control policies implemented in a single jurisdiction can generate positive health externalities in neighboring areas through reduced atmospheric transport of pollutants. Conversely, failure to regulate emissions in one county may undermine air quality improvements in adjacent counties.

This spatial interdependence argues for multi-jurisdictional coordination mechanisms, such as regional air quality management districts or interstate compacts, that can internalize cross-boundary health effects and optimize pollution control investments across broader geographic scales (Lin et al. 2023; Ge et al. 2023). Third, the findings inform targeted intervention strategies by revealing spatial heterogeneity in both exposure and health outcomes. Counties with high local $PM_{2.5}$ concentrations should be prioritized for direct emissions controls, while counties in high-pollution regions (even if their own levels are moderate) may still face elevated cancer risks due to spillover and thus warrant attention in regional planning efforts.

Several limitations and directions for future research should be noted. While our method handles spatial confounding effectively, identifying causal effects still relies on the unconfoundedness assumption, which may be violated in many practical settings. Future work could incorporate sensitivity analyses designed explicitly for spatial settings to assess the robustness of causal estimates to unmeasured confounding. For example, we have implicitly assumed that ϕ_Y, the spatial error component in the outcome model, is independent of the spatial components of the local and spillover exposures. If some spatial confounders are not taken into account, such components would show some degree of correlation. Additionally, extending our approach to handle temporal dynamics in spatial processes would enable causal inference for spatio-temporal data, an increasingly common data structure in environmental and policy research.

In conclusion, our INLA-based spatial causal inference methodology offers a powerful and computationally efficient approach to estimating causal effects in the presence of spatial interference and confounding. By explicitly modeling spatial dependence structures while maintaining computational tractability, our method enables researchers to draw more reliable causal conclusions from spatial data, with potential applications across diverse fields including environmental epidemiology, economics, and policy evaluation.

Acknowledgments

Open access publishing facilitated by Universita degli Studi di Palermo, as part of the Wiley - CRUI-CARE agreement.

Funding

The work of Chiara Di Maria was supported by PRIN 2022: Spatio-temporal Functional Marked Point Processes for probabilistic forecasting of earthquakes. P.I.Giada Adelfio, CUP B53C24006340006.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available in the National Cancer Institute at https://www.statecancerprofiles.cancer.gov/. These data were derived from the following resources available in the public domain: National Cancer Institute, https://www.statecancerprofiles.cancer.gov/.

References

Acharjee, M. K. 2018. *Air Quality and Lung Cancer: Analysis via Local Control*. Lamar University-Beaumont.

Akbari, K., S. Winter, and M. Tomko. 2023. "Spatial Causality: A Systematic Review on Spatial Causal Inference." *Geographical Analysis* 55: 56–89.

Aronow, P. M., and C. Samii. 2017. "Estimating Average Causal Effects Under General Interference, With Application to a Social Network Experiment." *Annals of Applied Statistics* 11: 1912–1947.

Bhattacharya, R., D. Malinsky, and I. Shpitser. 2020. "Causal Inference Under Interference and Network Uncertainty." In *Uncertainty in Artificial Intelligence*, 1028–1038. PMLR.

Christiansen, R., M. Baumann, T. Kuemmerle, M. D. Mahecha, and J. Peters. 2022. "Toward Causal Inference for Spatio-Temporal Data: Conflict and Forest Loss in Colombia." *Journal of the American Statistical Association* 117: 591–601.

Ciabattini, M., E. Rizzello, F. Lucaroni, L. Palombi, and P. Boffetta. 2021. "Systematic Review and Meta-Analysis of Recent High-Quality Studies on Exposure to Particulate Matter and Risk of Lung Cancer." *Environmental Research* 196: 110440.

Cressie, N. A. C. 1993. *Statistics for Spatial Data*, Wiley Series in Probability and Statistics. John Wiley & Sons, Inc. https://doi.org/10.1002/9781119115151.

Datta, A., S. Banerjee, J. S. Hodges, and L. Gao. 2019. "Spatial Disease Mapping Using Directed Acyclic Graph Auto-Regressive (Dagar) Models." *Bayesian Analysis* 14: 1221–1244. https://doi.org/10.1214/19-BA1177.

Davis, M. L., B. Neelon, P. J. Nietert, et al. 2017. "Addressing Geographic Confounding Through Spatial Propensity Scores: A Study of Racial Disparities in Diabetes." *Statistical Methods in Medical Research* 28: 734–748.

Davis, M. L., B. Neelon, P. J. Nietert, et al. 2021. "Propensity Score Matching for Multilevel Spatial Data: Accounting for Geographic Confounding in Health Disparity Studies." *International Journal of Health Geographics* 20: 10.

Gao, B., J. Wang, A. Stein, and Z. Chen. 2022. "Causal Inference in Spatial Statistics." *Spatial Statistics* 50: 100621.

Ge, T., X. Chen, Y. Geng, and K. Yang. 2023. "Does Regional Collaborative Governance Reduce Air Pollution? Quasi-Experimental Evidence From China." *Journal of Cleaner Production* 419: 138283.

Giffin, A., B. Reich, S. Yang, and A. Rappold. 2023. "Generalized Propensity Score Approach to Causal Inference With Spatial Interference." *Biometrics* 79: 2220–2231.

Hodges, J. S., and B. J. Reich. 2010. "Adding Spatially-Correlated Errors Can Mess Up the Fixed Effect You Love." *American Statistician* 64: 325–334.

Huber, M., and A. Steinmayr. 2021. "A Framework for Separating Individual-Level Treatment Effects From Spillover Effects." *Journal of Business & Economic Statistics* 39: 422–436.

Hudgens, M. G., and M. E. Halloran. 2008. "Toward Causal Inference With Interference." *Journal of the American Statistical Association* 103: 832–842.

Liao, W.-B., K. Ju, Q. Zhou, Y.-M. Gao, and J. Pan. 2020. "Forecasting Pm 2.5-Induced Lung Cancer Mortality and Morbidity at County Level in China Using Satellite-Derived Pm 2.5 Data From 1998 to 2016: A Modeling Study." *Environmental Science and Pollution Research* 27: 22946–22955.

Lin, W., K. Lin, L. Du, and J. Du. 2023. "Can Regional Joint Prevention and Control of Atmospheric Reduce Border Pollution? Evidence From China's 12th Five-Year Plan on Air Pollution." *Journal of Environmental Management* 342: 118342.

Lindgren, F., H. Rue, and J. Lindström. 2011. "An Explicit Link Between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach." *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 73: 423–498.

Paciorek, C. J. 2010. "The Importance of Scale for Spatial-Confounding Bias and Precision of Spatial Regression Estimators." *Statistical Science* 25: 107–125.

Papadogeorgou, G., C. Choirat, and C. M. Zigler. 2019. "Adjusting for Unmeasured Spatial Confounding With Distance Adjusted Propensity Score Matching." *Biostatistics* 20: 256–272.

Pearl, J. 2009. *Causality*. Cambridge University Press.

Reich, B. J., S. Yang, Y. Guan, A. B. Giffin, M. J. Miller, and A. G. Rappold. 2021. "A Review of Spatial Causal Inference Methods for Environmental and Epidemiological Applications." *International Statistical Review* 89: 605–634.

Rosenbaum, P. R. 2023. "Propensity Score." In *Handbook of Matching and Weighting Adjustments for Causal Inference*, vol. 18, 1st ed. Chapman and Hall/CRC.

Rubin, D. B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66: 688–701.

Rue, H., and L. Held. 2005. *Gaussian Markov Random Fields: Theory and Applications*. 1st ed. Chapman and Hall/CRC. https://doi.org/10.1201/9780203492024.

Rue, H., S. Martino, and N. Chopin. 2009. "Approximate Bayesian Inference for Latent Gaussian Models by Using Integrated Nested Laplace Approximations." *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 71: 319–392.

Schnell, P. M., and G. Papadogeorgou. 2020. "Mitigating Unobserved Spatial Confounding When Estimating the Effect of Supermarket Access on Cardiovascular Disease Deaths." *Annals of Applied Statistics* 14: 2069–2095.

Sinclair, B., M. McConnell, and D. P. Green. 2012. "Detecting Spillover Effects: Design and Analysis of Multilevel Experiments." *American Journal of Political Science* 56: 1055–1069.

Smith, M. J., M. E. Charlton, and J. J. Oleson. 2023. "Causal Decomposition Maps: An Exploratory Tool for Designing Area-Level Interventions Aimed at Reducing Health Disparities." *Biometrical Journal* 65: 2200213.

Thaden, H., and T. Kneib. 2018. "Structural Equation Models for Dealing With Spatial Confounding." *American Statistician* 72: 239–252.

Thoemmes, F. J., and E. S. Kim. 2011. "A Systematic Review of Propensity Score Methods in the Social Sciences." *Multivariate Behavioral Research* 46: 90–118. https://doi.org/10.1080/00273171.2011.540475.

U.S. Environmental Protection Agency. 2024. *Final Reconsideration of the National Ambient Air Quality Standards for Particulate Matter (PM). Final Rule*. U.S. Environmental Protection Agency. https://www.epa.gov/pm-pollution/final-reconsideration-national-ambient-air-quality-standards-particulate-matter-pm.

VanderWeele, T. J., and P. Ding. 2017. "Sensitivity Analysis in Observational Research: Introducing the e-Value." *Annals of Internal Medicine* 167: 268–274.

VanderWeele, T. J., and E. J. T. Tchetgen. 2011. "Effect Partitioning Under Interference in Two-Stage Randomized Vaccine Trials." *Statistics & Probability Letters* 81: 861–869.

Wang, T.-H., K.-Y. Huang, C.-C. Chen, et al. 2023. "Pm2. 5 Promotes Lung Cancer Progression Through Activation of the Ahr-tmprss2-il18 Pathway." *EMBO Molecular Medicine* 15: e17014.

Zigler, C. M., and G. Papadogeorgou. 2020. "Bipartite Causal Inference With Interference." *Statistical Science: A Review Journal of the Institute of Mathematical Statistics* 36: 109.

Appendix A

Additional Simulation Results

In addition to the simulation scheme described in Section 4, we ran another simulation study using the same setting described in Giffin et al. (2023), with a single confounding variable. Results are shown in Table A1.

TABLE A1 | Simulation bias multiplied by 1000 with standard errors (Giffin et al. 2023 setting).

$h(Z_s)$	Model	δ_1	δ_2	τ
W_s	JPS	32.10 (24.27)	10.34 (7.76)	0.00 (0.00)
	Naive	238.62 (43.67)	22.08 (15.94)	100.00 (0.00)
$-0.5(W_s)^3$	JPS	32.18 (24.41)	10.34 (8.17)	0.00 (0.00)
	Naive	382.64 (64.33)	126.59 (52.38)	111.60 (32.67)
$\exp(W_s)$	JPS	32.12 (24.08)	9.10 (6.80)	0.00 (0.00)
	Naive	202.42 (57.00)	220.44 (96.01)	37.00 (48.33)