CAUSAL EFFECT UNDER SPATIAL INTERFERENCE

A GENERALISED PROPENSITY SCORE WEIGHTED CONFORMAL APPROACH

Jing Zhang
School of Geographical Sciences
University of Bristol
Bristol, UK
jing.zhang.2021@bristol.ac.uk

ABSTRACT

Estimating causal effect under spatial interference is an important topic with relevance for both policy decisions and geographic research. Causal interference refers to the existence of dependency of a given unit's outcome on the treatments of other units. If unadjusted, interference leads to biased causal effect estimates. In this paper, the focus is on interference from the spillover of treatment. This is a type of interference structure suitable for representing a wide range of real world scenarios. For this type of spatial interference, existing methods are capable of estimating average causal effects but not individual level causal effects. This is a significant limitation, for example for cases involving areal data where we may be interested in how an intervention affects each geographic area. This paper proposes a new method for estimating individual causal effects under spatial interference. This is achieved with a generalized propensity score weighted conformal prediction approach. The proposed method is distribution-free. It is well suited for scenarios of spatial causal interference where the underlying causal process may be non-Gaussian or non-linear. Simulation experiments show that the method can achieve target performance in a range of test scenarios, including under unmeasured spatial confounding.

CCS CONCEPTS

• Applied computing → Physical sciences and engineering; • Earth and atmospheric sciences → Environmental sciences.

KEYWORDS

Causal inference, counterfactual prediction, spatial interference, generalized propensity score, conformal prediction

ACM Reference format:

Jing Zhang. 2023. Causal effect under spatial interference: A generalized propensity score weighted conformal approach. *In Proceedings of The 31th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '23), November, 2023, ACM, Hamburg, Germany, 10 pages.*

1 Introduction

1.1 Background and research objective

In the context of causal inference, 'interference' refers to the dependency of an observational unit's outcome on the treatments of other units. In Rubin's [1] original articulation of the Potential Outcome (PO) framework, no interference is one of the basic assumptions. No interference is widely known as one component of the Stable Unit Treatment Value Assumption (SUTVA), unadjusted violation of which leads to biased causal effect estimates.

Spatial interference is, broadly speaking, scenarios of causal interference due to spatial interaction among the units. Spatial interference is common in observational studies and also in randomised controlled experiments. Some real world examples are: air pollution treatments where pollution travels to non-treated areas; neighbourhood policing where policing in one area may affect crimes in others; vaccination where an increased vaccination rate could protect the unvaccinated, and survey experiments where the treatment information may propagate through participants' social network. In these cases, SUTVA cannot be sufficiently or efficiently guaranteed even through standard randomisation designs, because fundamentally the underlying causal mechanisms cannot be isolated to the individual units. And therefore, statistical adjustments are needed.

Since interference describes a statistical dependency, it can take many forms and arise from multiple mechanisms. For example, a statistical dependency between one unit's outcome and its neighbouring units' treatments may arise from spatial mechanisms such as: (1) unmeasured spatial confounding; (2) mediation by a spatial factor; (3) spillover of treatment among units. Under these scenarios, the statistical dependence manifested may present as spatial interference. Scenarios (2) and (3) are non-trivial, in the

sense that, the interference may point to factors of substantial significance to the causal relation under study. This paper focuses on spatial interference due to the spillover of treatments. The aim is to extend existing methods for estimating causal effects under spatial interference. A generalized propensity score (GPS) weighted conformal prediction approach is developed for estimating individual treatment effects (ITE). The proposed method is flexible. It is well suited for scenarios of spatial causal interference where the underlying causal process may be non-Gaussian, non-stationary, or non-linear.

1.2 Relation to prior work

The paper is closely related to several strands of literature. Section two will introduce them in more detail. Here is a brief sketch of how the paper relates to prior work:

The topic of spatial interference sits within a wider range of literature on causal interference. To define and estimate causal effects, this paper follows the PO framework with the assumption of conditional ignorability; the generalised propensity score is adopted as a component in the estimation procedure. These are guided by theoretical work of Manski [2], Aronow and Samii [3] and Forastiere et.al. [4]. In the causal inference literature, the most well studied estimand is the average treatment effect (ATE), also the conditional average treatment effect (CATE). This paper focuses on the ITE however, because it better accounts for treatment effect heterogeneity among individuals and response variability within individuals. Under spatial interference, the ITE could be more informative than the ATE or the CATE. The ITE could potentially support policy decision making where counterfactuals and causal effects in a few locations are of interest.

The focus on ITE also connects this paper to the literature of uncertainty quantification. Unlike ATE, the ITE aims to capture the variability in individual responses to interventions. This requires methods capable of uncertainty quantification on the point basis instead of at the population level. A range of techniques have been devised to this end (e.g. Wager & Athey, [5]used tree-based models). In this paper, ITE predictors are further wrapped within a conformal prediction framework. The conformal method is non-parametric. It wraps around existing prediction models and calibrate their point predictions into prediction sets with guaranteed finite sample coverage probability. Closely related to this paper are recent work on weighted conformal prediction and its application on ITE by Tibshirani et.al. [6], Lei et.al. [7], Jin et.al.[8].

Overall, the paper leverages conformal prediction and GPS to estimate causal effects under spatial interference. It makes an original contribution by integrating techniques from the spatial interference literature with the conformal approach and reconfiguring the prediction procedure. There are some limitations to the proposed procedure and promising directions for future work. (1) The measurement of coverage adopted here is average coverage. It is an indicator of average performance across all test points. Alternatively, conditional coverage can be used. Conditional coverage is measured locally conditional on covariate values, which better guarantees pointwise model performance. (2) Further efforts are needed to define and quantify prediction uncertainty related to the uncertainty of the spatial interference structure.

2 Literature

This section introduces literature on a few topics closely related to the proposed method. First, the definitions of causal effects under interference, the key estimation assumptions, and then the weighted conformal prediction technique.

2.1 Causal effects under interference

In the causal interference literature, most work set population level average effects as the inference target. Under the PO framework, causal effects under interference are typically defined in this set up: For individuals $i = 1, \ldots, n$, a binary treatment Z_i is randomly assigned. The treatment status of $i's$ neighbours is referred to as the treatment program G_i. The presence of interference means that the potential outcome of unit i depends on not only its own treatment Z_i but also its neighbours' treatment program G_i. If each combination of neighbour treatment status is considered a different treatment level, then there could be as many as 2^n potential outcomes for unit i. To keep the problem tractable, the neighbourhood treatment program is mapped onto a lower dimension $f(G_i)$ by a chosen function (e.g. mapped as a binary treatment). The conditional direct effect can then be stated as: $E[Y_i(Z_i = 1; f(G_i = g)) - Y_i(Z_i = 0; f(G_i = g))$ the conditional expectation of the difference in outcome with and without treatment Z [9, 10]. If the magnitude of direct effect does not depend on neighbourhood treatment *f(G),* then the conditional direct effect is equal to the average direct effect. More generally, the average direct effect can be obtained by marginalizing out *f(G)* [11]. Similarly, the indirect effect corresponding to the spillover treatment can be expressed as: $E[Y_i(Z_i = 0; f(G_i = g)) - Y_i(Z_i = 0; f(G_i = 0))]$ for the spillover effect unit *i* receives or $\sum_{j \neq i} E[Y_j(Z_i = 1; f(G_i = g)) - Y_j(Z_i = 0; f(G_i = g))]$ for the spillover effect unit *i* exerts on all other units[12].

2.2 Spatial interference and estimation

Depending on the extent of dependence, causal interference is categorized into partial interference and global interference (also known as network interference) [11]. Spatial interference may be either partial or global. Partial interference is characterized by disjoint observation groups within which interference is contained. It is the relatively well studied type of interference (e.g. [9, 13] [14]). The clustering could potentially be leveraged as an

analogy to random assignment (e.g. [15] [16] [17]), which benefits average effect estimation. For spatial interference, this applies when observations are naturally clustered into disconnected locations. Global interference does not place such restrictions on the interference structure; dependence can exist between arbitrary unit pairs. In spatial interference, global interference may apply to areal studies or observations sampled from a continuous spatial field (e.g. [18]). This study works under the conceptual framework of global interference.

The identifiability of causal effects under interference has been thoroughly discussed, among others, by Manski [2], Sussman and Airoldi [19]. Most of the principles apply for global interference. Short of fieldwork-based exposure mapping to obtain true exposure levels, the estimation relies on strong restriction assumptions about the structure of causal interaction. One consideration is the potential endogeneity in outcome and treatment levels. If such endogeneity exists, causal effect is generally unidentifiable. If we believe a more restraint form of spatial interaction can plausibly represent the causal process, then we can resume working within the PO framework. Apart from interaction restrictions, two basic assumptions inherited from the PO framework are also necessary for estimating the causal effect: the (conditional) ignorability assumption and the modified SUTVA [1] [3]. The two assumptions guarantee no unmeasured confounding and no unmeasured interference respectively.

Currently some of the most promising estimation frameworks are based on GPS (e.g. [20] [21] [22]). The GPS is a generalization from the classic binary treatment propensity score to accommodate multiple treatment levels. In spatial interference settings, the multiple treatment levels arise from the joint treatment of direct and neighborhood exposures. This paper adopts the GPS formula developed by Forastiere [4]. It defines a unit's probability of receiving a joint treatment *(Z=z , G=g)* as a function of the unit's self and neighbourhood covariates. Besides individual level covariates, network properties can also be included. Recent developments based on the Forastiere framework have further explored continuous exposures [23] and the incorporation of physical distance [24]. In this study, the GPS will be a component of the conformal model. In essence, the GPS will be used as a distribution balancing score.

2.3 Weighted conformal prediction

Conformal prediction is a simple way to generate prediction sets. The method was pioneered by Vovk [25, 26] to establish statistical guarantees for learning algorithms. It can wrap around arbitrary prediction models to form instance-wise error bounds. The full conformal procedure requires shuffling of data points and is computationally expensive. Alternatively, there has been major improvements on efficiency, including the split conformal (also known as the inductive conformal) [25], CV-conformal [27], jack-knife-conformal technique [28] [29], and conformalized quantile regression [30]. Succinctly put, the split conformal prediction follows such a procedure: Split data into training and calibration sets. A model is trained on the training set. It is then applied to predict labels in the calibration set. A nonconformity score *V* (typically the absolute residual) is calculated from the calibration set true and predicted labels. To guarantee coverage with error rate a, the *1-a* quantile of *V* is calculated, noted as *Vhat*. The model then makes an initial estimate on the test point. And the initial estimate adjusted by *Vhat* forms the output prediction interval. More details see Shafer and Vovk [31] .

One notable issue for classic conformal prediction is potential distribution shift between the training and target data distributions. To work around this issue, a weighted conformal approach has been developed over a series of work by Shimodaira [32], Tibshirani [6], Qiu [33]. Using the likelihood ratio between the target and training data distributions as a balancing weight *w(x,y)* to characterise the distribution shift, the weighted conformal produces a prediction interval with guaranteed coverage and robustness against the distribution shift.

Lei et al. [7] applied the weighted conformal techniques for counterfactual prediction and individual effect estimation. In a counterfactual prediction setting where the goal is to infer the potential outcome of control units had they been assigned the treatment (binary treatment, assuming strong ignorability and SUTVA). The training distribution $\boldsymbol{P}_{\boldsymbol{X},\boldsymbol{Y}(\boldsymbol{T}=\boldsymbol{1})|\boldsymbol{T}=\boldsymbol{1}}$ is from the observed treated population. The target distribution is from the untreated population, and it is an unobserved counterfactual joint distribution $\boldsymbol{P}_{\boldsymbol{X},\boldsymbol{Y}(\boldsymbol{T}=\boldsymbol{1})|\boldsymbol{T}=\boldsymbol{0}}$. Intuitively, the shift between the two distributions reflects the unbalance between treatment and control groups. In this case, as demonstrated by Lei et al., the balancing weights *w(x,y)* can be simplified as the inverse treatment propensity. And therefore, the inverse propensity scores can be used as weights in a standard weighted conformal prediction procedure to obtain counterfactual predictions. More recently, Jin et al. [8] studied how the conformal ITE method responds to potential violations of ignorability.

So far weighted conformal prediction has not been applied for causal inference under interference. The conformal prediction technique has some desirable properties when applied for causal inference. It is distribution-free and does not require specification of a statistical model. Therefore, it can avoid common types of misspecification bias such as due to erroneously assuming linear effect or the Gaussian distribution of data. By exploring this topic, this study makes an original contribution to the causal inference literature. This study also contribute to the conformal prediction

[1] Note that the modified SUTVA is stronger than the no endogeneity assumption. By restricting the neighborhood structure, it has already excluded endogeneity in treatment levels.

techniques by relaxing the SUTVA assumption underlying the current application of the conformal ITE procedure.

3 Methods

The main innovation of this paper is extending the weighted conformal prediction approach to causal interference conditions. This is realised via a new configuration of distribution balancing weights and restructuring the prediction procedure for four potential outcomes. The propensity score estimation adopts the Forastiere [4] approach introduced earlier in 2.1.2. This section mainly presents the adaptation of the weighted conformal procedure.

3.1 Estimands for individual effects

In this study, the target is individual direct and spillover effects. So far, there is no convention on the causal estimands in the literature. To start with, let's define the estimands and clarify necessary assumptions for estimation. Under interference, the total exposure T is a function of the unit's direct treatment Z_i and neighbourhood treatment program G_i. For simplicity, let's start with four levels of joint exposure, $T^{11} = (Z_i = 1, G_i = 1), T^{10} = (Z_i = 1, G_i = 0), T^{01} = (Z_i = 0, G_i = 1), T^{00} = (Z_i = 0, G_i = 0)$. A unit has four potential outcomes $(Y^{11}, Y^{10}, Y^{01}, Y^{00})$ corresponding to these treatment levels, at most one of which is observed. The estimand for individual direct effect τ_{ITE_z} is the difference between the unit's outcome under direct treatment levels 1 and 0, conditional on the neighbourhood treatment:

$$\tau_{ITE_z}(x) = Y_i^{1g}(x) - Y_i^{0g}(x) \quad (1)$$

Similarly, the estimand for individual spillover effect τ_{ITE_g} is the difference between the unit's outcome under neighbourhood treatment levels 1 and 0, conditional on the direct treatment:

$$\tau_{ITE_g}(x) = Y_i^{z1}(x) - Y_i^{z0}(x) \quad (2)$$

Under the assumption that the effects from Z and G are additive, the total effect is the sum of individual direct and spillover effects:

$$\tau_{ITE}(x) = \tau_{ITE_z}(x) + \tau_{ITE_g}(x) = Y_i^{11}(x) - Y_i^{00}(x) \quad (3)$$

The total effect is related to and yet distinct from CATE. The latter can be obtained from the former by taking expectation over the distribution of Z and G in the subpopulation strata. Existing methods for estimating CATE can produce unbiased mean. But unbiased variance estimator for CATE is still an open question. And our method tangentially contributes to solving this problem.

The estimation of individual effects relies on conditional ignorability under bivariate treatment. That is, a unit's joint treatment levels are independent of its outcomes conditional on the confounders:

$$(Y^{11}, Y^{10}, Y^{01}, Y^{00}) \perp (Z, G) | (X^z, X^g) \quad (4)$$

X^z, X^g represents a unit's individual and neighborhood covariates including all confounders. The ignorability condition implies no unmeasured confounding. Conditional on the confounders, the joint treatment levels are as if randomly assigned. A modified version of SUTVA is also assumed, namely there is no unmeasured interference beyond what has already been captured in the neighborhood treatments [3]. Violation of either of these assumptions leads to bias in the conditional mean estimation of the counterfactual outcomes, which could compromise the accuracy of the prediction intervals.

3.2 Outcome prediction

Since the estimands are based on four potential outcomes, the existing conformal prediction framework needs to be adapted. For a given counterfactual treatment level T_i, an outcome model is trained on data points with observed outcomes under T_i. There will be as many outcome models as treatment levels. It is also straight forward to accommodate more than four treatment levels, as long as the training data is of sufficient size.

Another issue to consider when configuring the conformal prediction pipeline is the choice of outcome predictor. In the spatial ITE application, we have to be aware of the predictor's performance under possibly spatially dependent training data. Take the example of spatial autocorrelation between covariates: Due to information sharing between adjacent observations, the effective sample size of the training data is smaller than the total number of observations [34]. For the trained outcome prediction models, this leads to an underestimated variance when making future predictions. As a result, the conformity scores (in our case the absolute residuals) from spatial data will be biased upwards compared with non-spatial data, and the length of conformalized prediction intervals could be biased upwards. This means that the proposed method may tend to produce conservative results with spatial data. While spatial prediction model is besides the scope of this paper, some useful references can be found from the recent review paper by Reich et.al. [35].

3.3 Conformal weights under interference

This study follow Forastiere's [4] formulation of GPS to derive distribution balancing weights for the conformal prediction. For the four levels of exposure, the joint treatment propensities are denoted as $e_{11}(x) = P(Z = 1, G = 1|X = x)$, $e_{10}(x) = P(Z = 1, G = 0|X = x)$, $e_{01}(x) = P(Z = 0, G = 1|X = x)$, $e_{00}(x) = P(Z = 0, G = 0|X = x)$. Sufficient overlap between different treatment groups' covariate distributions is assumed.

In a setting with causal interference, the weights can be derived in the following way: Suppose the goal is to infer counterfactual outcome Y^{11} for a unit with observed Y^{10}. An outcome model is

trained on observations from units with exposure $T = 11$. This produces a model $f_{11}(.) = Y^{11}|X^{11}$. The model $f_{11}(.)$ is applied on the test point to make a prediction for the counterfactual Y^{11}, that is $Y^{11}hat = f_{11}(X = Xtest)$. The training distribution is $P_{X^{11},Y^{11}}$ and the target distribution is $P_{X^{10},Y^{11}}$. To account for the distribution drift from X^{11} to X^{10} , a balancing weight is calculated based on the treatment propensities associated with X^{11} and X^{10}. The weight is:

$$wt_{11}(x,y) = \frac{dP_{target}}{dP_{train}}(x,y) = \frac{dP_{X^{10},Y^{11}}}{dP_{X^{11},Y^{11}}}(x,y) \quad (5)$$

Assuming invariant causal process and no confounding, the numerator $P_{X^{10},Y^{11}} = P_{Y^{11}|X^{10}} * P_{X^{10}} = P_{Y^{11}|X^{11}} * P_{X^{10}}$. The denominator $P_{X^{11},Y^{11}} = P_{Y^{11}|X^{11}} * P_{X^{11}}$. The weight simplifies to:

$$wt_{11}(x) = \frac{dP_{X^{10}}}{dP_{X^{11}}}(x) \propto \frac{p(T=10|X=x)}{p(T=11|X=x)} = \frac{e_{10}(x)}{e_{11}(x)} \quad (6)$$

More generally, for a unit with observed exposure level T_{obs}, the weight to estimate counterfactual outcome under exposure level T_{target} is:

$$wt_{target}(x) = \frac{e_{T_{obs}}(x)}{e_{T_{target}}(x)} \quad (7)$$

The generalized propensity score based conformal weights are compatible with standard propensity score weights and an intuitive understanding. When the goal is prediction (e.g. interpolation) within the same treatment level, the weight simplifies to 1 and no weighting is needed.

3.4 The estimation procedure

With the adapted procedure, the estimation process can be summarised as follows:

Input: Dataset consisting of unit level covariates, observed outcomes and direct treatments $D\{X_i, Y_i^{obs}, Z_i\}$. A target error rate a.

Step 1: For each unit, construct neighbourhood covariates X^g, neighbourhood exposure $G_i \in \{0,1\}$, joint treatment $T_i \in \{00, 01, 10, 11\}$. Use them to compute the joint propensity scores $e_{11}(x_i), e_{10}(x_i), e_{01}(x_i), e_{00}(x_i)$.

Step 2: Partition data into training and calibration sets. Train outcome model $f(.)$ on training set. Apply $f(.)$ on calibration set X_{cal} to obtain non-conformity scores V_{cal}. Apply $f(.)$ on test point X_{test} to obtain initial prediction interval for the counterfactual outcome $[Yhat_{low}, Yhat_{up}]$.

Step 3: For each point in the calibration set, compute the balancing weights. Use the weights to compute a weighted distribution of the non-conformity scores V_{cal_wt}. Take the 1-a quantile of V_{cal_wt} as the adjustment η .

Step 4: Construct intervals for the counterfactual outcomes $Y = [Yhat_{low} - \eta, Yhat_{up} + \eta]$

Step 5: Take the difference between observed outcome and predicted counterfactuals to output ITE estimate $C_{ITE} = [Yhat_{low} - \eta - Y_{obs}, Yhat_{up} + \eta - Y_{obs}]$.

Output: Counterfactual outcomes *Y*, ITE estimate C_{ITE}.

4 Simulation experiments

Model performance is evaluated on simulation experiments under a series of test scenarios. The main criterions of evaluation are: (1) the probability at which predicted counterfactual outcome intervals correctly cover true potential outcomes, and (2) the length of predicted ITE intervals. Ideally, the achieved coverage and interval length should closely track target coverage and true interval length. For both objectives, the proposed method performs well under all test scenarios. This section presents the test scenarios, discuss the results, and follow up with reflections on future refinements.

4.1 Experiment setup

Simulation experiments are arranged by five test scenarios. This will help us to understand the model's robustness to common estimation challenges in spatial causal inference. Besides spatial interference, the test scenarios cover non-Gaussian data, non-linear effects, spatial confounding, and spatially non-stationary effects. The spatial patterns of variables are illustrated in Figure.1. Specifics of the five test scenarios are as follows:

(1) Scenario 1: basic spatial interference setup

There are *n* observational units characterised by *k* covariates drawn from Uniform distribution $X^z \sim Unif[0,3]^k$. Each unit inhabits a random location on a unit square. Its neighbours are defined as the set of points within a certain distance band. Its neighbourhood attributes X^g are represented by the average values of its neighbours' covariates. The assignment of direct treatment Z is independently determined by a unit's attributes x_i^z based on propensity $e_z(X^z) = P(Z = 1|X^z = x_i^z)$. The neighbourhood exposure *G* is determined by a unit's neighborhood covariate levels based on propensity $e_g(X^g) = P\left(G = 1\middle|X^g = x_i^g\right)$. $e_z(X^z)$ and $e_g(X^g)$ are derived from Logit functions. In this scenario the covariates and the *Z* treatment levels are spatially randomly distributed (Figure.1. a. b.). The G levels are clustered (Figure.1.c) because nearby units have overlapping neighbourhoods and therefore correlated X^g levels. For all units, the marginal potential outcomes corresponding to direct treatment *Z* are $Y^z(\mathrm{Z}=1) = X^z * \beta^\top + \varepsilon$, $Y^z(\mathrm{Z}=0) = \varepsilon$, $\varepsilon^{i.i.d.} \sim N(0,1)$. The marginal potential outcomes corresponding to neighbourhood exposure *G* are $Y^g(\mathrm{G}=1) = 0.5 * \mathrm{Y}^z(\mathrm{Z}=1)$, $Y^g(\mathrm{G}=0) = 0$. A unit's potential outcomes are the aggregate effect of direct and indirect treatments. Each unit has four potential outcomes, one of which is observed.

(2) Scenario 2: non-linear effect

Scenario 2 inherits the setup of scenario 1, while using a non-linear function to generate the direct treatment outcomes $Y^z(Z = 1)$.

(3) Scenario 3 and 4: spatial confounding

Building on top of scenario 2, to introduce spatial confounding, parts of the X^z covariates are spatially smoothed (Figure.1.d). In Scenario 3, all spatial confounders are adjusted for. In contrast, one of the spatial confounders is left unadjusted in Scenario 4, simulating the case of unmeasured spatial confounding. Spatial confounding is a complicated topic on its own. For example, the scale of spatial smoothing matters. This simulation study does not fully address these facets of spatial confounding.

(4) Scenario 5: spatially non-stationary causal effects.

The β parameters is set to follow a spatial trend (Figure.1.g). Model performance is not guaranteed under this scenario. In practice, the prediction can still be approximately valid without adjusting for this spatial non-stationarity.

For each scenario, tests are run under different configurations of sample size n={1000,5000}, covariate dimension k={5,20}, and target counterfactual outcome coverage {0.3,0.5,0.7,0.9}. For a single run of an experiment, 100 test points are randomly selected. Based on these test points, average coverage is computed for each type of potential outcome and for all types on average. The achieved coverage is benchmarked on the target coverage. Average ITE prediction interval lengths are computed for the direct and spillover effects. The estimated interval lengths at each target coverage level are benchmarked on ground truth interval lengths. True lengths are approximated from the corresponding quantiles of the test points' true potential outcomes. Also, our method is compared with generalised random forest (GRF) [36, 37] and quantile Bayesian additive regression trees (BART) [38] on counterfactual outcome coverage. These tree-based models are popular options for ITE estimation in clinical trials.

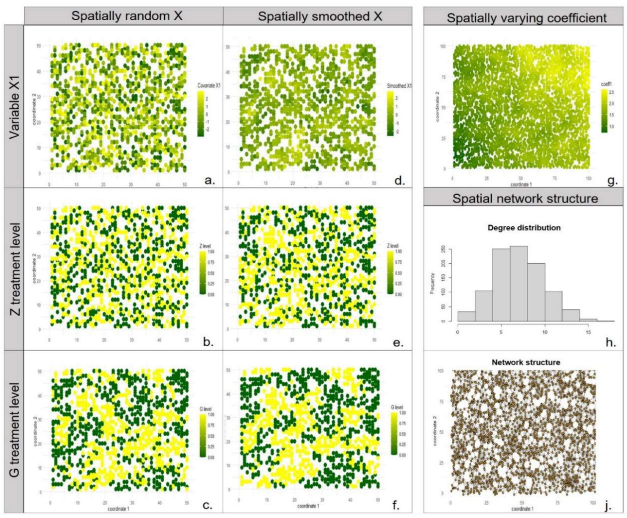


Figure.1 Illustration of simulation data spatial patterns.

4.2 Main results

Figure.2. charts the main results from the simulation experiments. Overall, the method performs well under all test scenarios. Figure. 2.1 and 2.2 are a summary of model performance. Figure.2.2 reports the lengths of ITE prediction intervals, benchmarked against approximated true lengths. (The reported ITE intervals are calculated assuming both counterfactual outcomes missing. In future applications, tighter bounds can be produced by using the predicted counterfactual outcome together with the observed outcome, if such an observation exists.) Figure.2.1. reports coverage probability averaged over all potential outcomes, benchmarked against target coverage. In most cases, the model approximately achieves optimal levels. A few insights can be drawn from the results:

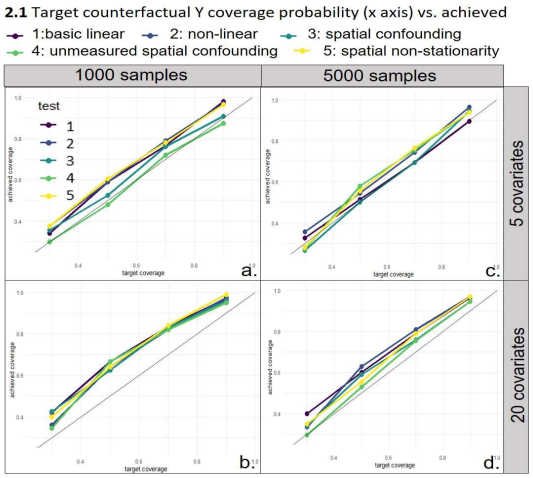
4.2.1 Regarding test scenarios

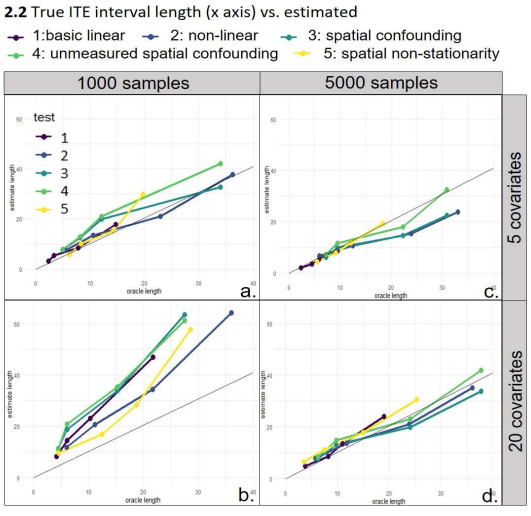
Scenarios 1 and 2 are based on spatially random non-Gaussian X variables, with no spatial confounding. The direct treatment conditions are independent between all units after adjusting for the confounders, satisfying conditional ignorability. The results demonstrate that our method can produce satisfactory counterfactual predictions and ITE estimation under spatial causal interference. Also, results from scenarios 1 and 2 tend to overshoot on coverage. This suggests that our method is conservative in variance estimation, which is a preferable property in causal inference when the topic is sensitive. Also, the results do not differ significantly between the two scenarios, suggesting that our method accommodates non-linearity well.

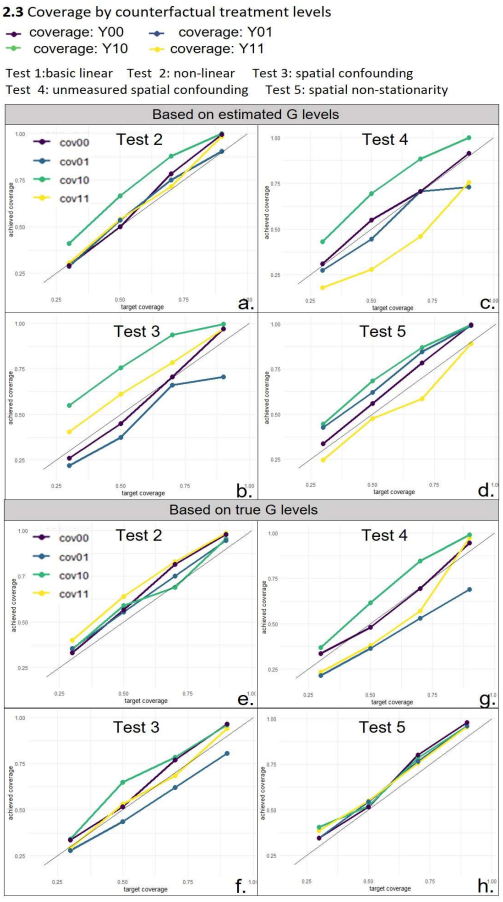
Scenarios 3 and 4 are based on spatially smoothed X variables. The spatial autocorrelation in X manifests as spatially correlated Z and G treatment levels. In theory, spatial autocorrelation in covariates should not affect model performance so long as conditional ignorability holds. Between the two scenarios, scenario 3 consistently yields higher coverage at shorter interval lengths. This reflects the influence of unmeasured spatial confounding, which has resulted in the relatively poorer performance in scenario 4. In principle, the predictions will be approximately correct under a moderate level of unmeasured spatial confounding, while prediction accuracy is negatively associated with the strength of unmeasured confounding. That is, a stronger magnitude of unmeasured confounding in scenario 4 would have led to worse results. To follow up on this, it is possible to design sensitivity analyses to quantify the extent to which unmeasured confounding biases the counterfactual prediction.

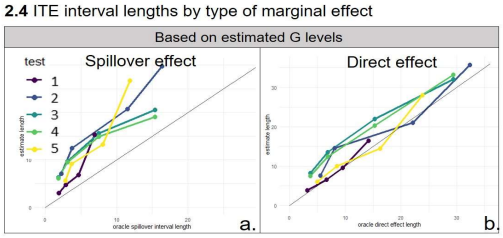
Scenario 5 simulates the case of spatially non-stationary causal process, while spatial heterogeneity is also present in scenario 3 with the combination of non-liner effect and spatially smoothed X. Scenario 5 is a case of heterogeneous causal effect where location is a mediator. The causal parameter is directly defined as a

function of location, and its spatial pattern is not encoded in any of the observed variables. In this case, to achieve unbiased inference, spatial location information should enter as an input. Otherwise, the individual estimates will suffer from bias due to the misspecification of outcome models and therefore only approximately correct. Respectively, an accurate interpretation of the results from scenario 5 would be a global average over the spatially heterogenous causal effects. For scenario 3, the spatial smoothing leads to a locally varying mean of X. And as the causal effect is a non-linear function of X, it will also present as a non-stationary function of space by extension. Relying on observations only, although our method performs well under both scenario 3 and 5, it does not distinguish between these two forms of observed non-stationarity. This is to caution that our method is designed for counterfactual prediction. It has limited powers in aiding the reasoning of causal mechanisms.









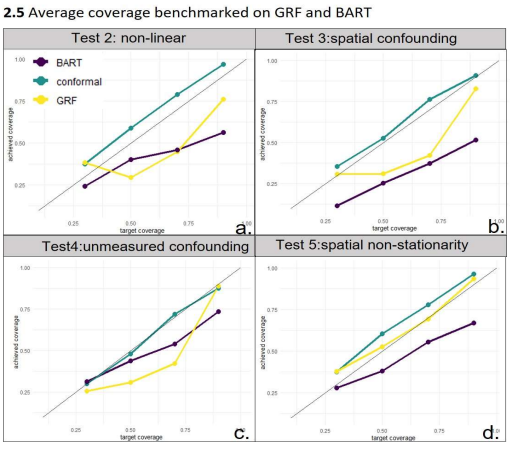


Figure.2 Main results of simulation experiments.

4.2.2 Regarding treatment levels and types of marginal effects.
As illustrated in Figure. 2.3 a-d. counterfactual outcome coverage varies depending on the treatment level of inquiry. In particular, coverage probabilities for treatment levels containing a positive spillover component (*G=1*) are lower than others. Figure. 2.3 e-h show that, holding everything else constant, the disparity in coverage across treatment levels greatly diminish when the ground truth *G* levels are supplied. Figure 2.4 a-b. show that, estimated ITE interval lengths on direct treatment effect are close to the true lengths, while the intervals on spillover effects are considerably larger than true lengths. To explain this model behaviour, a primary factor is the construction of neighbourhood treatment level *G* in the conformal pipeline. Inaccurate estimation of *G* affects downstream propensity score estimation and outcome model training. This could also be attributed to the outcome models: Theoretically, there is an uncertainty associated with exposure mapping when it comes to treatment spllovers. Recall that in our problem setup, only the direct treatment status Z is observed, while spillover treatment *G* has to be estimated based on properties of the neighbourhood. There are two untestable assumptions underlying the estimation of *G*: (1) It is assumed that the spatial structure of interference; (2) it is assumed that a simple mapping between neighbourhood properties and the spillover treatment *G*. This makes the estimated *G* levels a noisy approximation of true neighbourhood exposures. Since some observations may be labelled with the wrong *G* level, the elements of training sets for each treatment level may not be exactly faithful to ground truth. Consequently, the conditional outcome predictions could be biased. And this will affect spillover effect estimation more than it does direct effect estimation. While it is possible that using continuous *G* levels may help remedy oversimplification from assumption (2), interference network uncertainty is still a problem. This topic has only recently started to attract research attention (e.g. on network uncertainty [39] [40]; on misspecification of exposure mapping [41] [42]. Quantifying the uncertainty in spillover effect estimation due to our incomplete knowledge of causal interference network would be an interesting topic for future research.

4.2.3 Additional comments.
Besides the set of main results, some additional findings from a few additional tests: (1) Regarding outcome prediction, all results reported in Figure.2 are based on BART and the split conformal procedure. In comparison, Quantile Random Forest (QRF) more than halves the computing time but tends to produce very conservative results. (2) Model performance is stable over iterations of any single experiment. (3) Benchmarked on alternative methods (Figure.2.5), our method achieves better average coverage for counterfactual outcome predictions. (4) To verify the extent to which our results rely on accurate estimation of propensity scores, model performance based on true propensity scores, estimated propensity scores and null scores are compared. Although using true scores produce the best results, the estimated scores work reasonably well. This suggests that our method is robust to propensity score misspecification.

4.2.4 Remaining issues and future research.
A few issues remain, and this study has not been able to fully unpack them. First, causal interference is a phenomenon that could have resulted from multiple mechanisms potentially unidentifiable from observational studies. As has been stated at the beginning, the study only considers spatial interference arising from treatment spillover. This simplifies the inference problem to one of inferring the actual dosage of treatment and attributing its effect. Among other assumptions, endogeneity between neighbouring units' treatment levels is also excluded to simplify the construction of joint exposure levels. For other forms of spatial interference, for example interference due to spatial mediation, such simplifications could merit less plausibility. A related issue is network uncertainty in spatial causal interference. The uncertainty in causal interaction structures may come from several sources. For example, the links between units may be spatially and/or temporally dynamic due to the nature of social interactions; there may be heterogeneity and/or non-stationarity in the strength of the links. It is in principle possible to quantify this uncertainty, which we consider a worthy topic of further investigation. Finally, in our simulation, the points are spatially random, which satisfies the exchangeability assumption underlying conformal prediction. This may not apply for all types of spatial observations. If sample size permits, it is possible to build the estimation procedure on local sub-sampling of observations, assuming that points are locally exchangeable (e.g. [43] [44]. This would also yield more accurate estimates when the causal process is spatially dependent.

5 Summary

Causal effect under spatial interference is an important topic with relevance for both policy decisions and geographic research. Despite the well documented phenomenon of causal interference, methods to address it, especially the global interference, are still in their infancy. Interference significantly complicates the Potential Outcomes approach and requires strong assumptions about the nature and the structure of the causal interactions. This paper focuses on the spillover of treatments, which is a common type of non-trivial spatial interference. As an original contribution to existing techniques on estimating causal effects, this study proposes a generalized propensity score weighted conformal prediction approach for estimating individual treatment effects under spatial interference. The proposed method is distribution-free. It is well suited for scenarios of spatial causal interference where the underlying causal process may be non-Gaussian or non-linear. Simulation experiments show that the method can achieve target performance in a range of test scenarios, including under unmeasured spatial confounding. Overall, this study contributes to

the causal inference literature by extending counterfactual prediction based methods of estimating individual level treatment effects under global spatial interference.

ACKNOWLEDGMENTS

This project is funded by the SWDTP, ESRC UK.

REFERENCES

[1] Donald B Rubin, 1990. "Formal mode of statistical inference for causal effects," *Journal of statistical planning and inference,* vol. 25, no. 3, pp. 279-292.

[2] Charles F Manski, 2013. "Identification of treatment response with social interactions," *The Econometrics Journal,* vol. 16, no. 1, pp. S1-S23.

[3] Peter M Aronow and Cyrus Samii, 2017. "Estimating average causal effects under general interference, with application to a social network experiment," .

[4] Laura Forastiere, Edoardo M Airoldi, and Fabrizia Mealli, 2021. "Identification and estimation of treatment and interference effects in observational studies on networks," *Journal of the American Statistical Association,* vol. 116, no. 534, pp. 901-918.

[5] Stefan Wager and Susan Athey, 2018. "Estimation and inference of heterogeneous treatment effects using random forests," *Journal of the American Statistical Association,* vol. 113, no. 523, pp. 1228-1242.

[6] Ryan J Tibshirani, Rina Foygel Barber, Emmanuel Candes, and Aaditya Ramdas, 2019. "Conformal prediction under covariate shift," *Advances in neural information processing systems,* vol. 32.

[7] Lihua Lei and Emmanuel J Candès, 2021."Conformal inference of counterfactuals and individual treatment effects," *Journal of the Royal Statistical Society Series B: Statistical Methodology,* vol. 83, no. 5, pp. 911-938.

[8] Ying Jin, Zhimei Ren, and Emmanuel J Candès, 2023. "Sensitivity analysis of individual treatment effects: A robust conformal inference approach," *Proceedings of the National Academy of Sciences,* vol. 120, no. 6, p. e2214889120.

[9] Eric J Tchetgen Tchetgen and Tyler J VanderWeele, 2012. "On causal inference in the presence of interference," *Statistical methods in medical research,* vol. 21, no. 1, pp. 55-75.

[10] Tyler J VanderWeele, Eric J Tchetgen Tchetgen, and M Elizabeth Halloran, 2014. "Interference and sensitivity analysis," *Statistical science: a review journal of the Institute of Mathematical Statistics,* vol. 29, no. 4, p. 687.

[11] Michael E Sobel, 2006. "What do randomized studies of housing mobility demonstrate? Causal inference in the face of interference," *Journal of the American Statistical Association,* vol. 101, no. 476, pp. 1398-1407.

[12] Yuchen Hu, Shuangning Li, and Stefan Wager, 2022. "Average direct and indirect causal effects under interference," *Biometrika,* vol. 109, no. 4, pp. 1165-1172.

[13] Michael G Hudgens and M Elizabeth Halloran, 2008. "Toward causal inference with interference," *Journal of the American Statistical Association,* vol. 103, no. 482, pp. 832-842.

[14] Georgia Papadogeorgou, Fabrizia Mealli, and Corwin M Zigler, 2019. "Causal inference with interfering units for cluster and population level treatment allocation programs," *Biometrics,* vol. 75, no. 3, pp. 778-787.

[15] Ye Wang, Cyrus Samii, Haoge Chang, and PM Aronow, 2020. "Design-based inference for spatial experiments with interference," *arXiv preprint arXiv:2010.13599*.

[16] Sarah Baird, J Aislinn Bohren, Craig McIntosh, and Berk Özler, 2018. "Optimal design of experiments in the presence of interference," *Review of Economics and Statistics,* vol. 100, no. 5, pp. 844-860.

[17] Kosuke Imai, Zhichao Jiang, and Anup Malani, 2021. "Causal inference with interference and noncompliance in two-stage randomized experiments," *Journal of the American Statistical Association,* vol. 116, no. 534, pp. 632-644.

[18] Natalya Verbitsky-Savitz and Stephen W Raudenbush, 2012. "Causal inference under interference in spatial settings: a case study evaluating community policing program in Chicago," *Epidemiologic Methods,* vol. 1, no. 1, pp. 107-130.

[19] Daniel L Sussman and Edoardo M Airoldi, 2017. "Elements of estimation theory for causal effects in the presence of network interference," *arXiv preprint arXiv:1702.03578*.

[20] Guido W Imbens, 2000. "The role of the propensity score in estimating dose-response functions," *Biometrika,* vol. 87, no. 3, pp. 706-710.

[21] Keisuke Hirano, Guido W Imbens, and Geert Ridder, 2003. "Efficient estimation of average treatment effects using the estimated propensity score," *Econometrica,* vol. 71, no. 4, pp. 1161-1189.

[22] Kosuke Imai and David A Van Dyk, 2004. "Causal inference with general treatment regimes: Generalizing the propensity score," *Journal of the American Statistical Association,* vol. 99, no. 467, pp. 854-866.

[23] Justin R Williams and Catherine M Crespi, 2020. "Causal inference for multiple continuous exposures via the multivariate generalized propensity score," *arXiv preprint arXiv:2008.13767*.

[24] Andrew Giffin, BJ Reich, Shu Yang, and AG Rappold, 2022. "Generalized propensity score approach to causal inference with spatial interference," *Biometrics*.

[25] Vladimir Vovk, Ivan Petej, Paolo Toccaceli, Alexander Gammerman, Ernst Ahlberg, and Lars Carlsson, 2020. "Conformal calibrators," in *Conformal and probabilistic prediction and applications*, 2020: PMLR, pp. 84-99.

[26] Vladimir Vovk, Alexander Gammerman, and Glenn Shafer,2005. *Algorithmic learning in a random world*. Springer.

[27] Matteo Sesia and Emmanuel J Candès, 2020. "A comparison of some conformal quantile regression methods," *Stat,* vol. 9, no. 1, p. e261.

[28] Bradley Efron, 1992. "Jackknife-after-bootstrap standard errors and influence functions," *Journal of the Royal Statistical Society: Series B (Methodological),* vol. 54, no. 1, pp. 83-111.

[29] Rina Foygel Barber, Emmanuel J Candes, Aaditya Ramdas, and Ryan J Tibshirani, 2021. "Predictive inference with the jackknife+," .

[30] Yaniv Romano, Evan Patterson, and Emmanuel Candes, 2019. "Conformalized quantile regression," *Advances in neural information processing systems,* vol. 32.

[31] Glenn Shafer and Vladimir Vovk, 2008. "A Tutorial on Conformal Prediction," *Journal of Machine Learning Research,* vol. 9, no. 3.

[32] Hidetoshi Shimodaira, 2000. "Improving predictive inference under covariate shift by weighting the log-likelihood function," *Journal of statistical planning and inference,* vol. 90, no. 2, pp. 227-244.

[33] Hongxiang Qiu, Edgar Dobriban, and Eric Tchetgen Tchetgen, 2022. "Distribution-free prediction sets adaptive to unknown covariate shift," *arXiv preprint arXiv:2203.06126*.

[34] Daniel A Griffith, 2005. "Effective geographic sample size in the presence of spatial autocorrelation," *Annals of the Association of American Geographers,* vol. 95, no. 4, pp. 740-760.

[35] Brian J Reich, Shu Yang, Yawen Guan, Andrew B Giffin, Matthew J Miller, and Ana Rappold, 2021. "A review of spatial causal inference methods for environmental and epidemiological applications," *International Statistical Review,* vol. 89, no. 3, pp. 605-634.

[36] Nicolai Meinshausen and Greg Ridgeway, 2006. "Quantile regression forests," *Journal of machine learning research,* vol. 7, no. 6.

[37] Susan Athey, Julie Tibshirani, and Stefan Wager, 2019. "Generalized random forests,".

[38] Adam Kapelner and Justin Bleich, 2013. "bartMachine: Machine learning with Bayesian additive regression trees," *arXiv preprint arXiv:1312.2171*.

[39] Rohit Bhattacharya, Daniel Malinsky, and Ilya Shpitser, 2020. "Causal inference under interference and network uncertainty," in *Uncertainty in Artificial Intelligence*: PMLR, pp. 1028-1038.

[40] Shuangning Li and Stefan Wager, 2022. "Random graph asymptotics for treatment effect estimation under network interference," *The Annals of Statistics,* vol. 50, no. 4, pp. 2334-2358.

[41] Fredrik Sävje, Peter Aronow, and Michael Hudgens, 2021. "Average treatment effects in the presence of unknown interference," *Annals of statistics,* vol. 49, no. 2, p. 673.

[42] F Sävje, 2023. "Causal inference with misspecified exposure mappings: separating definitions and assumptions," *Biometrika,* p. 019.

[43] Rina Foygel Barber, Emmanuel J Candes, Aaditya Ramdas, and Ryan J Tibshirani, 2023. "Conformal prediction beyond exchangeability," *The Annals of Statistics,* vol. 51, no. 2, pp. 816-845.

[44] Huiying Mao, Ryan Martin, and Brian J Reich, 2022. "Valid model-free spatial prediction," *Journal of the American Statistical Association,* no. just-accepted, pp. 1-28.