Spatial Statistics 50 (2022) 100621

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

Causal inference in spatial statistics

Bingbo Gao [a], Jinfeng Wang [b,*], Alfred Stein [c], Ziyue Chen [d]

[a] *College of Land Science and Technology, China Agricultural University, 17 Tsinghua East Road, Beijing 100083, China*

[b] *State Key Laboratory of Resources and Environmental Information Systems, Institute of Geographic Sciences and Nature Resources Research, Chinese Academy of Sciences, A11 Datun Road, Beijing 100101, China*

[c] *Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Hengelosestraat 99, 7514 AE Enschede, The Netherlands*

[d] *College of Global and Earth System Sciences, Beijing Normal University, 19 Xinjiekou Street, Haidian, Beijing 100875, China*

ARTICLE INFO

Article history:
Received 25 November 2021
Received in revised form 12 January 2022
Accepted 20 January 2022
Available online 31 January 2022

Keywords:
Causal inference
Spatial statistics
Convergent-cross mapping
Granger test
Structural causal models
Perspective

ABSTRACT

Finding cause–effect relationships behind observed phenomena remains a challenge in spatial analysis. In recent years, much progress in causal inference has been made in statistics, economics, epidemiology and computer sciences, but limited progress has been made in spatial statistics due to the nonrandom, nonrepeatability and synchronism of spatial data. In this paper, we investigate the problem. We first refine the issues of causal inference, then discuss the causal inference issue in spatial statistics, next review the causal inference methods in other disciplines and analyze their potential to be used with cross-sectional data, and finally we look forward prospect of causal inference in spatial statistics.

1. Introduction

From the early history of human evolution, humans have asked 'why?', attempting to find cause–effect relationships behind observed phenomena. After understanding those, one is better equipped to conduct interventions to the physical world to reap benefits or avoid harm. Aristotle summarized four types of causes: material, formal, efficient and final causes (Falcon, 2019; Licata, 2019). Starting from a philosophical perspective, however, he did not provide scientific methods to derive those.

* Corresponding author.
E-mail address: wangjf@lreis.ac.cn (J. Wang).

https://doi.org/10.1016/j.spasta.2022.100621

Table 1
Corresponding relationships of terms for cause and effect variable.

Effect variable	Cause variable
⇕	⇕
Dependent variable	Independent variable
⇕	⇕
Response variable	Explanatory variables

Galileo Galilei opened the door to natural experiments in modern science in his famous experiment to determine whether heavier objects fall faster than lighter objects (Galilei, 1954; Gendler, 1998). By carefully designing the experiments and ensuring that all other conditions remain the same for the control and experimental groups except for the target condition, cause–effect relationships can be found from natural experiments after comparing the results of the two groups. When conditions in the system of interest cannot be all controlled, intervention experiments are still possible. Randomized controlled trials (RCTs) are the most famous intervention experiment widely adopted in medicine, biology and agriculture (Pearl, 2000; Imbens and Rubin, 2015). Although RCTs are treated as the gold standard in most causation research, it is impractical in earth system science studies where intervention experiments are usually impossible, expensive, time-consuming or even unethical (Camps-Valls et al., 2019; Runge et al., 2019). Therefore, causal inference for observational (nonexperimental) data urgently needed but is a great challenge in earth system science.

For more than a century, there have been many attempts to build a mathematical framework for causal inference from observational data. Statistical correlation is a start to this end. Although correlation does not imply causation, causation leads to correlation. Therefore, considerable efforts have been made to identify causality from correlation.Yule (1897) proposed estimating cause–effect relationships using regression analysis but is subject to criticism (Corrado and Fingleton, 2012; Gibbons and Overman, 2012). Sewall Wright (1918, 1920, 1921) is considered a pioneer for causal inference, and his path diagrams inspired the development of structural equation models (SEMs) in the 1930s (Hoover, 2017) and algorithms in computer science including causal network learning algorithms, structural causal models (SCMs), and the causal inference framework including the do-calculus and causal diagram of Pearl (2009). Granger (1969) proposed Granger causality based upon the predictability between two time series. Following this, convergent-cross mapping (CCM) was developed by Sugihara et al. (2012) for nonlinear relationships based upon the theories of dynamic systems. Rubin (1974) developed the potential outcomes framework for causal inference based on the works of Jerzey Neyman. Some of the causal inference methods seem to be applicable to spatial data.

Because the terms for cause and effect variable are diverse in different disciplines, and it is not proper to unify them due to their different strength in implying causation, some different terms are used in certain context in this study, and their corresponding relationships are listed Table 1.

2. Causal inference in spatial statistics

2.1. Related techniques in spatial statistics

In the presence of spatial autocorrelation, spatial statistical models aim to utilize location and distance to characterize the spatial pattern, gain accurate estimation or prediction, and build robust correlation relationships among variables. To date, numerous statistical methods have been proposed for spatial point process data, geostatistical data, and lattice data.

To measure the global spatial autocorrelation, indicators such as Moran's I, Geary's c, and the General G were developed (Moran, 1950; Geary, 1954; Getis and Ord, 1992; Fischer and Getis, 2010). Most take the general form of Formula (1) (Cliff and Ord, 1981):

$$\Gamma = \sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}Y_{ij} \tag{1}$$

where i and j are labels of spatial units, W is a spatial weight matrix that reflects the neighborhood between different spatial units based on their locations or distances, and Y is a matrix reflecting the relationship of values observed at the spatial units. In geostatistical studies, the semivariogram measures global spatial autocorrelation (Matheron, 1963; Isaaks and Srivastava, 1989).

$$\gamma(h) = \mathrm{E}\left\{[Y(s) - Y(s+h)]^2\right\} \tag{2}$$

where $\gamma(h)$ is the semivariogram, s and s+h are a pair of points separated at a distance $h$and$Y(s)$ and $Y(s+h)$ are observation values of the target variable. In addition to global measures, local spatial heterogeneity is measured by removing the outer summation in Eq. (1) (Cliff and Ord, 1981; Getis and Ord, 1992; Anselin 1995):

$$\Gamma_i = \sum_{j=1}^{n} W_{ij} Y_{ij} \tag{3}$$

more statistics were developed to test spatial point process (Ripley, 1976).

Spatially stratified heterogeneity (SSH) refers to the phenomena that the within strata is more similar than the between strata, e.g. landuse types and climate zones. Geographical detector is proposed to measure and attribute SSH (Wang et al., 2010, 2016; www.geodetector.cn) By dividing the study area into several homogenous parts (strata), the geographical detector q statistic is defined in (4).

$$q = 1 - \frac{\sum_{h=1}^{L} N_h \delta_h^2}{N\delta^2} \tag{4}$$

where L is the number of homogenous strata divided, N is the number of units in the study area, N_h is the number of units in stratum h, δ_h^2 is the variance in the target variable within strata h, and δ^2 is the global variance over the whole study area.

To process and utilize spatial autocorrelation, spatial regression models include spatial dependence on nearby observations and have a general form (5):

$$\begin{aligned} Y &= \rho WY + X\beta + WX\theta + u \\ u &= \lambda Wu + \epsilon \end{aligned} \tag{5}$$

where Y is the dependent variable, X denotes the independent variables, W is the spatial weight matrix, residual u has spatial effect, ϵ is the error, and ρ, β, θ and λ are coefficients. 0 or not of the coefficients lead to the specifications of spatial regression. To deal with spatial heterogeneity, Geographically weighted regression builds local models and uses local spatial autocorrelation (6) (Fotheringham et al., 2003).

$$y_i = \beta_{i0} + \sum_{k=1}^{p-1} \beta_{ik} x_{ik} + \epsilon_i \tag{6}$$

where y_i is the dependent variable at location i, x_{ik} is the kth independent variable at location i and β_{ik} is its coefficient, ϵ_i is the error at location i.

None of the models mentioned above state that they can infer causality. Bayesian hierarchical structural models can also be used to achieve spatial pattern analysis, spatial estimation/prediction and relationship modeling (Waller, 2005). They have the advantage of modeling the spatial process, with the potential to realize spatial causal inference.

2.2. Causal inference in spatial statistics

2.2.1. Causal inference studies with spatial statistics

Although few spatial statistics methods claim that they are for causal inference, many are extensively used to identify and measure cause–effect relationships. Instead of using causation or causality, those studies use terms such as "influence", "impact", "effect", "lead to", "driver", "promote" and "determinant". For example, Chi (2010) measured the direct and indirect effects of highway expansion on the population change in urban and rural areas by spatial regression, Chun and Guldmann (2014) identified the urban heat island determinants with spatial regression, Yu et al. (2007) analyzed the influencing factors of Milwaukee housing prices through spatial regression and geographically weighted regression, and Vaneckova et al. (2010) measured the effect of physical and sociodemographic risk factors on mortality in the elderly population in Sydney.

2.2.2. Key issues of causal inference in spatial statistics

When applying spatial regression for causation inference, spatial autocorrelation is taken into account and considered as spatial interaction, in addition to the relationship between explanatory and dependent variables. Like nonspatial regression, spatial regression fails to infer the influence of explanatory variables when unmeasured common causes exist either between explanatory and dependent variables or between explanatory variables. Such spatial confounding, or multicollinearity, has also been noticed in spatial statistics (Hanks et al., 2015) and may result in biased estimation of the effects of explanatory variables (Hanks et al., 2015). Restricted spatial regression, which forces the spatial random effects to be independent of the fixed effects of explanatory variables, was proposed to alleviate spatial confounding (Hughes and Haran, 2013). However, without a clear cause–effect relationship among the explanatory and dependent variables, it is difficult to avoid confounding. Yet, spillover is frequently used to explain the spatial autocorrelation in spatial regression (Golgher and Voss, 2016). Spillover is often explained as the dependent variable influence at one location on that of its spatial neighbors. This explanation is criticized, as causality is a relation between different variables and not of the same variable in different spatial units (Herrera Gomez et al., 2014). Partridge et al. (2012) considered spatial dependence to be unclear in the context of causality, as spillover can also be the effect of explanatory variables in spatial neighbors. For example, in the spatial (lag of) X model (SLX), spatial dependence is treated as the effect of explanatory variables within a neighborhood instead of that of the dependent variable (LeSage and Pace, 2009). Spillover may therefore contain either kind of causality, i.e., interaction of the dependent variable at different locations, and interaction of the dependent and explanatory variables at different locations. For instance, a causal effect between the dependent variable at two nearby locations is likely to exist in infectious disease and air pollution. What remains challenging is how to distinguish whether the spatial dependence is caused by dependent or explanatory variable in the neighborhood (Gibbons and Overman, 2012). To address this issue, Gibbons and Overman (2012) designed an idempotent spatial weighting matrix including prior knowledge about the mechanism, Corrado and Fingleton (2012) proposed integrating weighting into hierarchical modeling, while Haarsma and Qiu (2017) utilized the first-differenced spatial regression model to attribute spillover effects to both explanatory and dependent variables when studying the cause–effect relationship between population growth and agricultural land conversion activities.

Another major difficulty in spatial causal inference is that cross-sectional data in spatial statistics do not contain apparent information for identifying the direction of cause and effect. This is in contrast to time-series studies, where cause precedes effect in time. Cross-sectional data are considered to be observed contemporaneously or do not contain discernible time lags, which is also known as the mirroring effect (Jesus, 2013; Herrera et al., 2016). Therefore, more techniques are needed to infer causation from cross-sectional data.

2.2.3. Two new causal inference methods in spatial statistics

In the information increment concept of Granger and Wiener, if X is the cause of Y, the uncertainty of the distribution of Y can be reduced by adding the distribution information of X (Jesus, 2013; Herrera et al., 2016). Based on this concept, Herrera et al. (2016) developed a nonparametric method to test the causality between two variables of cross-sectional data. In this method, a multiple (larger than two and smaller than the total number of spatial observations) dimensional space is built for the cause variable X and the effect variable Y, respectively, as formulas (7) and (8). Causality can then be tested with the null hypothesis in Eq. (9). If the null hypothesis cannot be denied, X is not the cause of Y.

$$X_m(s_0) = \left(X_{s_0}, X_{s_1}, \ldots, X_{s_{m-1}}\right) \quad for \quad s_0 \in S \tag{7}$$

where S represents the study area, m is the dimension of the space, m-1 is also the number of neighbors included, s_0 is a location in S, and s_1, s_2 and s_{m-1} are the nearest neighbors to s_0 in ascending order of distance.

$$Y_m(s_0) = \left(Y_{s_0}, Y_{s_1}, \ldots, Y_{s_{m-1}}\right) for s_0 \in S \tag{8}$$

$$H_0 : h_m(Y) = h_m \{Y|\chi_W\} or h_m (Y) - h_m \{Y|\chi_W\} = 0 \tag{9}$$

where $h_m()$ is the symbolic entropy function using Shannon's entropy and χ_W is the causal spatial structure from X to Y. In (9), $h_m \{Y|\chi_W\}$ represents the symbolic entropy of Y conditioned on χ_W. If $h_m(Y)$ equals $h_m \{Y|\chi_W\}$, it implies that χ_W does not contain extra information about Y, i.e., the uncertainty of the distribution of Y is not reduced by adding the information of χ_W. It can then be stated that X is not the cause of Y under the causal spatial structure χ_W.

Geographical detectors have been widely used to measure causality using cross-sectional data in fields, such as the environment (Luo et al., 2019), social economics (Liu et al., 2019), disaster (Liu et al., 2018), and public health (Chien et al., 2018). The axiom for the causality is that if X causes Y, their spatial patterns (measured by SSH) tend to be matched. Which is measured by the Geodetector q statistic defined in Formula (4) when the strata are divided according to X, 1 if fully matched, 0 irrelated, and X explains 100*q% variance in Y.

3. Causal inference in other fields

Due to incomplete observation data, causal inference is an ill-posed problem. In contrast to limited causation studies in spatial statistics, causal inference has been researched extensively in statistics, social, biomedical and computer sciences, and much progress has been made. Among these, statistical causal inference methods based upon prediction properties, graph structure and potential outcomes have drawn much attention. To infer causality from insufficient observation data, different assumptions were set in those methods.

3.1. Causal inference based on prediction property

Causal inference based upon the prediction property utilizes the unidirectionality of time to predict the effect variable with a causal variable. Two important methods are the Granger test (Granger, 1969) and CCM (Sugihara et al., 2012; Tsonis et al., 2015).

3.1.1. The Granger test

The Granger test is based upon a notion in Wiener (1956): if the prediction accuracy of Y can be improved by using the past information of X compared with the prediction without it, then X can be taken as a Granger cause of Y. Any prediction can be defined by the vector autoregressive full model in (10). If the null hypothesis (11) cannot be denied, then (10) reduces to (12).

$$Y_t = \varphi_0 + \sum_{\tau=1}^{\tau_{max}} \varphi_\tau Z_{t-\tau} + \sum_{\tau=1}^{\tau_{max}} \omega_\tau X_{t-\tau} + \epsilon_t \tag{10}$$

where τ is the time lag, τ_{max} is the maximum time lag, Z represents other variables except X and the past of Y, φ_τ and ω_τ are corresponding coefficients, and ϵ_t is the error. If the null hypothesis H_0

$$H_0 : \omega_\tau = \omega_2 = \cdots = \omega_{\tau_{max}} = 0, \tag{11}$$

cannot be rejected, then

$$Y_t = \varphi_0 + \sum_{\tau=1}^{\tau_{max}} \varphi_\tau Z_{t-\tau} + \epsilon_t', \tag{12}$$

The assumption of the Granger test includes temporal precedence, information completeness, temporal invariance and separability. Temporal precedence requires that cause and effect cannot be contemporaneous, i.e., the cause precedes the effect. Information completeness means that we should include all common causes of X and Y in the full model in Formula (10). Temporal invariance assumes that a cause–effect relationship remains constant over time. Finally, separability requires that the information of a cause variable is not contained in the time series of the effect variable.

3.1.2. Convergent-cross mapping

When separability cannot be satisfied as a deterministic condition, then the Granger test becomes problematic (Granger, 1969). Sugihara et al. (2012) provided an example to demonstrate that if the weather is a cause of fish production, then weather information from the past is contained in the time-series data of fish production, which can thus be used to estimate past weather. In this way, weather information would be redundant when predicting fish production, and thus, the weather would be denied as the cause of fish production. To address this nonseparability and infer causation between weakly associated variables, they developed the CCM method based upon dynamic systems theory, which states that if X causes Y, both being time-series variables, then they belong to the same dynamic system and the states of X can be predicted from Y, but the converse is not necessarily the case.

The time-series data of X and Y can be used to reconstruct shadow manifolds as in (13),

$$\begin{aligned} M_{x,t} &= \left[x_t, x_{t-\tau}, x_{t-2\tau}, \ldots, x_{t-(E-1)\tau}\right] \\ M_{y,t} &= \left[y_t, y_{t-\tau}, y_{t-2\tau}, \ldots, x_{t-(E-1)\tau}\right], \end{aligned} \tag{13}$$

where $M_{x,t}$ and $M_{y,t}$ are the reconstructed shadow manifolds of X and Y, respectively, and E is the maximum time lag or the number of dimensions.

The reconstructed shadow manifolds are diffeomorphic to the original manifolds. With the reconstructed shadow manifolds, the state of Y can be predicted with the state of X through (14):

$$\widehat{M}_{x,k_0}|M_y = \sum_{i=1}^{E+1} \frac{d\left(M_{y,k_i}, M_{y,k_0}\right)}{\sum_{j=1}^{E+1} d\left(M_{y,k_j}, M_{y,k_0}\right)} M_{x,k_i} \tag{14}$$

where $\widehat{M}_{y,k_0}|M_y$ is the prediction of x at time k_0 and $d\left(M_{y,k_i}, M_{y,k_0}\right)$ is the distance between two points in the shadow manifold reconstructed from Y, defined in (15):

$$d\left(M_{y,k_i}, M_{y,k_0}\right) = \exp\left(-\frac{\left\|M_{y,k_i} - M_{y,k_0}\right\|}{\left\|M_{y,k_1} - M_{y,k_0}\right\|}\right), \tag{15}$$

Finally, as shown in (16), CCM uses the correlation between the predicted values and observed values of X after convergence to measure the strength of the causality:

$$\rho_{x\to y} = \lim_{L\to+\infty} \operatorname{cor}\left(M_x, \widehat{M}_x \left|M_y\right.\right), \tag{16}$$

where $\rho_{x\to y}$ is the correlation after convergence used to measure the causation effect of x on y, L is the sample size, and cor is the correlation function.

Causal inference methods based on prediction property can identify causality from temporal changes and could be used in spatio-temporal context. When applied to panel data, they could infer causality for each spatial units from the time-series data. For example, Chen et al. (2018) adopted CCM to infer the causation effect of meteorological factors on PM2.5 concentrations of 188 monitoring cities and revealed the spatial variation of the causation effects over China. It needs be noticed that those causation effects inferred are only local causation effects depending on the spatial location, and may not hold globally. For example, Gao et al. (2021) demonstrated that when the time-series observations do not catch significant changes of the casual and effect variables, causality could not be detected locally, but if examined from the perspective of spatial variation, significant global cause–effect relationship can be identified. However, both Granger test and CCM cannot be used to cross-sectional data. In future, efforts could be paid to develop new methods based on the prediction property to infer causality from spatial variation, and the main challenge is the how to overcome of the mirroring effect.

3.2. Causal inference based upon graph structure

If too many variables depend upon each other, the Granger test often fails to identify causation. If the target system is stochastic rather than deterministic, it is difficult to reconstruct the attractor manifolds using lagged embedding. In this case, CCM loses the power to identify causation (Runge, 2018). Following the pioneering works of Wright (1920), causal inference based on graph structure has been developed in economic and computer science. In economic science, the main achievement is structural equation modeling (SEM), while in computer science, inference methods can be divided into two types, i.e., causal network learning and the structural causal model (SCM) (Camps-Valls et al., 2019). Causal network learning algorithms identify the causation links as an iterative procedure. They require time-series data to obtain directed acyclic graphs, while for cross-sectional data, they identify several possible causal graphs within a Markov equivalence class. The SCM can be divided into two types according to whether a prior causation graph is needed: SCM with a prior graph and SCM without a prior graph. SEM, which depends upon a relationship structure given by prior knowledge and then fits equations with observation data, belongs to the first type. The second type can identify causation without the need for a prior graph, as causal network learning algorithms. The difference is that by setting more assumptions on the observation data and models, SCM can identify causal directions from cross-sectional data. In addition to equation fitting, many methods are adopted to identify causation structure from observation data, such as independence-based methods, score-based methods, additive noise models and models with known causal ordering (Peters et al., 2017).

3.2.1. SCMs with prior graph

An SCM is composed of a set of functions, and a directed graph called a causal diagram reflects the cause–effect relationships among variables built with prior knowledge. The causal diagram in Fig. 1 and equations in (17) are an example of SCM. It originated in economics and was called an SEM, in which variables are separated into explanatory and response variables. Response variables are those that we are interested in, and these are influenced by explanatory variables, which in turn are influenced by other outside factors (Wright, 1921; Haavelmo, 1944; Bollen, 1989). SEM has two important intersections with computer science. The first intersection occurred when a software package called LISREL was developed to realize SEMs in the 1970s, which could, in some cases, automatically determine and estimate path coefficients. It facilitated the solution of equations of SEMs but also brought some negative impact. Partly due to the development of computer software, many researchers in economics began to ignore the graph and only pay attention to the algebraic part. However, ignoring the graph caused SEM to become a simultaneous equation model without direction, which cannot distinguish cause and effect (Pearl and Mackenzie, 2018).

$$\begin{aligned} X_1 &:= f_1(X_3, N_1) \\ X_2 &:= f_2(X_1, N_2) \\ X_3 &:= f_3(N_3) \\ X_4 &:= f_4(X_2, X_3, N_4) \end{aligned} \tag{17}$$

The second intersection occurred with the development of machine learning when Pearl (2009) combined a Bayesian network with an SEM to give it the ability for causal inference. Inspired by SEM, Pearl proposed the inclusion of causal diagrams (Peters et al., 2017). Causal diagrams are important tools for formalizing causal inference and allow one to model interventions and counterfactuals in SCMs. Based on this, Pearl (2009) proposed the do-operator and the interventional probability $P(Y|do(X))$ to measure the causation effect of X on Y. Pearl thus defined confounding to be anything that makes $P(Y|do(X)) \neq P(Y|X)$. In further, Pearl and Mackenzie (2018) introduced front-door and backdoor adjustments, which adjust variables on the front-door path and backdoor path, respectively, to revise SCMs to deduce cause–effect relationships. The front-door path is a path pointing out *from* the cause variable, while the backdoor path is a path pointing *to* the cause variable. Pearl and Mackenzie (2018) provided three rules guiding the adjustment so that the causation effect can be estimated unbiasedly: information prevention, variable randomization and statistical adjustment.

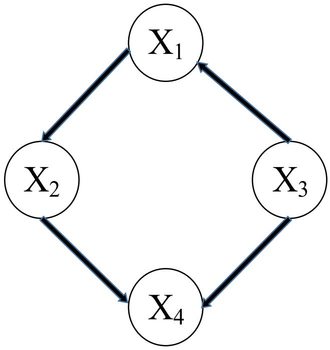

Fig. 1. A causal diagram.

SCMs with prior graph can work with cross-sectional data, by first assuming a causal diagram and then employing observed data to identify and estimate the cause–effect relationships. However, in spatial context, extra effort should be paid to handle the confounding brought by spatial autocorrelation, and overcome the lack of observation data with identical distribution due to spatial heterogeneity.

3.2.2. Causal network learning

Causal network learning is based upon the Markov condition and the faithfulness assumption. The Markov condition states that in a causal graph G with vertices X whose probability distribution is P, any variable X_i in X is independent of other variables, excluding the descendants and parents of X_i, given the direct causes of X_i. It associates the causal graph and conditional independences and guarantees that separation in G implies independence in P. The faithfulness assumption states that all conditional independencies in P are entailed by G. It guarantees that G entails all conditional independence in P. Based upon the Markov condition and the faithfulness assumption, causal network learning aims to recover the causal graphs from observational data.

The PC algorithm named after the inventors Peter and Clark is a causal network learning algorithm widely used in processing high-dimensional stochastic variables (Spirtes and Glymour, 1991). It is a basic step for many improved methods, such as the FCI algorithm, the RFCI algorithm and the CCD algorithm. In the absence of confounders, the PC algorithm first builds a complete undirected graph with all variables as vertices and with all pairs of vertices linked by undirected edges. It then iteratively deletes the edges according to the conditional independence defined in (18):

$$X \perp Y | S \tag{18}$$

where $\perp$ denotes conditional dependence, S is a subset of parent variables and all combinations are iterated and tested in an inner loop.

When time-series data are used, the time lag can be used to obtain the direction for the derived graph based upon the law that the cause precedes the effect in time. Concerning the causal graph for time-series data in Fig. 2, we note that the directions are definite and that there are no arrows from the present to the past. PCMCI is a PC-based algorithm for time-series data (Runge et al., 2019). It is robust in identifying cause–effect relationships from high-dimensional and temporally autocorrelated time-series data with both linear and nonlinear dependencies (Runge, 2018). It is realized in two steps using the momentary conditional independence (MCI) test in Formula (19). In the first step, an approximate set of parents is obtained for each variable with a fast variant of the PC algorithm, in which only a subset S is tested that has the largest association, instead of all arrows in an inner loop. In the second step, the parents retained after the first step are used as conditions to perform the MCI test.

$$MCI : X_{t-\tau} \perp \mathrm{Y}_t | P(Z_t) \tag{19}$$

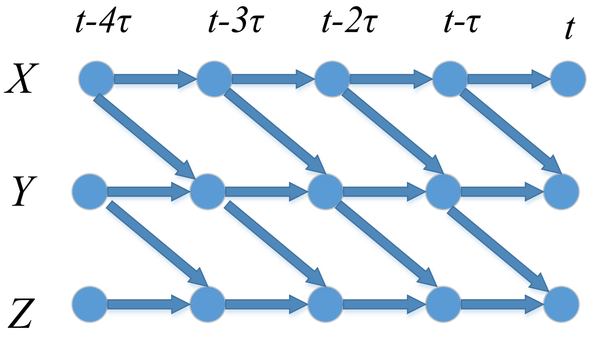

Fig. 2. A causal graph for time-series data.

where $P(Z_t)$ are the retained parents. If the MCI test cannot be denied, the corresponding arrow is removed. When no arrow can be removed in the causal graph, the second step terminates, and the causal inference result is obtained.

Like Granger test and CCM, Causal network learning methods for time series data like PCMCI could be used in spatio-temporal context, but cannot be applied to cross-sectional data. Additional assumptions and techniques are needed to enable them to identify the direction of cause–effect relationship from cross-sectional data due to the mirroring effect.

3.2.3. SCMs without prior graph

Similar to causal network learning, SCMs without prior graphs aim to recover the causal graph from the observed data. A major aim is to eliminate spurious Markov equivalent graphs and retain the correct graphs by adding new assumptions together with the Markov condition and the faithfulness assumption. This characteristic makes them more suitable for cross-sectional spatial statistics. Here we introduce some of them.

(1) Additive Noise Model (ANM):

The ANM assumes that the noise is additive to the fixed part, as in (20). Based upon this assumption, an ANM can identify the correct causal graph for most types of functions and noise distributions but cannot be guaranteed to be successful in all cases. For example, the causal graph cannot be identified when the functions are linear and the noise distribution is Gaussian, as in (21).

$$X_j := f_j\left(PA_j\right) + N_j, \quad j = 1, \ldots, d \tag{20}$$

where PA_j represents the parents of X_j, and N_j is the noise.

$$X_j := \sum_{k \in PA_j} \beta_{jk} X_k + N_j, \quad j = 1, \ldots, d \tag{21}$$

For a more comprehensive example of cases where ANM fails, we refer to Peters (2014).

For the linear Gaussian structure in (21), Peters and Bühlmann (2012) proved that if we assume that the noise variables have equal variance, the causal structure can be recovered with the penalized maximum likelihood score and a greedy search algorithm.

(2) The Linear Non-Gaussian Acyclic Model (LiNGAM):

Shimizu et al. (2006) developed LiNGAM to identify a full causal graph. As in (21), LiNGAM assumes that the functions are linear and that independent noise terms are represented as continuous variables with a non-Gaussian distribution. They proved that if these assumptions are valid, then the full causal graph can be identified using LiNGAM without the need to know the causal ordering of variables beforehand. Additionally, an improved version of LiNGAM named DirectLiNGAM was developed, which can converge to the right solution in a finite number of iterations (Shimizu et al., 2011).

(3) Nonlinear Gaussian additive noise models (nonlinear Gaussian ANMs):

Nonlinear Gaussian ANMs assume that the functions f_j in (20) are nonlinear and three times differentiable, and that the noise terms are variables with a Gaussian distribution. If this assumption is satisfied, then the corresponding graph can be identified from the distribution of variables (Peters et al., 2014).

Methods of SCMs without prior graph have the potential to be applied to cross-sectional data. However, due to the spatial autocorrelation, the assumption of independent noise distribution would be violated most often. Therefore, before applying those method to spatial data, the assumptions should be carefully examined, and spatial autocorrelation needs to be properly handled using techniques of spatial regression models.

3.3. Causal inference based upon potential outcomes

The cause–effect relationship between two variables is defined here as follows: setting different values (treatment) in a causal variable causes different outcomes to occur in the effect variable. Rubin (1974) noted that causal inference from observation data is a missing data problem: for each unit, we can only observe the outcome of one treatment, while the outcomes associated with other alternative treatments are absent. The potential outcomes are then used to define the outcomes of all treatments assigned to a given individual. They are "potential" because only one treatment was assigned and only one corresponding outcome was observed. Based upon the notion of potential outcomes and randomization, Rubin (1974) developed the Rubin causal inference framework for observational studies. An important assumption for causal inference is the stable unit treatment value assumption (SUTVA), which assumes that (1) the potential outcomes of different units do not interfere with each other and (2) all treatment levels are included so that for a fixed unit if we assign the same treatment, the outcome should remain the same.

The most important component in causal inference based upon potential outcomes is the assignment mechanism. This is defined as the process that determines how the treatments are bound to different units. It can be described as a function containing all covariates and potential outcomes. To make the causal inference operable, three constraints are added on the assignments: individualistic, probabilistic and unconfounded. Individualistic assignment disallows the probability of binding one specific treatment to depend upon potential outcomes and covariates of other units; probabilistic assignment limits the probability of binding any treatment to individuals to the range (0,1); unconfounded assignment requires that the assignment mechanism is independent of the potential outcome. Imbens and Rubin (2015) divided causal inference problems into three kinds according to the assignment mechanism: classical randomized experiments, regular assignment mechanisms and irregular assignment mechanisms. In classical randomized experiments, the assignment mechanisms satisfy the abovementioned three constraints, and are designed by researchers and well understood. The regular assignment mechanisms also satisfy the abovementioned three constraints but are not controlled by researchers and are not known beforehand. Finally, the irregular assignment mechanisms do not satisfy the probabilistic or unconfounded constraints. Both regular and irregular assignment mechanisms belong to observational studies.

There are four strategies for causal inference in regular assignment mechanisms: model-based imputation, weighting, blocking, and matching strategy. The inference approaches can be roughly divided into five classes, with the first four classes corresponding to the abovementioned four strategies and the fifth class being a combination of those strategies. Model-based imputation builds a model to predict the missing outcomes of a specific unit as (22):

$$Y^{mis}|Y^{obs}, X, W \tag{22}$$

where Y^{mis} are the missing outcomes, Y^{obs} are the observed outcomes, X represents the covariates and W represents the assignment. It focuses on the distribution estimation of conditional outcomes, while the other three strategies focus on the propensity score estimation, i.e., the conditional probability of assignment defined in (23).

$$e(x) = \frac{1}{N(x)} \sum_{i:X_i=x}^{N(x)} p_i(X, Y(0), Y(1)) \tag{23}$$

where $N(x)$ represents the number of units when $X_i = x$.

The weighting strategy obtains the propensity score by weighting the observations, while the blocking strategy divides the observations into several subclasses and treats individuals in the same subclass as completely randomized. Instead of estimating an unknown function, the matching strategy directly compares the observed outcomes between case units and control units searched based on the similarity of covariates. In irregular assignment mechanisms where the unconfoundedness assumption cannot be satisfied or guaranteed, sensitivity and bounds analyses may be used to obtain an interval estimation of the treatment effects. By making additional assumptions and collecting more information, we can also obtain an unbiased point estimation of the treatment effect.

If the observation data before and after the treatment are available, the Difference-in-difference (DID) is a useful method to estimate the treatment effect (Delgado and Florax, 2015). It is implemented in two steps as formula (24) and (25), by calculating the change (difference) of all units before and after the treatment, then estimating the difference on the changes between the treatment group and control group.

$$\Delta Y_i = Y_i(after) - Y_i(before) \tag{24}$$

where $Y_i(after)$ and $Y_i(before)$ are the observed outcomes before and after the treatment of the unit i, and ΔY_i is their difference.

$$e = E[\Delta Y_i(1)] - E[\Delta Y_i(0)] \tag{25}$$

where e is the treatment effect, $E[\Delta Y_i(1)]$ is the expectation of difference of units in the treatment group, and $E[\Delta Y_i(0)]$ is that of the control group. Two important assumption in DID is the common trend and ignorability. The common trend means that the units in treatment group and control group have the same trend. And ignorability imply that treatment assignments to units are independent of their outcomes.

Since Angrist et al. (1996) proposed embedding instrumental variables into causal inference, causal inference leveraging instrumental variables has been developed. A simple example using an instrumental variable is illustrated in Fig. 3, in which the treatment effect of X on Y needs to be estimated with an unobserved confounder U, and Z is an instrumental variable according to prior knowledge that satisfies three conditions: Z is the direct cause of X, all effects of Z on Y pass through X, and no backdoor path from Z to Y exists. Here, the instrumental variable is a secondary treatment, assigned to the original treatment, and is assumed to be unconfounded. Then, the classical method in the randomized experiment can be used to estimate the effect of Z on X, and finally, the effects of Z on X can be used to estimate the effects of X on Y. In the randomized experiment with instrumental variables, noncompliance should be considered to conduct the effect estimation.

Imbens and Rubin (2015) introduced inference methods for one-sided noncompliance and two-sided noncompliance. In one-sided noncompliance, only units assigned to active treatment defy the treatment assigned, but in two-sided noncompliance, the units assigned to the control treatment can also deny the treatment assigned. Frangakis and Rubin (2002) proposed the principal stratification in which they stratify the data into four types, i.e., compliers, always-takers, never-takers and defiers, and formulated the estimand for the causal effects.

The potential outcomes framework is closely tied to statistical methods (Imbens, 2020) and time-series data are not a necessary condition for it; thus, it can be used with cross-sectional data. For example, to estimate the effect of home purchase restriction policy on the housing price with the matching strategy, we can (1) divide cities taking the restriction policy into treatment group and others in the control group, (2) match cities in the treatment group to ones in the control group according to the propensity score which measures the similarity between cities, (3) estimate the

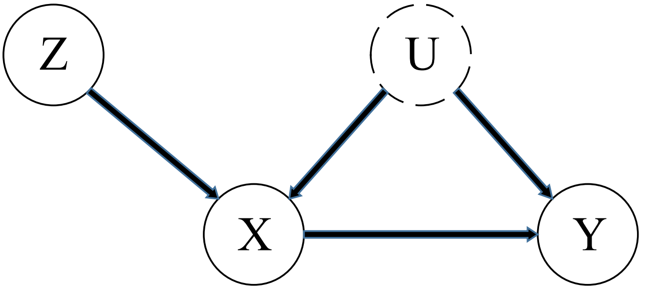

Fig. 3. An example of using instrumental variables.

difference on the housing price between matched pairs. If the housing price before the treatment are available, the DID could also be used to estimate the effect of the restriction policy. When used with spatial data, the SUTVA of potential outcomes framework many be violated due to the spatial autocorrelation and heterogeneity (Akbari et al., 2021). For example, the housing price of one city may be influenced by the restriction policy of a neighbor city, and the intensity of the restriction policy may differ among cities. However, we believe that, just like the way spatial statistics develop statistics to handle and leverage the spatial autocorrelation and heterogeneity, spatial causation inference method will be developed in near future in potential outcomes framework. Spatial location and distance could be employed to build measuring indices and spatial interaction would be processed to separate the direct and indirect effects. Another big difficulty to develop spatial causation inference methods is how to deal with continuous cause variables. They violate SUTVA and it is impossible to include all treatment levels. Present methods in the potential outcomes framework only works on category cause variables, thus we are not sure that it could be overcome.

4. Causal inference in spatial statistics: the future

With the surge of spatiotemporal data at all scales from the universe evolution to the spatial structure of DNA, spatial statistics have received much attention. But due to the nonrandom, nonrepeatability and synchronism of spatial data, limited progress on causal inference has been made in spatial statistics, when much progress has been made in statistics, economics, epidemiology and computer sciences. In those disciplines, a large number of causal inference methods based upon prediction properties, graph structure and potential outcomes have been developed. Among them, Granger test, CCM and Causal network learning methods require time-series data to infer causality, while SCMs and methods in potential outcomes framework have the potential to be applied to spatial cross-sectional data. Before adopting those method, their assumptions should be carefully examined. Spatial autocorrelation and heterogeneity are the main obstacles for direct application of those methods. The spatial autocorrelation and heterogeneity indices in spatial statistics could be adopted to measure the spatial pattern, and techniques of spatial regression could be employed to develop spatial causation inference methods. Also, spatial variation is another perspective to remove potential confounding factors to infer cause–effect relationships. Prediction property is an important way which we can learn from Granger test and CCM. And besides it, some inherited characteristics of geospace, such as the matching between spatial patterns and the occurrence of spatial abnormalities, go beyond a pair-point correlation, may imply causation. The corresponding spatial statistics have already been developed, and their implications for causation deserve to be investigated in depth.

Acknowledgment

This work was funded by the National Key R&D Program of China through grant 2021YFE0102300.

References

Akbari, K., Winter, S., Tomko, M., 2021. Spatial causality: A systematic review on spatial causal inference. Geogr. Anal..
Angrist, J.D., Imbens, W.G., Rubin, B.D., 1996. Identification of causal effects using instrumental variables. J. Amer. Statist. Assoc. 91, 444–455.
Bollen, K.A., 1989. Structural equation models with observed variables. In: Structural Equations with Latent Variables. John Wiley & Sons, New York, pp. 80–150.
Camps-Valls, G., Sejdinovic, D., Runge, J., Reichstein, M., 2019. A perspective on Gaussian processes for earth observation. Natl. Sci. Rev. 6, 616–618.
Chen, Z., Xie, X., Cai, J., Chen, D., Gao, B., He, B., Cheng, N., Xu, B., 2018. Understanding meteorological influences on PM2.5 concentrations across China: a temporal and spatial perspective. Atmos. Chem. Phys 18, 5343–5358.
Chi, G., 2010. The impacts of highway expansion on population change: An integrated spatial approach. Rural Sociol. 75, 58–89.
Chien, L.C., Lin, R.T., Liao, Y., Sy, F.S., Pérez, A., 2018. Surveillance on the endemic of zika virus infection by meteorological factors in Colombia: a population-based spatial and temporal study. BMC Infect. Dis. 18, 180.
Chun, B., Guldmann, J.M., 2014. Spatial statistical analysis and simulation of the urban heat island in high-density central cities. Landsc. Urban Plan. 125, 76–88.
Cliff, A.D., Ord, K., 1981. Spatial Processes: Models and Application. Pion, London.
Corrado, L., Fingleton, B., 2012. Where is the economics in spatial econometrics?* J. Reg. Sci. 52, 210–239.
Delgado, M.S., Florax, R.J.G.M., 2015. Difference-in-differences techniques for spatial data: Local autocorrelation and spatial interaction. Econom. Lett. 137, 123–126.
Falcon, A., 2019. Aristotle on causality. In: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University, California.
Fischer, M.M., Getis, A., 2010. Handbook of Applied Spatial Analysis Software Tools, Methods and Application. Springer-Verlag, Berlin, Heidelberg and New York.
Fotheringham, A.S., Brunsdon, C., Charlton, M., 2003. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. Wiley.
Frangakis, C.E., Rubin, D.B., 2002. Principal stratification in causal inference. Biometrics 58, 21–29.
Galilei, G., 1954. Dialogues Concerning Two New Sciences. Dover Publications, Inc, New York, Translated By Henry Crew & Alfonso de Salvio, with an Introduction By Antonio Favaro.
Gao, B., Li, M., Wang, J., Chen, Z., 2021. Temporally or spatially? Causation inference in earth system sciences. Sci. Bull..
Geary, R.C., 1954. The contiguity ratio and statistical mapping. Incorp. Stat. 5, 115–146.
Gendler, T.S., 1998. Galileo and the indispensability of scientific thought experiment. British J. Philos. Sci. 49, 397–424.
Getis, A., Ord, J.K., 1992. The analysis of spatial association by use of distance statistics. Geograph. Anal. 24, 189–206.
Gibbons, S., Overman, H.G., 2012. Mostly pointless spatial econometrics?* J. Reg. Sci. 52, 172–191.
Golgher, A.B., Voss, P.R., 2016. How to interpret the coefficients of spatial models: Spillovers, direct and indirect effects. Spat. Demogr. 4, 175–205.
Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37, 424–438.
Haarsma, D., Qiu, F., 2017. Assessing neighbor and population growth influences on agricultural land conversion. Appl. Spat. Anal. Policy 10, 21–41.
Haavelmo, T., 1944. The probability approach in econometrics. Econometrica 12.
Hanks, E.M., Schliep, E.M., Hooten, M.B., Hoeting, J.A., 2015. Restricted spatial regression in practice: geostatistical models, confounding, and robustness under model misspecification. Environmetrics 26, 243–254.
Herrera, M., Mur, J., Ruiz, M., 2016. Detecting causal relationships between spatial processes. Pap. Reg. Sci. 95, 577–594.
Herrera Gomez, M., Marín, Ruiz M., Mur, J., 2014. Testing spatial causality in cross-section data (with matlab code). Munich personal RePEc archive. pp. 1–14.
Hoover, K.D., 2017. Causality in Economics and Econometrics. The New Palgrave Dictionary of Economics. Macmillan UK, London. Palgrave, pp. 1–13.
Hughes, J., Haran, M., 2013. Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. J. R. Stat. Soc. Ser. B Stat. Methodol. 75, 139–159.
Imbens, G.W., 2020. Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics. J. Econ. Lit. 58, 1129–1179.
Imbens, G.W., Rubin, D.B., 2015. Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. Cambridge University Press, Cambridge.
Isaaks, E.H., Srivastava, R.M., 1989. An Introduction to Applied Geostatistics. Oxford University Press, New Yorks.
Jesus, M., 2013. Causality, uncertainty and identification: Three issues on the spatial econometrics agenda. Sci. Reg. 2013, 5–27.
LeSage, J., Pace, R.K., 2009. Introduction to Spatial Econometrics. CRC Press, New York.
Licata, G., 2019. Aristotle's doctrine of causes and the manipulative theory of causality. Axiomathes 29, 653–666.
Liu, J., Jin, X., Xu, W., Sun, R., Han, B., Yang, X., Gu, Z., Xu, C., Sui, X., Zhou, Y., 2019. Influential factors and classification of cultivated land fragmentation, and implications for future land consolidation: A case study of Jiangsu province in eastern China. Land Use Policy 88, 104185.
Liu, Y., Yang, Z., Huang, Y., Liu, C., 2018. Spatiotemporal evolution and driving factors of China's flash flood disasters since 1949. Sci. China Earth Sci. 61, 1804–1817.

Luo, L., Mei, K., Qu, L., Zhang, C., Chen, H., Wang, S., Di, D., Huang, H., Wang, Z., Xia, F., Dahlgren, R.A., Zhang, M., 2019. Assessment of the geographical detector method for investigating heavy metal source apportionment in an urban watershed of Eastern China. Sci. Total Environ. 653, 714–722.

Matheron, G., 1963. Principles of geostatistics. Econ. Geol. 58, 1246–1266.

Moran, P.A.P., 1950. Notes on continuous stochastic phenomena. Biometrika 37, 17–23.

Partridge, M.D., Boarnet, M., Brakman, S., Ottaviano, G., 2012. Introduction: whither spatial econometrics? J. Reg. Sci. 52, 167–171.

Pearl, J., 2000. Causality: Models, Reasoning and Inference. Cambridge University Press.

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

Pearl, J., Mackenzie, D., 2018. The Book of Why: The New Science of Cause and Effect. Basic Books, New York.

Peters, J., 2014. On the intersection property of conditional independence and its application to causal discovery. J. Causal Inference 3, 108–197.

Peters, J., Bühlmann, P., 2012. Identifiability of Gaussian structural equation models with equal error variances. Biometrika 101.

Peters, J., Janzing, D., Scholkopf, B., 2017. Elements of Causal Inference: Foundations and Learning Algorithms. MIT Press, Cambridge, MA.

Peters, J., Mooij, J.M., Janzing, D., Schölkopf, B., 2014. Causal discovery with continuous additive noise models. J. Mach. Learn. Res. 15, 2009–2053.

Ripley, B.D., 1976. The second-order analysis of stationary point processes. J. Appl. Probab. 13, 255–266.

Rubin, D.B., 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. J. Educ. Psychol. 66, 688–701.

Runge, J., 2018. Causal network reconstruction from time series: From theoretical assumptions to practical estimation. Chaos 28, 075310.

Runge, J., Kretschmer, M., Flaxman, S., Sejdinovic, D., 2019. Detecting and quantifying causal associations in large nonlinear time series datasets. Sci. Adv. 5, eaau4996.

Shimizu, S., Hoyer, P.O., Hyvärinen, A., Kerminen, A., 2006. A linear non-Gaussian acyclic model for causal discovery. J. Mach. Learn. Res. 7, 2003–2030.

Shimizu, S., Inazumi, T., Sogawa, Y., Hyvärinen, A., Kawahara, Y., Washio, T., Hoyer, P.O., Bollen, K., 2011. DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model. J. Mach. Learn. Res. 12, 1225–1248.

Spirtes, P., Glymour, C., 1991. An algorithm for fast recovery of sparse causal graphs. Soc. Sci. Comput. Rev. 9, 62–72.

Sugihara, G., May, R., Ye, H., Hsieh, C.-h., Deyle, E., Fogarty, M., Munch, S., 2012. Detecting causality in complex ecosystems. Science 338, 496–500.

Tsonis, A.A., Deyle, E.R., May, R.M., Sugihara, G., Swanson, K., Verbeten, J.D., Wang, G., 2015. Dynamical evidence for causality between galactic cosmic rays and interannual variation in global temperature. Proc. Natl. Acad. Sci. 112, 3253–3256.

Vaneckova, P., Beggs, P.J., Jacobson, C.R., 2010. Spatial analysis of heat-related mortality among the elderly between 1993 and 2004 in Sydney, Australia. Soc. Sci. Med. 70, 293–304.

Waller, L., 2005. BayesIan thinking in spatial statistics. In: Handbook of Statistics, Vol. 25.

Wang, J., Li, X., George, C., Liao, Y., Zhang, T., Gu, X., Zheng, X., 2010. Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun region, China. Int. J. Geograph. Inf. Sci. 24, 107–127.

Wang, J.-F., Zhang, T.-L., Fu, B.-J., 2016. A measure of spatial stratified heterogeneity. Ecol. Indic. 67, 250–256.

Wiener, N., 1956. Modern Mathematics for Engineers. McGraw-Hill, New York.

Wright, S., 1918. On the nature of size factors. Genetics 3, 367–374.

Wright, S., 1920. The relative importance of heredity and environment in determining the piebald pattern of guinea-pigs. Proc. Natl. Acad. Sci. 6, 320–332.

Wright, S., 1921. Correlation and causation. J. Agric. Res. 20, 557–585.

Yu, D., Wei, Y.D., Wu, C., 2007. Modeling spatial dimensions of housing prices in Milwaukee, WI. Environ. Plan. B: Plan. Des. 34, 1085–1102.

Yule, G.N., 1897. On the theory of correlation. J. R. Stat. Soc. 812–854.