Environmental Modelling and Software 172 (2024) 105929

Contents lists available at ScienceDirect

Environmental Modelling and Software

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

A review of causal analysis methods in geographic research

Zhixiao Zou [a,b,c], Changxiu Cheng [a,b,c,*]

[a] Key Laboratory of Environmental Change and Natural Disaster, Beijing Normal University, Beijing, China
[b] State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing, China
[c] Faculty of Geographical Science, Beijing Normal University, Beijing, China

ARTICLE INFO

Keywords:
Causal analysis
Causal theory
Geographic big data
Statistical theory
Information theory
Time series

ABSTRACT

In the age of big data, identifying causality has become the focus of different scientific disciplines. This importance is particularly acute in geography, where understanding complex systems is a tough task. This study provides an in-depth review of causal analysis techniques, assessing their strengths, assumptions, and limitations in geographic research. Using case studies of precipitation impacts on vegetation and runoff, we compare three key approaches: granger causality, the PC algorithm, and LiNGAM. Our findings reveal that (1) causal analysis is evolving from linear to nonlinear, and from bivariate to multivariate, despite challenges such as uncertainty testing. (2) Establishing causal direction is crucial, but distinguishing between direct and indirect causation is equally important. (3) While detailed assumptions enhance the refinement of causal approaches, they may limit their generalizability. Our results support the broader use of causal analysis in geographic research.

1. Introduction

In both pre-Qin poetry and ancient Greek philosophy, the exploration of cause-and-effect relationships first began at the philosophical level, such as Qu Yuan's inquiry into the causes of the separation of heaven and earth, the change of day into night, and the movement of the sun, moon, and stars. In his *Metaphysics*, Aristotle emphasized that the only way to truly understand something is to know the cause of that thing. Later, scientists often constructed the causal relationship of natural phenomena through systematic observation, controlled experimental design, and mechanistic models. For example, Galileo's falling body experiments in the 16th century (Gendler, 1998) and Fisher's experiments in 1970 (Fisher, 1970) discovered causality by implementing controlled or randomized experiments. For a long time, controlled experiments and mechanistic reasoning have played a very important role in advancing geographic research, but they also have some problems. This is mainly because geographic systems are open and complex giant systems, and it is often difficult to satisfy some basic assumptions of controlled experiments and mechanistic models (Cheng et al., 2018); in addition, there may be problems, such as infeasibility, high cost and long period of randomized experiments (Varian, 2016), difficulty of controlled trials in the field (Adamovic and Leibbrandt, 2023), and humanities-related ethics (Spirtes and Zhang, 2016).

In recent decades, analytical methods for identifying causal relationships from large amounts of data (i.e., statistical causal analysis) have undergone rapid development due to advances in statistics and computer science. In the middle of the last century, the concept of Granger causality and the test method were first proposed (Granger, 1969), which laid the foundation for the subsequent time-series causal analysis. In 1960, Thistlethwaite and Campbell proposed a natural experimental method for analyzing causality, i.e., regression discontinuity design (Thistlethwaite and Campbell, 1960). In 1972, Joreskog and VanThillo proposed a structural equation model (Joreskog and Van Thillo, 1972) for estimating the causal relationship between variables that are directly observable and those that are not, that is, variables and causal relationships between observed variables. Since 2000, scholars, such as Pearl, Spirtes, and Glymour, have demonstrated the feasibility of finding causal maps from observed data (Pearl, 2009; Spirtes et al., 2000; Angrist et al., 1996). Meanwhile, Schreiber (2000) proposed the analysis method of using transfer entropy to portray causal relationships among variables from an information-theoretic perspective. With the advent of the big data era, a large number of new methods have also emerged. Sugihara et al. (2012) proposed a convergent cross-mapping method based on Takens' theorem and nonlinear state space, which is different from the Granger causality test and is applicable to nonlinear and nonstationary time series data. In 2018, Runge et al. proposed using the causal graph model PCMCI algorithm, which is applicable to multivariate time series data (Runge et al., 2019a).

* Corresponding author. Key Laboratory of Environmental Change and Natural Disaster, Beijing Normal University, Beijing, China.
E-mail address: chengcx@bnu.edu.cn (C. Cheng).

https://doi.org/10.1016/j.envsoft.2023.105929
Received 21 September 2023; Received in revised form 14 November 2023; Accepted 22 November 2023
Available online 14 December 2023

Meanwhile, statistical causal analysis (hereafter referred to as causal analysis) has become a frontier and hot spot in geographic research methods, and the number of related studies has grown rapidly. Causal analysis is an important tool for deciphering complex geographic phenomena and processes. It can predict under what conditions the geographic event of interest is likely to occur in the future and derive control means in interventions, providing a basis for formulating countermeasures and promoting human development. Thus, causal analysis is key to geographic research. For example, Yang et al. (2021) used the structural equation modeling approach to examine the interactions of natural and anthropogenic drivers to vegetation changes in Jiangsu province; Wang et al. (2020) applied convergent cross mapping methods to quantify the causal effects of age structure, abundance, and environment on the spatial variability of marine fishes; Zou et al. (2021) used convergent cross-mapping methods to obtain the causal coupling between PM2.5 concentration and multiple meteorological elements in different seasons by the convergent cross-mapping method.

To better promote the application of causal analysis in geographic research, Runge et al. (2019b) sorted out the application of causal analysis methods in the Earth system from the perspective of time series, but they lacked a summary and review of the spatial perspective. Gao et al. (2022a) compared the difference between spatial and temporal causal analysis methods by analyzing the influencing factors of NPP but lacked an analysis and summary of the advantages and disadvantages of various methods. Gao et al. (2022b) summarized the application of causal analysis methods in spatial statistics but have not yet systematically sorted out the practice of these methods in geographic research. This paper systematically composes the causal analysis methods commonly used in geographic research from the theoretical basis of each type of method, condenses the core issues, basic assumptions, geographic research practices, advantages and disadvantages of each type of causal analysis method, and systematically analyzes the differences and connections of the three core methods by relying on contextual cases. These conclusions are important for promoting the application and practice of causal analysis in geographic research.

2. Causal analysis methods in geographic research

The core of causal analysis methods is mainly causal discovery and causal effect estimation. The former seeks to infer the causal graph from observed data, while the latter focuses on estimating causal effects from observed data, given the causal graph (Runge et al., 2023). According to the core theoretical foundation, causal analysis methods can be divided into two categories based on statistical theory and information theory, as shown in Fig. 1.

From the perspective of theoretical foundation, regression in statistical theory is the early causal analysis method, and later, with the development of computer science, dynamics and graph theory, a series of method branches emerged. The approach based on the information theory perspective mainly defines causality from the perspective of entropy.

From the perspective of data types, some methods are applicable to time series data, and some are applicable to spatial data. Time series causal analysis methods are simpler than spatial methods (O'Sullivan and Unwin k, 2010; Akbari et al., 2021; Gao et al., 2022a, 2022b). However, time series causal analysis often requires long time series and high-frequency sampling data, which also severely limits the application of the methods. Taking the global climate evolution process as an example, the identification of its cause and effect often requires longer time-series observation data. The current meteorological observations are relatively short to support the causal inference of related studies (Shichi et al., 2007). In addition, problems, such as cloud obscuration and long satellite revisit cycles, can also lead to insufficient sampling density, thus limiting causal analysis based on time series data. Spatial regression discontinuity design (Wuepper et al., 2020), structural equation modeling (Ao and Chang, 2020; Wu et al., 2010; Liu et al., 2022), the PC algorithm (Spirtes and Glymour, 1991), and LinGAM (Shimizu et al., 2006) are common causal analysis methods based on spatial data. These methods have played an important role in deciphering the interaction between geographic elements, but they also have some limitations. For example, although the effects of multiple variables on the target variable are taken into account, the spatial spread and spillover effects of the variables have not been considered.

It is worth noting that these causal analysis methods we mentioned are typically applied to either time series or spatially distributed data. With the development of Earth observation technology, geographic data often varies simultaneously in both time and space, yet there currently exists no causal analysis method specifically designed for spatio-temporal data.

3. The core theory of causal analysis and its application

The choice of causal analysis method depends on the scientific problem and the characteristics of the research object. The following is a detailed systematic review of several theories related to causal analysis methods to help the selection of causal analysis methods in geographic research.

3.1. Granger causality test

The idea of the Granger causality test (GC) first originated from the work of Wiener (1956). Later, Granger proposed a concept and corresponding time-series model for the Granger causality test (Granger, 1969). The classical Granger causality test is based on a linear autoregressive model and is mainly used for linear causal systems. Granger causality emphasizes a chronological sequence where the cause precedes the effect. The core idea of Granger causality test is to compare the accuracy of predicting the current value of Y with or without the historical value of X. For example, if X_t is the cause of Y_t, the historical value of X and the historical value of Y can be utilized to predict the current value of Y, and it will be better than using the historical value of Y only. This is because historical values are thought to contain missing information from the past that would be useful in predicting current values. Although the classical Granger causality test model is bivariate, it can be extended to a multivariate model based on the same idea, such as the VAR (Sims, 1980).

The Granger causality test was originally developed in the field of economics, but its application has expanded to the study of geography, with a particular emphasis on economic geography, such as the effects of human activities on land expansion, climate change, and air pollution (Wang and Liu, 2013; Zhang et al., 2011; Li et al., 2016). In recent years, it has also been applied to the fields of economy and carbon emissions,

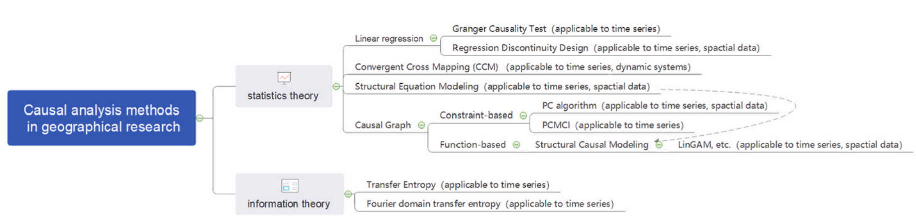


Fig. 1. Classification of causal analysis methods in geosciences.

economy-energy relationships (Dogan and Turkekul, 2016; Bello et al., 2018), and earthquakes (Ning et al., 2021).

3.2. Regression discontinuity design

Regression discontinuity design was first proposed by Thistlethwaite and Campbell in 1960 (Joreskog and Van Thillo, 1972), has only gradually gained attention and popularity since 1990, and is now an important research method in the field of causal analysis and policy evaluation (Imbens and Lemieux, 2008). The regression discontinuity design is a quasinatural experiment (Dinardo,2010; Van Der Klaauw, 2010), i.e., it is similar to a natural experiment, and its basic idea is that there exists a continuous variable X, which determines the probability of an individual receiving a policy intervention on either side of a certain threshold, and since the variable X is continuous on both sides of the threshold, it is random for an individual to fall into either side of the threshold, i.e., there is no artificial manipulation to make the probability of an individual falling into a certain side greater than a quasi-natural experiment constituted around the threshold. The continuous variable X is generally referred to as the grouping variable. When a policy depends on a threshold, one can estimate the causal effect by comparing the results of experiments on both sides of the threshold.

The regression discontinuity design is divided into spatial regression discontinuity and temporal regression discontinuity, which are mainly applicable to policy evaluation; thus, related studies mainly focus on human geography. For example, Wuepper et al. used a spatial regression discontinuity design to study the boundary characteristics of different countries and the effects of policies on the natural environment (Ao and Chang, 2020); some scholars also used a temporal regression discontinuity design to study the effects of heating policies on the life expectancy of residents in northern China (Chen et al., 2013) and to study the effects of regional policies on economic growth in Europe (Pellegrini et al., 2013).

3.3. Convergent cross mapping

Convergent Cross Mapping (CCM) belongs to the state-space model. It models time series data from a dynamical system perspective using the state-space reconstruction of variables and is used to infer causal relationships between two variables (Runge et al., 2019a). The state-space model usually assumes that the causal relationship between two variables occurs in an underlying dynamical system (Deyle and Sugihara, 2011), and reveals the causal relationship based on Takens' theorem and the reconstruction of the nonlinear state-space. It is mainly applicable to nonlinear, nonstationary and dynamic systems, but not to stochastic systems.

GC considers that a cause is a significant predictor of the result in linear systems. That is, if the historical record of X and Y value can accurately predict the current value of Y, then X is a cause for Y, emphasizing that the information in X and Y should be separable. On the other hand, CCM suggests that in nonlinear dynamic systems, while the cause may not be a significant predictor of the result, the result must include all information from the causative process. Thus, if the historical and current values of Y can predict those of X, then X is the cause, and Y is the effect. It is not necessary for the separability of information in causative variables. Importantly, although Granger employed linear functions in the introduction of GC, it can also be utilized to test causal relationships of strongly coupled variables in nonlinear systems (Sugihara et al., 2012).

Since CCM is applicable to nonlinear systems, it is mainly applied to climate-meteorological effects on land use and ecology, such as the relationship between climate change and land degradation (Zhang et al., 2018), the relationship between ENSO and coral diseases (Cramer et al., 2017), and the relationship between meteorology and air pollution (Zou et al., 2021).

3.4. Structural equation modeling

Structural equation modeling (SEM) is a multivariate analysis method for building, estimating, and testing causal models, which includes two parts: a graphical model and a mathematical model (Angelini et al., 2020). In a mathematical model, it consists of the variable set and the equation set. (1) The variable set includes observed variables and latent variables that cannot be directly observed (hereafter: latent variables). Wu et al. (2010) pointed out that the distance between residential areas and public squares and parks and the distance between residential areas and subways and bus stations are observed variables, while environmental convenience and transportation convenience are latent variables that cannot be directly observed. (2) The equation is a functional relationship that portrays the generation mechanism between variables, including the measurement model and structural model. Among them, the equations that describe the generative mechanism between observed and latent variables are called "measurement models". The equations that describe the generative mechanism between dependent and latent variables are called "structural models" (Hau, Wen, and Cheng, 2004). The structural model is used to characterize the relationship between the price of a residential place and its environmental convenience and accessibility, while the measurement model is used to characterize the relationship between the environmental convenience of a residential place and its distance from public squares and parks or the accessibility of a residential place and its distance from subway and bus stations. These equations can be linear or nonlinear. Linear equations are often used in geographic studies. In a graphical model, the schematic representation illustrates the interrelationships between system variables. These variables are connected by arrows, indicating either causality between the variables or correlation in their errors.

The specific process of this method is as follows: (1) the researcher selects the observed and latent variables related to the dependent variable according to the a priori knowledge and sets the corresponding measurement model and structural model, i.e., determines the functional logic relationship between the variables, but the coefficients of the equation are to be determined at this time; (2) the specific observed sample data are substituted into the set model structure, and the coefficients of the equation are solved, and if a unique solution can be obtained, the structural equation model is considered identifiable; and (3) the identifiable model is evaluated and tested to determine whether the hypothetical model proposed by the researcher can characterize the factual causality among variables.

This method is commonly used in the fields of human geography, the environment, and ecology. Examples include the impact of ecosystems on human well-being (Liu et al., 2022), the impact of ecosystems on population aging (Ao and Chang, 2020), the relationship between biodiversity, productivity, and the external environment (Grace et al., 2007), and the impact of storm events on forest diversity and food web structure in coastal zones (Byrnes et al., 2011).

3.5. Causal graph

A causal graph is an important means of expressing the causal relationship between variables. It is a typical directed acyclic graph. For example, Fig. 2 illustrates a simple causal graph in which nodes represent random variables and edges with arrows represent causal relationships, with the variable to which the arrow points being the effect and the opposite being the cause. For example, variable X is a cause of variable Z. In addition, causal graphs do not allow for directional loops, i.e., no variable can be its own cause or indirectly its own cause through other variables. Currently, there are two types of causal graph analysis methods based on constraints and functions.

3.5.1. Constraint-based causal graph methods

The method is based on the availability of a large amount of sample

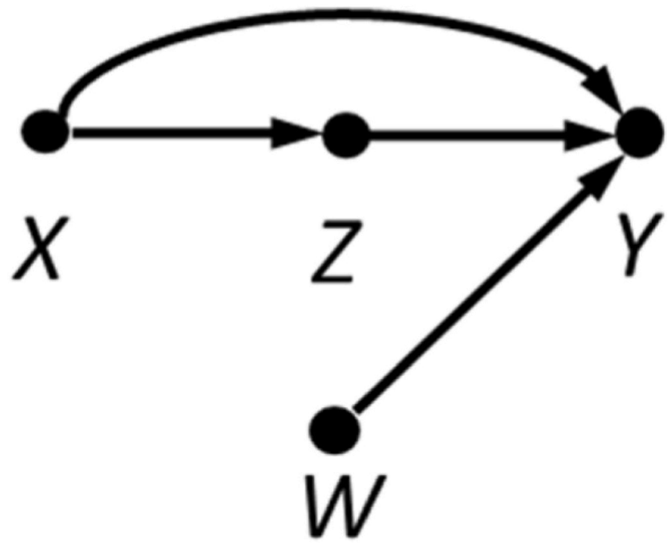


Fig. 2. A simple causal graph.

data, and the causal graph is generated by determining the causal relationship between variables through conditional independence tests (Glymour et al., 2019). There are two basic types of causal inference methods in causal graphs: (1) if variables X, Y, and Z meet the type I conditional independence test (as shown in Fig. 3a), their causal diagrams may belong to the chain structure or the cocause structure, but it is impossible to determine which one they belong to (referred to as the "not unique type I inference" problem). (2) If the variables X, Y and Z meet the type II conditional independence (as shown in Fig. 3b), it can be decided that they belong to the collision structure. There is only a one-to-one correspondence between the type II conditional independence and the causal graph.

A series of causal analysis methods that use the PC algorithm are typical causal graph methods based on conditional constraints. The PC algorithm tests the relationship between any three variables for all variables using the above method, and a complex causal graph structure is formed (Spirtes and Glymour, 1991). The PCMCI method (Runge et al., 2019b) is an improvement of the PC algorithm, i.e., it adds the instantaneous conditional independence test (MCI) to the PC algorithm, which is used to eliminate the pseudocausality. PCMCI works better than PC in geographic processes with strong temporal autocorrelation of variables. Similar to the PC algorithm, PCMCI cannot solve the problem of "first class inference is not unique".

Since PCMCI was developed in 2018, its application has been relatively rare, and it is only found in the study of biosphere-atmosphere interactions (Krich et al., 2020) and the study of the interaction of elements of terrestrial surface systems (Uereyen et al., 2022).

3.5.2. *Function-based causal graph methods*

The function-based causal graph method is based on structural equation modeling and generates a full causal graph by introducing appropriate functional assumptions, so it is also called "structural causal modeling". Specifically, the causal mechanism between variables is obtained by structural equation modeling, i.e., the functional relationship between the causal outcome variable Y and the cause variable X and the noise term E_Y is assumed, but the model alone cannot distinguish the causal direction, so some suitable assumptions have to be introduced to generate the full causal diagram. Taking the LiNGAM method as an example, assuming that the system is linear and has at least one non-Gaussian distributed noise term (E_Y) obeying nonzero variance, the following inference can be made from the sample data: if the causal variable (X) → the outcome variable (Y) and (E_Y) is independent of X, then the causal direction holds; otherwise, the causal direction does not hold (Shimizu et al., 2006).

Compared with SEM, the LiNGAM method introduces additional function assumptions between variables, which enable it to identify the causal direction. Compared with the PC series algorithm, the method can identify the full causal graph, i.e., it can solve the "not unique type I inference" problem by the assumed functions. The LiNGAM method was proposed in 2006 and is currently applied only in the remote sensing field (Pérez-Suay and Camps-Valls, 2019; Runge et al., 2019a). For example, Xiong et al. (2022) combined the LinGAM with the multi-head attention mechanism to generate causal multi-attention maps, which can explore the true causality between attention maps and predicted labels for fine-grained ship image classification.

3.6. *Transfer entropy*

Transfer entropy is a quantitative study of the information transfer relationship between variables from an information-theoretic perspective. When the time series X and Y are both smooth Markov processes, if the transfer entropy of X to Y is greater than the transfer entropy of Y to X, i.e., the degree to which the information of X can reduce the uncertainty of Y is greater than the degree to which the information of Y can reduce the uncertainty of X, then X is the cause of Y and Y is the effect of X, and the causality between the two variables is established in this way.

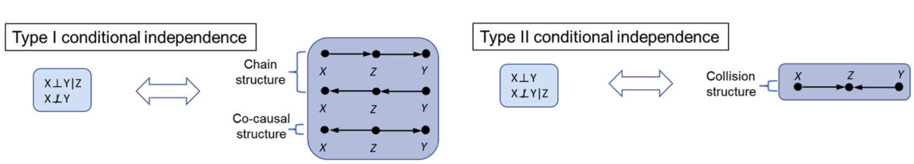
Unlike GC and SEM, transfer entropy does not require prior model construction. After adding the variables' historical information, it can measure the reduction of prediction uncertainty of the target variable to identify cause and effect. Thus, transfer entropy is applicable to nonlinear systems.

At present, transfer entropy is mainly applied to detect the main drivers of drought in future socioeconomic scenarios (Rajsekhar et al., 2015) and to study the interaction between precipitation extremes and ocean temperature anomalies in the Amazon River basin (Ramos et al., 2018).

4. Scope of causal analysis methods application

4.1. *A review of different causal analysis methods*

Different causal analysis methods have different conditions of application, as shown in Table 1. The assumptions of the Granger test include temporal precedence, information completeness, temporal invariance, and separability. Temporal precedence is the requirement that cause and effect cannot occur contemporaneously, i.e., cause precedes effect. Information completeness implies that we should include all common causes of X and Y in the full model. Temporal invariance



(a) Type I conditional independence test **(b)** Type II conditional independence test

Fig. 3. Two types of basic causal graph inference methods.

Table 1
Core theories, key strengths and shortcomings of different causal analysis methods.

Methodology	Core theories	Key strengths	Shortcomings
GC and its extended version	Granger causality theory (temporal precedence, information completeness, temporal invariance and separability)	Well-established significance test; good generalizability	Not applicable to high-dimensional data; difficult to detect contemporaneous effects
CCM	State-space Theory	For state-dependent nonlinear systems; contemporaneous effects can be found; no need to construct causal equations in advance	Not applicable to multivariate data
SEM	Structural equation modeling	For latent variables that cannot be directly observed	Need to construct causal equations in advance high demand for data quality
PC	Causal graph	For multivariate nonlinear systems; well-established test method; no need to construct causal equations in advance	Difficult to detect contemporaneous effects; the "not unique type I inference" problem
PCMCI	Causal graph	For multivariate nonlinear systems; well-established significance test; no need to construct causal equations in advance; pseudocausality due to time lag can be eliminated	Difficult to detect contemporaneous effects; the "not unique type I inference" problem
LiNGAM	Structural equation modeling + Causal graph	For linear systems; dealing with the "not unique type I inference" problem	Not applicable to multivariate nonlinear systems
Regression discontinuity design spatial and temporal	Regression discontinuity	Studies applicable to policy evaluation	Study results difficult to generalize
Transfer entropy, Fourier domain transfer entropy	Transfer entropy	For multivariable, time-lagged nonlinear systems; no need to construct causal equations in advance	Imperfect significance test

assumes that causality remains constant over time. Separability requires that information about the causative variable is not included in the time series of the effect variable (Gao et al., 2022b). In addition, due to the fact that two-variable Granger causality limits the interpretation of findings, multivariate Granger causality test based on a VAR model was later developed (Shojaie and Fox, 2022). Although VAR can be applied to multivariate Granger causality test, only one variable can be tested at a time. Meanwhile, to maintain degrees of freedom, standard VAR rarely use more than six to eight variables (Bernanke et al., 2005). Therefore, it is necessary to explore how to implement Granger tests in high-dimensional time series settings. Therefore, it is necessary to explore how to implement Granger tests in high-dimensional time series settings. Moreover, GC has a low requirement for data quality and a perfect significance test method, so it is well promoted.

CCM is suitable for dynamic systems, and it is difficult to generalize to multivariate data because the inference is performed between bivariate variables. It focuses more on contemporaneous effects because the current period of the outcome variable is introduced in predicting the current period of the cause variable. The method can precisely distinguish between (1) simultaneous changes in X and Y caused by Z and (2) changes in Y caused directly by X (Sugihara et al., 2012). CCM measure the causal effect by correlation coefficient and can test the statistically significance of cross mapping results (Tsonis et al., 2015). Additionally, Gao et al. (2023) extends the method to space to better relate it to earth data.

SEM is suitable for quantitatively describing the causal relationship between observed and latent variables but requires the researcher to assume the model in advance and high-quality observation data to add to it. If the number of variables is too large, it not only increases the difficulty of model assumptions but also increases the difficulty of model solving.

The PC and PCMCI methods are inferred across three variables, and their algorithms are relatively simple, so they can be easily extended to multivariate data. The family of methods is applicable to nonlinear systems, and there are methods applicable to nonlinear conditional independence tests (Runge, 2018a). In addition, PCMCI can effectively eliminate pseudocausality due to time lags. The analysis results of the PC series methods are directed acyclic graphs and, therefore, cannot express the feedback of contemporaneous effects.

LiNGAM can effectively solve the problem of "not unique type I inference", and it is difficult to generalize to multivariate data as SEM.

Among all causal analysis methods, temporal regression discontinuity design and spatial regression discontinuity design are special methods that are similar to natural experiments and suitable for policy evaluation at temporal and spatial breakpoints, but the external validity of regression discontinuity design is weak, i.e., the study results are difficult to generalize. This is because regression discontinuity design can only achieve local randomness near the breakpoints. Transfer entropy is an effective indicator to quantify nonlinear causality, but early transfer entropy can only measure the causal relationship between bivariate variables. To overcome this difficulty, Tian et al. (2021) proposed the Fourier domain transfer entropy spectrum, which is applicable to all kinds of nonlinear and multivariate stochastic processes, but its significance test method needs to be improved.

This shows that (1) causal analysis methods show the development trend from linear to nonlinear and bivariate to multivariate, for example, GC to CCM and CCM to PCMCI, but it may also bring problems, such as imperfection of significance test methods; (2) in the development of causal analysis methods, in addition to determining the direction of causality between variables, it is also necessary to distinguish direct causality, indirect causality and others. For example, CCM can distinguish whether two variables are directly causal, or the result of the same variable together; (3) if the method is more detailed assumptions about the system, the more refined and richer its expressive power; for example, the LiNGAM method is through the detailed assumptions of "the system is linear and the noise term follows a non-Gaussian distribution with nonzero variance". For example, the LiNGAM method solves the problem of "not unique type I inference" and realizes the fine expression of the causal relationship through the detailed assumptions of "the system is linear and the noise term follows a non-Gaussian distribution with nonzero variance"; in addition, the detailed assumptions necessarily require high data requirements, which restricts the generalizability of the method.

4.2. Verification of the difference and connection between GC, PC and LiNGAM

Taking the effect of spring and summer precipitation on vegetation and runoff in a region as an example, the causal scenario is assumed to be shown in Fig. 4a, where the weekly cumulative precipitation (X_2) affects the weekly cumulative runoff (X_1) in the same period, the weekly cumulative precipitation (X_2) affects the weekly average NDVI (X_3) with a lag, and the weekly average leaf area index (X_4) in the same period through the weekly average NDVI (X_3). To better determine the causal

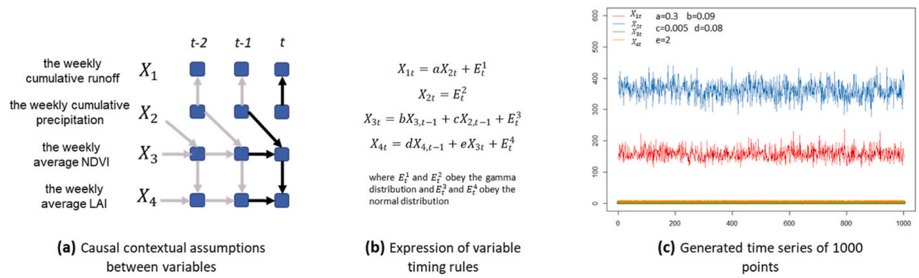


Fig. 4. Construction of causal situations for time series[16].

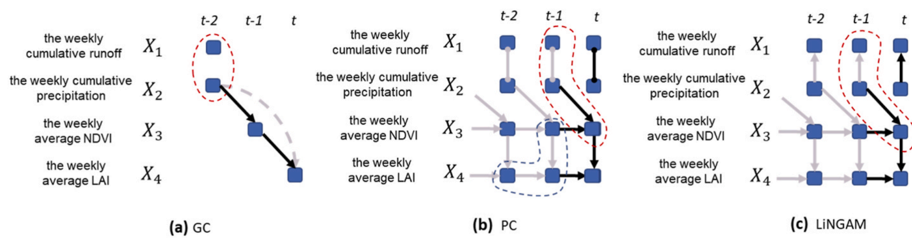


Fig. 5. The results of different causal methods.

recognition ability of GC, PC, and LiNGAM and avoid the influence of other factors in the observation, according to the causal scenario in Fig. 4a, the simulation rules of the above four variables are constructed (shown in Fig. 4b), where X_{2t} and X_{1t} are the weekly cumulative precipitation and runoff of the region, respectively, which usually obey the gamma distribution; X_{3t} is the weekly average NDVI of the region, which is mainly influenced by the previous period's weekly average NDVI, the previous period's weekly cumulative precipitation and random white noise; and X_{4t} is the weekly average leaf area index of the region, which is mainly influenced by the previous period's weekly average leaf area index, the current period's weekly average NDVI and random white noise. The equation in Fig. 4b portrays the characteristics of time lag and contemporaneity that this group of time series has. The time series curves for 1000 moments of the above four variables can be generated according to the equation in Fig. 4b, as shown in Fig. 4c.

Using the above 4-variable time series data as input, the causal relationships were identified by GC, PC, and LiNGAM, and the results are shown in Fig. 5. The differences and connections of each method are further verified by comparing the differences between each subplot of Figs. 5 and 4a.

(1) GC can only identify the causal relationships among weekly accumulated precipitation, weekly average NDVI and weekly average leaf area index. The core idea of GC is that the cause occurs before the effect, and it can effectively identify the causal relationship between variables with a time lag, for example, $X_2 \rightarrow X_3$, but it is difficult for GC to identify the contemporaneous causal relationship between variables (Runge, 2018b); for example, the contemporaneous causal relationship of $X_2 \rightarrow X_1$ is not identified at the red dashed circle in Fig. 5a. In addition, since the GC cannot distinguish whether the causal relationship is caused by the mediating variable or directly by itself, although Fig. 4a shows that $X_{2,t-1}$ indirectly influences $X_{4,t}$ through $X_{3,t}$, the GC is unable to distinguish, and the GC establishes both a causal link from $X_{2,t-1} \rightarrow X_{3,t} \rightarrow X_{4,t}$ and an additional causal link from $X_{2,t-1} \rightarrow X_{4,t}$, as shown in Fig. 5a.

(2) The PC algorithm can accurately identify all relationships except the causal relationship between weekly cumulative precipitation and weekly cumulative runoff. The method can find the collision structure in Fig. 3b; for example, the collision structure of $X_{3t} \rightarrow X_{4,t} \leftarrow X_{4,t-1}$ shown in the blue dashed circle in Fig. 5b. Of course, the PC algorithm has the problem of "first class inference is not unique", so it cannot identify the direction between X_2 and X_1, as shown in the red dashed circle in Fig. 5b.

(3) The LiNGAM method can identify all causal relationships in the context. Among them, there are three causal directions between X_1, X_2, and X_3. Since the method assumes that the model is linear and at least one noise term is non-Gaussian (Xiong et al., 2022), the method can accurately identify the causal directions between X_1, X_2, and X_3, as shown by the red dashed circle in Fig. 5c, i.e., identifying $X_2 \rightarrow X_1$, which solves the "not unique type I inference" problem.

Through verification, it can be found that (1) GC can effectively identify the causal relationship between variables with a time lag but ignore the contemporaneous causal relationship; (2) compared with GC, the PC series algorithm and LiNGAM method can not only identify the direction of causality but also distinguish direct and indirect causality; and (3) the PC series algorithm and LiNGAM can finely delineate the causal relationship at different moments, but the PC series algorithm cannot solve the problem of "not unique type I inference", while LiNGAM can identify and solve the problem through function setting.

5. Conclusion and outlook

5.1. Conclusion

This paper starts from the theoretical basis of various causal analysis methods, systematically composes the causal analysis methods commonly used in geographic research, condenses the core problems, basic assumptions, geographic research applications, advantages and disadvantages of various methods, and systematically analyzes the three types of causal analysis methods based on the case of "the effect of spring and summer precipitation on vegetation and runoff in a certain region". The differences and connections of the core methods are analyzed systematically. The conclusions are as follows: (1) the development trend of causal analysis methods is from linear to nonlinear and bivariate to multivariate, but it may bring problems, such as imperfect significance test of the methods; (2) in the process of development of causal analysis methods, in addition to determining the direction of causality among variables, it is also necessary to distinguish direct and indirect causality among variables; (3) the more detailed the preliminary assumptions of causal analysis methods, the expressive power is more refined and richer, but the degree of generalizability may be constrained.

The development of causal analysis is beneficial for identifying more geographically relevant causal relationship. In turn, the development of geoscience will increasingly require causal analysis methods. Therefore, causal analysis methods have become an important approach in the geosciences.

5.2. Outlook

The current causal analysis methods are oriented either to time series data or spatial data, and there is a lack of causal analysis for spatio-temporal integrated data. However, with the increase in panel data in the field of geographic research, the development of spatiotemporal causal analysis models should receive more attention. It can integrate the spatial distribution and temporal variation of variables to better discover the complex causal relationships of geographic elements in space and time (Pearl, 2009).

Most of the spatial causal analysis methods share similarities with temporal causal analysis methods in that they primarily assess the influences of various factors on the dependent variable, often without accounting for the effects of the dependent variable within the immediate neighborhood. For causal analysis of complex systems, such as geographic systems with strong self-interactions, it is necessary to add spatial spillover effects or address spatial error autocorrelation to existing spatial causal analysis models (Akbari et al., 2021).

Computer code availability.

No software/code was developed or used in this paper.

Authorship Statement

Zhixiao Zou: Conceptualization, Methodology, Interpretation, Writing – original draft. Changxiu Cheng: Supervision, Methodology, Funding acquisition, Writing – review & editing. All the authors have approved the manuscript.

Declaration of competing interest

We declared that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted. Moreover, this manuscript has not been published and is not yet under consideration for publication elsewhere. All the authors have approved the manuscript that is enclosed.

Data availability

No data was used for the research described in the article.

Acknowledgments

This work was jointly supported by the National Natural Science Foundation of China, Grant No. 42041007.

References

Adamovic, M., Leibbrandt, A., 2023. A large-scale field experiment on occupational gender segregation and hiring discrimination. Ind. Relat.: A Journal of Economy and Society 62, 34–59. https://doi.org/10.1111/irel.12318.

Akbari, K., Winter, S., Tomko, M., 2021. Spatial causality: a systematic review on spatial causal inference. Geogr. Anal. https://doi.org/10.1063/1.5025050.

Angelini, M.E., Kempen, B., Heuvelink, G.B.M., Temme, A.J.A.M., Ransom, M.D., 2020. Extrapolation of a structural equation model for digital soil mapping. Geoderma 367, 114226. https://doi.org/10.1016/j.geoderma.2020.114226.

Angrist, J.D., Imbens, G.W., Rubin, D.B., 1996. Identification of causal effects using instrumental variables. J. Am. Stat. Assoc. 91 (434), 444–455. https://doi.org/10.1080/01621459.1996.10476902.

Ao, Rongjun, Chang, Liang, 2020. Influencing mechanism of regional ageing in China based on the Structural Equation Model. Acta Geograph. Sin. 75 (8), 1572–1584.

Bello, M.O., Solarin, S.A., Yen, Y.Y., 2018. The impact of electricity consumption on CO2 emission, carbon footprint, water footprint and ecological footprint: the role of hydropower in an emerging economy. J. Environ. Manag. 219, 218–230. https://doi.org/10.1016/j.jenvman.2018.04.101.

Bernanke, B.S., Boivin, J., Eliasz, P., 2005. Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. Q. J. Econ. 120, 387–422.

Byrnes, J.E., Reed, D.C., Cardinale, B.J., et al., 2011. Climate-driven increases in storm frequency simplify kelp forest food webs. Global Change Biol. 17 (8), 2513–2524. https://doi.org/10.1111/j.1365-2486.2011.02409.x.

Chen, Y., Ebenstein, A., Greenstone, M., et al., 2013. Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy. Proc. Natl. Acad. Sci. USA 110 (32), 12936–12941. https://doi.org/10.1073/pnas.1300018110.

Cramer, K.L., O'Dea, A., Clark, T.R., et al., 2017. Prehistorical and historical declines in Caribbean coral reef accretion rates driven by loss of parrotfish. Nat. Commun. 8 (1), 1–8. https://doi.org/10.1038/ncomms14160.

Deyle, E.R., Sugihara, G., 2011. Generalized theorems for nonlinear state space reconstruction. PLoS One 6, e18295. https://doi.org/10.1371/journal.pone.0018295.

Dinardo, J., 2010. Natural experiments and quasi-natural experiments. In: Durlauf, S.N., Blume, L.E. (Eds.), Microeconometrics, the New Palgrave Economics Collection. Palgrave Macmillan UK, London, pp. 139–153. https://doi.org/10.1057/9780230280816_18.

Dogan, E., Turkekul, B., 2016. CO2 emissions, real output, energy consumption, trade, urbanization and financial development: testing the EKC hypothesis for the USA. Environ. Sci. Pollut. Control Ser. 23 (2), 1203–1213. https://doi.org/10.1021/acs.est.6b02562.

Fisher, F., 1970. A correspondence principle for simultaneous equation models. Econometrica 38, 73–92. https://doi.org/10.2307/1909242.

Gao, B., Li, M., Wang, J., et al., 2022a. Temporally or spatially? Causation inference in earth system sciences. Sci. Bull. 67 (3), 232–235. https://doi.org/10.1016/j.scib.2021.10.002.

Gao, B., Wang, J., Stein, A., et al., 2022b. Causal inference in spatial statistics. Spatial statistics, 100621. https://doi.org/10.1016/j.spasta.2022b.100621.

Gao, B., Yang, J., Chen, Z., et al., 2023. Causal inference from cross-sectional earth system data with geographical convergent cross mapping. Nat. Commun. 14, 5875. https://doi.org/10.1038/s41467-023-41619-6.

Gendler, T., 1998. Galileo and the indispensability of scientific thought experiment. Br. J. Philos. Sci. 49 (3), 397–424. https://doi.org/10.1093/bjps/49.3.397.

Glymour, C., Zhang, K., Spirtes, P., 2019. Review of causal discovery methods based on graphical models. Front. Genet. 10, 524. https://doi.org/10.3389/fgene.2019.00524.

Grace, J.B., Michael Anderson, T., Smith, M.D., et al., 2007. Does species diversity limit productivity in natural grassland communities? Ecol. Lett. 10 (8), 680–689. https://doi.org/10.1111/j.1461-0248.2007.01058.x.

Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica: J. Econom. Soc. 37 (3), 424–438. https://doi.org/10.2307/1912791.

Imbens, G.W., Lemieux, T., 2008. Regression discontinuity designs: a guide to practice. J. Econom. 142 (2), 615–635. https://doi.org/10.1016/j.jeconom.2007.05.001.

Joreskog, K.G., Van Thillo, M., 1972. LISREL: A General Computer Program for Estimating a Linear Structural Equation System Involving Multiple Indicators of Unmeasured Variables.

Krich, C., Runge, J., Miralles, D.G., et al., 2020. Estimating causal networks in biosphere–atmosphere interaction with the PCMCI approach. Biogeosciences 17 (4), 1033–1061. https://doi.org/10.5194/bg-17-1033-2020.

Li, G., Fang, C., Wang, S., et al., 2016. The effect of economic growth, urbanization, and industrialization on fine particulate matter (PM2. 5) concentrations in China. Environ. Sci. Technol. 50 (21), 11452–11459. https://doi.org/10.1021/acs.est.6b02562.

Liu, D., Chen, H., Zhang, H., Shi, Q., Geng, T., 2022. The impact of ecosystem services on human well-being and its group differences in the loess hilly and gully region. Geogr. Res. 41 (5), 1298–1310.

Ning, L., Hui, C., Cheng, C., 2021. Exploring the dynamics of global plate motion based on the granger causality test. Appl. Sci. 11 (17), 7853. https://doi.org/10.3390/app11177853.

O'Sullivan, D., Unwin, D.J., 2010. Geographic Information Analysis, second ed. J. Wiley, Hoboken.

Pearl, J., 2009. Causality. Cambridge university press.

Pellegrini, G., Terribile, F., Tarola, O., et al., 2013. Measuring the effects of European Regional Policy on economic growth: a regression discontinuity approach. Pap. Reg. Sci. 92 (1), 217–233.

Rajsekhar, D., Singh, V.P., Mishra, A.K., 2015. Integrated drought causality, hazard, and vulnerability assessment for future socioeconomic scenarios: an information theory perspective. J. Geophys. Res. Atmos. 120 (13), 6346–6378. https://doi.org/10.1002/2014JD022670.

Ramos, A.M.T., Zou, Y., de Oliveira, G.S., et al., 2018. Unveiling non-stationary coupling between Amazon and ocean during recent extreme events. Clim. Dynam. 50 (3), 767–776. https://doi.org/10.1007/s00382-017-3640-y.

Runge, J., 2018a. Conditional independence testing based on a nearest-neighbor estimator of conditional mutual information. International Conference on Artificial Intelligence and Statistics. PMLR, pp. 938–947. In: https://proceedings.mlr.press/v84/runge18a.html.

Runge, J., 2018b. Causal network reconstruction from time series: from theoretical assumptions to practical estimation. Chaos: An Interdisciplinary Journal of Nonlinear Science 28 (7), 075310. https://doi.org/10.1063/1.5025050.

Runge, J., Bathiany, S., Bollt, E., et al., 2019a. Inferring causation from time series in Earth system sciences. Nat. Commun. 10 (1) https://doi.org/10.1038/s41467-019-10105-3.
Runge, J., Nowack, P., Kretschmer, M., et al., 2019b. Detecting and quantifying causal associations in large nonlinear time series datasets. Sci. Adv. 5 (11), u4996. https://doi.org/10.1126/sciadv.aau4996.
Runge, J., Gerhardus, A., Varando, G., et al., 2023. Causal inference for time series. Nat. Rev. Earth Environ. 4, 487–505. https://doi.org/10.1038/s43017-023-00431-y.
Schreiber, T., 2000. Measuring information transfer. Phys. Rev. Lett. 85 (2), 461. https://doi.org/10.1103/PhysRevLett.85.461.
Shichi, K., Kawamuro, K., Takahara, H., et al., 2007. Climate and vegetation changes around lake Baikal during the last 350,000 years. Palaeogeogr. Palaeoclimatol. Palaeoecol. 248, 357–375. https://doi.org/10.1016/j.palaeo.2006.12.010.
Shimizu, S., Hoyer, P.O., Hyvärinen, A., et al., 2006. A linear non-Gaussian acyclic model for causal discovery. J. Mach. Learn. Res. 7 (10).
Shojaie, A., Fox, E.B., 2022. Granger causality: a review and recent advances. Annual Review of Statistics and Its Application 9, 289–319. https://doi.org/10.1146/annurev-statistics-040120-010930.
Sims, C.A., 1980. Macroeconomics and reality. Econometrica: J. Econom. Soc. 1–48. https://doi.org/10.2307/1912017.
Spirtes, P., Glymour, C., 1991. An algorithm for fast recovery of sparse causal graphs. Soc. Sci. Comput. Rev. 9 (1), 62–72. https://doi.org/10.1177/089443939100900106.
Spirtes, P., Zhang, K., 2016. Causal discovery and inference: concepts and recent methodological advances. Appl. Inf. 3, 3. https://doi.org/10.1186/s40535-016-0018-x.
Spirtes, P., Glymour, C.N., Scheines, R., et al., 2000. Causation, Prediction, and Search. MIT press.
Sugihara, G., May, R., Ye, H., et al., 2012. Detecting causality in complex ecosystems. Science 338 (6106), 496–500. https://doi.org/10.1126/science.1227079.
Thistlethwaite, D.L., Campbell, D.T., 1960. Regression-discontinuity analysis: an alternative to the ex post facto experiment. J. Educ. Psychol. 51 (6), 309. https://psycnet.apa.org/doi/10.1037/h0044319.
Tian, Y., Wang, Y., Zhang, Z., et al., 2021. Fourier-domain transfer entropy spectrum. Phys. Rev. Res. 3 (4), L042040. https://doi.org/10.1103/PhysRevResearch.3.L042040.
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. USA 112 (10), 3253–3256. https://doi.org/10.1073/pnas.1420291112.
Uereyen, S., Bachofer, F., Kuenzer, C., 2022. A framework for multivariate analysis of land surface dynamics and driving variables—a case study for indo-gangetic river basins. Rem. Sens. 14 (1), 197. https://doi.org/10.3390/rs14010197.
Van Der Klaauw, W., 2010. In: Durlauf, S.N., Blume, L.E. (Eds.), Regression-Discontinuity Analysis, Microeconometrics. Palgrave Macmillan UK, London, pp. 214–220. https://doi.org/10.1057/9780230280816_26.
Varian H R. Causal inference in economics and marketing. Proc. Natl. Acad. Sci. U.S.A., 2016113, 7310–7315. https://doi.org/10.1073/pnas.1510479113.
Wang, F., Liu, Y., 2013. Panel Granger Test on urban land expansion and fiscal revenue growth in China's prefecture-level cities. Acta Geograph. Sin. 68 (12), 1595–1606. https://doi.org/10.11821/dlxb201312001.
Wang, J.Y., Kuo, T.C., Hsieh, C., 2020. Causal effects of population dynamics and environmental changes on spatial variability of marine fishes. Nat. Commun. 11 (1), 1–10. https://doi.org/10.1038/s41467-020-16456-6.
Wiener, N., 1956. The theory of prediction. In: Beckenbach, E. (Ed.), Modern Mathematics for Engineers. McGraw-Hill, New York.
Wu, W., Liu, Z., Zhang, W., 2010. Determinants of residential land price: structure equation model analysis using land-leasing parcel data in beijing. Acta Geograph. Sin. 65 (6), 676–684.
Wuepper, D., Borrelli, P., Finger, R., 2020. Countries and the global rate of soil erosion. Nat. Sustain. 3 (1), 51–55. https://doi.org/10.1038/s41893-019-0438-4.
Xiong, W., Xiong, Z., Cui, Y., 2022. An explainable attention network for fine-grained ship classification using remote-sensing images. IEEE Trans. Geosci. Rem. Sens. 60, 1–14. https://doi.org/10.1109/TGRS.2022.3162195.
Yang, L., Shen, F., Zhang, L., et al., 2021. Quantifying influences of natural and anthropogenic factors on vegetation changes using structural equation modeling: a case study in Jiangsu Province, China. J. Clean. Prod. 280, 124330 https://doi.org/10.1016/j.jclepro.2020.124330.
Zhang, D.D., Lee, H.F., Wang, C., et al., 2011. The causality analysis of climate change and large-scale human crisis. Proc. Natl. Acad. Sci. USA 108 (42), 17296–17301. https://doi.org/10.1073/pnas.1104268108.
Zhang, J., Zhang, Y., Qin, S., et al., 2018. Effects of seasonal variability of climatic factors on vegetation coverage across drylands in northern China. Land Degrad. Dev. 29 (6), 1782–1791. https://doi.org/10.1002/ldr.2985.
Zou, Z., Cheng, C., Shen, S., 2021. The complex nonlinear coupling causal patterns between PM2. 5 and meteorological factors in Tibetan plateau: a case study in xining. IEEE Access 9, 150373–150382. https://doi.org/10.1038/s41467-020-16456-6.
Cheng, C., Shi, P., Song, C., et al., 2018. Geographic big-data: a new opportunity for geography complexity study. Acta Geograph. Sin. (8), 1397–1406.