Physica A 541 (2020) 123245

Contents lists available at ScienceDirect

Physica A

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

A novel Granger causality method based on HSIC-Lasso for revealing nonlinear relationship between multivariate time series

Weijie Ren, Baisong Li, Min Han *

Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116023, China

ARTICLE INFO

Article history:
Received 7 January 2019
Received in revised form 30 September 2019
Available online 19 October 2019

Keywords:
Multivariate time series
Nonlinear system
Granger causality analysis
Hilbert–Schmidt independence criterion

ABSTRACT

The causality analysis is an important research topic in time series data mining. Granger causality analysis is a powerful method that determines cause and effect based on predictability. However, the traditional Granger causality is limited to analyzing linear causality between bivariate time series, because it is based on vector autoregressive models. In this paper, we propose a novel method, named Hilbert–Schmidt independence criterion Lasso Granger causality (HSIC-Lasso-GC), for revealing nonlinear causality between multivariate time series. Firstly, for each time series, we perform stationarity test and state space reconstruction to extract the historical information. Then, we build a HSIC-Lasso model of all input variables and output variable, where the optimal model is selected by generalized information criterion. Finally, according to the significance test, we get the causality analysis results from all input variables to output variable. In the simulations, we use two benchmark datasets and two actual datasets to test the effectiveness of the proposed method. The results show that the proposed method can effectively analyze nonlinear causality between multivariate time series.

1. Introduction

Multivariate time series widely exist in complex systems among various fields, such as nature [1,2], industry [3,4], medicine [5,6], society [7–10], etc. There are complex and unknown relationships between multiple variables. Thus, it is of great practical significance to mine potential useful information and explain the unknown dynamics of complex systems. Causality analysis has become a research hotspot in time series data mining, which analyzes the cause and effect between multivariate time series [11]. Different from traditional correlation analysis methods, causality analysis can identify direct and asymmetric relationship. For example, in the natural sciences, there are complex drive-response relationships within and between the systems such as meteorology, hydrology and environment. By means of time series causality analysis, it is helpful to reveal the operation law of the system, and to simulate and predict future trends of natural phenomena. Zhu et al. [12] investigated the impact of urban dynamics on air pollution through spatio-temporal Granger causality analysis and visually showed the relationships among them. The results of causality analysis can determine the main influencing factors of air pollution, thus providing a theoretical basis for decision-making [13]. In the medical field, causality analysis is mainly used for the analysis of medical signals and plays an important role in medical decision-making. Wu et al. [14] proposed a multivariate Granger causality analysis method to analyze the network connectivity of scalp and depth-EEG signals in patients with epilepsy. The causality network of the brain can effectively locate the epileptic lesion, which is of great value in the study of pathogenesis [15]. In the social sciences, the analysis and prediction of finance and

* Corresponding author.
E-mail addresses: renweijie@mail.dlut.edu.cn (W. Ren), lbscomeon@mail.dlut.edu.cn (B. Li), minhan@dlut.edu.cn (M. Han).

https://doi.org/10.1016/j.physa.2019.123245
0378-4371/

energy are of great strategic importance and they are influenced by many factors such as politics, economy and climate change [16]. Based on causality analysis technology, accurate qualitative analysis and quantitative prediction of financial time series are important research contents. Rafindadi et al. [17] conducted the causality analysis of renewable energy consumption on German economic growth and found bidirectional influence relationship between them. The causality between energy consumption and economic growth can provide theoretical guidance for the formulation of energy conservation and emission reduction policy [18,19], which is of great significance in sustainable development and national economic construction. In summary, causality analysis technique has become an important means of multivariate time series analysis and has broad application prospects.

At present, there are many causality analysis models for multivariate time series. The earliest one was Granger causality analysis, proposed by Nobel Prize winner C.W.J. Granger in 1969 [20]. It is a method for judging whether there is linear causality between bivariate time series, and has been widely concerned by researchers since its introduction [21,22]. The traditional Granger causality analysis is a type of model-based approach that only gives qualitative analysis results. The causality analysis based on information measures is a kind of model-free method, including transfer entropy (TE) [23], symbolic transfer entropy [24], partial symbolic transfer entropy [25], conditional entropy (CE) [26], etc. These methods can quantitatively analyze the strength of causality [27]. Besides, the causality model based on state space [28–30], Bayesian network [31,32] and other graph models [33] can also analyze various types of causality. Compared with other methods, Granger causality analysis has the advantages of easy operation and strong interpretability, thus having the widest range of applications. Since the traditional Granger causality analysis is based on linear regression models, it can only analyze linear causality between bivariate time series, so a series of improved models have emerged [34].

For multivariate systems, various multivariate Granger causality analysis models have been introduced. They can be divided into two types, namely model-based and model-free approaches. For model-based approach, an improved scheme is to establish vector autoregressive (VAR) models with conditional variables. For example, Geweke [35] proposed conditional Granger causality index (CGCI) based on VAR models, which could make a distinction between direct and indirect causality. Recently, Siggiridou et al. [36] suggested limiting the order of VAR models using modified backward-in-time selection (mBTS) and proposed mBTS-CGCI, which was successfully applied to the causality analysis of high-dimensional time series. Moreover, in order to reduce the computational complexity, Arnold et al. [37] proposed Lasso Granger causality (Lasso-GC) model for high-dimensional time series. After this, multiple improved models were proposed, such as truncating Lasso Granger causality model [38], group Lasso Granger causality model [39] and group Lasso nonlinear conditional Granger causality model [40]. In summary, model-based methods are computationally efficient and therefore suitable for solving causality analysis problems in high-dimensional time series. In addition, for the model-free methods, multivariate Granger causality analysis is established based on the conditional probability density functions (CPDFs) [41]. For example, Diks et al. [42] extended the bivariate Diks–Panchenko nonparametric causality test [43] to multivariate case and provided reliable estimates of the CPDFs. Besides, the copula-based Granger causality model, proposed by Hu and Liang [44], can also be extended to multivariate case by introducing conditional variables into the marginal probability density functions. Obviously, the model-free multivariate Granger causality analysis is based on predictability and needs to estimate the CPDFs. Due to the low computational efficiency and precision of the high-dimensional probability density functions, model-free methods are limited in multivariate time series causality analysis.

With the deepening of research, people have found that nonlinear causality is common in a large number of actual systems [45], so many nonlinear Granger causality models have emerged. For model-based approach, establishing nonlinear parameter model is a common strategy. For example, Ancona et al. [46] extended the Granger causality analysis to nonlinear bivariate time series by means of radial basis functions. Furthermore, Marinazzo et al. [47] put forward a nonlinear Granger causality model based on kernel method (KGC), which performs linear Granger causality in reproducing kernel Hilbert spaces (RKHS). The key of KGC lies in the choice of kernel function and it is easy to realize the causality analysis of high-dimensional variables through the inner product of kernel function. Apart from the kernel-based models, there are some other nonlinear Granger causality models, such as Granger causality based on neural networks [48] and Gaussian process regression [49]. Besides, several model-free methods have been put forward for nonlinear time series, i.e. nonparametric Granger causality test. Hiemstra and Jones [50] first proposed a nonparametric test for nonlinear Granger causality, which was used for testing the stock price–volume relation. Then, in order to overcome the over-rejection problem in Hiemstra–Jones nonparametric causality test, Diks and Panchenko [43] raised a new bivariate nonparametric test for nonlinear Granger causality. In addition, nonparametric tests based on copula [44,51] are also used for nonlinear Granger causality analysis.

With the continuous increase of data dimensions, the relationship between variables becomes more and more complex. Thus, the causality analysis method develops to analyze multivariate complex systems with nonlinear relationships. As mentioned above, the model-free methods can deal with the bivariate nonlinear Granger causality well, but the curse of dimensionality may occur for multivariate causality analysis. In contrast, model-based nonlinear causality methods have obvious advantages in dealing with multivariate time series. Therefore, we propose a novel model-based approach, named Hilbert–Schmidt independence criterion Lasso Granger causality (HSIC-Lasso-GC) model, which achieves nonlinear causality analysis between multivariate time series. The proposed method can simultaneously achieve causality analysis results from multiple input variables to output variable. Furthermore, the input and the output variables are mapped to RKHSs, by which the nonlinear causality can be revealed. In addition, the time complexities of the proposed model is analyzed and discussed in detail. The rest of this paper is organized as follows. Section 2 describes several classical causality analysis models, which are also the comparison methods of this paper. Section 3 introduces the proposed HSIC-Lasso-GC model. Section 4 presents the experiments on benchmarks and real-world systems. Finally, the conclusion is given in Section 5.

2. Related Granger causality tests

Granger causality model has played an important role in time series analysis. In this section, we will analyze the basic principles and several extensions of Granger causality test.

2.1. Granger causality index

The concept of Granger causality was first proposed by C.W.J. Granger [20], the Nobel Laureate in Economics. Consider two time series X and Y with length n, and they are assumed to be jointly stationary. Time series X is said to Granger-cause (G-cause) time series Y if the prediction results of the future information of Y based on the past information of X and Y are better than the results only based on the past information of Y. In order to achieve Granger causality, we establish the VAR models as follows

$$Y_t = \sum_{i=1}^{p} \alpha_i Y_{t-i} + \varepsilon_{Y,t} \tag{1}$$

$$Y_t = \sum_{i=1}^{p} a_i X_{t-i} + \sum_{i=1}^{p} b_i Y_{t-i} + \varepsilon_{Y|X,t} \tag{2}$$

where p is the order of VAR model, α_i, a_i and b_i are coefficients of the models, and $\varepsilon_{Y,t}$ and $\varepsilon_{Y|X,t}$ are residual errors of the models. According to the basic idea of Granger causality, we can infer the Granger causality from X to Y, expressed as $X \to Y$, by comparing the residual errors of (1) and (2). The definition of Granger causality index (GCI) is as follows

$$\mathrm{GCI}_{X\to Y} = \ln \frac{\mathrm{var}\left(\varepsilon_Y\right)}{\mathrm{var}\left(\varepsilon_{Y|X}\right)} \tag{3}$$

If $\varepsilon_{Y|X} < \varepsilon_Y$, that is $\mathrm{GCI}_{X\to Y} > 0$, it indicates that X G-causes Y. However, since the traditional Granger causality analysis is based on linear models and only analyzes the causality between two variables, it is greatly limited in practical applications. Therefore, a large number of improved Granger causality models have been proposed.

2.2. Conditional Granger causality index

As the traditional Granger causality model does not consider the confounding effects, the analysis results are prone to false causality when dealing with multivariate system. In order to solve the above problem, Geweke [35] proposed an improved method, conditional Granger causality model, which creates VAR models with conditional variables Z.

$$Y_t = \sum_{i=1}^{p} \alpha_i Y_{t-i} + \sum_{i=1}^{p} \beta_i Z_{t-i} + \varepsilon_{Y|Z,t} \tag{4}$$

$$Y_t = \sum_{i=1}^{p} a_i X_{t-i} + \sum_{i=1}^{p} b_i Y_{t-i} + \sum_{i=1}^{p} c_i Z_{t-i} + \varepsilon_{Y|XZ,t} \tag{5}$$

Similarly, by comparing the residual errors of (4) and (5), we can determine whether there is Granger causality of $X \to Y$. The definition of conditional Granger causality index (CGCI) is as follows

$$\mathrm{CGCI}_{X\to Y|Z} = \ln \frac{\mathrm{var}\left(\varepsilon_{Y|Z}\right)}{\mathrm{var}\left(\varepsilon_{Y|XZ}\right)} \tag{6}$$

If $\mathrm{CGCI}_{X\to Y|Z} > 0$, we can conclude that X G-causes Y. In addition, Z is the representative of conditional variables, which may be more than one variable. Therefore, CGCI can effectively deal with the causality analysis of multivariate systems.

2.3. Lasso-Granger causality model

For multivariate systems, GCI and CGCI need to detect causality between any two variables, resulting in high computational complexity. In order to efficiently analyze high-dimensional time series, Arnold et al. [37] proposed the Lasso-Granger causality (Lasso-GC) model. The method first establishes the Lasso regression model of all input variables, and then identifies the causality based on the model coefficients. The Lasso regression model [52] is based on L1-regularization and its objective function is as follows

$$\min \left\{ \frac{1}{n} \left\| \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\alpha} \right\|_2^2 + \lambda \left\| \boldsymbol{\alpha} \right\|_1 \right\} \tag{7}$$

where $\boldsymbol{Y} \in \mathbb{R}^n$ represents the output vector of length n, $\boldsymbol{X} \in \mathbb{R}^{n\times d}$ represents input vectors of d dimensions, $\boldsymbol{\alpha} \in \mathbb{R}^d$ is the model coefficients, and λ is the regularization parameter. If the coefficient $\boldsymbol{\alpha}_j$ of time series X_j is zero or close to zero, there is no causality from X_j to Y. The major advantage of Lasso-GC is that it achieves causality between all input variables and output variable by establishing one regression model, which greatly reduces the amount of calculation.

3. Methodology

In order to achieve nonlinear causality analysis of multivariate time series, we propose a novel Granger causality method, named Hilbert–Schmidt independence criterion Lasso Granger causality (HSIC-Lasso-GC) model. In this section, we first introduce the HSIC-Lasso model. Then, we describe the proposed model in detail.

3.1. Hilbert–Schmidt independence criterion

Hilbert–Schmidt independence criterion (HSIC) [53] is a kernel independence measure that evaluates the statistically independence between two sample sets. Gretton et al. [53] introduce cross-covariance operator in the RKHSs, and define HSIC as the Hilbert–Schmidt norm of cross-covariance operator. Let X and Y be two random variables, of which the marginal probability distributions are P_x, P_y and the joint probability distribution is P_{xy}. Consider two feature maps $\phi: X \rightarrow F$ and $\psi: Y \rightarrow G$, where F, G are RKHSs and $k\left(x, x'\right) = \left\langle \phi(x), \phi(x')\right\rangle$, $l\left(y, y'\right) = \left\langle \psi(y), \psi(y')\right\rangle$ are the kernel functions. The cross-covariance operator $C_{xy}: G \rightarrow F$ is defined as follows

$$C_{xy} = \mathrm{E}_{xy}\left[\left(\phi(x) - \mu_x\right) \otimes \left(\psi(y) - \mu_y\right)\right] \tag{8}$$

where $\otimes$ is the tensor product, $\mu_x = \mathrm{E}_x\phi(x)$, $\mu_y = \mathrm{E}_y\psi(y)$, and E_x, E_y, E_{xy} denote the expectations. Then, the HSIC [54] is defined as the squared Hilbert–Schmidt norm of C_{xy}:

$$\begin{aligned} \mathrm{HSIC}\left(P_{xy}, F, G\right) &= \left\|C_{xy}\right\|_{\mathrm{HS}}^2 \\ &= \mathrm{E}_{xx'yy'}\left[k\left(x, x'\right) l\left(y, y'\right)\right] + \mathrm{E}_{xx'}\left[k\left(x, x'\right)\right] \mathrm{E}_{yy'}\left[l\left(y, y'\right)\right] \\ &\quad -2E_{xy}\left[\mathrm{E}_{x'}\left[k\left(x, x'\right)\right] \mathrm{E}_{y'}\left[l\left(y, y'\right)\right]\right] \end{aligned} \tag{9}$$

where $\mathrm{E}_{xx'yy'}$ is the expectation over independent pairs (x, y) and (x', y'). Given a dataset $Z = \{(x_i, y_i) \,|i = 1, \ldots, n\}$, the empirical estimator of HSIC is as follows

$$\mathrm{HSIC}\,(Z, F, G) = \frac{1}{n^2}\mathrm{tr}\,(\boldsymbol{KHLH}) \triangleq \mathrm{HSIC}\,(\boldsymbol{K}, \boldsymbol{L}) \tag{10}$$

where $\mathrm{tr}\,(\cdot)$ is the trace operator, $\boldsymbol{K}, \boldsymbol{L} \in \mathbb{R}^{n\times n}$ are the kernel matrices with elements $K_{ij} = k\left(x_i, x_j\right)$ and $L_{ij} = l\left(y_i, y_j\right)$, $\boldsymbol{H} = \boldsymbol{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^{\mathrm{T}} \in \mathbb{R}^{n\times n}$ is the centering matrix, $\boldsymbol{I} \in \mathbb{R}^{n\times n}$ is the identity matrix and $\mathbf{1} \in \mathbb{R}^n$ is a vector of ones. HSIC is a popular measure of dependence. The value of HSIC is 0 if and only if X and Y are independent, and a large value indicates high dependence between them.

3.2. HSIC-Lasso model

Suppose we have a dataset $\{(\boldsymbol{x}_i, y_i) \,|i = 1, \ldots, n\}$, where $\boldsymbol{x}_i \in \mathbb{R}^d$ is the input vector and $y_i \in \mathbb{R}$ is the output value. The goal of feature selection is to find m $(m < d)$ related features from d input features, which can explain and predict the output well. The Lasso model is widely used for feature selection between input features and output, which is based on linear regression model. On the basis of Lasso regression model, Yamada et al. [55] proposed a feature-wise nonlinear Lasso based on HSIC for feature selection, called HSIC-Lasso model. First, map the original input and output samples to the RKHSs and get the Gram matrices $\boldsymbol{K}^{(k)} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{L} \in \mathbb{R}^{n\times n}$, where $K_{ij}^{(k)} = K\left(x_{ki}, x_{kj}\right), k = 1, 2, \ldots, d$, $L_{ij} = L\left(y_i, y_j\right)$, and $K\left(x, x'\right)$ and $L\left(y, y'\right)$ are the kernel functions. The objective function of HSIC-Lasso model is as follows

$$\begin{aligned} &\min_{\boldsymbol{\alpha}\in\mathbb{R}^d} \frac{1}{2}\left\|\overline{\boldsymbol{L}} - \sum_{k=1}^{d}\alpha_k\overline{\boldsymbol{K}}^{(k)}\right\|_{Frob}^2 + \lambda\left\|\boldsymbol{\alpha}\right\|_1 \\ &\text{s.t.}\quad \alpha_1, \ldots, \alpha_d \geq 0 \end{aligned} \tag{11}$$

where $\left\|\cdot\right\|_{Frob}$ is the Frobenius norm, $\overline{\boldsymbol{K}}^{(k)} = \boldsymbol{HK}^{(k)}\boldsymbol{H}$, $\overline{\boldsymbol{L}} = \boldsymbol{HLH}$, and $\boldsymbol{H} \in \mathbb{R}^{n\times n}$ is the same central matrix as before. $\boldsymbol{\alpha} \in \mathbb{R}^d$ are the unknown parameters to be solved, which are limited to non-negative values. If $\alpha_k = 0$, it indicates that the corresponding input feature is irrelevant and needs to be removed. The first term in (11) can be rewritten as follows

$$\begin{aligned} \frac{1}{2}\left\|\overline{\boldsymbol{L}} - \sum_{k=1}^{d}\alpha_k\overline{\boldsymbol{K}}^{(k)}\right\|_{Frob}^2 &= \frac{1}{2}\left\|\overline{\boldsymbol{L}}\overline{\boldsymbol{L}}\right\|_{Frob} - \sum_{k=1}^{d}\alpha_k\left\|\overline{\boldsymbol{K}}^{(k)}\overline{\boldsymbol{L}}\right\|_{Frob} + \frac{1}{2}\sum_{k,l=1}^{d}\alpha_k\alpha_l\left\|\overline{\boldsymbol{K}}^{(k)}\overline{\boldsymbol{K}}^{(l)}\right\|_{Frob} \\ &= \frac{1}{2}\mathrm{tr}\left(\overline{\boldsymbol{L}}\overline{\boldsymbol{L}}\right) - \sum_{k=1}^{d}\alpha_k\mathrm{tr}\left(\overline{\boldsymbol{K}}^{(k)}\overline{\boldsymbol{L}}\right) + \frac{1}{2}\sum_{k,l=1}^{d}\alpha_k\alpha_l\mathrm{tr}\left(\overline{\boldsymbol{K}}^{(k)}\overline{\boldsymbol{K}}^{(l)}\right) \\ &= \frac{n^2}{2}\mathrm{HSIC}\left(\overline{\boldsymbol{L}}, \boldsymbol{L}\right) - n^2\sum_{k=1}^{d}\alpha_k\mathrm{HSIC}\left(\overline{\boldsymbol{K}}^{(k)}, \boldsymbol{L}\right) + \frac{n^2}{2}\sum_{k,l=1}^{d}\alpha_k\alpha_l\mathrm{HSIC}\left(\overline{\boldsymbol{K}}^{(k)}, \boldsymbol{K}^{(l)}\right) \end{aligned} \tag{12}$$

where HSIC $\left(\overline{\boldsymbol{L}}, \boldsymbol{L}\right)$ is a constant that can be ignored. $\sum_{k=1}^{d} \alpha_k \text{HSIC}\left(\overline{\boldsymbol{K}}^{(k)}, \boldsymbol{L}\right)$ indicates the relevance of all input features to the output. If the kth input feature and output are independent, the value of HSIC is close to zero and the corresponding coefficient α_k also tends to be zero. $\sum_{k,l=1}^{d} \alpha_k \alpha_l \text{HSIC}\left(\overline{\boldsymbol{K}}^{(k)}, \boldsymbol{K}^{(l)}\right)$ represents the redundancy between all input features. The redundant features get a large value of HSIC, which will be eliminated by HSIC-Lasso. It can be concluded that HSIC-Lasso is related to the popular feature selection method of minimum redundancy maximum relevancy (mRMR) [56]. For the solution of the HSIC-Lasso model, we can use the dual augmented Lagrangian method.

3.3. HSIC-Lasso-GC model

3.3.1. Stationarity test

As we all know, the precondition of Granger causality analysis is that the time series is a stationary process. The most common method in the stationarity test of time series is the unit root test. It has been shown that if the time series possesses a unit root, it is a non-stationary time series. In this paper, augmented Dickey–Fuller (ADF) [57] is adopted to test whether the time series is stationary or not, and its formulation is as follows

$$\Delta X_t = \alpha + \beta t + \delta X_{t-1} + \sum_{i=1}^{p} \beta_i \Delta X_{t-i} + \varepsilon_t \tag{13}$$

where α is a constant, β is the coefficient of time trend, p is the order of autoregressive process, ε_t is an error term. The null hypothesis of ADF is H_0: $\delta = 0$, indicating that the time series is non-stationary. Under this condition, the time series needs to be differenced to make it stationary. If the null hypothesis is rejected and the alternative hypothesis H_1: $\delta < 0$ is accepted, it indicates that the time series is stationary.

3.3.2. State space reconstruction

The state space reconstruction theory proposed by Takens [58] provides a theoretical basis for analyzing the dynamic characteristics of nonlinear systems. Takens' theorem pointed out that if finding the lower bound of the embedding dimension of state space, the dynamics of the original system can be recovered in the state space. The reconstructed state space contains most of the information of the original system. Let $X(t)$ be an observable time series, the reconstructed state space based on Takens' theorem is as follows

$$\mathrm{X}(t) = [X(t), X(t-\tau), \ldots, X(t-(m-1)\tau)] \tag{14}$$

where τ and m are delay time and embedding dimension, respectively. The key issue is how to choose the appropriate delay time and embedding dimension. Kugiumtzis [59] believed that delay time and embedding dimension are related and proposed the concept of the time window $\tau_w = (m-1)\tau$. Kim et al. [60] proposed the C–C method for solving the time window, which can calculate delay time and embedding dimension simultaneously. In this paper, we use the C–C method to determine the delay time and embedding dimension.

3.3.3. Granger causality analysis based on HSIC-Lasso

After stationarity test and state space reconstruction, we get a set of input features $\left[\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_q\right]^{\mathrm{T}} \in \mathbb{R}^{d\times n}$, where q is the number of original time series, d is the number of features after state space reconstruction and n is the number of samples. Then, the input features and the output $\boldsymbol{Y} = [y_1, y_2, \ldots, y_n] \in \mathbb{R}^n$ are mapped to RKHSs, where the kernel functions are selected as Gaussian kernels:

$$K\left(x, x'\right) = \exp\left(-\frac{\left(x-x'\right)^2}{2\sigma_x^2}\right) \tag{15}$$

$$L\left(y, y'\right) = \exp\left(-\frac{\left(y-y'\right)^2}{2\sigma_y^2}\right) \tag{16}$$

Next, we build the HSIC-Lasso regression model, of which the objective function is shown in Eq. (11). In the Lasso regression, the choice of regularization parameter λ severely affects the final result. If the value of regularization parameter is too large, the related features may be removed. Conversely, if the value of regularization parameter is too small, it may result in irrelevant or redundant features being preserved. Therefore, we choose the appropriate regularization parameter based on generalized information criterion (GIC) [61]. GIC is commonly used for model selection, which is defined as follows

$$\mathrm{GIC} = \upsilon M - 2\ln(\mathrm{L}) \tag{17}$$

where L is the likelihood function of the model, and M is the number of parameters. v is a positive number that controls the properties of variable selection and the values of v in [2,6] have been found to obtain good performance [62]. If the error of the model obeys an independent normal distribution, GIC can be calculated as follows

$$\mathrm{GIC} = \upsilon M + n\ln(\mathrm{RSS}) \tag{18}$$

where RSS represents the residual error of the model and n denotes the sample size. To simplify the calculation, we determine the regularization parameter λ by Eq. (18). Then, HSIC-Lasso regression model is built according to the optimal regularization parameter, and the parameters $\boldsymbol{\alpha}$ is obtained for analyzing the Granger causality from X_j ($j = 1, \dots, q$) to Y.

3.3.4. Significance test

In order to analyze the validity of the results of causality analysis, a statistical significance test was introduced. First, the original time series X_j is randomly shifted to obtain permutation time series.

$$\tilde{X}_j = \{x_{s+1}, \dots, x_n, x_1, \dots, x_s\} \tag{19}$$

where s is the shift factor and n is the number of samples. The permutation time series has the same marginal distribution as the original time series and it is independent. Then, the test statistics of the original time series and permutation time series are calculated separately. Finally, a significant test is performed and its specific steps are as follows:

Step 1 According to the time shift method, N sets of permutation time series are obtained.

Step 2 Calculate the test statistics I_j for the original time series X_j and the test statistics $\tilde{I}_j(\cdot)$ for all permutation time series $\tilde{X}_j$ separately.

Step 3 Construct the H_0 hypothesis: there is Granger causality from X_j to Y.

Step 4 Calculate p value, i.e. $p_{value,j} = \frac{1+\left(I_j \geq \tilde{I}_j(\cdot)\right)}{1+N}$.

Step 5 Select a significance level of $\alpha = 0.95$. If $p \leq \alpha$, reject the H_0 hypothesis; otherwise, accept the H_0 hypothesis.

3.3.5. Summary and discussion

We first summarize the implementation steps of the proposed method and then compare the computational complexity of the related algorithms. The realization of HSIC-Lasso-GC model is as follows.

Algorithm 1: HSIC-Lasso-GC

Input: Time series dataset $\{\boldsymbol{X},\boldsymbol{Y}\}$, where $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_q]^{\mathrm{T}} \in \mathbb{R}^{q\times n}$ and $\boldsymbol{Y} = [y_1, y_2, \cdots, y_n] \in \mathbb{R}^n$, q is the number of input variables, and n is the number of samples.

Output: Granger causality analysis results $X_j \to Y, j = 1,2,\cdots,q$.

1. Stationarity test and difference stationary for time series dataset $\{\boldsymbol{X},\boldsymbol{Y}\}$.
2. State space reconstruction: calculate delay time and embedding dimension using C-C method, and obtain the reconstructed state spaces for all input variables $\{\boldsymbol{X}_j(t) \mid j = 1,2,\cdots,q\}$.
3. Build the HSIC-Lasso regression model:
    (a) Map the input and output samples to the RKHSs and get the Gram matrices $\boldsymbol{K}^{(k)}$ and $\boldsymbol{L}$.
    (b) **for** $\lambda = 0:0.02:1$
        a) Solve the parameter $\boldsymbol{\alpha}$ based on objective function (11), and get the residual error $\mathrm{RSS}(\lambda)$;
        b) Calculate the value of $\mathrm{GIC}(\lambda)$ using (18).
    (c) **end for**
    (d) Select the optimal model with smallest GIC and obtain the parameter $\boldsymbol{\alpha}$.
4. Analyze Granger causality based on significance test:
    (a) **for** $j = 1:q$
        a) **if** $p_{value,j} > 0.95$
            Accept the H_0 hypothesis, i.e. there is Granger causality from X_j to Y.
        b) **else if** $p_{value,j} \leq 0.95$
            Reject the H_0 hypothesis, i.e. there is no Granger causality from X_j to Y.
        c) **end if**
    (b) **end for**
5. Output the results of Granger causality analysis.

In practice, we need to deal with large-scale time series dataset with hundreds or thousands of features. Therefore, the computational complexity of the causality analysis algorithm is an important evaluation standard. Next, we analyze the computational complexity of the proposed method and compare it with CGCI and Lasso-GC. Consider the time series dataset $\{\boldsymbol{X}, \boldsymbol{Y}\}$ with n samples, where q is the number of input time series. Our goal is to achieve causality analysis between q input time series and output. After state space reconstruction, we can get d-dimensional input features. For a d-order VAR model, the computational complexity is between $O\left(d^2n^2\right)$ and $O\left(dn\right)$ [37]. Since we need to do model selection, the computational complexity of the model selection is between $O\left(Cd^2n^2\right)$ and $O\left(Cdn\right)$, where C is the number of model selection.

CGCI: For multivariate time series, Granger causality requires analysis between each pair of time series, a total of q times. Thus, the computational complexity is between $O\left(Cqd^2n^2\right)$ and $O\left(Cqdn\right)$.

Lasso-GC: In the Lasso model, the least angle regression is used to solve the model. As it achieves causality analysis between all input features and output by establishing one regression model, the computational complexity of Lasso-GC is between $O\left(Cd^2n^2\right)$ and $O\left(Cdn\right)$.

HSIC-Lasso-GC: Firstly, the computational complexity of Gram matrices is $O\left(dn^2\right)$. Next, the computational complexity of solving HSIC-Lasso model is $O\left(d^3\right)$ [63]. Therefore, the total computational complexity of HSIC-Lasso-GC is $O\left(dn^2 + Cd^3\right)$.

Based on the above analysis, the computational complexity of HSIC-Lasso-GC is significantly reduced compared to CGCI. Compared with the Lasso-GC model, the proposed method adds kernel mapping operation, resulting in an increase in computation. However, the computational complexity of HSIC-Lasso-GC is still within the allowable range, which can solve the causality analysis of high-dimensional time series.

4. Simulations

In this section, two benchmark datasets and two actual datasets are used for simulations. The proposed HSIC-Lasso-GC model is compared with the causality models CGCI [35], conditional Granger causality in the GCCA toolbox [64], Lasso-GC [37], KGC [47] and CE [26]. The implementation steps of the proposed method are shown in Algorithm 1, where the source code of the HSIC-Lasso regression model is derived from the project (https://github.com/jhwjhw0123/HSIC_Lasso_with_optimization) on GitHub.

4.1. Example I: Benchmark datasets

System 1 (example 1 in [31]) is a linear coupled system with 5 variables. The equations are as follows

$$\begin{aligned}
X_{1,t} &= 0.95\sqrt{2}X_{1,t-1} - 0.9025\,X_{1,t-2} + \varepsilon_{1,t} \\
X_{2,t} &= 0.5\,X_{1,t-2} + \varepsilon_{2,t} \\
X_{3,t} &= -0.4\,X_{1,t-3} + \varepsilon_{3,t} \\
X_{4,t} &= -0.5\,X_{1,t-1} + 0.25\sqrt{2}\,X_{4,t-1} + 0.25\sqrt{2}\,X_{5,t-1} + \varepsilon_{4,t} \\
X_{5,t} &= -0.25\sqrt{2}\,X_{4,t-1} + 0.25\sqrt{2}\,X_{5,t-1} + \varepsilon_{5,t}
\end{aligned} \tag{20}$$

where ε is Gaussian white noise with zero mean and unit variance. From the equations, we can get the causalities $X_1 \to X_2, X_1 \to X_3, X_1 \to X_4$ and $X_4 \leftrightarrow X_5$, as shown in Fig. 1(a). Each square in Fig. 1 represents the causality from row to column, where white one indicates a causality.

System 2 (system 4 in [65]) is a nonlinear coupled system with 3 variables. The equations are as follows

$$\begin{aligned}
x_{1,t} &= 3.4x_{1,t-1}\left(1 - x_{1,t-1}\right)^2 \exp\left(-x_{1,t-1}^2\right) + 0.4\varepsilon_{1,t} \\
x_{2,t} &= 3.4x_{2,t-1}\left(1 - x_{2,t-1}\right)^2 \exp\left(-x_{2,t-1}^2\right) + 0.5x_{1,t-1}x_{2,t-1} + 0.4\varepsilon_{2,t} \\
x_{3,t} &= 3.4x_{3,t-1}\left(1 - x_{3,t-1}\right)^2 \exp\left(-x_{3,t-1}^2\right) + 0.3x_{2,t-1} + 0.5x_{1,t-1}^2 + 0.4\varepsilon_{3,t}
\end{aligned} \tag{21}$$

where ε is Gaussian white noise with zero mean and unit variance. From the equations, we can get the causalities $X_2 \to X_3$, $X_1 \to X_2$ and $X_1 \to X_3$, as shown in Fig. 1(b).

In the simulations, we generate 1000 samples for the two systems. The parameter settings for each causality analysis method are shown in Tables 1 and 2, where m represents the and embedding dimension, τ represents the delay time, σ represents the width of Gaussian kernel function and λ represents the regularization parameter. Among them, the model orders for CGCI, GCCA and Lasso-GC are selected by generalized information criterion. According to the selected parameters, the causality analysis results are shown in Figs. 2 and 3. The Figs. 2 and 3 reveal the causality indexes of different methods and each square represents the causality from row to column, where the values are represented in different colors. In order to verify the reliability of the causality analysis results, we generate 99 sets of permutation time series for significance test. The results of significance test are shown in Tables 3 and 4, where the bold content indicates the correct causality.

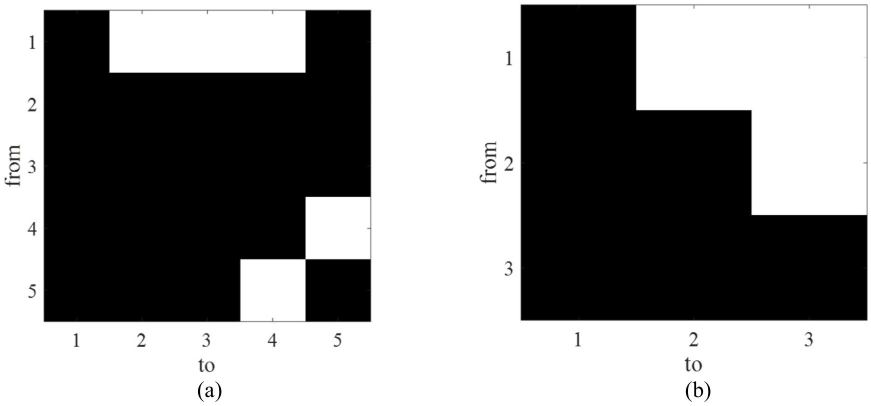


Fig. 1. The causalities of the simulated systems: (a) system 1, (b) system 2.

Table 1
Parameter settings for each causality analysis method of system 1.

Method	CGCI	GCCA	KGC	Lasso-GC	CE	HSIC-Lasso-GC
Parameters	$m = 3$ $\tau = 1$	$m = 3$ $\tau = 1$	$m = 4$ $\tau = 1$ $\sigma = 50$	$m = 6$ $\tau = 1$ $\lambda = 0.02$	$m = 4$ $\tau = 1$	$m = 3$ $\tau = 1$ $\sigma = 0.63$ $\lambda = 0.05$

Table 2
Parameter settings for each causality analysis method of system 2.

Method	CGCI	GCCA	KGC	Lasso-GC	CE	HSIC-Lasso-GC
Parameters	$m = 3$ $\tau = 1$	$m = 3$ $\tau = 1$	$m = 3$ $\tau = 1$ $\sigma = 48$	$m = 5$ $\tau = 1$ $\lambda = 0.1$	$m = 5$ $\tau = 1$	$m = 3$ $\tau = 1$ $\sigma = 2.2$ $\lambda = 0.1$

Table 3
Significance test for causality analysis results of system 1.

	CGCI	GCCA	KGC	Lasso-GC	CE	HSIC-Lasso-GC
$\boldsymbol{X_1 \rightarrow X_2}$	**100**	**100**	**99**	64	**100**	**100**
$\boldsymbol{X_1 \rightarrow X_3}$	**100**	**100**	31	**92**	**100**	**97**
$\boldsymbol{X_1 \rightarrow X_4}$	**100**	**100**	41	7	**100**	**100**
$X_1 \rightarrow X_5$	59	59	0	0	0	22
$X_2 \rightarrow X_1$	0	52	0	98	0	3
$X_2 \rightarrow X_3$	65	44	0	52	0	32
$X_2 \rightarrow X_4$	89	41	99	77	0	14
$X_2 \rightarrow X_5$	70	43	0	55	0	8
$X_3 \rightarrow X_1$	0	44	0	75	0	13
$X_3 \rightarrow X_2$	70	31	0	0	0	18
$X_3 \rightarrow X_4$	72	71	0	0	0	31
$X_3 \rightarrow X_5$	87	45	0	0	0	23
$X_4 \rightarrow X_1$	0	20	0	38	0	15
$X_4 \rightarrow X_2$	58	14	0	77	0	25
$X_4 \rightarrow X_3$	63	58	0	0	0	34
$\boldsymbol{X_4 \rightarrow X_5}$	**100**	**100**	90	0	**100**	**98**
$X_5 \rightarrow X_1$	72	52	0	65	0	22
$X_5 \rightarrow X_2$	63	78	0	51	0	0
$X_5 \rightarrow X_3$	90	96	0	0	0	8
$\boldsymbol{X_5 \rightarrow X_4}$	**100**	**100**	84	82	**100**	**100**

According to the causality analysis results of system 1, it is evident that the CGCI, CE and HSIC-Lasso-GC model obtain the correct results, which include all the true causalities without any false causality. Although the GCCA obtains all the true causalities, it identifies the false causality of $X_5 \rightarrow X_3$. Moreover, the KGC and Lasso-GC not only fail to identify the true causalities, but also obtain several false causalities. Considering the causality analysis results of system 2, it is clear that the GCCA, KGC and HSIC-Lasso-GC model can identify the true causalities and avoid false causalities. The CGCI and CE get the true causalities, but contain false causalities. In addition, Lasso-GC model fails to analyze nonlinear causality of

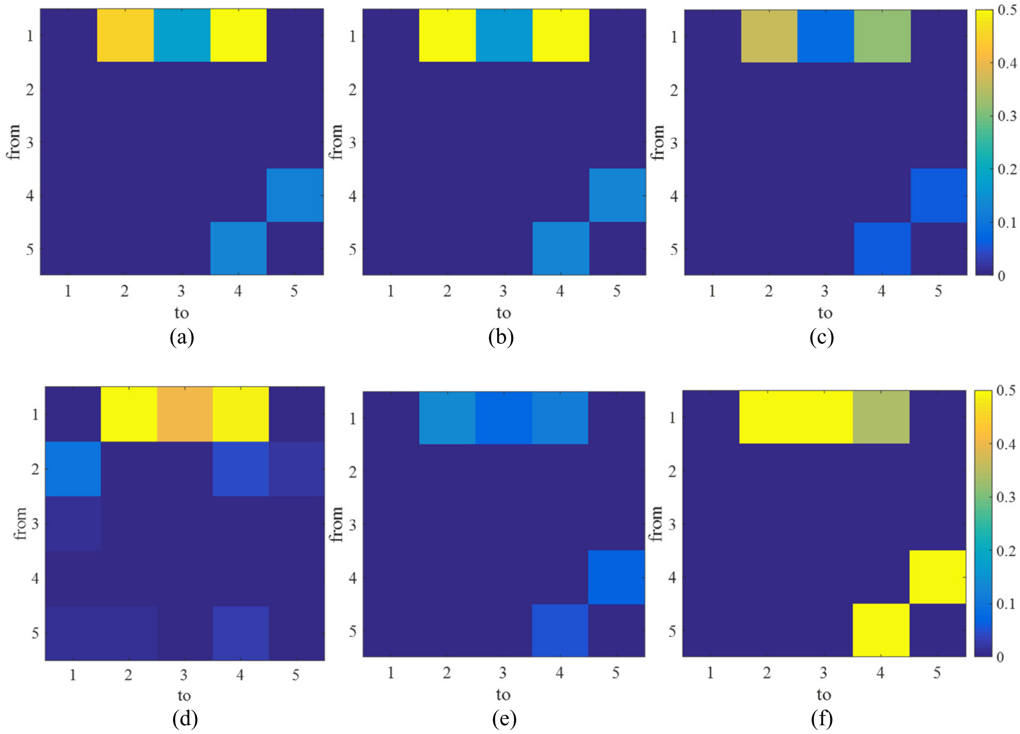


Fig. 2. The causality analysis results of system 1: (a) CGCI, (b) GCCA, (c) KGC, (d) Lasso-GC, (e) CE, (f) HSIC-Lasso-GC.

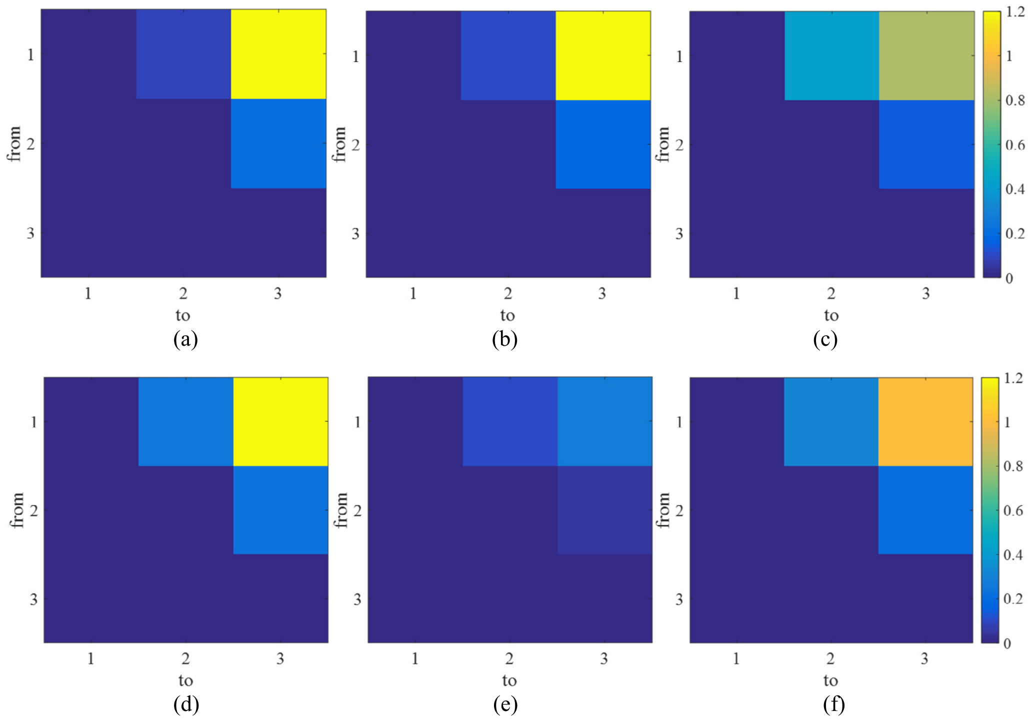


Fig. 3. The causality analysis results of system 2: (a) CGCI, (b) GCCA, (c) KGC, (d) Lasso-GC, (e) CE, (f) HSIC-Lasso-GC.

system 2. From the above analysis, we conclude that the proposed HSIC-Lasso-GC model can identify linear and nonlinear causality in multivariable systems and performs well in benchmark simulations.

Table 4
Significance test for causality analysis results of system 2.

	CGCI	GCCA	KGC	Lasso-GC	CE	HSIC-Lasso-GC
$\boldsymbol{X_1 \rightarrow X_2}$	**100**	**100**	**100**	**99**	**100**	**100**
$\boldsymbol{X_1 \rightarrow X_3}$	**100**	**100**	**100**	73	**100**	**100**
$X_2 \rightarrow X_1$	99	35	0	0	0	5
$\boldsymbol{X_2 \rightarrow X_3}$	**99**	**98**	**100**	60	**100**	**100**
$X_3 \rightarrow X_1$	100	42	0	0	100	6
$X_3 \rightarrow X_2$	56	26	0	0	100	16

Table 5
Embedding dimensions and time delays for AQI and meteorological time series.

Index	1	2	3	4	5	6	7	8	9	10	11
Features	PM2.5	PM10	O_3	NO_2	SO_2	CO	Temperature	Dew point	Pressure	Humidity	Wind speed
m	2	2	3	2	5	5	3	2	2	3	4
τ	4	4	4	2	2	2	1	4	3	2	3

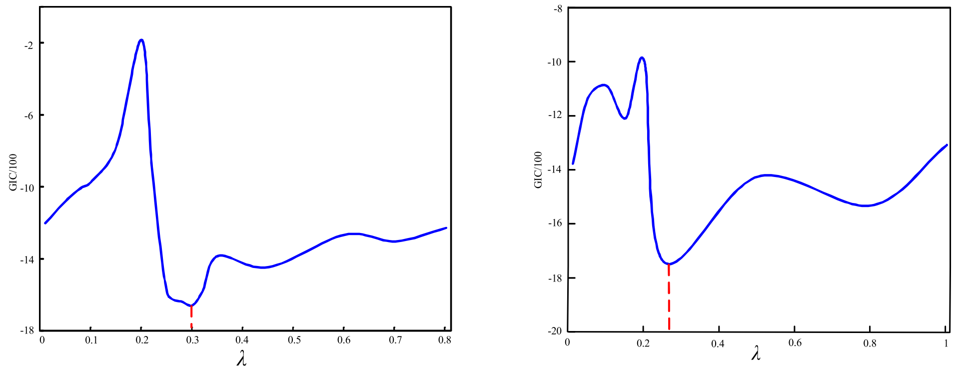


Fig. 4. Model selection of HSIC-Lasso: (a) PM2.5, (b) PM10.

4.2. Example II: AQI-meteorological time series of Haidian District, Beijing

With the rapid development of the economy, air pollution represented by haze has become one of the major environmental problems. However, the cause of haze is very complicated. Taking the air quality index PM2.5 as an example, its concentration is not only affected by atmospheric pollutants such as NO_2, CO, O_3 and SO_2, but also affected by meteorological factors such as temperature, pressure, humidity and wind speed. If we can analyze the causes and major pollutants of PM2.5, it will play an important role in controlling air pollution and developing management policies. The causality analysis method can solve the above problem by identifying the causality between PM2.5 and the dependent variables based on observed time series. Therefore, we apply the proposed method to study the causality between AQI and meteorological time series. The dataset is from Haidian District, Beijing, containing 646 samples from July 2014 to February 2015. The features of the dataset are shown in Table 5, including six-dimensional AQI time series and five-dimensional meteorological time series. In the simulation, we focus on the influence of other variables on PM2.5 and PM10 time series.

After stationarity test and difference stationary, we calculate time delay and embedding dimension for each time series using C–C method. The parameters of state space reconstruction are shown in Table 5. According to the implementation steps of HSIC-Lasso-GC model, we conduct model selection based on GIC and the results are shown in Fig. 4.

From Fig. 4, we select regularization parameters $\lambda = 0.3$ and $\lambda = 0.27$ to build the HSIC-Lasso models for PM2.5 and PM10, respectively. Then, we can obtain the causality analysis results based on the coefficient matrix $\boldsymbol{\alpha}$. The causality analysis results for all comparison methods are shown in Fig. 5. The causality analysis results shown in Fig. 5 express causalities from rows to columns, where the values are represented by different colors. According to the significance test, we get the causality analysis results of all comparison methods for PM2.5 and PM10 time series, as shown in Tables 6 and 7.

Next, we establish prediction models for the PM2.5 and PM10 time series to verify the effectiveness of causality analysis results. We choose the echo state network (ESN) [66] as the predictive model, which is commonly used for time series modeling. We select the first 75% of samples in the original dataset for training and the remaining samples for testing. The

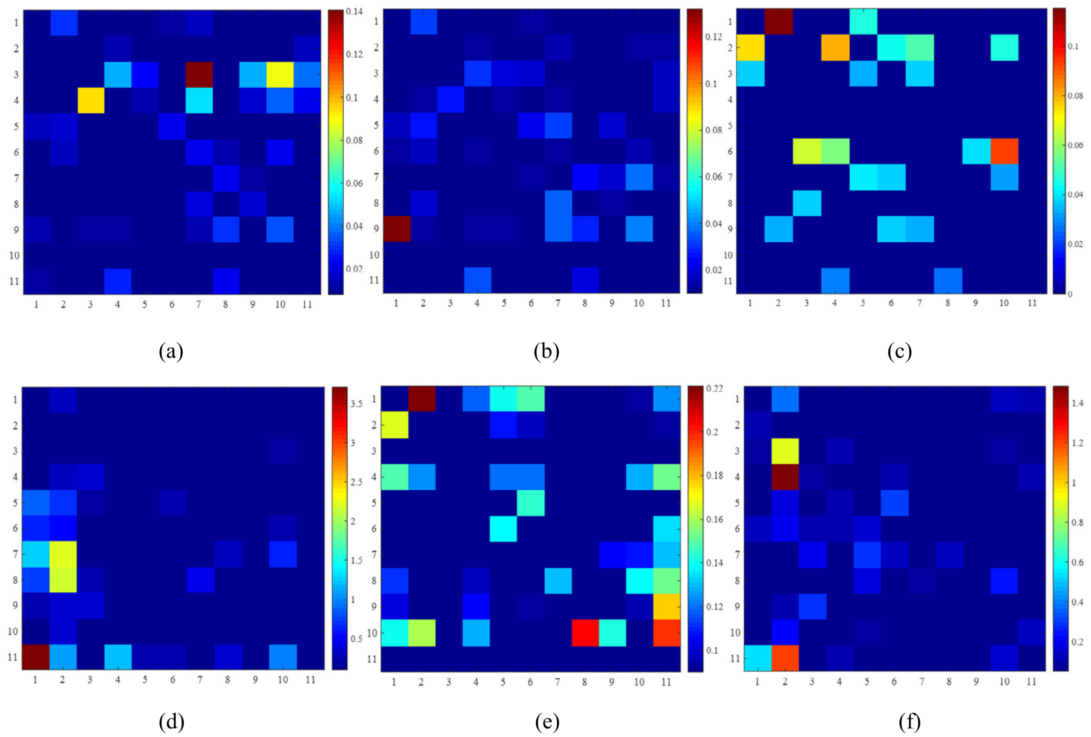


Fig. 5. The causality analysis results of AQI-meteorological time series: (a) CGCI, (b) GCCA, (c) KGC, (d) Lasso-GC, (e) CE, (f) HSIC-Lasso-GC.

Table 6
The causality analysis results of PM2.5 time series.

Methods	Dependent variables
CGCI	SO_2, Pressure, Wind speed
GCCA	SO_2, CO, Pressure
KGC	PM10, O_3
Lasso-GC	SO_2, Temperature, Wind speed
CE	PM10, NO_2, Dew point, Pressure, Humidity
HSIC-Lasso-GC	PM10, CO, Wind speed

Table 7
The causality analysis results of PM10 time series.

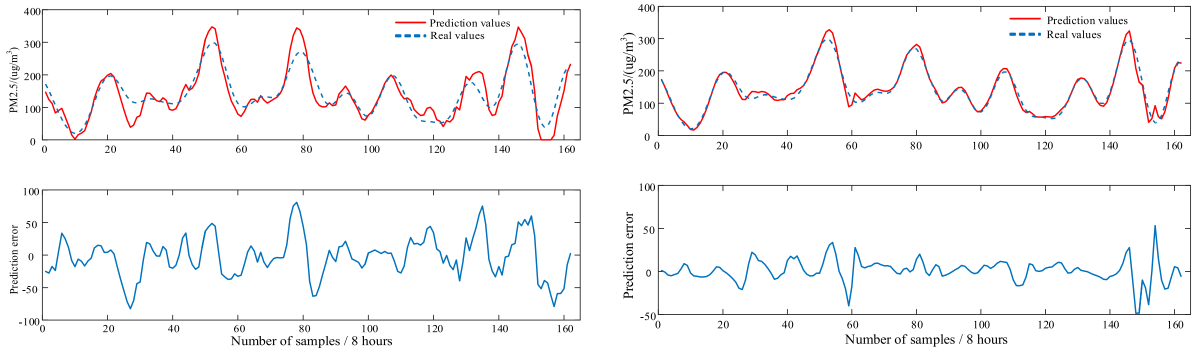
Methods	Dependent variables
CGCI	PM2.5, SO_2, CO
GCCA	PM2.5, NO_2, SO_2, CO, Pressure, Humidity
KGC	PM2.5, Pressure
Lasso-GC	SO_2, Temperature, Dow point, Wind speed
CE	PM2.5, NO_2, Humidity
HSIC-Lasso-GC	O_3, NO_2, Wind speed

prediction results are evaluated by two indicators, namely root mean square error (RMSE) and symmetric mean absolute percentage error (SMAPE):

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{22}$$

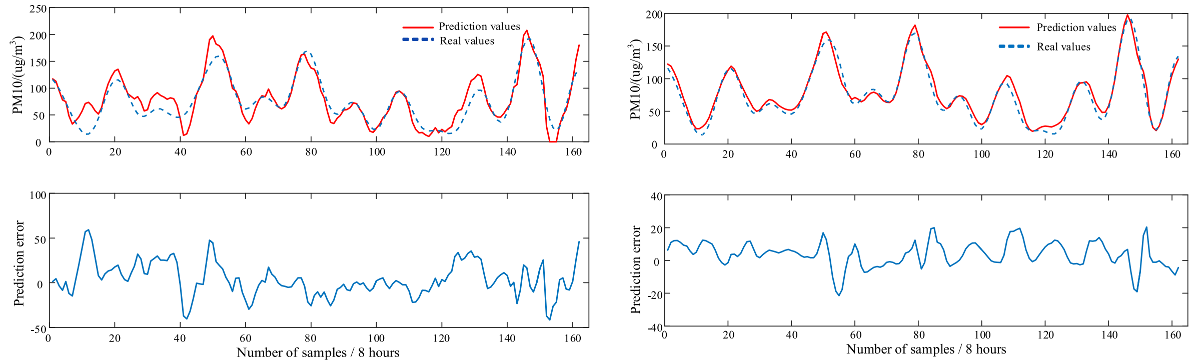
$$\mathrm{SMAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|y_i - \hat{y}_i\right|}{\left(\left|y_i\right| + \left|\hat{y}_i\right|\right)/2} \tag{23}$$

where y_i and $\hat{y}_i$ denote the real values and prediction values, respectively, n is the number of samples. 20 independent experiments are conducted for all comparison algorithms and the results are average values of 20 experiments.



(a) (b)

Fig. 6. The prediction results of PM2.5 time series: (a) All variables, (b) HSIC-Lasso-GC.



(a) (b)

Fig. 7. The prediction results of PM10 time series: (a) All variables, (b) HSIC-Lasso-GC.

Table 8
Comparison of prediction results of PM2.5 time series.

Methods	Input variables	RMSE	SMAPE
–	All variables	52.2034	0.2114
CGCI	1, 5, 9, 11	49.0525	0.1733
GCCA	1, 5, 6, 9	34.4513	0.1270
KGC	1, 2, 3	22.7686	0.0906
Lasso-GC	1, 5, 7, 11	48.3912	0.1801
CE	1, 2, 4, 8, 9, 10	26.2778	0.1098
HSIC-Lasso-GC	1, 2, 6, 11	12.7031	0.0566

Table 9
Comparison of prediction results of PM10 time series.

Methods	Input variables	RMSE	SMAPE
–	All variables	27.7567	0.2463
CGCI	1, 2, 5, 6	44.6675	0.2309
GCCA	1, 2, 4, 5, 6, 9, 10	30.7569	0.1792
KGC	1, 2, 9	12.7326	0.1150
Lasso-GC	2, 5, 7, 8, 11	24.2876	0.2278
CE	1, 2, 4, 10	15.0809	0.1080
HSIC-Lasso-GC	2, 3, 4, 11	10.8675	0.0857

Figs. 6 and 7 show the prediction and error curves of PM2.5 and PM10 time series. It is clear that the prediction error based on all variables is much larger than the prediction error based on the proposed causality method. The proposed method can select appropriate input variables for prediction model, thereby improving the prediction accuracy. Tables 8 and 9 show the prediction results of PM2.5 and PM10 time series based on all comparison algorithms. Since there are

Table 10
Embedding dimensions and time delays for meteorological and ENSO time series.

Index	1	2	3	4	5	6	7	8	9
Features	Temperature	Precipitation	Pressure	Humidity	Niño 1.2	Niño 3	Niño 3.4	Niño 4	SOI
m	2	2	2	2	5	5	2	2	3
τ	2	2	2	3	2	2	5	6	3

Table 11
The causality analysis results of temperature time series.

Methods	Dependent variables
CGCI	Precipitation, Pressure, Humidity, Niño 1.2, Niño 3, Niño 3.4, Niño 4, SOI
GCCA	Precipitation, Pressure, Humidity, Niño 1.2, Niño 3, Niño 3.4, Niño 4, SOI
KGC	Precipitation, Pressure, Niño 1.2, Niño 4
Lasso-GC	Pressure, Humidity, Niño 1.2, Niño 3, Niño 3.4, SOI
CE	Pressure
HSIC-Lasso-GC	Precipitation, Pressure, Humidity, Niño 1.2, Niño 3

complex nonlinear relationships between AQI and meteorological time series, causality methods based on linear models, i.e. CGCI, GCCA and Lasso-GC, are difficult to obtain ideal results, which cannot improve the prediction results. In contrast, the nonlinear causality methods, i.e. KGC, CE and HSIC-Lasso-GC, greatly improve the prediction results. Among them, the prediction errors of HSIC-Lasso-GC are smallest, indicating the effectiveness of the proposed method. Therefore, we conclude that the HSIC-Lasso-GC model can effectively identify nonlinear causality of multivariate systems and has a good prospect in practical applications.

Then, we further analyze the results of causality analysis. Many scholars have studied and concluded that pollutants and meteorological factors have an effect on PM2.5 concentration. For example, Chen et al. [67] investigated the causality of meteorological factors on PM2.5 concentration in the Jing-Jin-Ji region and found that wind, humidity and daily sunshine duration are the most influential meteorological factors for PM2.5 concentration in winter. In the comparison algorithms of this paper, some of them produce obvious false causalities, such as the KGC recognizes CO $\rightarrow$ Humidity and CE recognizes NO_2 $\rightarrow$ Wind speed, which do not match the actual situation. Compared with them, it is well known that wind speed plays an important role in the diffusion of atmospheric pollutants [67], and the proposed method can clearly analyze the influence of wind speed on the concentration of PM2.5 and PM10. According to the causality analysis results of our method, we discover that atmospheric pollutants and wind speed seriously affect the PM2.5 and PM10 concentration. We have demonstrated the effectiveness of the above results by the experiments of PM2.5 and PM10 concentration prediction.

4.3. Example III: ENSO and meteorological time series of Guangzhou, China

El Niño southern oscillation (ENSO) phenomenon is the abnormal change of sea surface temperature (SST) and pressure occurring in the equatorial Pacific Ocean, which has a direct impact on the global weather and climate. El Niño refers to abnormal warm SSTs along the equatorial Pacific region, whereas La Niña refers to abnormal cold SSTs in this region. In order to determine the El Niño and La Niña events, the SST indexes of four regions are generally used, which are Niño 1.2, Niño 3, Niño 3.4 and Niño 4. Southern oscillation index (SOI) is defined as the normalized pressure difference between Tahiti and Darwin. SOI is closely related to the occurrence of El Niño and La Niña. A negative SOI corresponds to an El Niño event, whereas a positive SOI corresponds to a La Niña event. There are many studies investigating the impact of ENSO on global climate. It has been proven that ENSO events can cause extreme weather in China, such as floods and droughts [68]. Therefore, while studying the prediction of meteorological time series, it is necessary to consider the impact of ENSO [69]. In the simulation, we select the surface meteorological dataset of Guangzhou, China, from the China Meteorological Data Service Center (http://data.cma.cn/). The meteorological dataset contains the monthly average time series of temperature, precipitation, pressure and humidity from July 1951 to May 2016, for a total of 779 samples. The ENSO dataset is derived from the National Oceanic and Atmospheric Administration (https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/) and contains five-dimensional time series, which are Niño 1.2, Niño 3, Niño 3.4, Niño 4 and SOI. Our goal is to analyze the causality of other variables on temperature and precipitation of Guangzhou, and then build predictive models for temperature and precipitation time series.

First of all, the stationarity test and difference stationary are performed on all the time series. Then, we calculate the time delays and embedding dimensions for the time series based on C–C method, as shown in Table 10. Thereafter, according to the algorithm flow of HSIC-Lasso-GC, we build the HSIC-Lasso regression models for temperature and precipitation time series and select the optimal models with the smallest GIC value. The model selection results are shown in Fig. 8. We select regularization parameters $\lambda = 0.2$ and $\lambda = 0.5$ to build the HSIC-Lasso models for temperature and precipitation time series. Finally, we obtain Granger causality results based on coefficient matrix $\boldsymbol{\alpha}$ and significance test. The causality analysis results of all comparison methods for temperature and precipitation time series are shown in Tables 11 and 12.

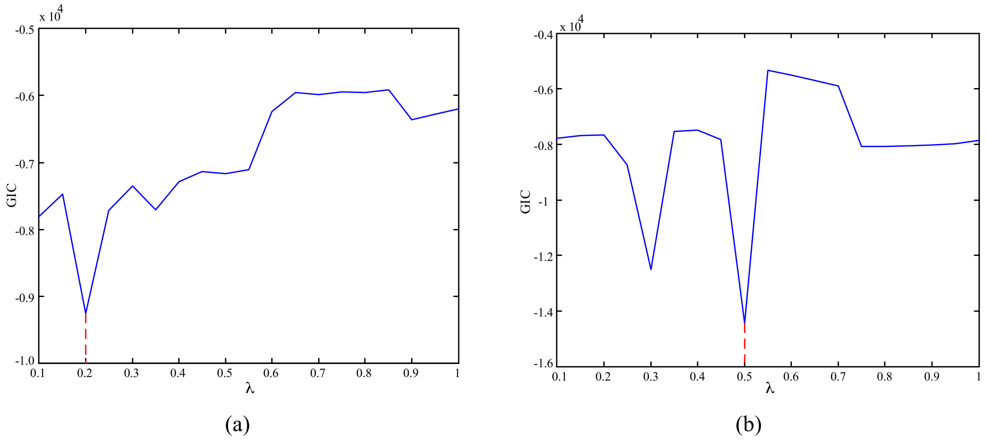


Fig. 8. Model selection of HSIC-Lasso: (a) temperature, (b) precipitation.

Table 12
The causality analysis results of precipitation time series.

Methods	Dependent variables
CGCI	Temperature, Pressure, Humidity, Niño 1.2, Niño 3, Niño 3.4, Niño 4, SOI
GCCA	Temperature, Pressure, Humidity, Niño 1.2, Niño 3, Niño 3.4, Niño 4, SOI
KGC	Humidity, Niño 1.2
Lasso-GC	Niño 3.4, Niño 4, SOI
CE	Temperature, Niño 3.4, SOI
HSIC-Lasso-GC	Pressure, Humidity, Niño 1.2, Niño 3

Table 13
Comparison of prediction results of temperature time series.

Methods	Input variables	RMSE	SMAPE
–	All variables	2.2549	0.0450
CGCI	All variables	2.2549	0.0450
GCCA	All variables	2.2549	0.0450
KGC	1, 2, 3, 5, 8	2.3476	0.0466
Lasso-GC	1, 3, 4, 5, 6, 7, 9	2.2027	0.0441
CE	1, 3	2.5497	0.0484
HSIC-Lasso-GC	1, 2, 3, 4, 5, 6	1.9331	0.0371

Next, based on the results of causality analysis, we separately establish prediction models for temperature and precipitation time series. We also choose ESN as the predictive model and select the first 80% of samples for training and the remaining samples for testing. We perform five-step prediction for temperature and precipitation time series, and the prediction results are shown in Figs. 9 and 10. It is obvious that the prediction results based on the selected variables of HSIC-Lasso-GC method are better than the prediction results based on all variables. Further, Tables 13 and 14 give the quantitative prediction results of all comparison methods. All experimental results are the average of 20 independent tests. Since CGCI and GCCA are linear causality analysis methods, they cannot identify direct nonlinear causalities and select all variables as the inputs of prediction model. After state space reconstruction, the variable dimension of the system has reached 25 dimensions. For the causality analysis method of CE, as the dimension of conditional variables increases, the estimation accuracy of the CE decreases, resulting in poor causality analysis results. Therefore, it can be discovered that CE is not suitable for analyzing high-dimensional systems. Compared with other methods, the proposed HSIC-Lasso-GC method achieves the best prediction results. It is a powerful tool for analyzing the nonlinear causality of multivariable systems.

In addition, combined with the actual background of the study area, we further discuss the application value of the causality analysis results. According to the causality analysis results of the proposed method, we find that not only the meteorological factors affect temperature and precipitation, but also Niño 1.2 and Niño 3 directly affect them. It is clear that ENSO has a great influence on weather changes. The prediction results of temperature and precipitation time series also confirm the validity of the above results. On the basis of the existing researches and news reports, it can be found that the El Niño event has an important impact on the temperature and precipitation in Guangzhou [70]. Most of the previous research results are based on statistical methods, which are consistent with the experimental results in this paper. Therefore, the proposed causality analysis model can provide a new and effective research methods for climate change and has broad prospects in natural disaster forecasting.

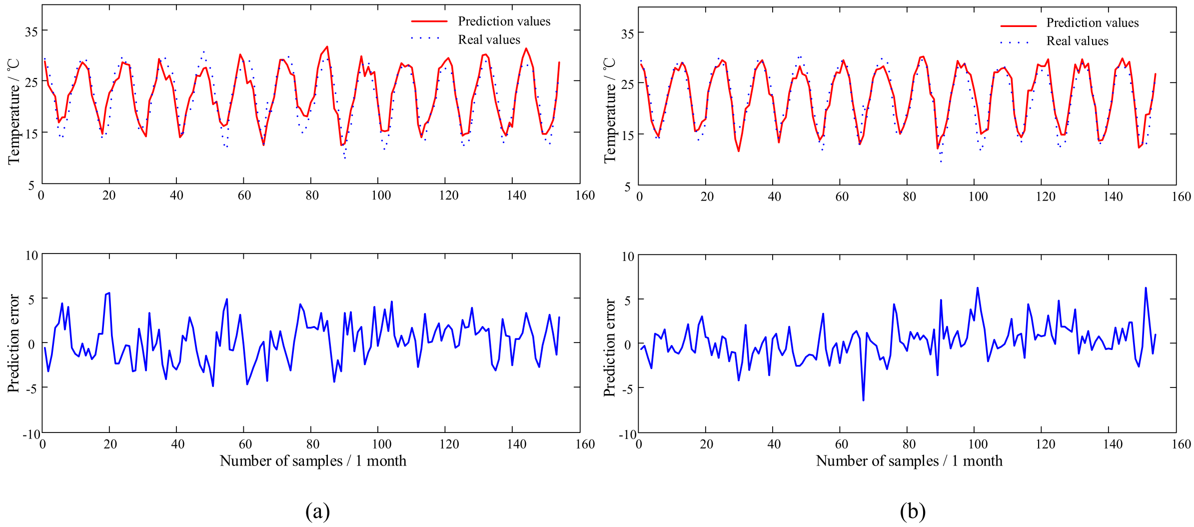


Fig. 9. The five-step prediction results of temperature time series: (a) All variables, (b) HSIC-Lasso-GC.

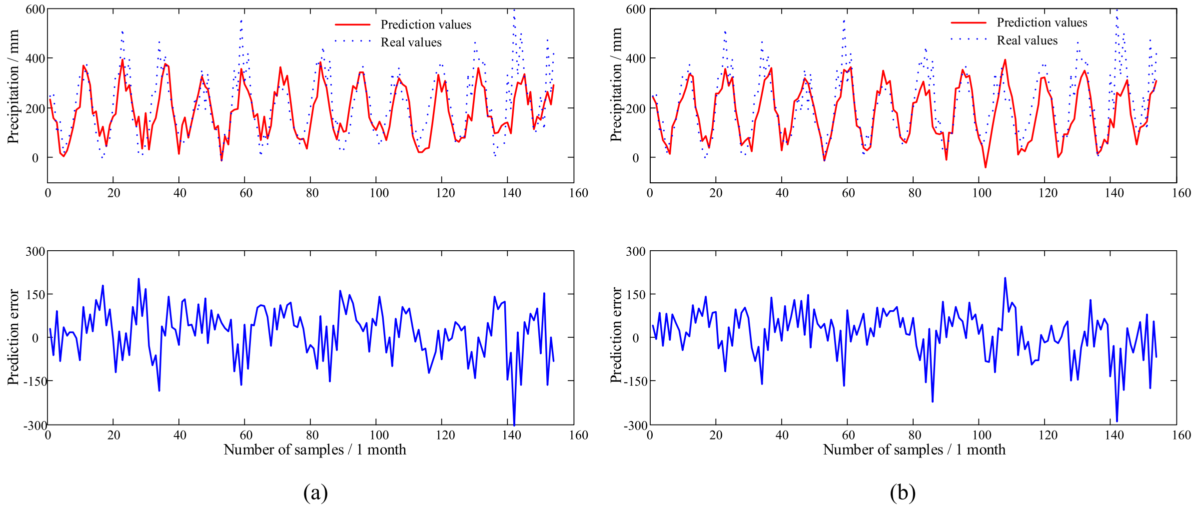


Fig. 10. The five-step prediction results of precipitation time series: (a) All variables, (b) HSIC-Lasso-GC.

Table 14
Comparison of prediction results of precipitation time series.

Methods	Input variables	RMSE	SMAPE
–	All variables	86.8495	0.5853
CGCI	All variables	86.8495	0.5853
GCCA	All variables	86.8495	0.5853
KGC	2, 4, 5	84.5805	0.4219
Lasso-GC	2, 7, 8, 9	94.8360	0.5061
CE	1, 2, 7, 9	83.9480	0.3694
HSIC-Lasso-GC	2, 3, 4, 5, 6	80.0451	0.3272

5. Conclusion

As big data analysis brings new challenges, causality analysis has become an important part of time series data mining. In this paper, we propose the Hilbert–Schmidt independence criterion Lasso Granger causality (HSIC-Lasso-GC) model, which can effectively analyze nonlinear causality between multivariate time series. Through theoretical analysis and experimental verification, we found that the proposed method has the following advantages:

(1) The proposed method can efficiently achieve causality analysis of multivariate time series. It performs state space reconstruction on multivariate time series, which can provide the potential historical information of all input variables for causality analysis. Through the above operation, it avoids the process of input variable selection and the determination of model order, which improves the computational efficiency of the regression model. Furthermore, by establishing the Lasso regression model, the proposed method can simultaneously conduct causality analysis of multiple input variables on the output variable, which is suitable for solving high-dimensional problems.
(2) The proposed method can successfully identify nonlinear causality. The HSIC-Lasso model maps the input and output samples to the RKHSs and finds the non-redundant input features with strong statistical dependence on output feature. Therefore, it is able to achieve nonlinear causality analysis of multivariate time series. Experimental results of benchmark datasets show that the proposed method identifies all the true causalities without any false causality based on significance test.
(3) The proposed method can explain the dynamic characteristics of unknown complex systems, which has been applied to actual multivariable systems with nonlinear relationships. According to the experimental results of AQI-meteorological time series, we discover that atmospheric pollutants and wind speed seriously affect the PM2.5 and PM10 concentration. In addition, the results of causality analysis can be used to select appropriate input variables, which conducive to establishing accurate prediction models for PM2.5 and PM10. In the simulation of ENSO-meteorological time series, we find that not only the meteorological factors affect temperature and precipitation, but also Niño 1.2 and Niño 3 directly affect them. Therefore, the HSIC-Lasso model is a powerful tool for analyzing the nonlinear causality of multivariable actual systems.

Although the proposed HSIC-Lasso-GC model can successfully identify nonlinear causality between multivariate time series, there are still some issues that need to be solved. First, as the choice of delay variables seriously affects the results of causality analysis, it is critical to study the appropriate input variable selection and non-uniform embedding method, which helps to improve the accuracy and computational efficiency of causality analysis. Second, the nonlinear causality model has many unknown parameters to be determined, such as the type and parameters of kernel function, regularization parameter, etc. Proposing efficient optimization strategies and optimization goals is important to ensure the effectiveness of causality models. Besides, most of the existing causality analysis models are based on the precondition of stationarity. Studying the causality analysis of non-stationary time series is one of the key research topics in the future. Finally, multivariate dynamic system usually has multiple temporal scales, so multiscale Granger causality analysis is also a future research trend.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61773087) and the Fundamental Research Funds for the Central Universities (DUT18RC(6)005). The authors want to thank the anonymous reviewers for their constructive comments, which have played an important role in the improvement of our paper. We also want to thank Prof. Makoto Yamada and Dr. Chen Wang for providing the source code of HSIC-Lasso model.

References

[1] Q. Wang, Multifractal characterization of air polluted time series in China, Physica A 514 (2019) 167–180.
[2] H. Hassani, X. Huang, R. Gupta, M. Ghodsi, Does sunspot numbers cause global temperatures? A reconsideration using non-parametric causality tests, Physica A 460 (2016) 54–65.
[3] S. Baek, D.Y. Kim, Empirical sensitivity analysis of discretization parameters for fault pattern extraction from multivariate time series data, IEEE Trans. Cybern. 47 (5) (2017) 1198–1209.
[4] B. Lindner, L. Auret, M. Bauer, J.W. Groenewald, Comparative analysis of granger causality and transfer entropy to present a decision flow for the application of oscillation diagnosis, J. Process Control 79 (2019) 72–84.
[5] W. Ren, M. Han, Classification of EEG signals using hybrid feature extraction and ensemble extreme learning machine, Neural Process. Lett. (2018) http://dx.doi.org/10.1007/s11063-018-9919-0.
[6] H. Bakhshayesh, S.P. Fitzgibbon, A.S. Janani, T.S. Grummett, K.J. Pope, Detecting connectivity in EEG: A comparative study of data-driven effective connectivity measures, Comput. Biol. Med. 111 (2019) 103329.
[7] A.A. Rafindadi, Does the need for economic growth influence energy consumption and CO2 emissions in Nigeria? Evidence from the innovation accounting test, Renew. Sustain. Energy Rev. 62 (2016) 1209–1225.
[8] A. Papana, C. Kyrtsou, D. Kugiumtzis, C. Diks, Financial networks based on granger causality: A case study, Physica A 482 (2017) 65–73.
[9] F. Wen, J. Xiao, C. Huang, X. Xia, Interaction between oil and US dollar exchange rate: nonlinear causality, time-varying influence and structural breaks in volatility, Appl. Econ. 50 (3) (2018) 319–334.
[10] A. Azqueta-Gavaldón, CaUsal inference between cryptocurrency narratives and prices: Evidence from a complex dynamic ecosystem, Physica A 537 (2019) 122574.
[11] A.A. Rafindadi, I. Ozturk, Natural gas consumption and economic growth nexus: Is the 10th Malaysian plan attainable within the limits of its resource?, Renew. Sustain. Energy Rev. 49 (2015) 1221–1232.

[12] J.Y. Zhu, C. Sun, V.O. Li, An extended spatio-temporal granger causality model for air quality estimation with heterogeneous urban big data, IEEE Trans. Big Data 3 (3) (2017) 307–319.
[13] Z. Chen, X. Xie, J. Cai, D. Chen, B. Gao, B. He, N. Cheng, B. Xu, Understanding meteorological influences on PM 2.5 concentrations across China: a temporal and spatial perspective, Atmos. Chem. Phys. 18 (8) (2018) 5343–5358.
[14] G.R. Wu, F. Chen, D. Kang, X. Zhang, D. Marinazzo, H. Chen, Multiscale causal connectivity analysis by canonical correlation: theory and application to epileptic brain, IEEE Trans. Biomed. Eng. 58 (11) (2011) 3088–3096.
[15] P. Van Mierlo, M. Papadopoulou, E. Carrette, P. Boon, S. Vandenberghe, K. Vonck, D. Marinazzo, Functional brain connectivity from EEG in epilepsy: Seizure prediction and epileptogenic focus localization, Prog. Neurobiol. 121 (2014) 19–35.
[16] A.A. Rafindadi, I. Ozturk, Effects of financial development, economic growth and trade on electricity consumption: Evidence from post-fukushima Japan, Renew. Sustain. Energy Rev. 54 (2016) 1073–1084.
[17] A.A. Rafindadi, I. Ozturk, Impacts of renewable energy consumption on the german economic growth: Evidence from combined cointegration test, Renew. Sustain. Energy Rev. 75 (2017) 1130–1141.
[18] A.A. Rafindadi, A.S. Mika'Ilu, Sustainable energy consumption and capital formation: Empirical evidence from the developed financial market of the United Kingdom, Sustain. Energy Technol. Assess. 35 (2019) 265–277.
[19] A.A. Rafindadi, O. Usman, Globalization, energy use, and environmental degradation in South Africa: Startling empirical evidence from the maki-cointegration test, J. Environ. Manag. 244 (2019) 265–275.
[20] C.W.J. Granger, Investigating causal relations by econometric models and cross-spectral methods, Econometrica (1969) 424–438.
[21] X. Huang, P.M. Maçaira, H. Hassani, F.L.C. Oliveira, G. Dhesi, Hydrological natural inflow and climate variables: Time and frequency causality analysis, Physica A 516 (2019) 480–495.
[22] A. Porta, L. Faes, Wiener–Granger causality in network physiology with applications to cardiovascular control and neuroscience, Proc. IEEE 104 (2) (2016) 282–309.
[23] T. Schreiber, Measuring information transfer, Phys. Rev. Lett. 85 (2) (2000) 461.
[24] M. Staniek, K. Lehnertz, Symbolic transfer entropy, Phys. Rev. Lett. 100 (15) (2008) 158101.
[25] A. Papana, C. Kyrtsou, D. Kugiumtzis, C. Diks, Detecting causality in non-stationary time series using partial symbolic transfer entropy: evidence in financial data, Comput. Econ. 47 (3) (2016) 341–365.
[26] L. Faes, G. Nollo, A. Porta, Information-based detection of nonlinear granger causality in multivariate processes via a nonuniform embedding technique, Phys. Rev. E 83 (5) (2011) 051112.
[27] K. Hlaváčková-Schindler, M. Paluš, M. Vejmelka, Causality detection based on information-theoretic approaches in time series analysis, Phys. Rep. 441 (1) (2007) 1–46.
[28] G. Sugihara, R. May, H. Ye, C.H. Hsieh, E. Deyle, M. Fogarty, S. Munch, Detecting causality in complex ecosystems, Science (2012) 1227079.
[29] B. Cummins, T. Gedeon, K. Spendlove, On the efficacy of state space reconstruction methods in determining causality, SIAM J. Appl. Dyn. Syst. 14 (1) (2015) 335–381.
[30] X. Huang, H. Hassani, M. Ghodsi, Z. Mukherjee, R. Gupta, Do trend extraction approaches affect causality detection in climate change studies?, Physica A 469 (2017) 604–624.
[31] C. Zou, J. Feng, Granger causality vs. dynamic Bayesian network inference: a comparative study, BMC Bioinformatics 10 (1) (2009) 122.
[32] L. Yin, X. Ma, Causality between oil shocks and exchange rate: A Bayesian, graph-based VAR approach, Physica A 508 (2018) 434–453.
[33] S. Kleinberg, G. Hripcsak, A review of causal inference for biomedical informatics, J. Biomed. Inform. 44 (6) (2011) 1102–1112.
[34] L. Barnett, A.K. Seth, The MVGC multivariate granger causality toolbox: a new approach to Granger-causal inference, J. Neurosci. Methods 223 (2014) 50–68.
[35] J. Geweke, Measurement of linear dependence and feedback between multiple time series, J. Amer. Statist. Assoc. 77 (378) (1982) 304–313.
[36] E. Siggiridou, D. Kugiumtzis, Granger causality in multivariate time series using a time-ordered restricted vector autoregressive model, IEEE Trans. Signal Process. 64 (7) (2016) 1759–1773.
[37] A. Arnold, Y. Liu, N. Abe, Temporal causal modeling with graphical granger methods, in: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining San Jose, California, USA, 2007, pp. 66–75.
[38] A. Shojaie, G. Michailidis, Discovering graphical Granger causality using the truncating lasso penalty, Bioinformatics 26 (18) (2010) i517–i523.
[39] A. Bolstad, B.D. Van Veen, R. Nowak, Causal network inference via group sparse regularization, IEEE Trans. Signal Process. 59 (6) (2011) 2628–2641.
[40] G.X. Yang, W. Lin, X.F. Wang, Reconstruction of complex directional networks with group lasso nonlinear conditional granger causality, Sci. Rep. 7 (1) (2017) 2991.
[41] Z. Bai, W.K. Wong, B. Zhang, Multivariate linear and nonlinear causality tests, Math. Comput. Simul. 81 (1) (2010) 5–17.
[42] C. Diks, M. Wolski, Nonlinear granger causality: guidelines for multivariate analysis, J. Appl. Econometrics 31 (7) (2016) 1333–1351.
[43] C. Diks, V. Panchenko, A new statistic and practical guidelines for nonparametric Granger causality testing, J. Econom. Dynam. Control 30 (9–10) (2006) 1647–1669.
[44] M. Hu, H. Liang, A copula approach to assessing granger causality, NeuroImage 100 (2014) 125–134.
[45] S. Nazlioglu, World oil and agricultural commodity prices: Evidence from nonlinear causality, Energy Policy 39 (5) (2011) 2935–2943.
[46] N. Ancona, D. Marinazzo, S. Stramaglia, Radial basis function approach to nonlinear Granger causality of time series, Phys. Rev. E 70 (5) (2004) 056221.
[47] D. Marinazzo, M. Pellicoro, S. Stramaglia, Kernel method for nonlinear Granger causality, Phys. Rev. Lett. 100 (14) (2008) 144103.
[48] A. Montalto, S. Stramaglia, L. Faes, G. Tessitore, R. Prevete, D. Marinazzo, Neural networks with non-uniform embedding and explicit validation phase to assess Granger causality, Neural Netw. 71 (2015) 159–171.
[49] P.O. Amblard, O.J. Michel, C. Richard, P. Honeine, A Gaussian process regression approach for testing Granger causality between time series data, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012, pp. 3357–3360.
[50] C. Hiemstra, J.D. Jones, Testing for linear and nonlinear Granger causality in the stock price-volume relation, J. Finance 49 (5) (1994) 1639–1664.
[51] T. Bouezmarni, J.V. Rombouts, A. Taamouti, Nonparametric copula-based test for conditional independence with applications to Granger causality, J. Bus. Econom. Statist. 30 (2) (2012) 275–287.
[52] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Ser. B Stat. Methodol. 58 (1) (1996) 267–288.
[53] A. Gretton, O. Bousquet, A. Smola, B. Schölkopf, Measuring statistical dependence with Hilbert–Schmidt norms, in: International Conference on Algorithmic Learning Theory, Springer, Berlin, Heidelberg, 2005, pp. 63–77.
[54] A. Gretton, K. Fukumizu, C.H. Teo, L. Song, B. Schölkopf, A.J. Smola, A kernel statistical test of independence, in: Advances in Neural Information Processing Systems, 2008, pp. 585–592, Vancouver, British Columbia, Canada.
[55] M. Yamada, W. Jitkrittum, L. Sigal, E.P. Xing, M. Sugiyama, High-dimensional feature selection by feature-wise kernelized lasso, Neural Comput. 26 (1) (2014) 185–207.
[56] H. Peng, F. Long, C. Ding, Feature selection based on mutual information criteria of max-dependency, IEEE Trans. Pattern Anal. Mach. Intell. 27 (8) (2005) 1226–1238.

[57] L.X. Yang, C. Lee, J.J. Su, Behavior of the standard Dickey–Fuller test when there is a fourier-form break under the null hypothesis, Econom. Lett. 159 (2017) 128–133.
[58] F. Takens, Detecting strange attractors in turbulence, in: Dynamical Systems and Turbulence, Springer-Verlag, Heidelberg, Germany, 1981, pp. 366–381.
[59] D. Kugiumtzis, State space reconstruction parameters in the analysis of chaotic time series—the role of the time window length, Physica D 95 (1) (1996) 13–28.
[60] H. Kim, R. Eykholt, J.D. Salas, Nonlinear dynamics, delay times, and embedding windows, Physica D 127 (1–2) (1999) 48–60.
[61] Y. Zhang, R. Li, C.L. Tsai, Regularization parameter selections via generalized information criterion, J. Amer. Statist. Assoc. 105 (489) (2010) 312–323.
[62] P. Stoica, Y. Selen, Model-order selection: a review of information criterion rules, IEEE Signal Process. Mag. 21 (4) (2004) 36–47.
[63] D. He, I. Rish, L. Parida, Transductive HSIC lasso, in: Proceedings of the 2014 SIAM International Conference on Data Mining, 2014, pp. 154–162.
[64] A.K. Seth, A MATLAB toolbox for granger causal connectivity analysis, J. Neurosci. Methods 186 (2) (2010) 262–273.
[65] A. Papana, C. Kyrtsou, D. Kugiumtzis, C. Diks, Simulation study of direct causality measures in multivariate time series, Entropy 15 (7) (2013) 2635–2661.
[66] M.A. Chitsazan, M.S. Fadali, A.M. Trzynadlowski, Wind speed and wind direction forecasting using echo state network with nonlinear functions, Renew. Energy 131 (2019) 879–889.
[67] Z. Chen, J. Cai, B. Gao, B. Xu, S. Dai, B. He, X. Xie, Detecting the causality influence of individual meteorological factors on local PM 2.5 concentration in the Jing-Jin-Ji region, Sci. Rep. 7 (2017) 40735.
[68] P. Zhai, R. Yu, Y. Guo, Q. Li, X. Ren, Y. Wang, W. Xu, Y. Liu, Y. Ding, J. Meteorol. Res. 30 (3) (2015) 283–297.
[69] F. Mekanik, M.A. Imteaz, A. Talei, Seasonal rainfall forecasting by adaptive network-based fuzzy inference system (ANFIS) using large scale climate signals, Clim. Dynam. 46 (9–10) (2016) 3097–3111.
[70] G. Li, J. Chen, X. Wang, Y. Tan, X. Jiang, Modulation of pacific decadal oscillation on the relationship of El Niño with southern China rainfall during early boreal winter, Atmos. Sci. Lett. 18 (8) (2017) 336–341.