



RESEARCH

A new metric for nonlinear causality and its application to the complex system interactions

Yujia Mi · Aijing Lin

Received: 3 January 2025 / Accepted: 23 April 2025 / Published online: 8 May 2025
© The Author(s), under exclusive licence to Springer Nature B.V. 2025

Abstract Identifying and detecting causal relationships in complex systems is a challenging task. In this paper, we propose a new method to quantify causality in complex systems. The basic idea of the method is that if there is a causal relationship between two variables, then historical information about one variable can be a good estimate of the other. The method establishes an attractor with the help of phase space reconstruction, fully considers the nearest neighbor points at each moment and their change patterns, and constructs a weighted change pattern matrix to portray the global properties of the system. Further, we predict the change by cross-prediction and use cosine similarity entropy to measure the similarity between the predicted and true values and quantify the causal strength in this way. In addition, we extend the method to the multivariate case, which helps us to quantify the direct coupling between the two variables, eliminating the effect of indirect coupling on the results of the process. Finally, we apply the method to the analysis of two physiological datasets, which further reveals the interactions in different phys-

iological systems, and also provides new perspectives for studying causality in complex systems and nonlinear time series.

Keywords Multivariate causality · Physiological interaction · Complex system · Signal classification and recognition

1 Introduction

Identifying causal relationships in complex systems can help us understand the internal connections and structure of the system, thus deepening our knowledge of complex systems [1]. Scholars in various fields have been working on causality in different systems, such as economy [2], climate [3], physiology [4] and so on [5].

To date, a variety of methods have been used to identify and quantify interdependencies between variables [6, 7]. Granger causality (GC) [8] for causality in linear systems by measuring the ability to use the prior values of one time series to predict the future values of another time series. From there, GC has also been generalized to nonlinear fields such as nonparametric regression methods [9], local linear prediction [10], kernel estimation methods [11], which are suitable for detecting causality in different situations. However, some special properties in complex systems in the real world make the performance of GC more lackluster. In 2000, the transfer entropy (TE) [12] proposed by Schreiber is based on information theory, which is more appli-

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11071-025-11294-x>.

Y. Mi · A. Lin (✉)
School of Mathematics and Statistics, Beijing Jiaotong University, Haidian District 100044, Beijing, China
e-mail: ajlin@bjtu.edu.cn

Y. Mi
e-mail: 23111518@bjtu.edu.cn

cable to the detection of causality in nonlinear systems. Although the initial TE was based on bivariate expansion, a series of studies have proposed partial TE for identifying indirect coupling [13] to better study the interactions in complex systems, Mao et al. [14] extended TE to multivariate systems, and Stramaglia et al. [15] promoted TE for identifying higher-order causality.

The above methods are undoubtedly effective, but their detection may fail when the variables do not satisfy the separability assumption [16]. Therefore, some methods based on phase space reconstruction have been proposed to remedy this deficiency. The initially proposed cross convergence mapping (CCM) [17] used the correlation coefficient between predicted and true values as a causality measure. Similarly, several studies have generalized CCM to multivariate situations for more accurate portrayal of interactions in complex systems [18–20]. The recently proposed cross convergent sorting (CCS) [21] builds on the CCM by defining causality based on the nature of the reconstructed attractor. The pattern causality proposed by Stavrogloou et al. [22] reveals the nature of the internal system from a new perspective, categorizing causality into positive, negative and dark causality for mining potential interactions. The embedding entropy (EE) [23] defined causality by combining phase space reconstruction and mutual information in a dynamics framework.

Besides that, there are many studies to develop the discussion of causality from different perspectives. Kathpalia [24] proposed a data-based, model-free Intervention Complexity Causality (ICC) to define Compression Complexity Causality (CCC). Later, he extended CCC to PCCC [25] for causality studies in multivariate systems. Moreover, Ying et al. [26] proposed a rigorous framework for detecting causality in nonlinear dynamical systems by defining causality through direct measurement of the continuity scale of the dynamical system under study. Ma et al. [27] proposed cross-mapping smoothing (CMS) for detecting the interactions between variables in a short-term time series. Krakovská [28] proposed a correlation dimension based approach to reveal causal relationships between and within systems.

In addition, some causal inference methods based on graphical models and information theory have shown good performance in different application scenarios. For example, the unique-redundant causal decomposition (SURD) [29] provides a new perspective for

causal analysis of complex systems by decomposing the synergy, uniqueness and redundancy effects among multiple variables. The Peter-Clark (PC) algorithm [30] based on causal structure learning and its extended variants fast causal inference (FCI) [31] and relaxed FCI (RFCI) [32] also play an important role in causal inference. The PC algorithm is a causal inference method based on conditional independence tests, and it is widely used in causal graph modelling. However, it requires the assumption that the data satisfy markovianity and is sensitive to measurement errors. The FCI algorithm takes into account the effect of potential confounding variables on top of PC, and is able to learn causal structures in the presence of potentially unobserved variables. RCFI further improves the computational efficiency of FCI, which makes it more feasible to be applied on large-scale datasets. These methods have been successfully applied to causal inference problems in different domains, especially in dealing with potential confounding variables, measurement errors and complex causal network structures.

In the process of causality research in complex systems, several main problems are faced. Firstly, we have to consider the challenges posed by data indivisibility [22, 23], in particular, TE and GC perform more sluggishly in the face of indivisible data [23]. For this reason, although CCM overcomes this problem to a certain extent, the quantification of causality using Pearson's correlation coefficient, a linear metric, may lead to some errors because the system itself is nonlinear [23]. Secondly, the quantification process of causality may face dimensionality disasters, such as the calculation of TE [33] and multivariate TE [34]. Therefore how to obtain reliable estimates based on high dimensional data is also a problem we need to face. Finally, real complex systems usually contain multiple variables with direct and indirect causal relationships between them, and we need to distinguish between the two in order to better quantify the direct links between variables. In summary, it is necessary to propose a new approach to better quantify causality in complex systems.

In this paper, we propose a new metric for causality studies called dispersion pattern cosine similarity entropy (DPCSE). The basic idea of the method is that if X is the cause of Y , then the historical information of Y can well estimate X . Firstly, in order to enhance the robustness of the model to noise, we process the sequences with dispersion, converting the original sequences into discrete sequences with a finite num-

ber of symbols. Secondly, we reconstruct the attractor using phase space reconstruction. We consider not only the points themselves, but also the change patterns of the neighborhood set of points, and use this to construct change pattern matrix. Further, the change pattern matrix is predicted using the idea of cross-prediction. Finally, the similarity between the predicted and true values is quantified by cosine similarity entropy (CSE). In addition, considering the complexity of multivariate systems, we extend the method to the multivariate case for measuring the direct coupling between two variables, effectively eliminating the influence of indirect coupling generated by other variables on the direct coupling. Overall, the contributions of this paper are as follows:

1. The dispersion pattern cosine similarity entropy (DPCSE) is proposed to measure the causal relationship between complex systems and extended to the multivariate case. On the basis of phase space reconstruction, not only focusing on the current moment, but also further considering the changing law of the nearest spatio-temporal neighbourhood of the current moment. This improvement better portray the global properties of the variables and the dynamic evolution over time, and also enhance the adaptability to nonlinear dynamic systems.
2. Preprocessing the data with discrete methods enhances the robustness of the model.
3. An improved cosine similarity entropy is used to quantify the similarity between predicted and true values, and thus to measure causality, which makes it better to be used in nonlinear dynamical systems.

The remainder of this paper is organized as follows: Sect. 2 describes in detail our methodology and the tools used, Sect. 3 shows the simulated dataset and the experimental results. Section 4 presents the real datasets and its results. Finally, we summarize the full paper.

2 Method

2.1 Cosine similarity entropy

At first, reconstructed the original $x(t)\{t = 1, \dots, N\}$ and obtain embedding vector $X_t = [x(t), \dots, x(t + (m - 1)\tau)]$ (m is embedding dimension and τ is time delay). Then, for two different moments i and j , define

the cosine similarity (cs) between all pairwise vectors X_i and X_j :

$$cs(X_i, X_j) = \frac{X_i \cdot X_j}{|X_i||X_j|}, \quad i \neq j \quad (1)$$

The range of cs is $[-1, 1]$. Cosine distance has a better performance in higher dimensional spaces compared to Euclidean distance. In high dimensional spaces, most of the vectors are concentrated in the boundary regions of the space, and the cosine similarity can handle these situations better and thus portray the similarity between multiple vectors. However, at this point the Euclidean distance may be affected by dimensional catastrophe, where many vectors become similar to each other in high dimensional space, and the difference in Euclidean distance may become small or indistinguishable. Table 1 shows a comparison of several common similarity metrics.

To further obtain the angular distance between two vectors from cs , the angle between the two vectors is first calculated:

$$\alpha(X_i, X_j) = \cos^{-1}(cs(X_i, X_j)) \quad (2)$$

Then, we can obtain angular distance according to cs :

$$AngDis(X_i, X_j) = \alpha(X_i, X_j)/\pi \quad (3)$$

The range of $AngDis$ is $[0, 1]$, it closer to 1 indicates the greater similarity between the vectors. A threshold r is selected and then the portion of the distance matrix greater than r is eliminated, name the new matrix as $R(r)$. The order of this step is screen all pairwise of similarity vectors. Then, compute the local probability of occurrences of similar patterns:

$$\Omega_i(r) = \frac{1}{(N - m - 1)} R_i(r) \quad (4)$$

$R_i(r)$ denotes the local similarity between the i -th point and its neighbouring points in a neighbourhood of a given radius r . Ω_i represents for a given X_i , the number of X_j whose distance from it is less than r . Moreover, the global probability of occurrences of similar

Table 1 Comparison of four classic similarity measures

	Interpretation	Advantages	Disadvantages
Euclidean distance	Geometric shortest distance	Intuitive, most commonly used	Failure in high dimensional space, sensitive to outliers
Manhattan distance	Distance at right angles between two objects	Robust to outliers	Poor intuition, poor results in high dimensional space
Chebyshev distance	Maximum difference between two vectors in any coordinate dimension	Low computational complexity	Ignore other dimensional information
Cosine similarity	Cosine of the angle between two vectors	Applicable to high-dimensional data, low computational complexity	Disregarding the vector size

patterns can obtain:

$$\Omega(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} \Omega_i(r) \quad (5)$$

Finally, CSE can be represented as:

$$\text{CSE}(m, \tau, r) = -[\Omega(r) \log_2 \Omega(r) + (1 - \Omega(r)) \log_2 (1 - \Omega(r))] \quad (6)$$

The definition of CSE uses shannon entropy to quantifying the complexity of recurrence matrices. There are two main uncertain parameters in the above method: embedding dimension m and distance threshold r . Theoretically, $m \geq 2$ because vectors are constructed in more than two dimensions. The range of r is $[0, 1]$, and it has been shown that the experimental results show regularity with r . It is necessary to choose the appropriate r according to different models. In this paper, we focus on measuring the similarity between two high-dimensional matrices with the help of cosine similarity entropy, which in turn quantifies the causal relationship.

2.2 Dispersion of series

Real sequences usually contain a lot of noise, and in order to weaken the effect of noise on the results as much as possible, we first perform dispersion on the original sequence. The source of the idea of dispersion is dispersion entropy [35], and we can understand

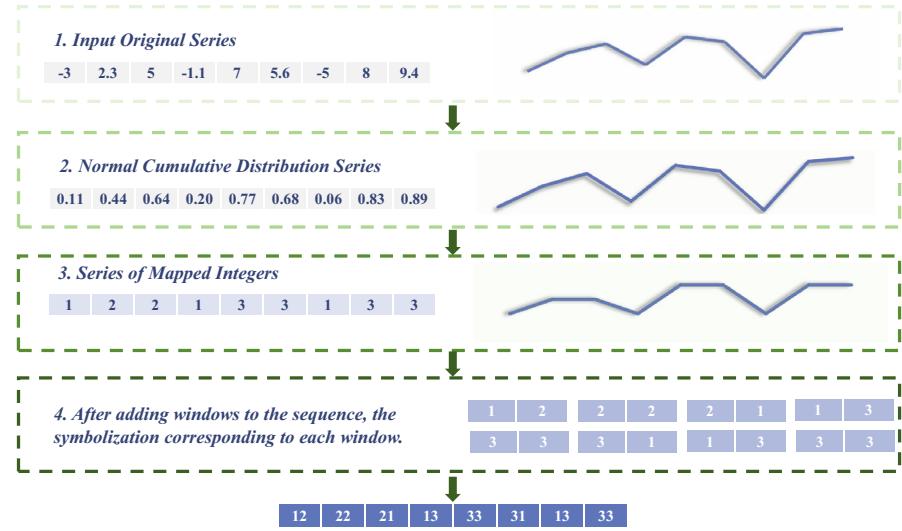
the process of dispersion as a process of symbolisation of sequences. Different from the traditional symbolisation based on ordering, this process fully considers the amplitude information of the original sequence and is more capable of overcoming the influence of noise.

First, we encode the series using the idea of dispersion to enhance stability of the model. At first, perform a normalized mapping of the original $x = \{x_t\}_{t=1}^N$ with the help of normal cumulative distribution function (NCDF).

In this way, x is mapped to x' which lies in the interval $[0, 1]$. In the second step, x' is encoded into q classes by a linear mapping $y_j^q = \text{round}(qx'_j + 0.5)$.

Next, for selected embedding dimension m and delay time l , y_j^q is reconstructed as $Y_i^{m,q} = \{y_i^q, y_{i+l}^q, \dots, y_{i+(m-1)l}^q\}$. After this, each element in $Y_i^{m,q}$ corresponding to a symbolic: $y_i^q = \pi_0, y_{i+l}^q = \pi_1, \dots, y_{i+(m-1)l}^q = \pi_{m-1}$. Combine them and obtain the dispersion series $\rho(\pi_0\pi_1\dots\pi_{m-1})$ of $Y_i^{m,q}$. Since the length of $Y_i^{m,q}$ is m and each element in it was divided into q classes in the previous step, there are theoretically a total of c^q possible cases for the dispersion sequence of $Y_i^{m,q}$. In Fig. 1, we have given an example. After the above steps, the original time series is encoded into symbols with a finite number of possibilities. In all subsequent analyses, we proceeded based on the post-dispersion sequence.

Fig. 1 The dispersion process for a sequence. For the input sequence, it is first transformed into a sequence between 0 and 1 by means of a normal cumulative distribution function and then mapped linearly into an integer sequence. For a given embedding dimension and delay time, the integer sequence is reconstructed to complete the discretisation process of the sequence



2.3 Pattern causality based on promote cosine similarity entropy

At first, we construct attractors based on phase space reconstruction to obtain more information from the original sequence. For given series $x(t)$ and $y(t)$ with the length of N , we first follow the steps in Sect. 2.2 to disperse it to obtain the symbolic series. For ease of description, we still denote as $x(t)$ and $y(t)$. We denote the embedding dimension of the phase space reconstruction as E and the time delay as τ . Then, the reconstructed attractor based on dispersion series is: $\mathcal{M}_X = [\widehat{x}(1); \dots; \widehat{x}(N - (E - 1)\tau)]$, $\widehat{x}(N - (E - 1)\tau) = \{x(N), x(N - \tau), \dots, x(N - (E - 1)\tau)\}$. Performing the same operation on $y(t)$ and yields \mathcal{M}_Y . Then define the distance matrix of attractors:

$$\begin{aligned} \mathcal{D}_X &= \{d_x(i, j)\} = \{|\widehat{x}(i) - \widehat{x}(j)|\}, \\ \mathcal{D}_Y &= \{d_y(i, j)\} = \{|\widehat{y}(i) - \widehat{y}(j)|\}. \end{aligned} \quad (7)$$

$$(i, j = 1, 2, \dots, N - (E - 1)\tau, i \neq j)$$

For each time t , find $E + 1$ nearest neighbor points and record the corresponding time indices. This process needs to be aborted at step h in advance, with in order to project them ahead to determine the future states:

$$\begin{aligned} \mathcal{N}(\widehat{x}(t)) &= \arg \min_{j=1}^{N-(E-1)\tau-h} \{d_x(t, t_j)\} \\ &= \{\mathcal{N}((\widehat{x}(t_1)), \mathcal{N}((\widehat{x}(t_2)), \dots, \mathcal{N}((\widehat{x}(t_{E+1})))\} \end{aligned} \quad (8)$$

For the recorded time, we need to advance h steps:

$$\mathcal{T}(\widehat{x}(t)) = \{t_i + h\}_{i=1}^{E+1} = \{\widehat{t}_{x_i}\}_{i=1}^{E+1} \quad (9)$$

In the next step, the distance of these nearest neighbor points from $\widehat{y}(t)$ needs to be calculated.

$$d(\widehat{y}_i) = d(\widehat{y}(t), \widehat{y}(\widehat{t}_{x_i})), i = 1, 2, \dots, E + 1 \quad (10)$$

Then, we take the properties of the attractor's nearest neighbor points into account, thus providing a more comprehensive picture of the system. Therefore, we consider the weighted variations pattern of each t :

$$\mathcal{V}(\widehat{x}(t)) = \sum_{j=1}^{E+1} \omega_j^x(t) v_j^x(t), \quad (11)$$

It can be seen that $\mathcal{V}(\widehat{x}(t))$ is a linear combination of $\omega_j^x(t)$ and $v_j^x(t)$, where $v_j^x(t)$ is a vector representing the change in x at different moments:

$$\begin{aligned} v_j^x(t) &= \left(\frac{\widehat{x}_j^{(2)} - \widehat{x}_j^{(1)}}{\widehat{x}_j^{(1)}}, \dots, \frac{\widehat{x}_j^{(E+1)} - \widehat{x}_j^{(E)}}{\widehat{x}_j^{(E)}} \right), \\ \widehat{x}(t_j) &= \left(x(t_{x_1}), \dots, x(t_{x_{E+1}}) \right) \\ &= \left(x_j^{(1)}, \dots, x_j^{(E+1)} \right) \end{aligned} \quad (12)$$

Further, the combination of $v_j^x(t)$ is weighted by $\omega_j^x(t)$. And the magnitude of the weight is determined by the distance between the vector and the nearest neighbour point:

$$\omega_j^x(t) = \frac{e^{-d(\hat{x}(t), \hat{x}(t_j))}}{\sum_j e^{-d(\hat{x}(t), \hat{x}(t_j))}} \quad (13)$$

$$\mathcal{V}'(\hat{x}(t)) = \sum_{j=1}^{E+1} \omega_j^y(t) v_j^{x'}(t) \quad (14)$$

where $v_j^{x'}(t)$ denotes the predicted value of the change pattern of x , and $\mathcal{V}'(\hat{x}(t))$ denotes the predicted value of the weighted change pattern of x . Here the calculation of ω_j^y is the same as in Eq. 13. The main difference from Eq. 14 is in the calculation of predicted change patterns. The idea of cross-projection is used in the prediction of the weighted change pattern of x . The predicted pattern of change in x is computed by using $\hat{x}'(t_j)$, which is the projection of the nearest neighbour time subscript of y onto x .

$$\begin{aligned} v_j^{x'}(t) &= \left(\frac{\hat{x}'_j^{(2)} - \hat{x}'_j^{(1)}}{\hat{x}'_j^{(1)}}, \dots, \frac{\hat{x}'_j^{(E+1)} - \hat{x}'_j^{(E)}}{\hat{x}'_j^{(E)}} \right), \\ \hat{x}'(t_j) &= (x'(t_{\hat{y}_1}), \dots, x'(t_{\hat{y}_{E+1}})) \\ &= (x'_j^{(1)}, \dots, x'_j^{(E+1)}) \end{aligned} \quad (15)$$

Then compare the pattern between $\mathcal{V}'(\hat{x}(t))$ and $\mathcal{V}(\hat{x}(t))$, we defined a symbolization function:

$$\mathcal{P}_{\mathcal{V}(\hat{x}(t))} = \text{pattern}(\mathcal{V}(\hat{x}(t))) = \begin{cases} 1, \mathcal{V}(\hat{x}(t)) > 0 \\ 0, \mathcal{V}(\hat{x}(t)) = 0 \\ -1, \mathcal{V}(\hat{x}(t)) < 0 \end{cases} \quad (16)$$

Note that the above operation needs to be performed on every element in $\mathcal{V}(\hat{x}(t))$. Similarly, we can also get $\mathcal{P}_{\mathcal{V}(\hat{x}'(t))}$. Before the next step, we need to remove those moments where $\mathcal{P}_{\mathcal{V}(\hat{x}'(t))} \neq \mathcal{P}_{\mathcal{V}(\hat{x}(t))}$. Finally, measure the causality between x and y by further inscribing the similarity of the prediction and true results. We denote the successful predictions $\mathcal{V}(\hat{x}')$ and $\mathcal{V}(\hat{y}')$ as \mathcal{S}'_x and \mathcal{S}'_y , respectively, and the corresponding true change patterns as \mathcal{S}_x and \mathcal{S}_y .

Further, we have made some improvements to CSE and quantify the similarity between \mathcal{S}'_x and \mathcal{S}_x , \mathcal{S}'_y and \mathcal{S}_y , respectively. The process of measuring similarity is as follows:

In order to study the interaction between x and y , we cross-predict x with the relevant information of y .

1. For every t , compute the angular distance between each pairwise of $S'_{x(t)}$ and $S_{x(t)}$, $S'_{y(t)}$ and $S_{y(t)}$ according to Eq. 1 - 2, and obtain \mathcal{R}^x and \mathcal{R}^y .
2. All distances in \mathcal{R}^x and \mathcal{R}^y are combined and normalized to $[0, 1]$, and the $r - th$ percentile of the sorted distances is used as a threshold to filter the points in the original distance matrix. That is, elements of \mathcal{R}^x and \mathcal{R}^y that are greater than $r\%$ of the merged distance are eliminated.
3. Compute the local probability of occurrences of similar patterns:

$$\Omega_i^x = \frac{1}{(N - E)} R_i^x, \Omega_i^y = \frac{1}{(N - E)} R_i^y \quad (17)$$

Please note that N is the length of series $x(t)$ and $y(t)$. This operation takes into account not only the distribution of the recurrence matrix but also the accuracy of the prediction.

4. Further, compute the global probability of occurrences of similar patterns:

$$\Omega^x = \frac{1}{N - E + 1} \sum_{i=1}^{N-E+1} \Omega_i^x, \\ \Omega^y = \frac{1}{N - E + 1} \sum_{i=1}^{N-E+1} \Omega_i^y \quad (18)$$

5. Finally, we can define the causality between x and y based on Eq. (6) and named DPCSE:

$$\text{DPCSE}_{x \rightarrow y} = -[\Omega^x \log_2 \Omega^x + (1 - \Omega^x) \log_2 (1 - \Omega^x)] \\ \text{DPCSE}_{y \rightarrow x} = -[\Omega^y \log_2 \Omega^y + (1 - \Omega^y) \log_2 (1 - \Omega^y)] \quad (19)$$

The larger value represents the stronger causality. We use an improved cosine similarity entropy to measure the similarity between predicted and true values, and thus quantify causality. The core idea behind this is that if x is the cause of y , then the historical information of y can well estimate x . Therefore, if the change pattern matrix from x and the change pattern matrix predicted with y have high similarity, it means that there is a strong causal relationship from y to x at this point.

In addition, we define directional causality for portraying the direction and strength of information flow between signals:

$$d\text{DPCSE}_{X \rightarrow Y} = \frac{\text{DPCSE}_{X \rightarrow Y} - \text{DPCSE}_{Y \rightarrow X}}{\text{DPCSE}_{X \rightarrow Y} + \text{DPCSE}_{Y \rightarrow X}} \quad (20)$$

The range of $d\text{DPCSE}$ is $[-1, 1]$. If the value is positive, X is the source variable, Y is the target variable, and the magnitude of its absolute value indicates the strength of the causal relationship. At last, In order to clarify the main parameters of the methodology more clearly, we have summarised and explained the meaning of each parameter in the Supplementary Material.

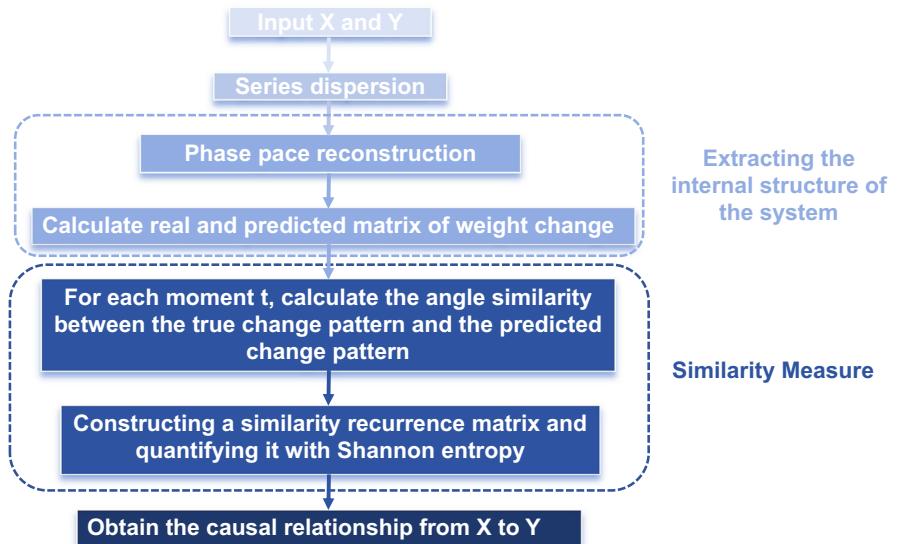
Finally, we run through the complete steps of the methodology, and the whole process is presented in Fig. 2.

1. Dispersion processing is performed on the sequences, which results in a symbolised sequence. The purpose of this step is to enhance the robustness of the model to noise.
2. Construct the change pattern matrix.
 - (a) Perform phase space reconstruction on the symbolised series and then calculate the distance matrix between each moment vector using Eq. 7;
 - (b) Find the nearest neighbour points based on the distance matrix;
 - (c) Construct the change pattern matrix according to Eq. 11;
3. Calculate the predicted value of the change pattern matrix using the idea of cross-prediction according to Eq. 14.
4. Calculate the similarity between the true and predicted values of the change pattern matrix at each moment using the improved cosine similarity entropy. The larger the entropy value, the stronger the similarity and thus the greater the causal strength at that moment.
5. Calculate the directional causal intensity according to Eq. 20 to further clarify the cause and effect in the system.

2.4 Extension of the multivariate case

Direct and indirect causality is widespread in the real world, and the remaining variables can also have an impact on the detection of causality between the target variables. Therefore, we extend DPCSE to the multivariate case to more precisely quantify the causal

Fig. 2 Experimental flow chart. For the input two sequences X and Y , the dispersion process is first performed and then the true and predicted values of the weighted change patterns are computed according to Eqs. 7–15. Finally, the similarity between the true and predicted values is calculated by the improved CSE, which in turn characterises the causal intensity



relationship between the target variables and named pDPCSE. For a multivariate system contain $x(t)$, $y(t)$ and $z(t)$ with the same length N . Assume at this point that there are two causal paths from x to y : direct causality $x \Rightarrow y$, and indirect causality $x \Rightarrow z \Rightarrow y$. In order to distinguish between these two relationships, we can distinguish between neighborhoods that contain indirect and direct information, and eliminate the effect of the third variable by eliminating neighborhoods that contain indirect information.

After constructing the attractor \mathcal{M}_x , \mathcal{M}_y and \mathcal{M}_z as above, for each $\widehat{y}(t) \in \mathcal{M}_y$, its nearest neighbors containing indirect information should also correspond to the neighbors of $\widehat{x}(t) \in \mathcal{M}_x$ and $\widehat{z}(t) \in \mathcal{M}_z$. Based on this, at first, we can find the nearest neighbors point of $\widehat{z}(t)$ and denoted as $\mathcal{N}(\widehat{z}(t))$. We can also obtain corrsponding time indeies $\mathcal{T}(\widehat{z}(t))$.

Then, find the corresponding set of cross-mappings $\widehat{\mathcal{N}}^{\ddagger}(\widehat{y}(t)) = \{\widehat{y}(t) | t \in \mathcal{T}(\widehat{z}(t))\}$ is the part of the \mathcal{M}_y that contains indirect information via z . Moreover, $\widehat{\mathcal{N}}^z(\widehat{x}(t)) = \{\widehat{x}(t) | t \in \mathcal{T}(\widehat{x}(t))\}$ should also be removed.

After above steps, we can rediscover the set of nearest neighbor points of y that need to be satisfied:

$$\begin{aligned} \widehat{\mathcal{N}}^{y|z}(\widehat{x}(t)) = \\ \{\widehat{x}(t) | t \in \mathcal{T}(\widehat{y}(t)), t \notin \tilde{\mathcal{T}}^z(\widehat{x}(t)) \cup \tilde{\mathcal{T}}^z(\widehat{y}(t))\} \end{aligned} \quad (21)$$

In the next step, we can continue to predict the weighted pattern change of y using $\widehat{\mathcal{N}}^{y|z}(\widehat{x}(t))$ like Eqs. 11–14 and obtain $\mathcal{S}_{y|z}$ and $\mathcal{S}'_{y|z}$. At last, the causality from x to y on condition z can be computed. Although z is one-dimensional in the case presented above, it can be extended to multiple dimensions. In this case, the influence of the nearest neighbor points in each z on y needs to be eliminated during the construction of the nearest neighbor points.

2.5 Significance test

To test the validity of the results, we used the time-shifted [36,37] series as the surrogate series for the significance test. For this, M sets of surrogate data are generated, denoted as $(p)DPCSE^{surr}$, and the result of the original sequence is denoted as $(p)DPCSE$. Then, the distributions of $(p)DPCSE^{surr}$ and $(p)DPCSE$ are compared by a permutation test [38,39]. If $(p)DPCSE$ is significantly greater than $(p)DPCSE^{surr}$, we assume that a causal relationship exists at this point, otherwise the hypothesis of a causal relationship is rejected.

3 Simulated experiment

3.1 Datasets

3.1.1 Bivariate model

In the simulate experiments, we verify the validity of the method with two sets of chaotic data. The first model is Henon mapping. This is a very classical model of chaos, often used to model complex systems with causal relationships.

$$\begin{aligned} X_t &= 1.4 - X_{t-1}^2 + 0.1X_{t-2} \\ Y_t &= 1.4 - (eX_{t-1} + (1-e)Y_{t-1})Y_{t-1} + 0.3X_{t-2} \end{aligned} \quad (22)$$

This is a one-way coupled model. Theoretically, we can only detect causality from X to Y . The causality from Y to X should converge to 0. Where the initial values are random values between 0 and 1. The coupling strength of the model is controlled by $e \in [0, 1]$. As e increases, the coupling from X to Y increases.

Next, we generated the following two-way coupling model:

$$\begin{aligned} X_{t+1} &= X_t [r_X - r_X X_t - \beta_{XY} Y_t] \\ Y_{t+1} &= Y_t [r_Y - r_Y Y_t - \beta_{YX} X_t] \end{aligned} \quad (23)$$

Where $r_X = 3.7$, $r_Y = 3.7$, β_{XY} and β_{YX} are the two tuning parameters used to vary the strength of the coupling between X and Y . When $\beta_{XY} = 0$ ($\beta_{YX} = 0$), the model degenerates into a one-way coupling model, at which time there is only causality in the $X(Y)$ to $Y(X)$ direction. In our experiments, we fix the causal intensity in one direction and let the other value vary between 0 and 0.4 to study the trend of the results of the three methods at this time.

3.1.2 Multivariate model

The third model we used is trivariate logistic model.

$$\begin{aligned} X_{t+1} &= \lambda_X X_t (1 - X_t) + \varepsilon_{X_t} \\ Y_{t+1} &= \lambda_Y Y_t [1 - (1 - \frac{C_{XY}}{\lambda_Y})Y_t - \frac{C_{XY}}{\lambda_Y} X_t] + \varepsilon_{Y_t} \end{aligned}$$

$$\begin{aligned} Z_{t+1} &= \lambda_Z Z_t [1 - (1 - \frac{C_{XZ} + \beta_{YZ}}{\lambda_Z})Z_t \\ &\quad - \frac{C_{XZ}}{\lambda_Z} X_t - \frac{C_{YZ}}{\lambda_Z} Y_t] + \varepsilon_{Z_t} \end{aligned} \quad (24)$$

where $\lambda_X = 3.7$, $\lambda_Y = 3.72$, $\lambda_Z = 3.78$, ε is Gaussian white noise with zero mean. The initial value of X , Y and Z are chosen randomly between $[0, 1]$. C_{XY} is the X to Y coupling regulation parameter, C_{XZ} is the X to Z coupling regulation parameter, and C_{YZ} is the Y to Z coupling regulation parameter. In the initial model, we set $C_{XZ} = 0.3$, $C_{XY} = 0.2$, $C_{YZ} = 0.3$. In terms of model structure, the initial model has both indirect causation of X to Z , when the causality is $X \Rightarrow Y \Rightarrow Z$. There is also direct causation of X to Z . We can change the strength of the direct causality by adjusting C_{XZ} .

3.2 Results of simulation data

3.2.1 Results of bivariate model

In this part, we evaluated the methods with our methods and compared with TE and CCM. Our method is mainly based on phase space reconstruction. According to the Takens' embedding theorem, the phase space reconstruction needs to satisfy the following three conditions: the data come from a deterministic dynamical system rather than a purely stochastic process, the time series needs to be long enough, and reasonable parameters. In our experiments, the simulated data are all from nonlinear dynamical systems and the amount of data is sufficient, so the reliability of the experimental results can be guaranteed. In the initial henon model, we set $N = 2000$ and $e = 0.6$. Since the Henon model is unidirectionally coupled, we are mainly concerned with the X to Y .

First, we examined the effect of embedding dimension E and distance threshold r , the results are shown in Fig. 3. We set E to 4–12 and r to 0.2–0.9. Initially, as E increases, the DPCSE tends to decrease and converge more closely to the true value, which indicates that too small an E will cause the method to amplify the original coupling, and the results are more affected by r when E is small. When E gradually increases, the DPCSE does not change significantly with r , and we can assume that the results tend to stabilize at this time. Therefore, in the selection of these two parameters, we should avoid smaller E and r values. Secondly, we investigated the

effect of two parameters (number of categories q and embedding dimension m) in the dispersion process on the results, which are shown in Fig. 4. Overall, different combinations of the two parameters have little effect on the results. Among them, the selection of m hardly affects the results, and the increase of q leads to a small increase in the results, which may be due to the fact that more categories carry more information about the original sequence. In summary, the selection of parameters in the dispersion process has little effect on the results, while the selection of the threshold r affects the results to some extent. However, when r is in the range of 0.6–0.8, the results tend to be stable, so we take it as the empirical interval of r .

Next, we compare the performance of the three methods in terms of series length and computational efficiency as the N changes, as illustrated in Fig. 5. As N increases, the intensity of detected causality and the intensity of directed causality show an increasing trend, while the increasing trend of directed causality intensity is more obvious. However, this increasing trend is not infinite, but tends to stabilise after N reaches a certain value. The reason for this may be because longer series carry more information and thus we can capture more causality from longer sequences, but when the sequences increase to a certain level, the information that can be captured tends to be saturated and thus the results become stable. In addition, the computational efficiency also decreases with increasing N , especially for TE. DPCSE is more efficient compared to TE and CCM and ensures high computational efficiency with increasing N . In conclusion, in practice, we should choose an appropriate value of N to ensure the computational efficiency as much as possible while the information in the sequence is effectively captured.

Further, we investigate the stability of the three methods under different coupling strengths and SNR, and the results are shown in Fig. 6. (a) and (b) shows the causal strength from X to Y and the directional causal strength respectively, and the subplot of each figure is the case under different noises. For the original series, three methods all increase uniformly as the e increases, while the TE decreases sharply when $e > 0.7$, this time it is difficult to detect the correct causal relationship. In addition, although the trend of CCM is consistent with the trend of coupling strength, the value of CCM is always stronger than the model setting. As the noise intensity increases, the detection ability of each model slightly decreases, and when $SNR = 10dB$,

the causal strengths detected by TE and CCM differ significantly from those in the original state, whereas DPCSE maintains high robustness. For the directional causal strength, when e is too large, TE and CCM get smaller directional causal strengths, and although the values detected by DPCSE at this time also show a decreasing trend, the decreasing trend is slower.

Further, we tested the methods using a two-way coupling model. In the initial experiment, we chose the sequence length $N = 2000$. In this part of the experiment, we mainly investigated the variation of causal strengths obtained by the three methods with the coupling strength, and the experimental results are shown in Fig. 7.

First, when $\beta_{XY} = 0$, let β_{YX} take values between 0 and 0.4. At this point, there is no causality from Y to X in the model, so the causal intensity on Y to X is theoretically detected to be small. As can be seen from the figure, on X to Y , the CCM shows a rapidly increasing trend with β_{YX} , which is stabilised at $\beta_{YX} = 0.16$, but its coupling strength is always much larger than the model preset. both DPCSE and TE show a growing trend, but the trend of DPCSE is more stable. On Y to X , TE and DPCSE barely detect causality, while CCM still gets a weak causality at this point. When $\beta_{YX} = 0$, let β_{XY} take values between 0 and 0.4, at which point the causal strength detected from Y to X should tend to 0. The figure shows that on X to Y , TE and DPCSE are more stable. On Y to X , CCM shows the same abrupt increase and tends to stabilise after $\beta_{XY} = 0.16$, while TE and DPCSE show more stable growth.

Next, we take $\beta_{XY} = 0.4$ and let β_{YX} vary between 0 and 0.4 and study the trend of β_{YX} . On X to Y , the results of the three methods show a similar phenomenon as before. While on Y to X , the value of CCM is more stable, DPCSE shows a slight increasing trend, but TE shows an irregular trend. Finally, take $\beta_{YX} = 0.4$ and let β_{XY} vary between 0 and 0.4. At this point, on X to Y , DPCSE shows the same trend of slow growth, while DPCSE shows a more stable performance and TE shows an overall decreasing trend. On Y to X , all three show a significant upward trend, but the performance of DPCSE is more stable and more in line with the model's presuppositions.

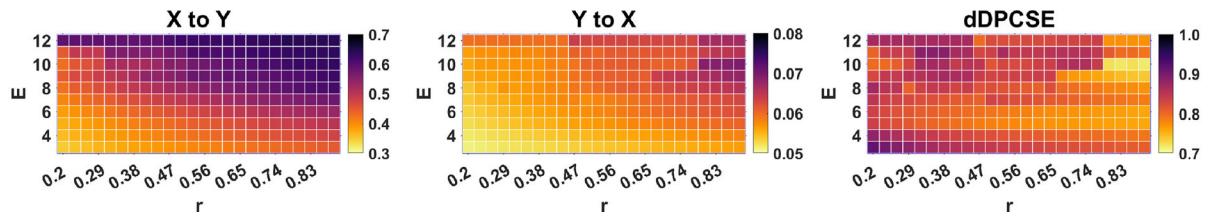


Fig. 3 Impact of embedding dimension and threshold selection on results. Embedding dimension E is the main parameter in phase space reconstruction, which determines the dimensionality of the weighted change pattern and the complexity of the

computation. Threshold r is the parameter involved in the CSE computation process, if r takes too large value, more data are considered similar, if r is too small, fewer pairs of similar data

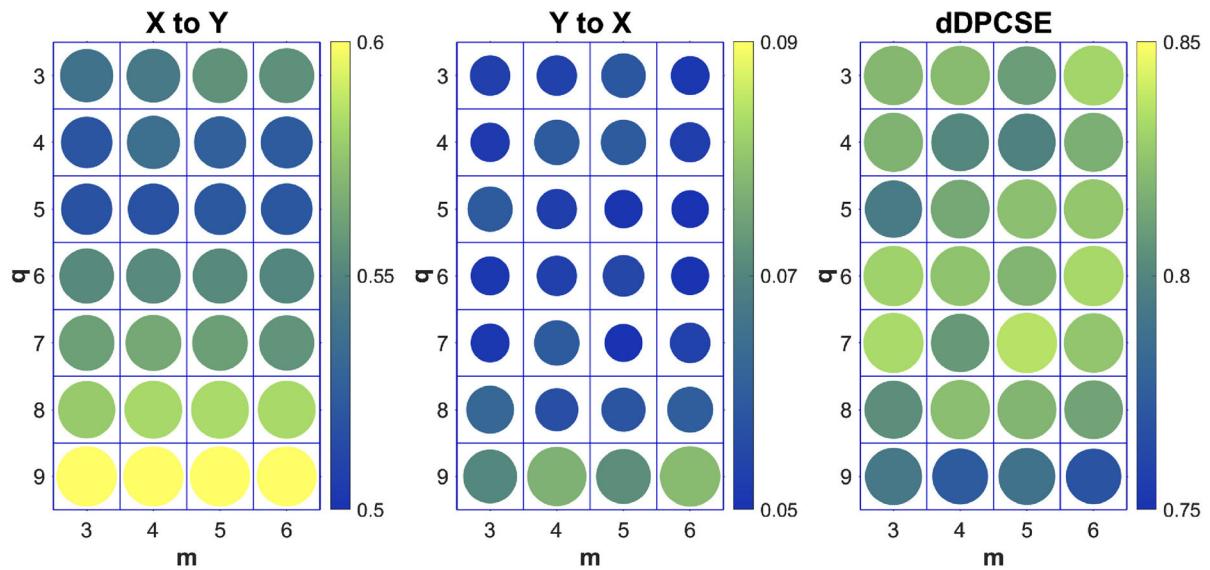


Fig. 4 Impact of embedding dimension m , and classic q in dispersion process on results. The selection of m hardly affects the results, and the increase of q leads to a small increase in the results

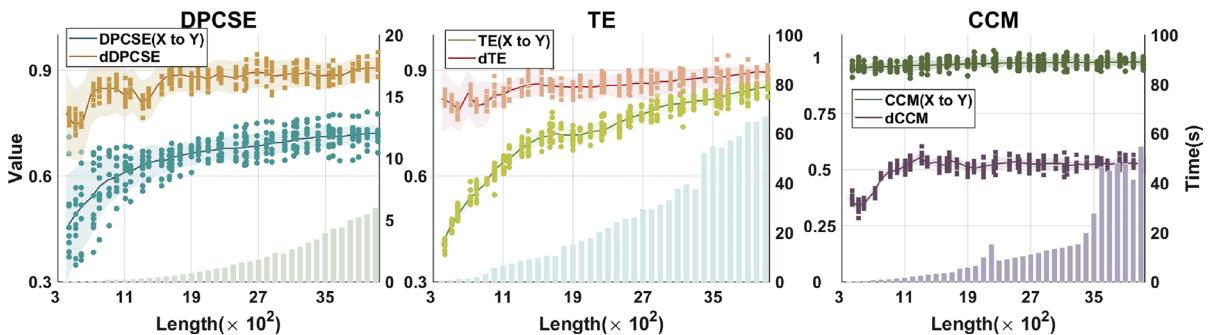
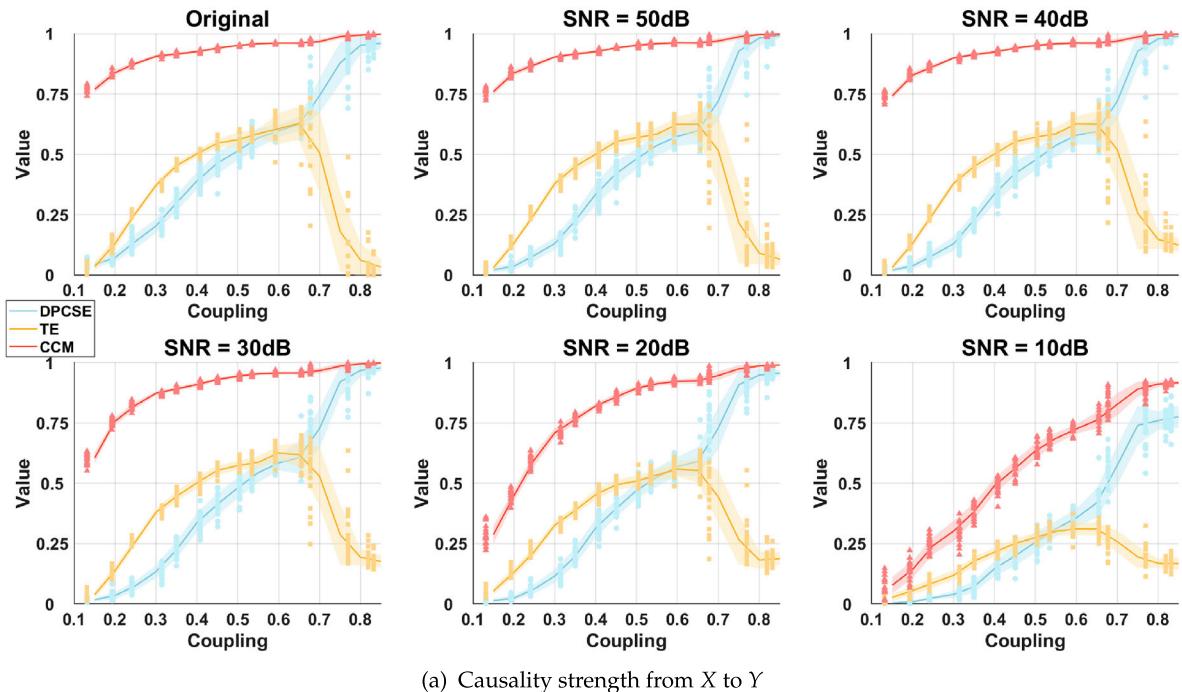


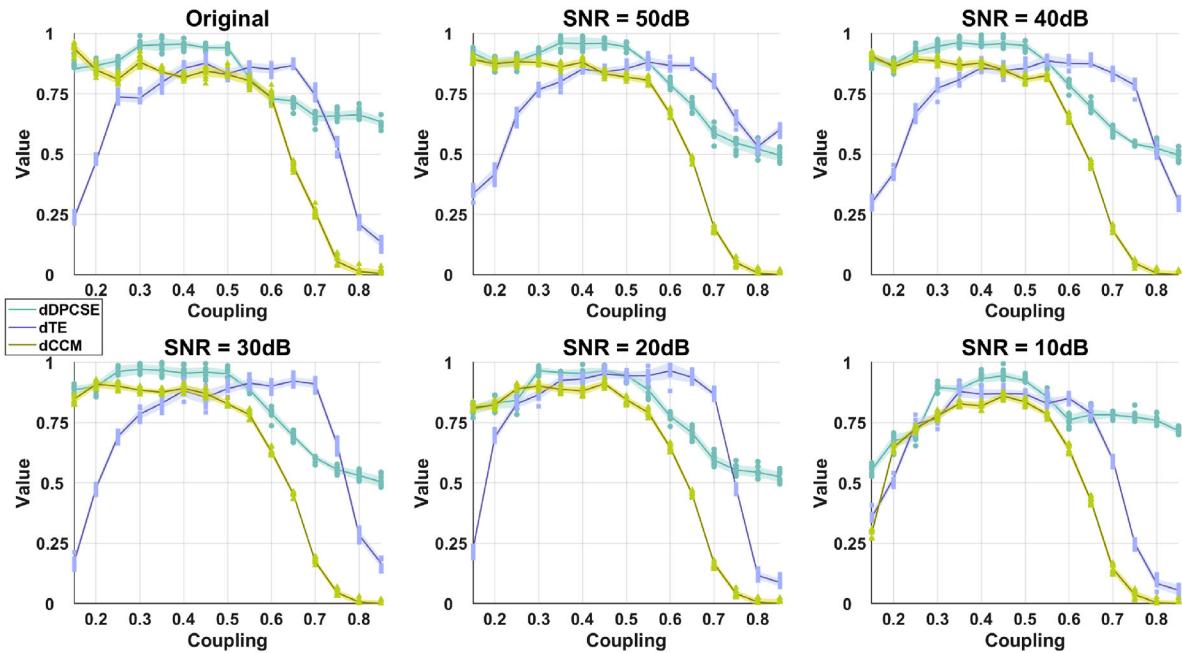
Fig. 5 The effect of series length on results as well as run efficiency. The left y-axis represents the causal intensity and the right y-axis represents the computational time (in s)

3.2.2 Results of multivariate model

We then use a trivariate logistic model to validate pDPCSE. in the modelling assumptions, we assume



(a) Causality strength from X to Y



(b) Direction causality strength

Fig. 6 The causal intensity of X to Y and the directional causal intensity are detected when C_{XY} varies under different noises. Different coloured lines in the diagram indicate different methods

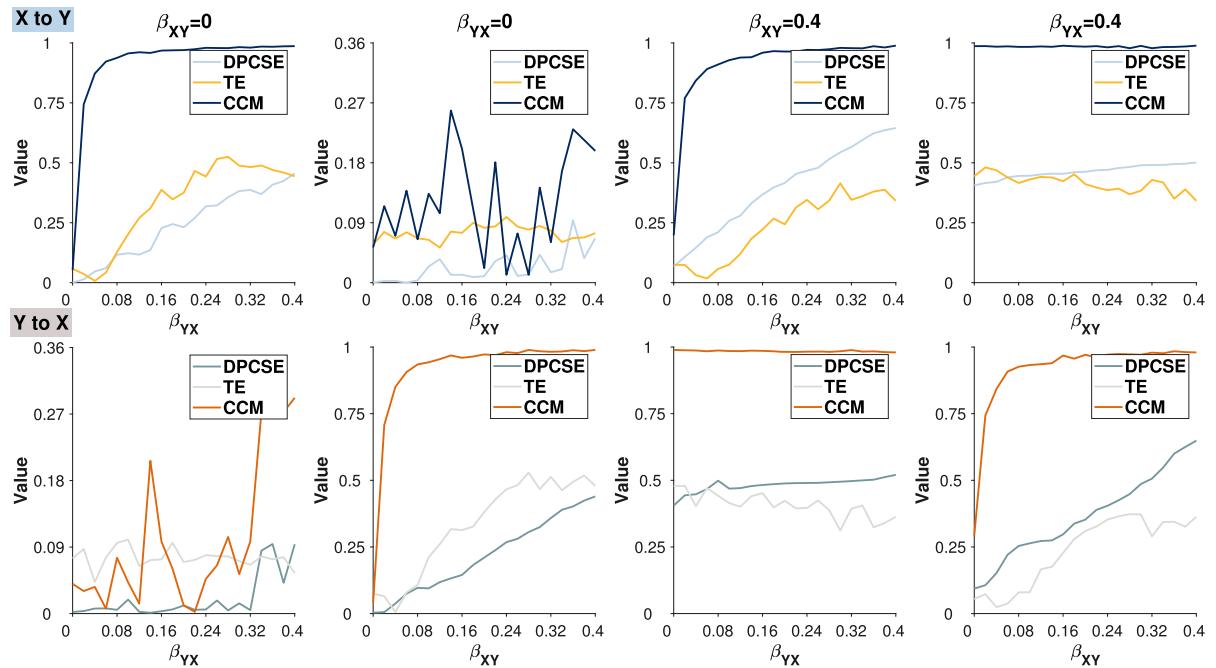


Fig. 7 Results from a two-way coupled model. The first row shows the results for X to Y and the second row shows the results for Y to X . We tested the change in results when fixing the cou-

pling strength in one direction and changing the coupling strength in the other direction, respectively

$X \Rightarrow Y \Rightarrow Z$, where Y is the intermediate variable, and we aim to investigate the direct causality from X to Z . First, we investigate the performance of the three methods as C_{XZ} varies under different noise intensities, as shown in (Fig. 8). In each subfigure, we compare the DPCSE and pDPCSE separately, with each line indicating a different noise intensity. According to the model assumptions, both direct and indirect causal intensities are greater than 0 at this point, and if a bivariate approach is used to detect the causal intensity from X to Z should include both direct and indirect causality. If the multivariate approach is used, the direct causal intensity from X to Z can be detected. It can be seen that pDPCSE is smaller than DPCSE in all noise intensities, indicating that pDPCSE can exclude the interference of indirect causality to a certain extent, thus better detecting direct causality. In addition, the values of pTE and PCCM are also smaller than the bivariate case. When no noise is added, the values of pDPCSE and pTE are approximately linear with the coupling strength, and although pCCM also shows a linear trend, its detected causal strength is larger than the model assumption. When the noise intensity is small, the three methods

have the same trend with the coupling intensity and can sensitively detect the change of the coupling intensity. And the multivariate model can detect the direct causality better. When $SNR = 10dB$, the value of pTE tends to 0, and pCCM is also smaller, and both of them are insensitive to the increase of coupling strength, which indicates that it is difficult for them to detect the correct causality at this time. Although the value of pDPCSE is smaller than the model preset at this point, overall its performance is more stable and it is able to detect changes in coupling strength.

In addition, we investigated the strength of direct causality detected by the three methods when the strength of indirect causality is varied. For this purpose, C_{XZ} was fixed and C_{XY} and C_{YZ} were allowed to vary, and the results are shown in Fig. 9. Theoretically, the direct causal intensity should be less affected by changes in C_{XY} and C_{YZ} , but in practice, it may be difficult to achieve a completely ideal result. It can be seen that the bivariate approach detected weaker causality than the trivariate approach as C_{YZ} and C_{XY} changed. Among them, pDPCSE and pTE showed the same trend. When C_{XY} and C_{YZ} were too large, the

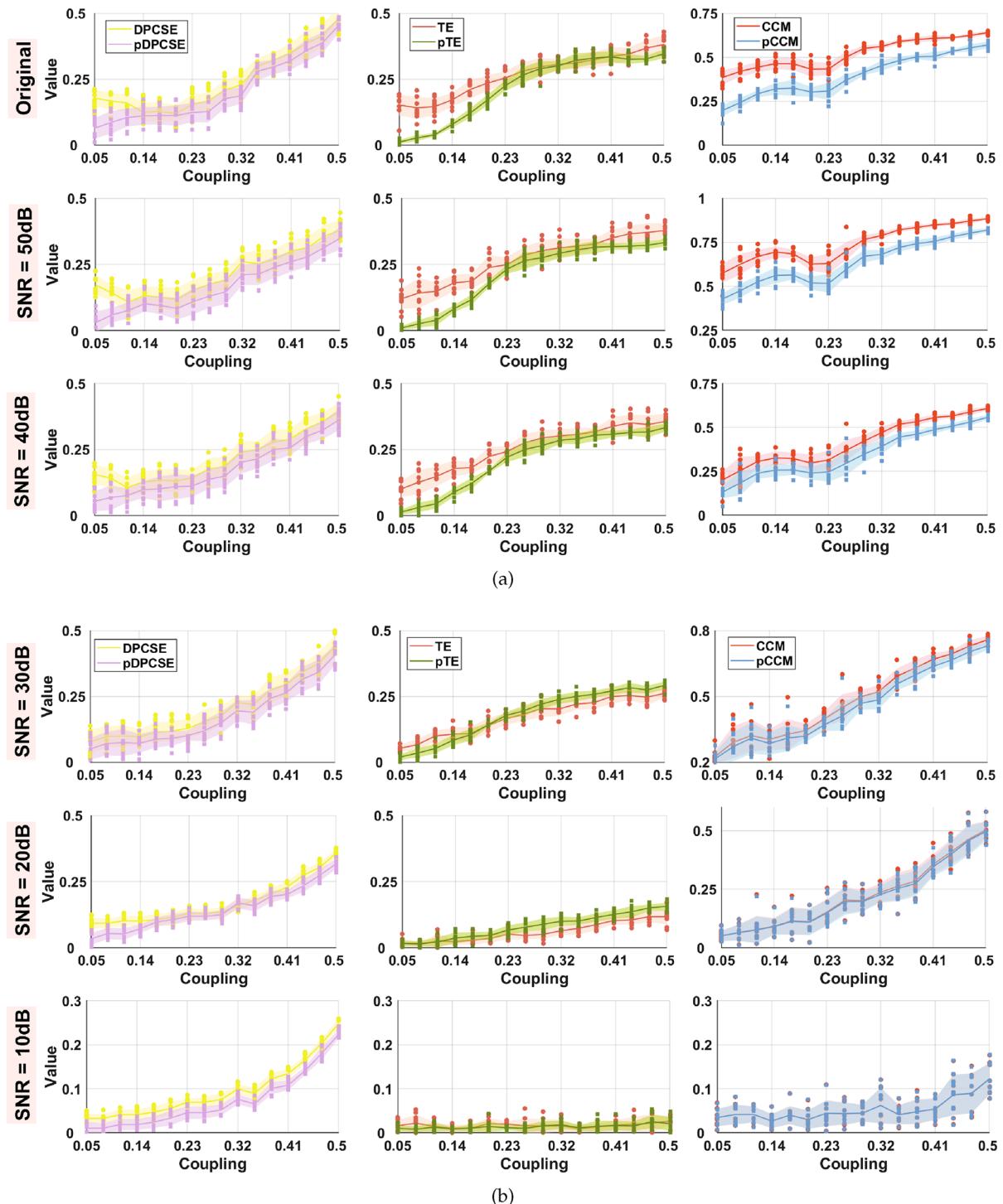


Fig. 8 The figure presents the change in the intensity of direct causation from X to Z detected when C_{XZ} is increased. Different rows present the results under different intensities of noise

strength of indirect causality detected by both tended to zero, indicating that the strength of direct causality received some degree of influence from stronger indirect causality. However, in contrast, the performance of pDPCSE is more stable than that of pTE because pDPCSE is less affected. On the other hand, pCCM performs poorly in this respect, as it is more affected by the strength of indirect causation, and pCCM is almost completely affected by it when C_{YZ} is large. And pCCM always detects a greater causal strength than the model assumptions.

Similarly, we vary the two indirect causal strengths C_{XY} and C_{YZ} , and observe the change of indirect causal strengths C_{YZ} and C_{XY} detected by the three methods, and the experimental results are shown in Fig. 10. For C_{YZ} , when C_{XY} changes, the causal strengths of Y to Z detected by the three methods do not change much. And when C_{YZ} increases, the causal intensity detected by the three methods increases. For C_{XY} , the results of pDPCSE are better than others. As C_{XY} increases, pDPCSE shows a uniform increasing trend, while pTE detects weaker coupling and pCCM detects overly strong coupling, which is not in line with reality. When C_{YZ} changes, the causal strength of X to Y detected by the three methods does not change much. In summary, our methods are able to sensitively detect changes in different parameters of the model and better distinguish between direct and indirect causality.

Finally, we investigated the effect of different sequence lengths on the results as well as computational efficiency (Fig. 11). Similar to the results above, the detected causal strengths show an increasing trend as the N increases, and the value of pDPCSE is always smaller than DPCSE, further indicating that our methods are able to exclude the interference of indirect variables on the results to a certain extent. In addition, the computational efficiency of the three methods decreases with the increase of N , but in general, the computational efficiency of pDPCSE is optimal.

4 Real experiments

4.1 Data description

DataSets1: The first dataset we used is EMG Physical Action Data Set [40] (<https://archive.ics.uci.edu/dataset/213/emg+physical+action+data+set>). This dataset has data from four samples (Age: 25–30)

including three males and one female. Each subject was required to complete twenty actions as instructed. Ten of the movements were in normal form, while the other ten were aggressive movements. All experiments were conducted in accordance with the Code of Ethics of the British Psychological Society, observing ethical regulations and safety precautions. Depending on the experimental setup and precautions taken, the risk of eventual injury is minimal. Subjects volunteered to participate in this series of experiments so they could withdraw from the study at any time.

The dataset contains EMG data for eight channels for each individual. These eight electrodes were distributed on the right and left legs and right and left arms of the subjects. The sampling frequency of the signal is 500Hz. The pre-processing process of the data and the experimental procedure are shown in detail in the Supplementary Material.

DataSet2: This is an EEG dataset from Yayasan Kita Tambu Ahati (YKBH), Jakarta, Indonesia (<https://data.mendeley.com/datasets/4r8hp2hmb4/>) [41]. A total of 14 subjects (5 females and 9 males, Age: 13–15) participated in the data collection process. Seven of them were pornography addicts and seven were healthy. EEG was recorded in the dataset for nine different states of the subjects, including the baseline phase (eyes open and closed), different emotional states (happy, sad, calm, and fearful), and during the performance of the task (memorization, thinking, and recall).

The collected EEG data consisted of 19 channels with a sampling frequency of 250 Hz. In this experiment, we used the EEGs of 14 subjects during the performance of a memory task as a sample for the study, aiming to investigate the differences in brain functional connectivity between pornography addicted patients and normal people during the performance of a memory task. A more detailed description of the dataset and the preprocessing procedure is shown in the Supplementary Material.

4.1.1 Results of real data

4.1.2 Results of Dataset1

We computed the interactions between muscles of the whole body of the subjects in different body states, sitting in the normal state and pushing in the aggressive state, were selected for comparison. We used a data length of 2000, and the results are shown in Figs. 12

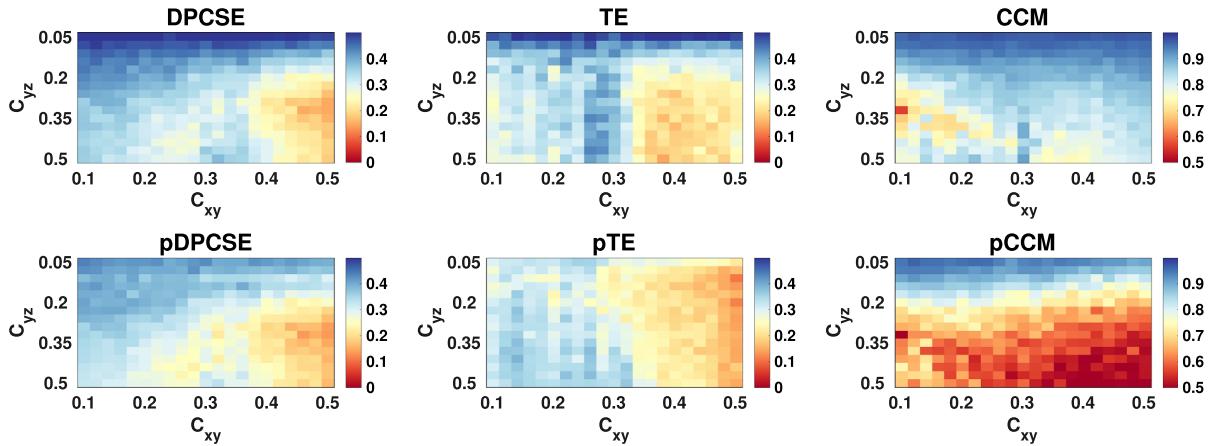


Fig. 9 Fixing C_{XZ} , the direct causality from X to Z detected by the three methods when C_{XY} and C_{YZ} vary. The horizontal axis represents the change in C_{XY} and the vertical axis represents the change in C_{YZ}

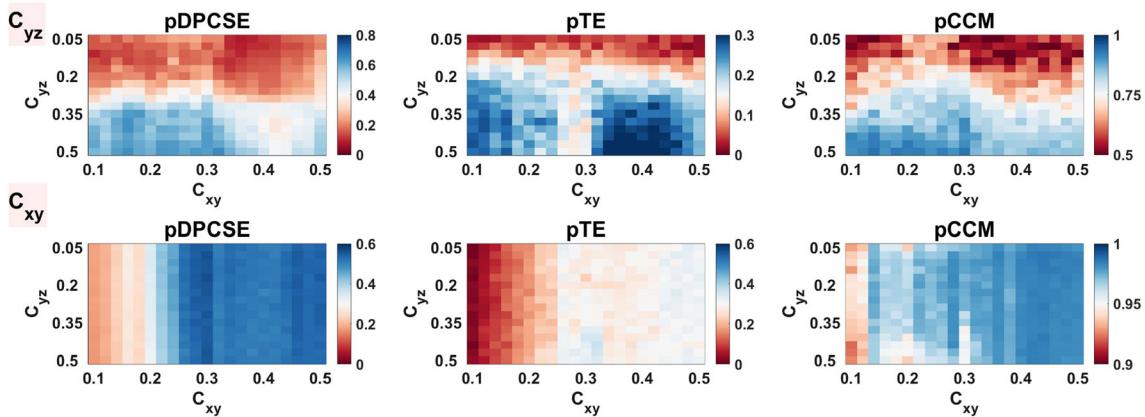


Fig. 10 Fixed C_{XZ} , the causality of X to Y and the causality of Y to Z detected by the three methods when C_{XY} and C_{YZ} vary. The horizontal axis represents the change in C_{XY} and the vertical axis represents the change in C_{YZ}

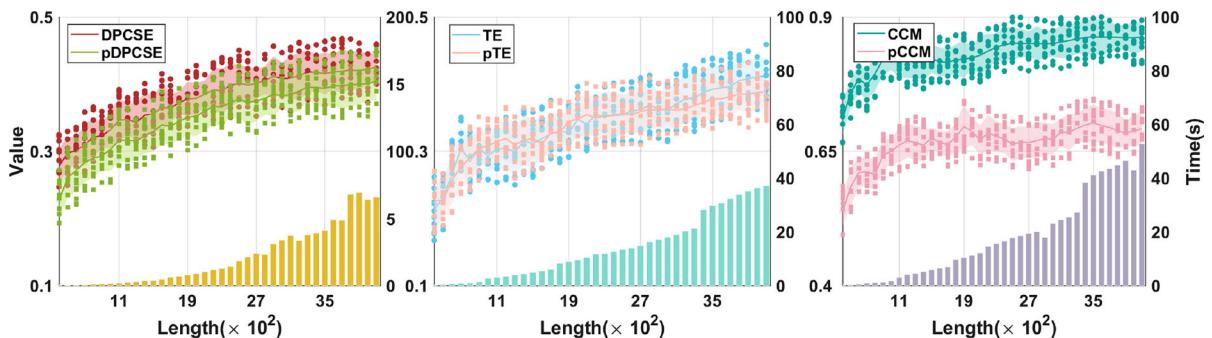


Fig. 11 The effect of sequence length on results as well as run efficiency. The figure shows that pDPCSE has a higher computational efficiency compared to the other two methods

and 13. For parameter selection, we chose $q = 6$, $m = 3$, $E = 8$, $r = 0.7$ for the experiment. Figure 12 demonstrates the causal and directional causal strengths between the EMG of each channel, respectively. Figure 13 demonstrates the portion of all connections that showed significant differences ($p < 0.05$). Overall, the strength of the interaction between the muscles was greater during sitting than pushing. In the pushing state, the upper limb muscles showed a weaker causal relationship between the muscles, while the lower limbs, especially between the right hamstring and the left thigh, showed a stronger interaction. In the seated state, the strength of the interactions between the various muscles was more evenly distributed. And the differences were mainly reflected in the interactions with left and right biceps as source signals and the rest of the parts as target signals. In the interaction with the lower limb as the source signal (C5-C8), the difference was not significant. In addition, the difference in information interaction between the left side of the body was greater than that between the right side of the body. Orientation causality was further compared between the two states. In the state of pushing, the directionality of the interactions between the parts was stronger because it presented a greater directional causal intensity in absolute value and a greater difference in the directional causal intensity between the channels, that is, the delineation of the source and target variables was clearer at this time. And the directional causal intensity at sitting tends to 0, which indicates that the interaction is less directional at this time, and the causal difference between the source and target variables is not obvious. At this time, the difference between the two is mainly concentrated in the interaction dominated by biceps and triceps.

Figure 13 presents a clearer and more intuitive presentation of the distribution of the significance difference across the channels. Obviously, the differences are mainly in the left biceps brachii, right biceps brachii, and right thigh, and the interactions with the channels of the above mentioned sites as source and target variables are more different in the two states.

Finally, we take the significant parts of the directed causation as feature vectors and input them into a KNN classifier to classify the different states. The obtained classification accuracy is 83.75%, precision is 98.48%, recall is 73.50%, and F1-score is 84.17%, which indicates that the metrics extracted by pDPCSE can classify the different states better, and it can also provide us with

a new idea to understand the interactions between the muscles in different states.

4.1.3 Results of Dataset2

In this section, we examine differences in brain functional connectivity between pornography-addicted subjects and normal individuals. We set $N = 2000$. Firstly, we computed the causality and the difference between the two groups of subjects in each channel separately, and the results are shown in Fig. 14. In the figure's with red boxes indicate the presence of significant differences ($p < 0.05$). In general, subjects had the weakest interaction in Delta and the strongest interaction in the Alpha. The differences between two groups of subjects were more in Delta and least in Gamma. In Delta, differences between them appeared mainly in the coupling with the left prefrontal, right temporal, and bilateral frontal lobes as the source variables, when normal subjects showed stronger causality than pornography addicts. In theta, the differences were mainly in the coupling with the left prefrontal lobe as the target variable as well as the temporal lobe as the source variable, and this phenomenon was also present in Alpha. In Beta, the area of distribution of the discrepancy was more concentrated, mainly in the interactions between the temporal, prefrontal, and frontal lobes. In Gamma, on the other hand, the coupling strength between the central occipital lobe and the left parietal lobe was significantly higher in pornography-addicted patients than in normal subjects. In summary, the differences were mainly in the interaction relationship between the prefrontal, frontal and temporal lobes. From the point of view of the left hemisphere and right hemisphere, in Delta, the differences were mainly centered in the interactions of individual electrodes within the left and right hemispheres. In the rest of the bands, except for Beta, the differences in the interactions between the left and right brains were not particularly pronounced.

Figure 15 illustrates the strength of the directional causality, with the red boxes indicating the portions where there were significant differences between them. Directional coupling strength showed more differences across the frequency bands. In the low and mid-frequency bands (Delta, Theta, and Alpha), there are differences almost everywhere in the directional coupling relationship with the central region electrode as the target variable and the rest of the region electrodes as the source variable, which is different from the above

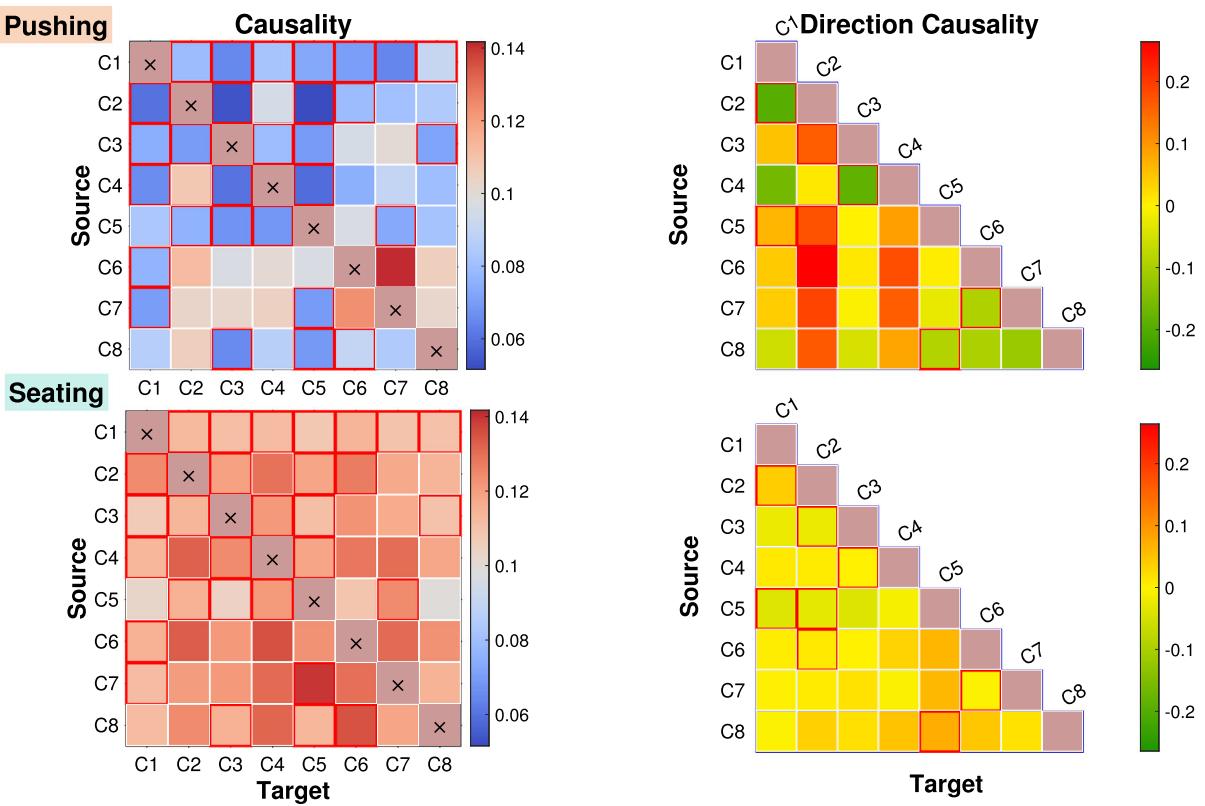
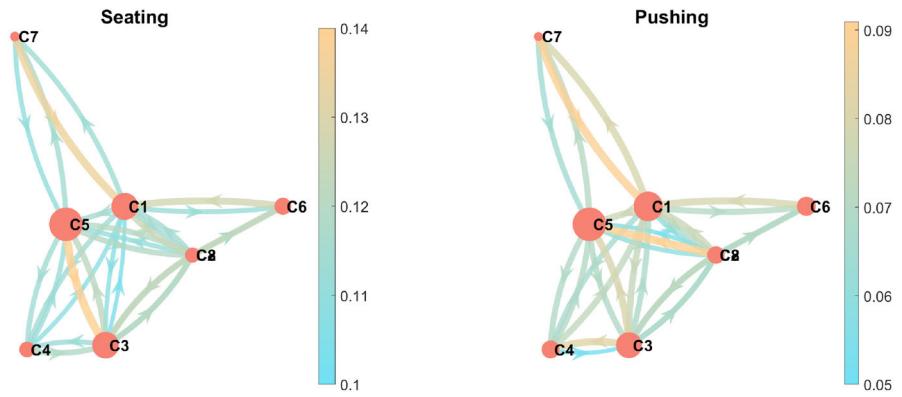


Fig. 12 Interactions between individual EMG signals in the sitting and pushing states. The figure on the right shows directional causation. The colour blocks circled by red boxes in the figure

indicate that there is a significant difference in causal intensity between the two states at this point

Fig. 13 The interaction of different parts of the muscle in two states. The direction of the arrow indicates the direction of causality. The colour of the arrow indicates the intensity of the causal relationship. The size of each node indicates the total causal intensity flowing into or out of that node, flanking the importance of the node



mentioned differences in the overall causal strength. In addition, in the middle and high frequency bands, they showed greater differences in the occurrence of directed coupling with the left occipital lobe as the source variable. Overall, pornography-addicted subjects showed a stronger directionality in the interactions

between channels, meaning that there was a clearer causal division.

Further, the significant portion of the causal intensity under each frequency band is extracted and used as a feature vector to classify patients and normal people. Due to the large number of features, we use principal

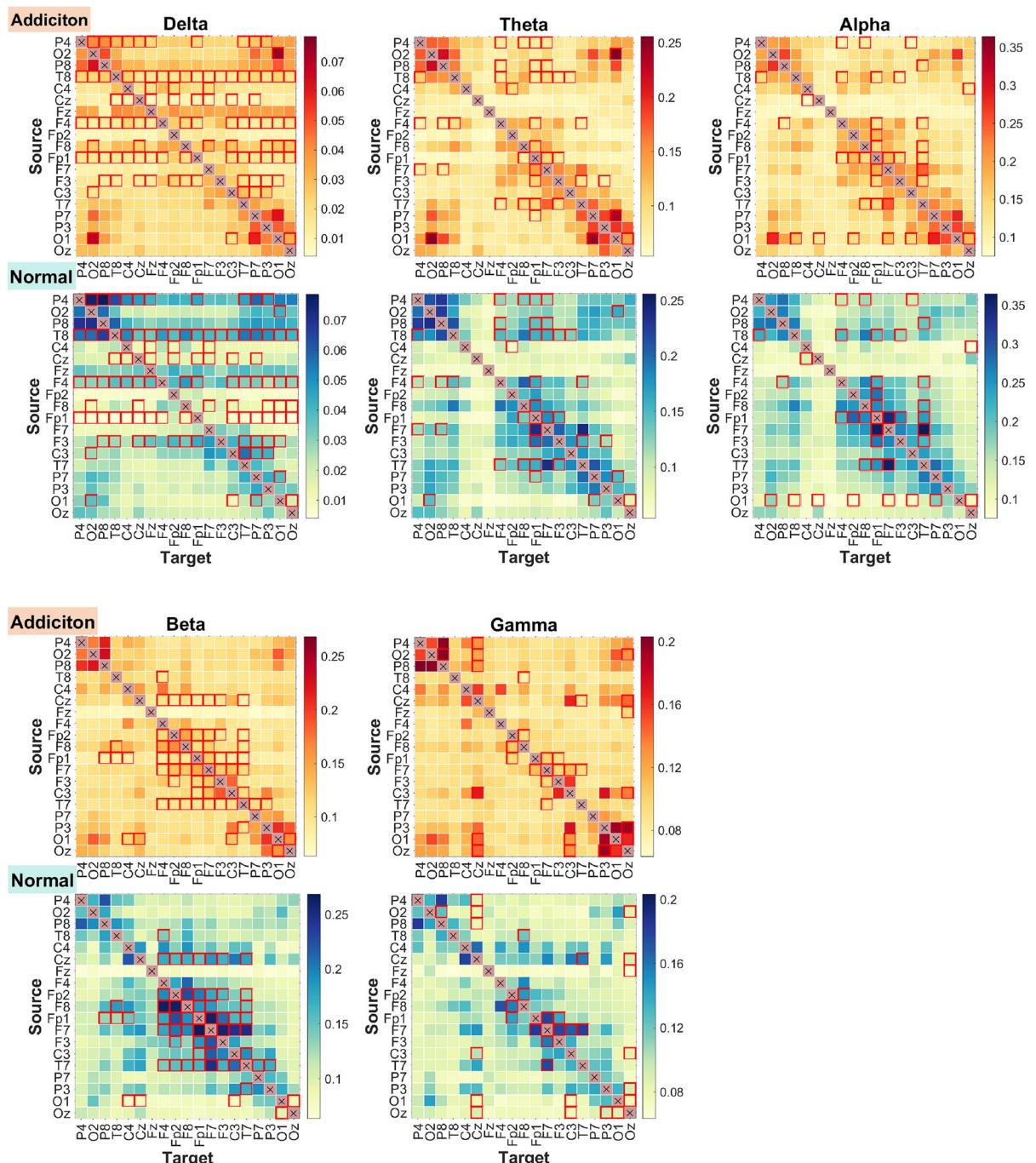


Fig. 14 Causal relationships between channels at different frequency bands in two groups of subjects. The colour blocks circled by red boxes in the figure indicate that there is a significant difference in causal intensity between the two states at this point

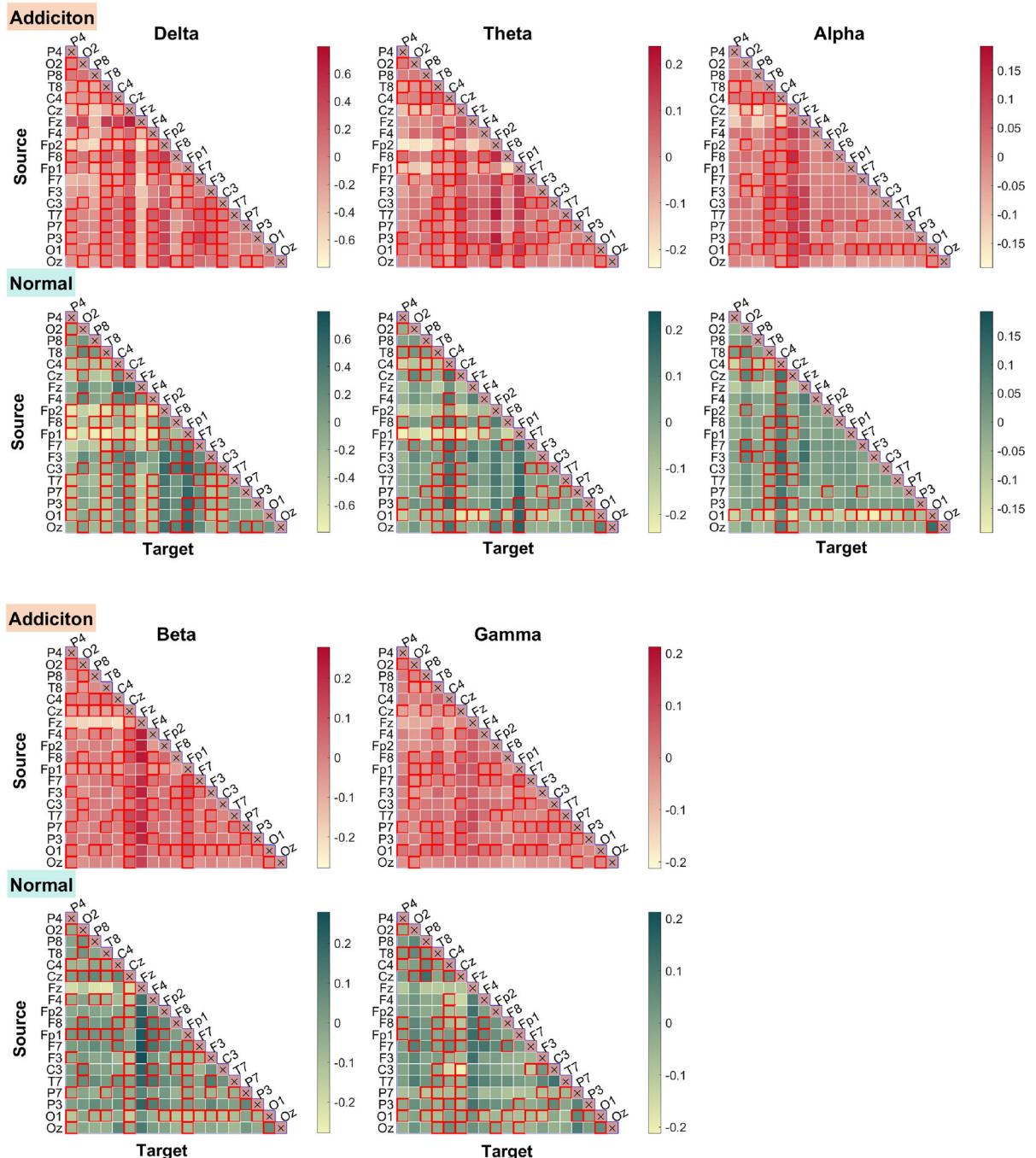


Fig. 15 Directional causality between channels at different frequency bands in two groups of subjects. Since the oriented causal intensity matrix is symmetric, only the lower triangular matrix

is shown. The colour blocks circled by red boxes in the figure indicate that there is a significant difference in causal intensity between the two states at this point

component analysis (PCA) for dimensionality reduction, after which we use KNN to classify and the classification results are shown in Table 2. Bolded numbers indicate the metric with the best classification results among the three metrics. The results show that the eigenvectors extracted with pDPCSE have good classification effects under each frequency band, with classification accuracy above 75%, and even 80%. To further show the classification effect, we compare with CCM and TE, and the results show that the metrics extracted by our method achieve the best classification effect. It shows that the pDPCSE method can better quantify the brain interaction between normal and pornography addicted subjects during the execution of the memory task, and the quantitative results can also provide a new reference basis and index for the classification of the two groups, which provides a new idea for the study of the brain interaction of the pornography addicted patients as well as for the identification.

Finally, we divided the individual electrodes into six regions (ROIs) according to their location, and compared the differences in the coupling effects on each region between the two subjects using *t*-tests. We plotted the parts where there were significant differences as in Fig. 16. After dividing the regions, both groups of subjects similarly show smaller differences in Gamma and larger differences in Delta. In terms of the strength of causality, in addition to Gamma, the interaction between regions in other bands was significantly weaker in pornography addicted subjects than in normal, whereas the opposite phenomenon was presented in the Gamma band. In addition, the differences between the two were smaller in the parietal and temporal lobes, and the differences were more concentrated in the interactions related to the occipital and prefrontal lobes.

5 Discussion and conclusion

In this paper, dispersion pattern cosine similarity entropy is proposed for detecting causal relationships in complex systems. The method constructs an attractor by using phase space reconstruction. Then, we consider not only the properties of each momentary point, but also the changes in the set of nearest neighbor points, thus constructing a weighted change pattern matrix, which portrays and captures the spatio-temporal properties of the original series. In the next step, the idea

of cross-prediction is used to quantify the causal relationship by measuring the similarity between the true and predicted values. The similarity measure process then adopts the cosine similarity entropy, which evaluates the similarity between vectors by angular distance, Shannon entropy, and has a better performance on short time series. In addition, this paper also extends the above method to the multivariate case, and this extension helps us to identify the direct and indirect causal relationships between variables. Compared with the existing methods, the breakthrough of this method is that it not only considers the attributes of each momentary point, but also fully takes into account the changes of the set of nearest neighbouring points, which better portrays and describes the dynamical system in a comprehensive way by constructing the change pattern matrix. Secondly, in CCM, Pearson's correlation coefficient is used to measure the correlation between the predicted and real values, and then quantify the causal relationship. While facing a nonlinear system, the performance of Pearson's coefficient, a linear measure, may be limited. Thus, we use cosine similarity entropy as a similarity measure and a quantification tool for causality. In particular, the stable performance of cosine distance in high-dimensional space contributes to the accurate quantification of causality and is more generalisable [42]. Finally, extending the method to multivariate systems can help us better distinguish between direct and indirect causality. When we want to describe the direct coupling between X to Y , if the indirect coupling from X to Z to Y exists in the system at the same time, then using the general method may lead to more couplings being detected, whereas pDPCSE effectively eliminates the influence of indirect couplings on the results.

A series of simulation experiments were conducted to test the validity of the method. The effect of the parameters on the results is first considered, and then the performance of the method is discussed for different noises, series lengths and coupling strengths, and compared with existing methods. Of course, we also test the performance of the method on nonseparable data. The results of the simulation experiments demonstrate the good applicability of our proposed method, which can detect causality in complex systems more accurately and efficiently. Finally, we used the method in the analysis of two physiological datasets. First, we used multichannel EMG data from healthy subjects in normal and aggressive states, and analysed and compared

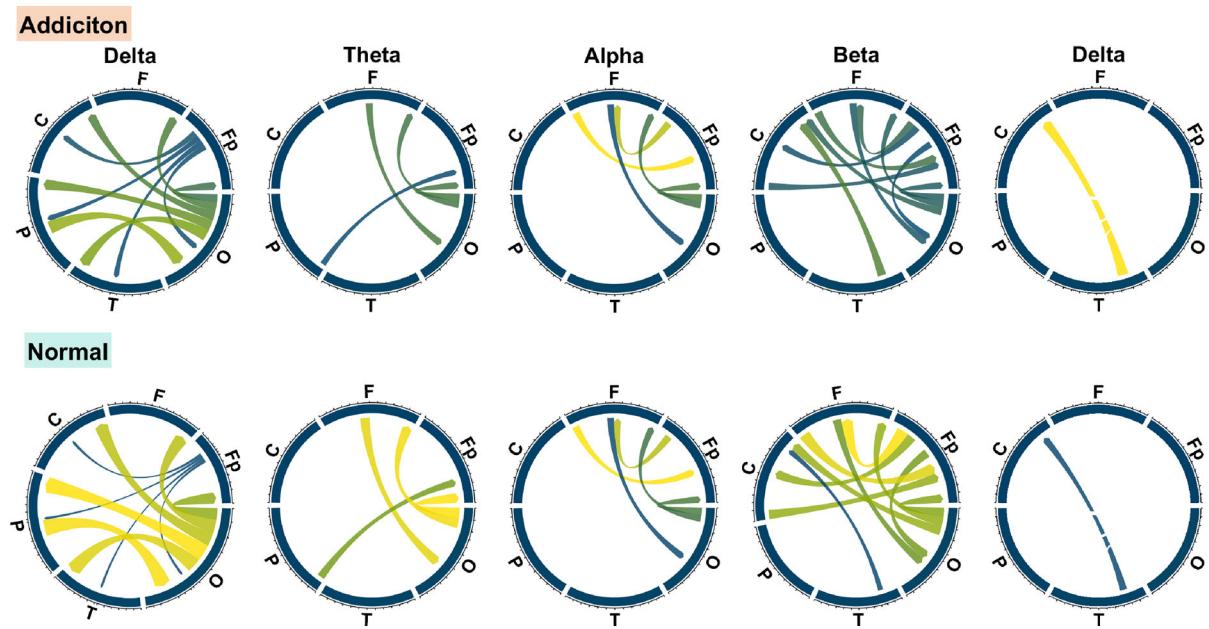


Fig. 16 The portions where there were significant differences between the two were extracted and plotted on a chord diagram. The darker colors indicate less causal strength

Table 2 The performance of the indicators created with the three methods in performing the classification task

		Delta (%)	Theta (%)	Alpha (%)	Beta (%)	Gamma (%)
pDPCSE	Precision	79.45	84.73	86.65	77.53	85.96
	Recall	84.01	77.53	83.08	80.08	80.54
	F1-score	78.08	79.45	84.21	75.68	80.41
	Accuracy	78.96	79.12	83.84	76.00	80.64
TE	Precision	76.25	80.40	83.72	74.07	83.42
	Recall	79.34	73.73	80.81	77.01	76.14
	F1-score	77.77	76.92	82.24	75.51	79.62
	Accuracy	75.56	76.72	80.84	71.53	78.44
CCM	Precision	75.72	82.00	82.92	74.07	81.76
	Recall	80.28	74.93	80.68	75.95	77.54
	F1-score	77.93	78.30	81.78	74.99	79.59
	Accuracy	76.56	75.85	81.31	73.93	76.77

Delta, Theta, Alpha, Beta and Gamma in the table represent five different frequency brain bands, each corresponding to different cognitive states and physiological functions

the differences in coupling between individual parts of the muscle in different states. This study provides new perspectives for understanding the mechanisms of muscle co-operation and contributes to motor control research and optimisation of rehabilitation training. In the second dataset, we analysed and compared the interactions and differences between individual electrodes

and regions of the brain during the performance of a memory task in pornography-addicted patients and normal subjects. This study helps to reveal the effects of pornography addiction on functional brain networks, provides new empirical support for cognitive neuroscience research in this area, and provides a scientific basis for the development of personalised inter-

vention strategies. In addition, we constructed metrics to perform a classification task aimed at distinguishing between subjects in different states and obtained high classification accuracy. In summary, our method not only provides an effective way to study the interactions of different physiological systems, but also offers new ideas for the detection of causality in nonlinear systems and shows broad application prospects in the fields of exercise rehabilitation, neuroscience and clinical diagnosis.

However, our study still has some problems. Firstly, at present, we only identify the low-order coupling relationship between variables, and the high-order causality is also an aspect we need to pay attention to, so the next step of our work will be centred on deeper causality. Secondly, there are a large number of parameters used in the method, and although the robustness of the method under different parameters has been discussed in the experimental section, it is still necessary to optimise the method to make it as independent of the parameters as possible. In addition, the computational efficiency of the method on high-dimensional or large datasets needs to be improved, and we will start with this in the next step to improve the computational efficiency of the method based on dimensionality reduction preprocessing and other strategies, and conduct experiments on larger datasets. Moreover, we will also focus on synergies and redundancies between variables, fully consider systems with feedback loops or multiple interactions, and try to integrate more information and develop more expressive models in order to improve the adaptability and resolution of the method to complex systems. In terms of the model's interpretability, although DPCSE demonstrates better performance in causality detection, there is still room for improvement in its interpretability compared to the traditional GC. We plan to further optimise the model in our future work, exploring more intuitive visualisation methods and easier-to-understand causal inference mechanisms to enhance its interpretability and transparency in practical applications. Finally, DPCSE is not sufficiently integrated with causal analysis methods from other related fields, which is one of our main focuses. In this regard, we will integrate our methods with those in other fields (e.g., economics, physics, engineering, etc.) in our next work. For example, structural equation modelling (SEM) in economics and a range of causal discovery methods. These combinations may provide new perspectives and useful

insights. The generalisation ability and effectiveness of the methods can be further enhanced through cross-domain knowledge integration. In terms of the application of the method, we will also apply the method to larger scale data and apply it to data from multiple domains to solve more practical problems [43, 44].

Funding This work is supported by the Fundamental Research Funds for the Central Universities (2023YJS071).

Data Availability The data that support the findings of this study are openly available in <https://data.mendeley.com/datasets/4r8hp2hmb4/5> and <https://archive.ics.uci.edu/dataset/213/emg+physical+action+data+set>.

Declarations

Conflict of interest The authors declare that they have no Conflict of interest.

References

- Pearl, J.: Causal inference in statistics: An overview. *Stat. Surv.* **3**, 96–146 (2009)
- Mirza, F.M., Kanwal, A.: Energy consumption, carbon emissions and economic growth in Pakistan: Dynamic causality analysis. *Renew. Sustain. Energy Rev.* **72**, 1233–1240 (2017)
- Atasoy, M., Atasoy, F.G.: The impact of climate change on tourism: A causality analysis. *Turkish J. Agricult.-Food Sci. Technol.* **8**(2), 515–519 (2020)
- Hu, S., Wang, H., Zhang, J., Kong, W., Cao, Y., Kozma, R.: Comparison analysis: Granger causality and new causality and their applications to motor imagery. *IEEE Trans. Neural Netw. Learn. Syst.* **27**(7), 1429–1444 (2015)
- Tu, C., Fan, Y., Fan, J.: Universal cointegration and its applications. *IScience* **19**, 986–995 (2019)
- Runge, J., Gerhardus, A., Varando, G., Eyring, V., Camps-Valls, G.: Causal inference for time series. *Nature Rev. Earth Environ.* **4**(7), 487–505 (2023)
- Nogueira, A.R., Pugnana, A., Ruggieri, S., Pedreschi, D., Gama, J.: Methods and tools for causal discovery and causal inference. *Wiley Interdiscip. Rev. Data Mining Knowled. Discovery* **12**(2), 1449 (2022)
- Shojaie, A., Fox, E.B.: Granger causality: A review and recent advances. *Annual Rev. Stat. Appl.* **9**, 289–319 (2022)
- Bell, D., Kay, J., Malley, J.: A non-parametric approach to non-linear causality testing. *Econom. Lett.* **51**(1), 7–18 (1996)
- Chen, Y., Rangarajan, G., Feng, J., Ding, M.: Analyzing multiple nonlinear time series with extended granger causality. *Phys. Lett. A* **324**(1), 26–35 (2004)
- Schiff, S.J., So, P., Chang, T., Burke, R.E., Sauer, T.: Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble. *Phys. Rev. E* **54**(6), 6708 (1996)
- Schreiber, T.: Measuring information transfer. *Phys. Rev. Lett.* **85**(2), 461 (2000)

13. Duan, S., Zhao, C., Wu, M.: Multiscale partial symbolic transfer entropy for time-delay root cause diagnosis in non-stationary industrial processes. *IEEE Trans. Ind. Electron.* **70**(2), 2015–2025 (2022)
14. Mao, X., Shang, P.: Transfer entropy between multivariate time series. *Commun. Nonlinear Sci. Numer. Simul.* **47**, 338–347 (2017)
15. Stramaglia, S., Faes, L., Cortes, J.M., Marinazzo, D.: Disentangling high-order effects in the transfer entropy. *Phys. Rev. Res.* **6**, 032007 (2024)
16. Yuan, A.E., Shou, W.: Data-driven causal analysis of observational biological time series. *Elife* **11**, 72518 (2022)
17. Sugihara, G., May, R., Ye, H., Hsieh, C.-H., Deyle, E., Fogarty, M., Munch, S.: Detecting causality in complex ecosystems. *Science* **338**(6106), 496–500 (2012)
18. Yang, L., Lin, W., Leng, S.: Conditional cross-map-based technique: From pairwise dynamical causality to causal network reconstruction. *Chaos: An Interdiscip. J. Nonlinear Sci.* **33**(6) (2023)
19. Leng, S., Ma, H., Kurths, J., Lai, Y.-C., Lin, W., Aihara, K., Chen, L.: Partial cross mapping eliminates indirect causal influences. *Nature Commun.* **11**(1), 2632 (2020)
20. Jiang, Q., Jiang, J., Wang, W., Pan, C., Zhong, W.: Partial cross mapping based on sparse variable selection for direct fault root cause diagnosis for industrial processes. *IEEE Trans. Neural Netw. Learn. Syst.* **35**(5), 6218–6230 (2024)
21. Breston, L., Leonardi, E.J., Quinn, L.K., Tolston, M., Wiles, J., Chiba, A.A.: Convergent cross sorting for estimating dynamic coupling. *Scientif. Rep.* **11**(1), 20374 (2021)
22. Stavroglou, S.K., Pantelous, A.A., Stanley, H.E., Zuev, K.M.: Hidden interactions in financial markets. *Proc. Nat. Acad. Sci.* **116**(22), 10646–10651 (2019)
23. Shi, J., Chen, L., Aihara, K.: Embedding entropy: A nonlinear measure of dynamical causality. *J. Royal Soc. Interf.* **19**(188), 20210766 (2022)
24. Kathpalia, A., Nagaraj, N.: Data-based intervention approach for complexity-causality measure. *PeerJ Comp. Sci.* **5**, 196 (2019)
25. Kathpalia, A., Manshour, P., Paluš, M.: Compression complexity with ordinal patterns for robust causal inference in irregularly sampled time series. *Scientif. Rep.* **12**(1), 14170 (2022)
26. Ying, X., Leng, S.-Y., Ma, H.-F., Nie, Q., Lai, Y.-C., Lin, W.: Continuity scaling: A rigorous framework for detecting and quantifying causality accurately. *Research* **2022**, 9870149 (2022)
27. Ma, H., Aihara, K., Chen, L.: Detecting causality from nonlinear dynamics with short-term time series. *Scientif. Rep.* **4**(1), 7464 (2014)
28. Krakovská, A.: Correlation dimension detects causal links in coupled dynamical systems. *Entropy* **21**(9), 818 (2019)
29. Martínez-Sánchez, Á., Arranz, G., Lozano-Durán, A.: Decomposing causality into its synergistic, unique, and redundant components. *Nature Commun.* **15**(1), 9296 (2024)
30. Nur, Y.S.R., Aldo, D., Sa'Adah, A., et al.: Implementation of pc algorithm to the incidence factor of stunting disease. In: 2024 International Conference on Information Technology Research and Innovation (ICITRI), pp. 93–98 (2024). IEEE
31. Entner, D., Hoyer, P.O.: On causal discovery from time series data using fci. *Probabilist. Graph. Models* **16** (2010)
32. Colombo, D., Maathuis, M. H., Kalisch, M., Richardson, T. S.: Learning high-dimensional directed acyclic graphs with latent and selection variables. *Ann. Stat.* 294–321 (2012)
33. Papana, A., Papana-Dagiasis, A., Siggiridou, E.: Shortcomings of transfer entropy and partial transfer entropy: Extending them to escape the curse of dimensionality. *Int. J. Bifurcation Chaos* **30**(16), 2050250 (2020)
34. Runge, J., Heitzig, J., Petoukhov, V., Kurths, J.: Escaping the curse of dimensionality in estimating multivariate transfer entropy. *Phys. Rev. Lett.* **108**(25), 258701 (2012)
35. Rostaghi, M., Azami, H.: Dispersion entropy: A measure for time-series analysis. *IEEE Signal Process. Lett.* **23**(5), 610–614 (2016)
36. Andrzejak, R., Ledberg, A., Deco, G.: Detecting event-related time-dependent directional couplings. *New J. Phys.* **8**(1), 6 (2006)
37. Wagner, T., Fell, J., Lehnertz, K.: The detection of transient directional couplings based on phase synchronization. *New J. Phys.* **12**(5), 053031 (2010)
38. Maris, E., Oostenveld, R.: Nonparametric statistical testing of EEG-and meg-data. *J. Neurosci. Methods* **164**(1), 177–190 (2007)
39. Lindner, M., Vicente, R., Priesemann, V., Wibral, M.: Trenntool: A matlab open source toolbox to analyse information flow in time series data with transfer entropy. *BMC Neurosci.* **12**, 1–22 (2011)
40. Theodoridis, T.: EMG Physical Action Data Set. UCI Machine Learning Repository. <https://doi.org/10.24432/C53W49> (2011)
41. Kang, X., Agastya, I.M.A., Handayani, D.O.D., Kit, M.H., Rahman, A.W.B.A.: Electroencephalogram (EEG) dataset with porn addiction and healthy teenagers under rest and executive function task. *Data in Brief* **39**, 107467 (2021)
42. Dubey, V.K., Saxena, A.K.: A cosine-similarity mutual-information approach for feature selection on high dimensional datasets. *J. Inform. Technol. Res.* **10**(1), 15–28 (2017)
43. Tu, C., D'Odorico, P., Suweis, S.: Critical slowing down associated with critical transition and risk of collapse in crypto-currency. *Royal Soc. Open Sci.* **7**(3), 191450 (2020)
44. Lin, H., Tu, C., Fang, J., Gioli, B., Loubet, B., Gruening, C., Zhou, G., Beringer, J., Huang, J., Dušek, J., et al.: Forests buffer thermal fluctuation better than non-forests. *Agricult. Forest Meteorol.* **288**, 107994 (2020)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.