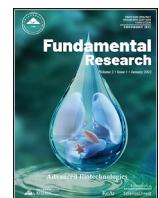


Contents lists available at ScienceDirect



Fundamental Research

journal homepage: <http://www.keaipublishing.com/en/journals/fundamental-research/>

Article

Detecting dynamical causality by intersection cardinal concavity

Peng Tao^{a,b,1}, Qifan Wang^{c,d,e,1}, Jifan Shi^{f,g,h,1}, Xiaohu Hao^b, Xiaoping Liu^{a,b}, Bin Minⁱ, Yiheng Zhang^{c,d,e}, Chenyang Li^{c,d,e}, He Cui^{c,d,e,*}, Luonan Chen^{a,b,j,*}

^a Key Laboratory of Systems Health Science of Zhejiang Province, School of Life Science, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Hangzhou 310024, China

^b Key Laboratory of Systems Biology, Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences, Shanghai 200031, China

^c Center for Excellence in Brain Science and Intelligent Technology, Institute of Neuroscience, Chinese Academy of Sciences, Shanghai 200031, China

^d Chinese Institute for Brain Research, Beijing 102206, China

^e University of Chinese Academy of Sciences, Beijing 100049, China

^f Research Institute of Intelligent Complex Systems, Fudan University, Shanghai 200433, China

^g State Key Laboratory of Medical Neurobiology and MOE Frontiers Center for Brain Science, Institutes of Brain Science, Fudan University, Shanghai 200032, China

^h Shanghai Artificial Intelligence Laboratory, Shanghai 200232, China

ⁱ Shanghai Center for Brain Science and Brain-inspired Technology, Shanghai 200031, China

^j School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China

ARTICLE INFO

Article history:

Received 17 September 2022

Received in revised form 21 December 2022

Accepted 31 January 2023

Available online xxx

Keywords:

Causal inference

Dynamical causality

Nonlinear causality

Cross mapping

Non-separability problem

False-negative problem

ABSTRACT

Discovering causality from observed time series data is of great importance in various disciplines but also a challenging task. In recent years, cross-mapping methods have been developed to solve the non-separability or false-negative problem that traditional methods, e.g., Granger causality or transfer entropy, cannot handle. However, these cross-mapping methods suffer still from nonlinearity and robustness problems on the noisy data. Here, we propose cross-mapping cardinality (CMC), which detects direct causality in a robust and nonlinear manner by quantifying the intersectional cardinality (IC) from the neighbors of the cause variable to the cross-mapping neighbors of the effect variable in the delay embedding space. We theoretically and computationally show the new causal concept “IC concavity”, i.e. concave IC curve against the neighbor size implies causality in the sense of dynamical causality, in contrast to the non-causality of linear IC curve. Thus, the causal strength is measured reliably by the IC curve, which exploits both IC continuity and information transfer of the cross-mapping function from effect to cause variables. Through verification on various simulated and real-world datasets, the accuracy and robustness of CMC are demonstrated significantly better than existing methods. In particular, we validated CMC with the pulse data from motor cortex neurons by training a rhesus monkey to conduct a flexible manual interception experiment. CMC effectively identified the causal relations between neurons while the traditional methods failed. In summary, our approach with the new concept of IC concavity provides a powerful data-driven tool for detecting dynamical causality in complex systems.

1. Introduction

Causality provides a deeper understanding and description of the relationship between variables in complex systems, compared with the statistical correlation. Causal inference algorithms have been widely used in many fields of natural science and engineering, such as earth science [1], economics [2], medicine [3], neuroscience [4] and machine learning [5]. Research work for quantifying causal relations can be traced back to the seminal work [6] of Neyman in the early 1920s, and the randomized controlled trial (RCT) has become the gold standard for causal inference. However, in practice, most RCTs are high-

cost or ethics-prohibited. Thus, inferring causality only by observational data is important from both theoretical and application viewpoints. The earliest work in this field, named the potential outcome model (POM), came from Rubin et al. [7–9], who calculated the average causal effect (ACE) based on the ignorability assumption (also known as *unconfoundedness*). Later, Pearl introduced the *do calculus* in the directed acyclic graph (DAG), and proved the equivalence between his structural causal model (SCM) and POM [10–12]. But SCM conducts causal inference based on known DAGs, which are usually unavailable in practical problems, and also there are Markov equivalence classes that cannot be distinguished [13]. Many related algorithms, such as the well-known PC

* Corresponding authors.

E-mail addresses: hecui@cibr.ac.cn (H. Cui), lnchen@sibs.ac.cn (L. Chen).

¹ These authors contributed equally to this work.

<https://doi.org/10.1016/j.fmre.2023.01.007>

2667-3258/© 2023 The Authors. Publishing Services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

[14], LiNGAM [15], GES [16], MMPC [17] and their variants [18–20], have been developed, mainly for time-independent data or intervention data. However, for the time-series data that is widely available in the real world, these algorithms do not take full advantage of the underlying dynamical information contained in the data [21].

In contrast, causal inference from the perspective of dynamics [22] can effectively overcome the above difficulties, such as the traditional Granger causality (GC) [23]. The core idea of GC is that if the past of the variable x is helpful to the prediction of the future of the variable y , then x is called the Granger cause of y . Transfer entropy (TE) [24] generalizes GC to nonlinear causality by introducing information entropy. Although these two types of methods have been continuously refined in recent studies [25–28], they suffer from the non-separability problem [2] (see an example in Supplementary Materials Note 5). Actually, according to the embedding theory, completely removing the information of a non-independent variable from a deterministic nonlinear system is generally impossible. The integration of phase space reconstruction and cross-mapping techniques can effectively solve the non-separability problem [2,29–36]. One of the representative methods is convergent cross mapping (CCM), which detects the consistency between the causal variable and the effect one by cross mapping. Recently, Leng et al. [37] combined CCM and partial correlation, and proposed partial cross mapping (PCM) to detect the direct causality in complex systems. However, the local linear operations in CCM/PCM limit their performance in detecting nonlinear causality [37].

In this paper, to robustly detect nonlinear causal relationships in time series data, we propose a model-free method, cross-mapping cardinality (CMC), based on a new causal concept, i.e. concavity of intersectional cardinality (IC). It quantifies IC from the neighbors of the cause variable to the cross-mapping neighbors of the effect variable in the delay embedding space, which avoids the linear operation in the existing methods (see Supplementary Materials Fig. S1 for the comparison of CMC and CCM). In particular, both theoretically and computationally, we show that concave-down IC curve against neighbor size actually implies dynamical causality [37], in contrast to non-causality case with linear IC curve. In addition, we also derive direct CMC (DCMC), which is able to further detect direct causality. Therefore, DCMC can handle both nonlinear and indirect situations while solving both the non-separability [37] and the strong association problems [22]. We show that CMC achieved state-of-the-art (SOTA) performances on different benchmark datasets. Furthermore, for real-world systems, the causal relations obtained by CMC are more interpretable than CCM and PCM, in particular for our interception experiment on rhesus monkey. In short, with the new concept of IC concavity, CMC is a powerful data-driven causal network reconstruction method that can help us better understand and model complex dynamical systems.

2. Material and methods

2.1. Cross mapping cardinality for dynamical causality

2.1.1. Original state space and its delay embedding space

For simplicity, we first consider a two-variable (x and y) dynamical system in the original state space, where the corresponding observed time series of length L are denoted as $x(t)$ and $y(t)$ with $t=1, 2, \dots, L$, respectively. According to the delay embedding theory [38], we can construct the corresponding two time series $X(t)$ and $Y(t)$ with $t=1, 2, \dots, N+1$, where $N=L-E$, in the delay embedding space (Fig. 1a), respectively, by:

$$\begin{cases} X(t) = [x(t), x(t+1), \dots, x(t+E-1)] \\ Y(t) = [y(t), y(t+1), \dots, y(t+E-1)] \end{cases} \quad (1)$$

where E is the embedding dimension and time delay interval is 1. $X(t)$ and $Y(t)$ are points on the manifold $M_X = \{X(t)\}$ and $M_Y = \{Y(t)\}$ respectively in the delay embedding space. For any point $X(t)$ on M_X , we can find its k nearest neighbors based on Euclidean distance in

the delay embedding space, denoted as $X^{kNN}(t) = \{X(t_i^x), i = 1, 2, \dots, k\}$, and their corresponding cross-mapping points on M_Y are $Y(X^{kNN}(t)) = \{Y(t_i^x), i = 1, 2, \dots, k\}$, both of which are sets with k points.

2.1.2. Linear causality in convergent cross mapping

To detect the causal links, the CCM algorithm examines the consistency of the original time series and the cross mapping reconstructed one with three steps:

- (1) Find the corresponding point set $Y(X^{kNN}(t)) = \{Y(t_i^x), i = 1, 2, \dots, k\}$ after the cross mapping of $X^{kNN}(t)$;
- (2) Perform a weighted sum of the points in $Y(X^{kNN}(t))$ to obtain the estimated $\hat{Y}(t), \hat{Y}_X(t)$;
- (3) Calculate the absolute value of the Pearson correlation coefficient (PCC) between $Y(t)$ and $\hat{Y}_X(t), \text{abs}(PCC(Y(t), \hat{Y}_X(t)))$.

The schematic diagram of the CCM algorithm is shown in Supplementary Fig. 1a. CCM uses two linear operations, i.e. weighted summation and PCC, in the last two steps, which limits its ability to detect nonlinear causality. To overcome this shortcoming, we propose the following CMC algorithm.

2.1.3. IC continuity

Here, to test the causal link from y to x in the state space, we consider the cross-mapping function $f : M_X \rightarrow M_Y$ in the embedding space, and define the intersectional cardinality (IC) between $Y(X^{kNN}(t))$ and $Y^{rNN}(t)$, i.e. cross-map from $X^{kNN}(t)$ to $Y^{rNN}(t)$, as:

$$IC(r) = \left| Y\left(X^{kNN}(t)\right) \cap Y^{rNN}(t) \right| \quad (2)$$

where $Y^{rNN}(t) = \{Y(t_i^y), i = 1, 2, \dots, r\}$ are the r nearest neighbors of $Y(t)$ and k is fixed, $r \leq N$ and $k \leq N$. Clearly, IC of Eq. 2 corresponds to cross-mapping function f . Then, we define IC continuity, which is used to derive the dynamical causality based on the continuity or smoothness of the cross-mapping function [37], as follows:

For a given positive integer k , there exists an integer r with $k \leq r \leq N$, such that $Y(X^{kNN}(t)) \subseteq Y^{rNN}(t)$ or $IC(r) = k$. IC continuity is then defined as the reciprocal of r , that is, $1/r$.

Intuitively, IC continuity measures the extent to which the cross-mapping neighbors $Y(X^{kNN}(t))$ deviate from the corresponding phase point $Y(t)$ on the Manifold M_Y . The closer the neighbors are to $Y(t)$, the stronger the IC continuity. In other words, IC can also be viewed to quantify the continuity of a dependent function from X to Y , which is also logically consistent with natural interpretation as functional dependency [39]. Specifically, considering the continuity of a function f at any observed point (X^*, Y^*) from X to Y , i.e. $Y = f(X)$, we can interpret the continuity stipulated by $\epsilon - \delta$ definition. If for every $\epsilon > 0$ there exists a δ such that $|f(X) - f(X^*)| < \epsilon$ holds for any X satisfying $|X - X^*| < \delta$, then f is continuous at point (X^*, Y^*) from X to Y . In such a sense in this work (with ϵ as k , and δ as r), we use IC continuity to approximately quantify the continuity. Since the continuity of the cross-mapping function derives the causal relation [2,22,34,37] in the sense of dynamical causality, next we use IC continuity to quantify the causality from time-series data.

2.1.4. IC causality and concavity

Fortunately, the area under the IC curve (AUC) against r provides a natural quantification of IC continuity (Fig. 1b), that is, $AUC(H_0) = 0.5kN$ (expectation) and $AUC(H_1) > 0.5kN$. According to the dynamical causality framework [22], the IC causality or causal strength (CS) for a causal link from y to x can be defined by the distance between the indexes of two contradictory H_1 and H_0 models, which can be described by the following CMC causal model,

- H_0 model assumes no causal link from y to x denoted as $y \not\rightarrow x$, then $AUC(H_0) = 0.5kN$.

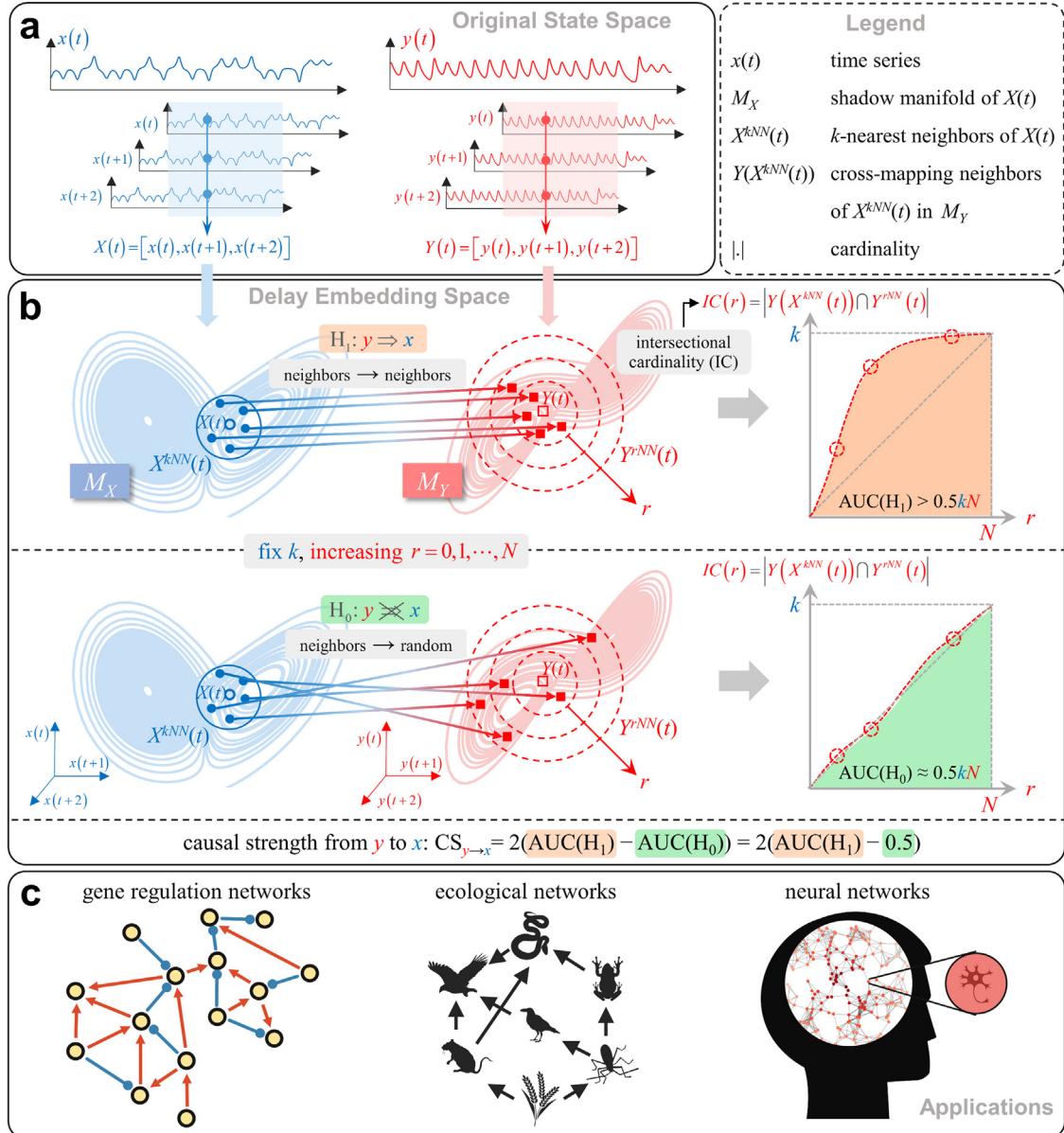


Fig. 1. The diagram of the cross-mapping cardinality (CMC) method. (a) $x(t)$ and $y(t)$ ($t = 1, 2, \dots, L$) represent the observed time series of variables x and y in original state space respectively. Two corresponding manifolds in delay embedding space can be reconstructed by the delay-coordinates embedding, where the embedded dimension E is set to 3 for display convenience. (b) In the delay embedding space, in order to define the causal strength (or score), we define two contradictory models H_1 (causal link from y to x) and H_0 (no causal link from y to x). For the H_1 model, after the k nearest neighbors of $X(t)$, $X^{kNN}(t)$, are mapped to M_Y , their corresponding points, set $Y(X^{kNN}(t))$, are also expected to be the neighbors of $Y(t)$. Therefore, with the increase of r ($r = 0, 1, \dots, N = L - E$), the intersectional cardinality (IC) between $Y(X^{kNN}(t))$ and $Y^{rNN}(t)$, $IC(r) = |Y(X^{kNN}(t)) \cap Y^{rNN}(t)|$, converges quickly to k and then saturates, which is a concave down curve in general. At this time, we use the area under the IC curve (AUC) as the index of the H_1 model, then $AUC(H_1) > 0.5kN$. For the H_0 model, since there is no causal relation from y to x , $Y(X^{kNN}(t))$ is randomly distributed on whole M_Y . For such a case, IC increases linearly/evenly with r , so $AUC(H_0) \approx 0.5kN$ (theoretically, $=$ should be taken). Based on the difference between the H_1 and H_0 models, the normalized causal strength (dividing kN) of the CMC method is defined as $2(AUC(H_1) - 0.5)$. (c) Three typical application scenarios of our method in this study.

- H_1 model assumes a causal link from y to x denoted as $y \Rightarrow x$, then $AUC(H_1) > 0.5kN$.
 - Causal strength from y to x is:
- $$CS_{y \rightarrow x} = AUC(H_1) - AUC(H_0) = AUC(H_1) - 0.5kN. \quad (3)$$

Note that, due to the IC continuity, for the H_1 model, it is possible to use only the information of x to represent or predict y , but not vice versa, which is different from GC and TE (see “Causality and predictability” below, or Supplementary Materials Note 1 for the comparison between CMC and other methods). Therefore, in the H_1 model, k points in

$Y(X^{kNN}(t))$ are expected to be neighbors of $Y(t)$, while in the H_0 model, they are expected to be randomly mapped to whole M_Y (Fig. 1b). For the extreme case of the H_1 model, $Y(X^{kNN}(t)) = Y^{rNN}(t)$ when $r = k$, we have the expectation $IC(r) = r$ for $r \leq k$, while $IC(r) = k$ for $r > k$, this is to say, IC increases to k quickly with r and then saturates. Clearly, due to the monotonous and saturational features of $IC(r)$ against r , it can be shown that IC is generally a concave down curve for H_1 model (exactly its envelope is a concave down curve due to discrete k , also see Fig. 2b,c). On the other hand, for the extreme case of the H_0 model, $Y(X^{kNN}(t))$ is uniformly randomly distributed on whole M_Y , thus $IC(r) = \frac{k}{N}r$ holds for both $r \leq k$ and $r > k$, which means that IC increases to k slowly or

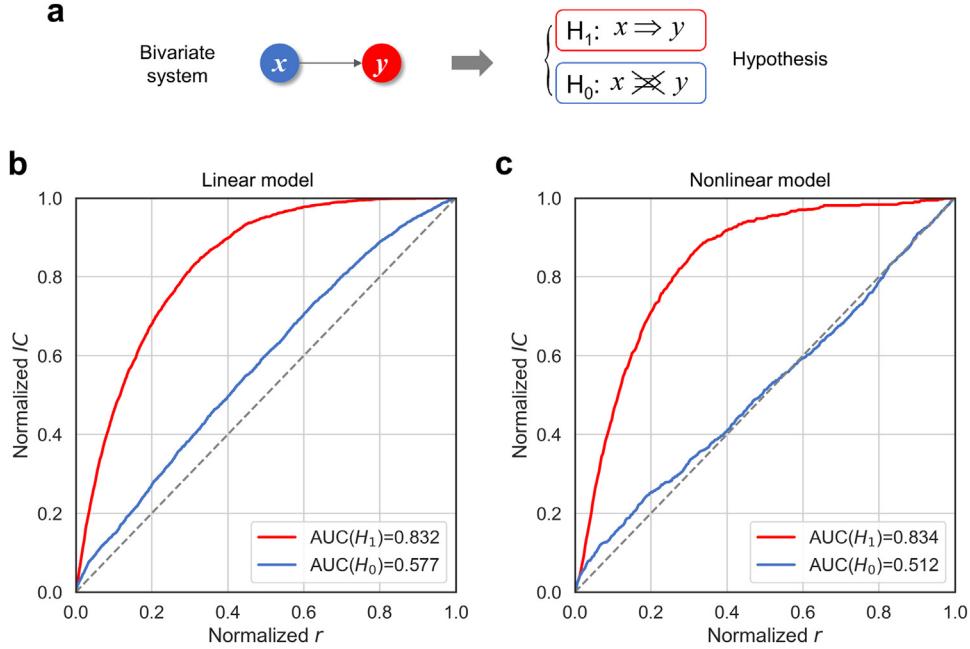


Fig. 2. Validation of CMC on two-variable systems. (a) The unidirectional case, that is, $x \Rightarrow y$, with the two contradictory H_1 and H_0 models. (b) The IC curves of the H_1 ($\beta_{yx} = 100$) and H_0 ($\beta_{yx} = 0$) of the linear model with $\varepsilon_{x,y}(t) \sim N(0, 0.1)$. (c) The IC curves of the H_1 ($\beta_{yx} = 0.3$) and H_0 ($\beta_{yx} = 0$) of the nonlinear model with $\varepsilon_{x,y}(t) \sim N(0, 0.001)$. The specific calculation process is given in Supplementary Materials Note 2.

linearly with r for the H_0 model, noting $N > k$. Thus, IC concavity for r during $N \geq r \geq k$ implies causality from y to x .

Hence, the degree of the IC curve concavity can be used as a measurement of causality. In addition to IC continuity, the area under the IC curve (AUC) also provides a natural quantitative index of the concavity (Fig. 1b). Note that k and N are constants, that is, they can be eliminated by normalization. For convenience, divided by kN , Eq. 3 can be normalized between 0 and 1 as:

$$CS_{y \rightarrow x} = \max(0, 2(AUC(H_1) - 0.5)) \quad (4)$$

It can be seen that CMC gets rid of linear operations, which can efficiently detect nonlinear causality without non-separability and strong-association problems [22,37] from both theoretical and computational viewpoints. In addition, the mathematical differences between CCM and CMC can be found in Supplementary Materials Fig. S1.

2.1.5. Causality and predictability

We take the two-variable dynamical system as an example to explain the causality and its measurement:

$$\begin{cases} y(t+1) = f(y(t), x(t)) \\ x(t+1) = g(x(t)) \end{cases} \quad (5)$$

• **Original state space.** From the first equation of Eq. 5, it can be seen that the future of y contains the past information of x , so the past of x is helpful to the prediction of the future of y in the original state space [34,37] (x, y), then $x \Rightarrow y$, which is called Granger causality. Therefore, those methods in the sense of Granger causality simply aim to construct f in Eq. 5 by linear regression (GC) or mutual information (TE) between x and y in the original state space, and then test the causality by the prediction from the constructed f . Thus, in the traditional GC method, causality is defined as predictability, that is, if the past of x is helpful to the prediction of the future of y , x is considered to be the cause of y .

• **Delay embedding space.** The cross mapping-based methods are somehow the opposite but in the delay embedding space. Intuitively, according to the implicit function theorem, there exists a function between x and y , namely, $x(t) = h(y(t), y(t+1))$, based on the first

equation in Eq. 5. Bringing it into the second equation in Eq. 5, we obtain $x(t+1) = g \circ h(y(t), y(t+1))$, which shows that $x(t+1)$ can be completely represented by $y(t)$ and $y(t+1)$, but not vice versa, i.e. $y(t+1)$ cannot be completely represented by $x(t+1)$ and its finite past. In other words, the past of y is helpful to the representation/prediction of the future of x in the delay embedding space (X, Y), then $x \Rightarrow y$, which is called dynamical causality [22,37]. The above conclusion can be extended to a more general case (see Eqs. 1-3). Therefore, cross-mapping methods in the sense of dynamical causality simply aim to construct a mapping $g \circ h$ (locally linear for CCM and PCM, while nonlinear for CMC and CME) between X and Y in two delay embedding spaces/manifolds, M_X and M_Y , and then test the causality by the prediction from the mapping $g \circ h$. As stated above, CMC in the sense of dynamical causality is theoretically able to solve the non-separability and nonlinearity problems in contrast to the traditional GC/TE. For Eq. 5, there may be some cases where the implicit function theorem does not hold, but these cases are beyond the scope of this paper (see Supplementary Materials Note 6 for a detailed discussion).

It should be noted that our CMC/DCMC method is based on the dynamical causality framework that is grounded on the assumptions of time order (causes precede effects). For example, when the function $x(t) = y(t)$ perfectly holds (without any perturbations), it is not possible to distinguish x from y from a dynamic perspective [1,40], which is beyond the scope of this work, and also such a case is rarely observed in a real system.

2.1.6. Eliminate indirect causality

So far, we have only considered any two-variable system. When there are more than two variables in a system, indirect causality may occur due to the transitivity of causality [37], for example, when $y \Rightarrow z \Rightarrow x$, there may be an indirect causality from y to x . To eliminate them, we further propose the direct CMC (DCMC) that subtracts indirect CS or information transfer from total CS in Eq. 4. To eliminate this indirect causality, various technologies can be adopted, such as conditional mutual information (CMI) [41] and partial correlation [37]. The core idea is to remove the indirect causality transmitting through intermediate vari-

ables from the total causality. Here, we propose the direct CMC (DCMC) that subtracts indirect CS or information transfer from total CS in Eq. 4. First, considering the case of only one intermediate variable z , we use the product of the causal strength of y to z and the causal strength of z to x to represent the indirect causal strength (ICS) of y to x . Specifically, by using CS of Eq. 4, we fix the number of nearest neighbors of $Z(t)$ and $X(t)$ as k , and gradually increase the number of nearest neighbors r of $Y(t)$, thus by removing the indirect effect from y to x :

$$ICS_{y \rightarrow x} = CS_{z \rightarrow x} \times CS_{y \rightarrow z} = \frac{1}{k} \left\langle \left| Z(X^{kNN}(t)) \cap Z^{kNN}(t) \right| \right\rangle_t \times CS_{y \rightarrow z}, \quad (6)$$

we have the direct causal effect from y to x as:

$$CS_{y \rightarrow x}^{DCMC} = CS_{y \rightarrow x} - ICS_{y \rightarrow x}. \quad (7)$$

Intuitively, it removes the indirect IC information transfer from total IC information transfer. It should be noted that the product form of indirect causation used in Eq. 6 is only an empirical scheme, and other more sophisticated schemes can also be adopted. Nevertheless, because of its simple form and good performance, it has been widely used in previous research [42].

For the case of multiple intermediate variables, the emergence of high-order indirect causality, e.g., $y \Rightarrow z^1 \Rightarrow z^2 \dots \Rightarrow z^s \Rightarrow x$, will complicate the calculation. For such a case, the computational cost increases exponentially. Therefore, in practice, we only consider the first-order case. In addition, as in the previous work [37], we select several nodes with the highest possibility of indirect causality as intermediate variables by maximizing $CS_{z_i \rightarrow x} + CS_{y \rightarrow z_i}$, and then eliminate the IDS transmitted through these nodes. A more detailed algorithm for DCMC to calculate the causal strength of a multivariate system can be found in Supplementary Materials Note 1.

2.2. Datasets and parameters

In the following, we will test our method on multiple simulated and real-world datasets, involving gene regulatory networks, ecological networks, and in particular, neural networks from our interception experiment on rhesus monkey (Fig. 1c).

2.2.1. Simulated datasets and real-world datasets

The efficiency of CMC was firstly validated on two simulated models, logistic map and Henon map, then on four widely used real-world datasets, gene regulatory network, COVID-19 in Japan, food chain network and cardiovascular disease. The introductions of these datasets and the corresponding preprocessing can be found in Results or Supplementary Notes 2 and 3. In addition to these publicly available datasets, we verified the CMC method using the neural dataset obtained by our manual interception experiment on rhesus monkey, which is explained further below.

2.2.2. Flexible manual intercept task and neural data acquisition

A rhesus monkey (male, 9.5 kg body weight) was trained to perform a flexible manual intercept task. A trial begins with a circular target appearing in the center of the screen, and by holding the circular target for a fixed amount of time (fixation epoch: 600 ms). After that, a moving or stationary target will appear on the periphery of the center, and the moving target will have four speeds, 120 and 240 degrees per second clockwise or counterclockwise. At this time, the monkey needs to hold the center for a variable amount of time until the center point goes dark, which prompts the monkey to leave the center point and reach the peripheral target (cue/delay epoch: 400–600 ms). Eventually, the monkey will complete the interception and the display will give feedback on the actual point and press position. The total movement time was required to be within 800ms (move and feedback epoch: <800ms). For the percentage of the task conditions, it is one-fifth of the condition for each of the five speeds. Two 128-channel Utah arrays (Blackrock Inc., USA)

were implanted in the premotor and primary motor cortex and single and multi-unit activity were recorded simultaneously. Neural activity was recorded at full bandwidth with a sampling frequency of 30 kHz and a resolution of 16 bits (Cerebus, Blackrock Inc. USA). Spikes were detected (threshold: $4.5 \times \text{std}$) and extracted and then spike sorting was processed by an in-house developed method based on NeuralEnsemble and done further visual inspection by Waveclus in MATLAB to detect and remove neural drift and artifacts. After spike sorting, spike trains were aligned to three different periods and binned by a 1ms time window for further analysis (TC: [100–600ms], GO: [-500–0ms] and MO: [-200–300ms]). Animal housing, care, and all experimental procedures were approved by the Animal Care Committee of Shanghai Institutes for Biological Sciences (ER-SIBS-221603P), Chinese Academy of Sciences.

2.2.3. Generate simulated neural networks

Following the previous work [43], the firing rate data of the real neural network at the MO period were used to generate simulation data with consistent firing patterns by an inhomogeneous Poisson process. After that, five Erdős-Rényi random networks [44] with 32 nodes and a connection probability of 0.05 were generated as the ground truths. Finally, the Gamma transfer function ($a=5$, b is randomly varied from 0 to 3) was used to add connections to the simulation data (with a time delay of 1ms, 400ms in total). Considering that the spiking time series data were discrete (0 or 1), the points on the manifold often coincide with each other, thus a weak noise ($N(0, 0.001)$) was added to the simulation data when calculating the causal strength. For each network model, 100 independent simulations were carried out, and the average of the causal strength of 100 simulations was taken as the result.

2.2.4. Determine the parameters

Similar to CCM and its variants, phase space reconstruction and cross mapping are two cornerstones of our CMC approach. For phase space reconstruction, theoretically, the original attractor (in the original state space) and reconstructed manifold (in the delay embedding space) are topologically equivalent when $E > 2d$, where d is the fractal dimension of the original attractor. There are already some mature approaches [36,45], and in this work, false nearest neighbor (FNN) and delayed mutual information (DMI) were applied to determine E and τ (time interval), respectively. Furthermore, the effects of different parameters on the results were tested on partial datasets (e.g., Fig. 3b and Fig. 6g), and the results show that our method is robust to the parameters. For cross mapping, without specific instructions, $E + 1$ nearest neighbors (namely, $k = E + 1$) was used since it is the minimum number of points needed for a bounded simplex in an E -dimensional space. The parameters used for each dataset are given in the corresponding captions.

3. Results

3.1. Performance on simulated datasets

3.1.1. 2-variable causality

To give the reader a general view of our method, we first consider the simplest unidirectional case, that is, $x \Rightarrow y$ (Fig. 2a). Two representative dynamical models are utilized in this work, namely, a linear vector autoregression (VAR) model:

$$\begin{cases} x(t+1) = 0.45x(t) + 0.45x(t-1) + \varepsilon_x(t) \\ y(t+1) = 0.45y(t) + 0.45y(t-1) + \beta_{yx}x(t) + \varepsilon_y(t) \end{cases} \quad (8)$$

and a nonlinear model:

$$\begin{cases} x(t+1) = x(t)(3.7 - 3.7x(t)) + \varepsilon_x(t) \\ y(t+1) = y(t)(3.7 - 3.7y(t) - \beta_{yx}x(t)) + \varepsilon_y(t) \end{cases} \quad (9)$$

where $\beta_{yx}, \varepsilon_x(t)$ and $\varepsilon_y(t)$ represents the coupling coefficient from x to y and the noise of x and y , respectively. As shown in Fig. 2b, the AUCs of the H_1 ($\beta_{yx} = 100$) and H_0 ($\beta_{yx} = 0$) of linear model with $\varepsilon_{x,y}(t) \sim N(0, 0.1)$ are 0.832 and 0.577, respectively. Similarly, the AUCs

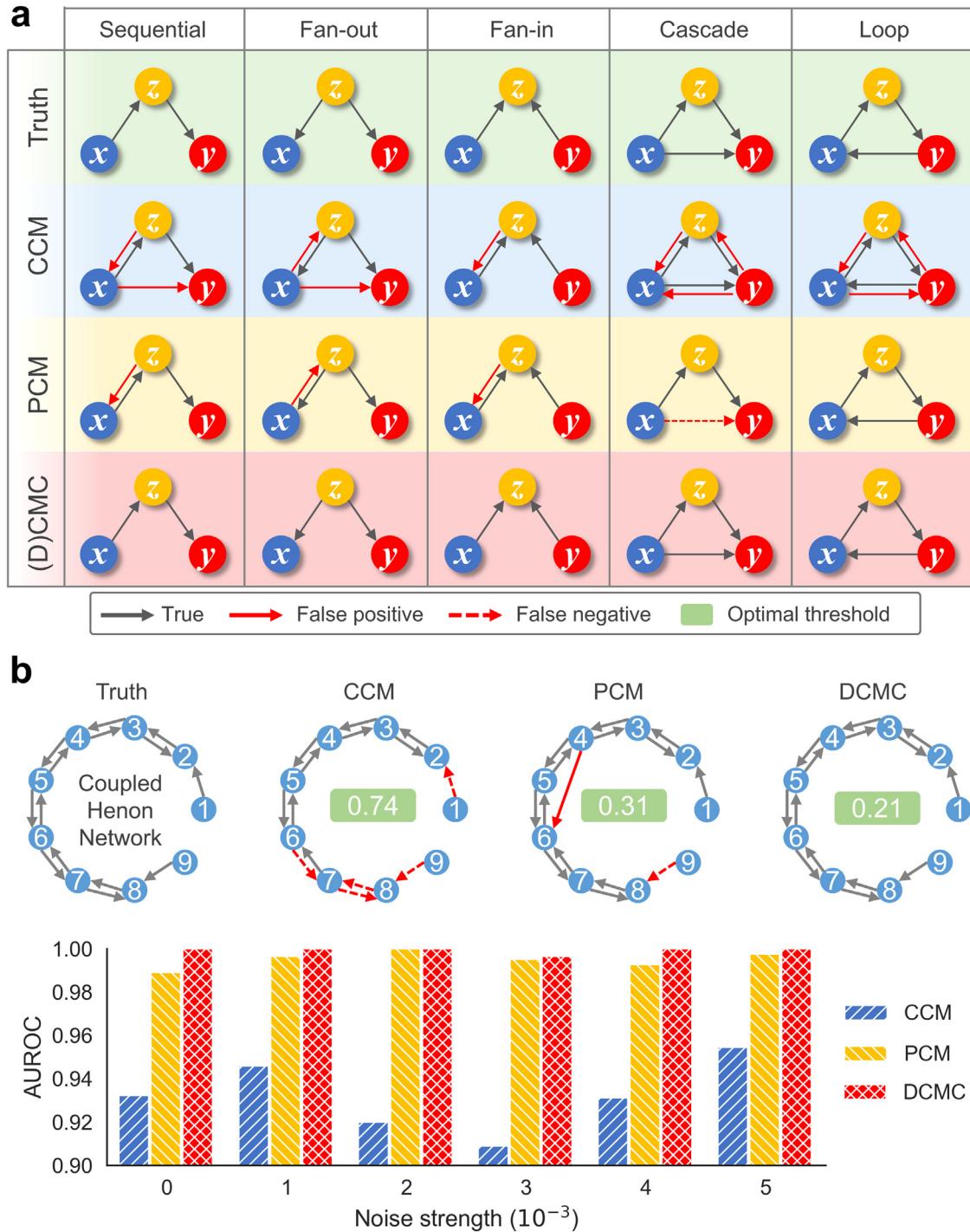


Fig. 3. Performance on simulated datasets. (a) The five columns from left to right represent five representative 3-variable logistic networks, namely, sequential, fan-out, fan-in, cascade and loop. The four rows from top to bottom represent the real network structure and the predicted structures of CCM, PCM and DDCMC, respectively. The red solid and dotted arrows represent the false positive and false negative causal links, respectively (for fairness, the threshold of causal strength is 0.5 for all). The parameters of phase space reconstruction of all the three methods are $E = 3$. (b) The four columns from left to right (upper panel) are the real structure of the 9-variable Henon network and the optimal predicted structures (with the least error links) of CCM, PCM and DDCMC, respectively. The meaning of the arrow is consistent with that in a). The parameters of phase space reconstruction are $E = 7$. The optimal threshold (green rectangle) is the threshold with the least false links when it continuously increases from 0 to 1 at an interval of 0.01. (lower panel) For the 9-variable Henon network, the statistical results of the area under the receiver operating characteristic curves (AUROCs) of CCM, PCM and DDCMC under different Gaussian noise strengths. For the 3-variable logistic networks and 9-variable Henon network, 100 and 10 simulations ($L = 1000$) are carried out under different Gaussian noises (strength are 0.003 and 0.004 in a and b), respectively. Finally, the average results of these simulation data are taken as the final causal strengths.

of the H_1 ($\rho_{yx} = 0.3$) and H_0 ($\rho_{yx} = 0$) of the nonlinear model with $\varepsilon_{x,y}(t) \sim N(0, 0.001)$ are 0.834 and 0.512, respectively (Fig. 2c). These results are in line with the definition of the causality from the concept of IC concavity. Furthermore, to verify the robustness of our method, we further calculated the AUCs of the H_1 and H_0 on linear and nonlinear models with different ρ_{yx} under different types and intensities of noise (Supplementary Materials Figs. S2, S3), and the results are also consistent with those in Fig. 2b,c. These simple examples verify that our method can indeed detect causality.

3.1.2. 3-variable causality

Below we illustrate the superiority of our method with three variables (x , y , and z) (see Supplementary Information (SI) for the data generation of a logistic system). As shown in Fig. 3a, five typical cases are considered. The first three are sequential, fan-out and fan-in cases, which correspond to the three basic V-structures in SCMs, namely chain, fork and collider, respectively. When the thresholds of all methods are set as 0.5, compared with the real causal network, the predicted networks of both CCM and PCM have false positive links (red solid arrow) from x to z , which may be caused by nonlinearity. In contrast, our CMC algorithm can eliminate these false positive results. In addition, in the case of sequential and fan-out, due to the existence of z , CCM wrongly includes an indirect link from x to y . PCM uses partial correlation to eliminate this spurious link, while CMC can eliminate this link without considering indirect causality. In addition to the above V-structures, two more complex cases are also considered, one is the cascade case where both direct and indirect causality coexist, and the other is the cyclic case that the SCM can not handle. In the two cases, CCM is unable to determine the direction of causality. Although PCM can remove all false positive links, it will also generate false negative links (red dashed arrow), which means that when there are both direct and indirect causal links, PCM may remove the direct one. Again, only the CMC method can correctly predict both causal networks. We also compared the above methods with the traditional GC and TC methods. Since both CG and TC have the problem of scale bias, when the threshold is set to 0.5, neither method can detect causality (see Supplementary Materials Figs. S4-S8 for detailed results). In Supplementary Materials Fig. S9, we further consider the effect of threshold and noise on the results, and our method is the most robust to both factors.

3.1.3. 9-variable coupling Henon network

Next, we consider a more complex situation, that is, a 9-variable coupled Henon map network (see Supplementary Materials Note 2 for the dynamical equations). The real causal network and the optimal predicted network of different methods are shown in Fig. 3b (upper panel), where neither CCM nor PCM can predict the correct network. By plotting the computed CS matrixes, it can be seen that CCM has a problem of false positives, and PCM can alleviate the problem, while our method can remove all false positive links (Supplementary Materials Fig. S10). To give a more quantitative analysis, the area under the receiver operating characteristic curve (AUROC) and the precision-recall curve (AUPRC) and structural Hamming distance (SHD) are applied to evaluate the performance of the prediction. As shown in Fig. 3b (lower panel), under different noise levels, the AUROCs of DCMC are better than those of CCM and PCM. Similar results can be obtained for AUPRC and SHD (Supplementary Materials Fig. S11). In addition, the effect of different embedding dimensions E on the result is considered, which shows that our method is the most robust (Supplementary Materials Fig. S12).

3.2. Validation on real-world datasets

The above results on simulated benchmark datasets demonstrate several advantages of our method: 1) accurate to detect nonlinear causality; 2) precise to distinguish direct and indirect causality; 3) robust to noise or parameters. Below we validate our method on several real networks including a flexible manual interception experiment on motor cortex

neurons by training a rhesus monkey. How to preprocess the raw data is described in Supplementary Materials Note 3.

3.2.1. Gene regulatory network

The time-series data are collected from the DREAM4 Network Challenge [46], which includes 10 gene regulation networks (GRNs), of which the first 5 networks have 10 genes and the last 5 networks have 100 genes. The network topologies are obtained by extracting subnetworks from transcriptional regulatory networks of *E. coli* and *S. cerevisiae* [47]. Fig. 4a shows one true network with 10 genes (other networks can be seen in Supplementary Materials Figs. S13, S14). Since the real structures of these networks are known, the ROC curves can be plotted for all methods (see Fig. 4b for the ROC curves of the network in Fig. 4a, and see Supplementary Materials Figs. S13 and S14 for the others) and the AUROC can be calculated. Fig. 4c-e compare the distributions of AUROCs of CCM, PCM and DCMC methods in the first 5 networks, the last 5 networks and all networks, respectively. It can be found that DCMC significantly outperforms CCM and PCM (paired t-test, p-values < 0.01 (**)). Similar results can be obtained for AUPRC and SHD Supplementary Materials Fig. S15).

3.2.2. COVID-19 in Japan

We collected the data on the number of cumulative confirmed COVID-19 cases in all 47 prefectures in Japan in 2020, where Tokyo has the earliest confirmed case and far more cumulative confirmed cases than the other 46 prefectures. Our result shows that Tokyo is a major source of the outbreak of COVID-19 in Japan. In Fig. 5a, we analyzed the relationship between the geodesic distances from 46 prefectures to Tokyo and the DCMC-inferred causal strengths from Tokyo to them. Clearly, the smaller the geodesic distance, the more the number of confirmed cases, and the larger the causal strength. Furthermore, when only considering the 10 prefectures (red circles) with the largest number of the confirmed cases except for Tokyo, there is a clear negative correlation between the geodesic distance and the causal strength (the coefficient of determination, R^2 , is 0.67), which is consistent with our common sense. In addition, the top 5 prefectures affected by Tokyo are Kanagawa, Osaka, Saitama, Chiba, and Aichi (Supplementary Materials Fig. S16 shows the cumulative and newly confirmed cases in these prefectures and the causal strength matrix of DCMC). As shown in Fig. 5b, among these 5 prefectures, Kanagawa, Saitama and Chiba are Tokyo's neighboring prefectures and are in the Tokyo metropolitan area (red dashed circle), while Aichi and Osaka are core prefectures of Nagoya and Osaka metropolitan areas (blue and green dashed circle), respectively. This result implies that the spread of the COVID-19 is related to both distance and population level or economic activity, which is quite reasonable due to their more advanced transportation systems, which are more conducive to the spread of the COVID-19. All detection results are verified using hypothesis testing (Supplementary Materials Note 4), and the corresponding p-values are given in Supplementary Materials Table S1.

3.2.3. Air pollution and cardiovascular disease

Supplementary Materials Fig. S17a, it shows the concentrations of four air pollutants, NO_2 , O_3 , SO_2 and respirable suspended particulates (RSP), and the number of cardiovascular patients in hospitals in Hong Kong from 1994 to 1997 [48]. According to the computed result of DCMC, both NO_2 and RSP are the causes of cardiovascular diseases (CVDs), which is consistent with previous works [36,49-51] (Fig. 5d). In contrast, other methods, while detecting these two causal relations, yield contradictory results, such as CCM and PCM finding that CVDs are the cause of SO_2 (Supplementary Materials Fig. S17b).

3.2.4. Food chain network

Next, we consider a food chain network consisting of four species, *Pico cyanobacteria*, *Rotifers*, *Nano flagellates* and *calanoid copepods*. The time-series data record the population of the four species during an 8-year ecological experiment [52] (Supplementary Materials Fig. S18a).

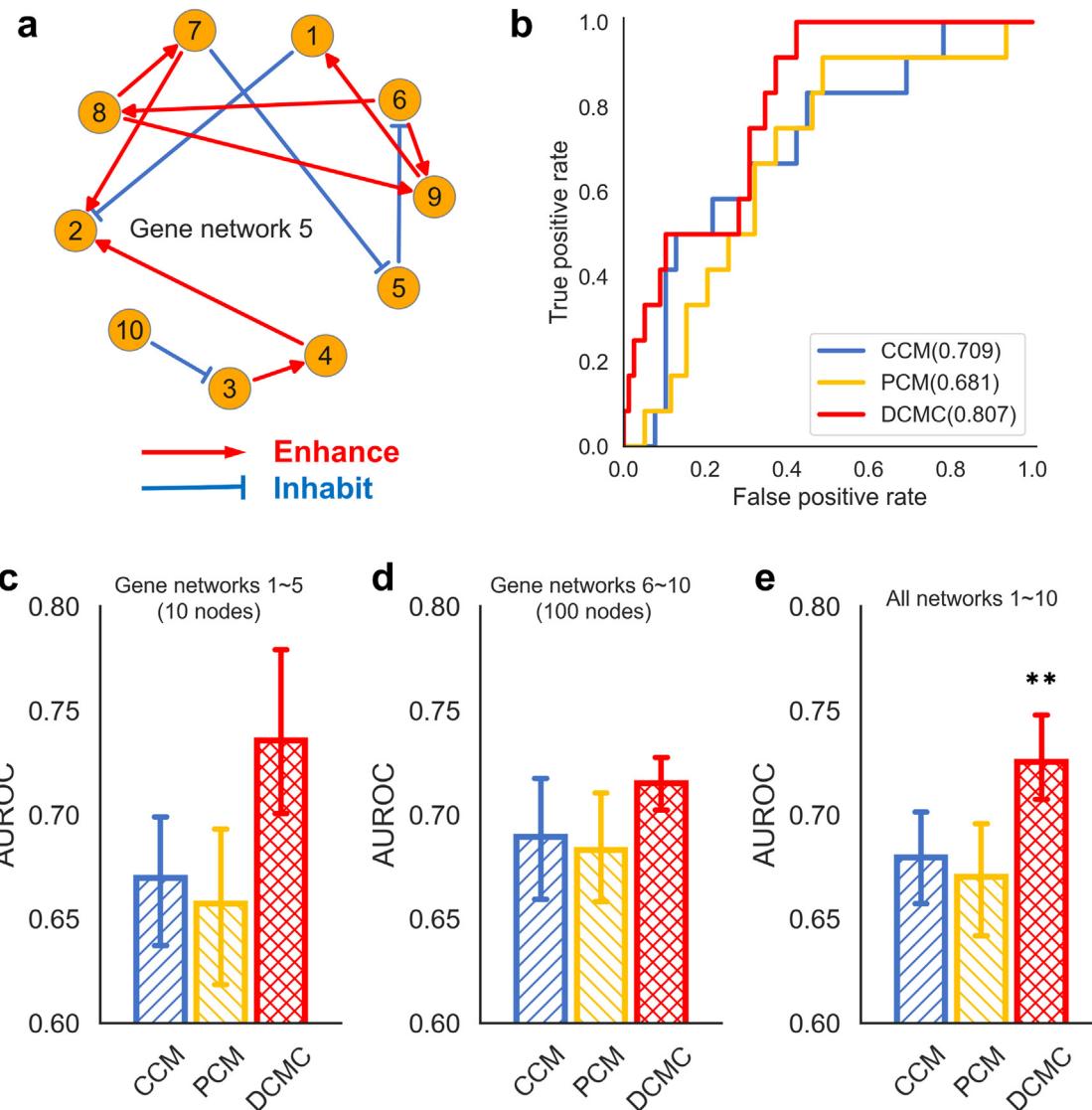


Fig. 4. Results of gene regulatory network (GRN) inference. (a) The fifth GRN with 10 interacting genes from the DREAM4 Network Challenge. The red and blue arrow represents an activating and inhibitory effect, respectively. (b) The ROC curves of CCM, PCC and DCMC for the GRN in a). The corresponding AUROCs are also indicated. The statistical results of AUROCs of CCM, PCC and DCMC on five 10-node GRNs, five 100-node GRNs and all GRNs are shown in c-e), respectively. The reconstruction parameters are $E = 9$.

Based on this data, the DCMC method successfully identified three out of four known predation relations, the only missing one is from *Pico cyanobacteria* to *calanoid copepods* (Fig. 5c). In contrast, other methods lead to at least five false positive links on the premise of ensuring that at least three causalities are correctly identified, which obviously violates the ecological rule (Supplementary Materials Fig. S18b).

3.2.5. Neural network of the rhesus monkey's motor cortex

The neural system is usually considered to be a dynamical system, which has also been used to understand the neural mechanisms of motor control [53–58], thus DCMC, in principle, can be applied to reconstruct the neural network generating motor commands. To assess the applicability of DCMC to the spiking time-series data from real neural networks, a rhesus monkey was trained to perform a flexible manual intercept task [59,60] (see Material and methods) in this work. As shown in Fig. 6a, three task periods, touch center (TC), delay (GO), and movement onset (MO), were selected in each correct trial. In the TC period, a circular target appears in the center of the screen, and the monkey was required to hold the target. Then in the GO period, a moving or stationary target appeared on the periphery of the screen and the monkey

was required to keep holding the central point. Finally, the darkening of the center point instructed the monkey to reach out to intercept the peripheral target in the MO period. In each trial, the neural activities of 136 single units located in the motor cortex were recorded. In total, 1378 correct trials were performed and Fig. 6b shows the corresponding endpoints that are approximately distributed on a circle. To analyze the differences in the neural network when the monkey learns different directions, all trials are divided into eight groups according to movement direction (denoted as D1~D8), and then we use DCMC to calculate the causal strength matrix based on the spike data in each direction (average the data in each direction, see Material and methods). Fig. 6c shows the reconstructed neural networks for three periods in the first direction (D1) with a threshold of 0.2. An intuitive result is that the tightness of the network in the MO period is higher than that of the TC and GO periods, that is, there are more nodes with large out-degrees (consistent results can be obtained with data in all directions, see Supplementary Materials Fig. S19). In order to further analyze the rationality of the reconstructed networks, we define the index distance between D_i and D_j as $\min(\text{abs}(i - j), 8 - \text{abs}(i - j))$, where $i, j = 1, 2, \dots, 8$ and $i \neq j$, and the root mean square errors (RMSEs) of neural networks in different peri-

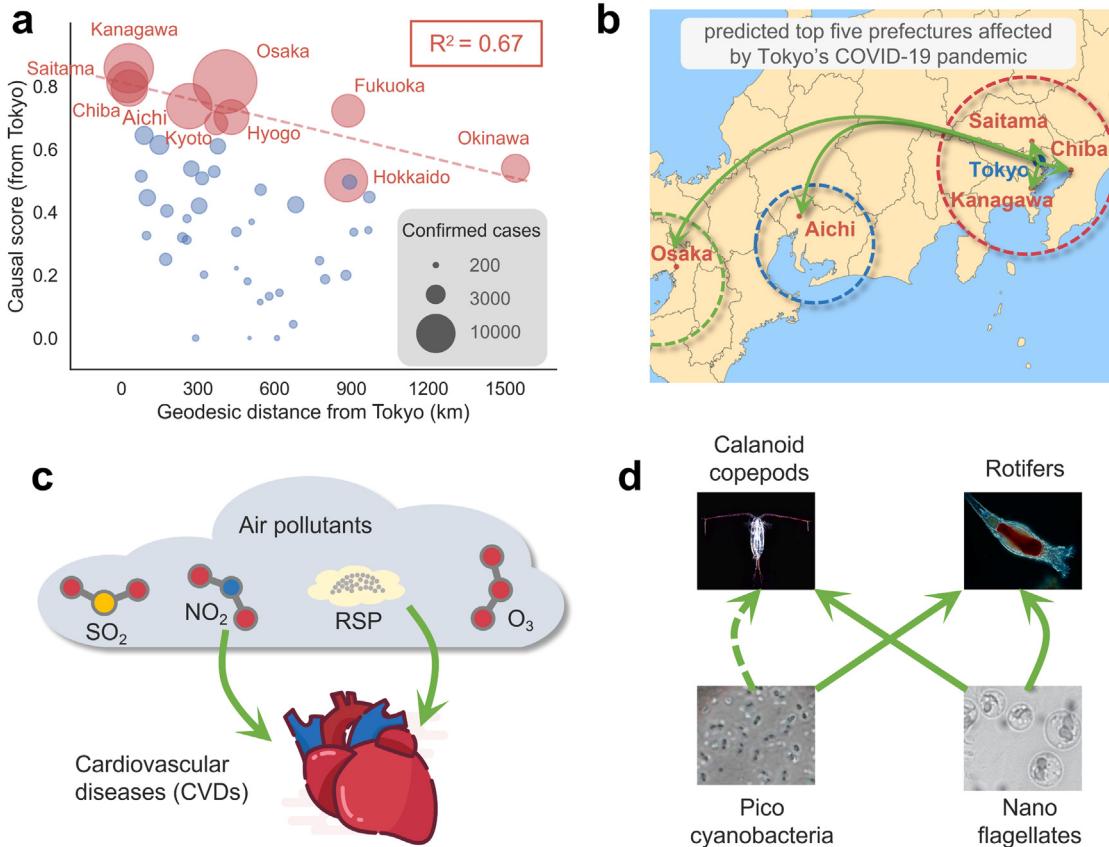


Fig. 5. Validation of CMC on real-world datasets. (a) The scatter diagram of the DCMC-predicted causal strengths from Tokyo to other 46 prefectures in Japan and the geodesic distances from Tokyo to these prefectures. The size of the point represents the cumulative number of confirmed cases of COVID-19 in that prefecture, and 10 prefectures greater than 3000 are marked with red. For these 10 prefectures, the linear regression curve (red dotted line) reveals an obvious negative correlation between the causal strength and the geodesic distance (the determination coefficient, $R^2 = 0.67$). (b) In particular, the five prefectures most affected by Tokyo, Kanagawa, Osaka, Saitama, Chiba, and Aichi, are located on the map. Red, blue and green dashed rings represent Tokyo, Nagoya and Osaka metropolitan areas, respectively. DCMC-detected causal interactions between air pollutants and cardiovascular diseases and a four-species food chain network are shown in (c) and (d), respectively. Real and missing causal links are represented by green solid and dashed arrows, respectively. For these datasets, the reconstruction parameters are $E = 20$.

ods and directions are compared. As shown in Fig. 6d, in both the TC and GO periods, the RMSEs between the 8 directions are relatively small (<0.05), while in the MO period, the RMSEs between them are significantly increased. Moreover, the increase of the index distance in the TC period does not cause a change in the RMSE, and in the GO period, the RMSE increases slightly with the index distance. In contrast, the RMSE increases significantly only in the MO period, indicating the completion of the motor preparation and the occurrence of the actual movement, and the causal network of neurons for different motor directions still reflects the property of cosine coding [61]. Notably, this coding mechanism, which allows for efficient encoding of the direction of the manual movement, is widely found at the single neuron level, while our results suggest that this mechanism is also present at the neural network level. In addition, we use the principal component analysis (PCA) to reduce the dimensionality of the causal strength matrix in different directions and construct a causal network state space of neural activity. To ensure that the space after the dimensionality reduction of three different periods is the same, the three corresponding causal strength matrixes are integrated and the first three principal components (PC1: 52.8%, PC2: 9.7%, PC3: 8.6%, 71.1% in total) are used to visualize the states of different movement directions. As shown in Fig. 6e, the 8 directions in the TC period are tangled with each other, and when it comes to the GO period, the difference between these 8 directions is still very small (can be enlarged to see a rough ring structure), while only in the MO period, the geometry of the eight directional states presents a distinct ring structure. It is worth mentioning that this geometric feature is also the

key for neural networks to store a sequence in memory [62]. Taken together, the causal strength matrix computed by the DCMC method can not only accurately characterize the temporal and spatial information of neuronal dynamics simultaneously, but also reveal the underlying geometric characteristics of the original spiking data in our rhesus experiment.

However, it is difficult to obtain real synaptic connectivity information while recording population activity with existing multichannel recording techniques *in vivo*. Direct verification of the causality in an anatomical sense will be considered in our future work. Here, we additionally constructed datasets with the same neural activity patterns as the real data for validating our method (see Material and methods for details). In short, we generated five 32-nodes neural network models based on experimental data. Fig. 6f compares the ROC curves of CCM, PCM and DCMC for the first model, and the AUROCs of CCM and PCM are less than 0.6, which means that both of them are unable to correctly detect causality (or neural connection) in neural networks. In contrast, the AUROC of DCMC is close to 0.9, which is much higher than CCM and PCM. The statistical AUROC results for all five networks under different embedding dimensions are shown in Fig. 6g, which shows that the performance of DCMC does not depend on its parameters, which strongly implies the efficiency of the DCMC method for neural networks. It is worth noting that this improvement comes from the new nonlinear detection ability, rather than removing the contribution of indirect causality (Supplementary Materials Fig. S20).

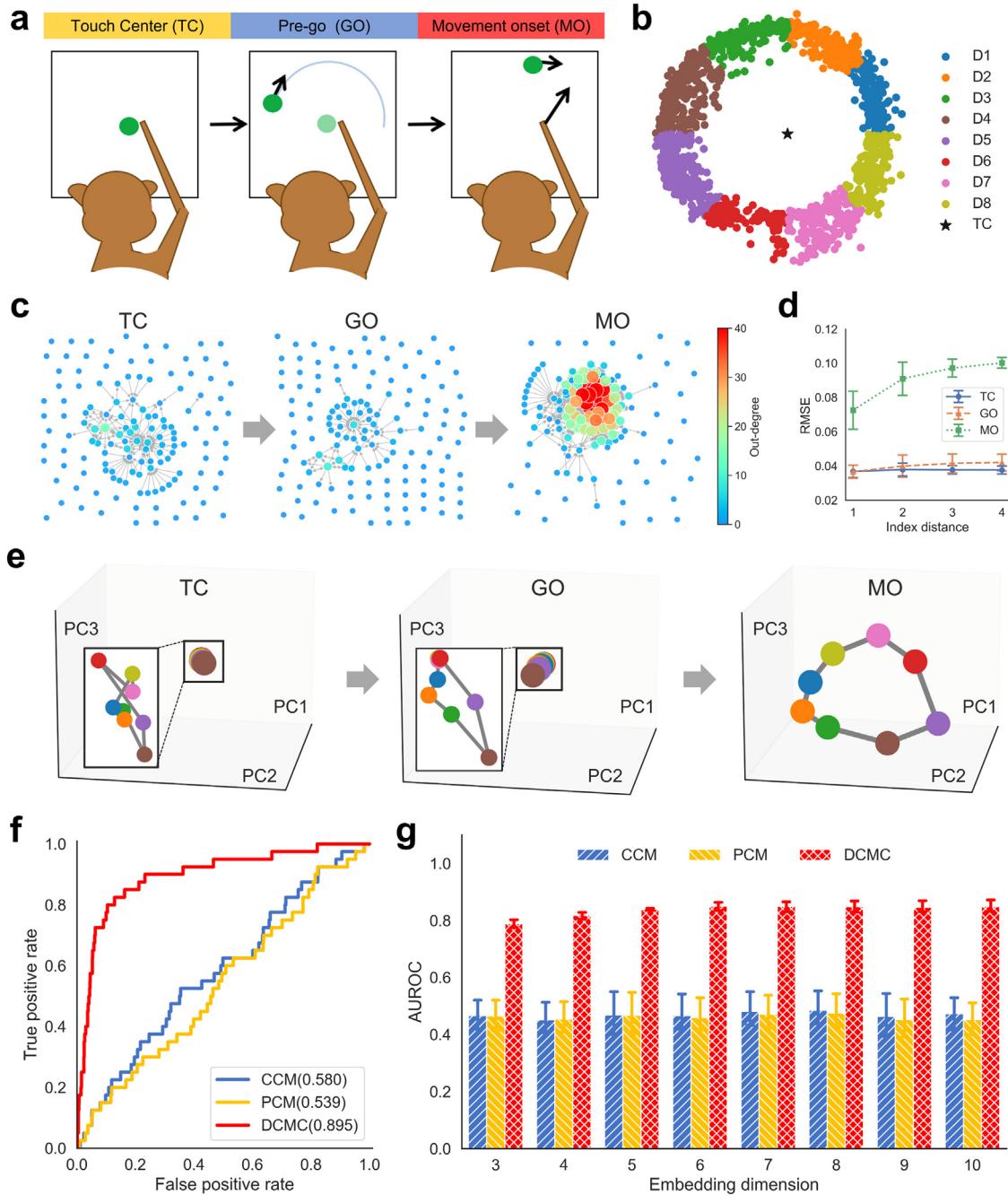


Fig. 6. Results on neural networks of rhesus monkey. (a) Schematic diagram of our neurophysiological experiment of monkey behavioral paradigm. Behavioral and neuronal activities are recorded during the flexible manual interception task. Three time periods, TC, GO and MO, are selected for subsequent causal analysis. (b) The endpoints of the reaching movement of the monkey, and each endpoint represents a trial (1378 in total). These endpoints are divided into eight parts (0 to 360°, 45° for each), representing the eight directions of movement. (c) The DCMC-inferred neural networks in the first direction (D1) for three periods (threshold is 0.2). The size of the node reflects its out-degree. (d) The change curves of RMSE with the index distance between different directions in the three cycles. (e) Visualization of dimensionality reduction of causal strength matrixes in the eight directions. The three periods share the same three-dimensional space. (f) ROC curves of CCM, PCM and DCMC on one of the five 32-node simulated neural networks. (g) Statistical results of AUROCs of CCM, PCM and DCMC on all five simulated neural networks under different embedding dimensions.

4. Discussion and conclusion

In this paper, we proposed the CMC method based on a new concept of IC concavity in the sense of dynamical causality, which mainly solves the non-separable problem against GC/TC and the nonlinear problem against CCM/PCM. In fact, to solve the nonlinear problem, another way is to introduce mutual information (MI) into CCM or PCM. Actually, we have constructed such entropy denoted as *cross-mapping entropy*

(CME) (see Supplementary Materials Note 1 for details of CME and also DCME), i.e.:

$$CS_{y \rightarrow x}^{CME} = MI\left(Y\left(X^{kNN}(t)\right), Y(t)\right) = H(Y(t)) - H\left(Y(t)|Y\left(X^{kNN}(t)\right)\right) \quad (10)$$

which is effective for dynamical causal inference, and also has a compact mathematical form in the delay embedding space (X, Y) by exploiting cross-neighbor-mapping information, which is different from *transfer*

entropy in the original state space (x, y) , i.e.:

$$\begin{aligned} CS_{y \rightarrow x}^{TE} = & H(x(t)|x(t-1), \dots, x(t-n)) \\ & - H(x(t)|x(t-1), \dots, x(t-n), y(t-1), \dots, y(t-n)) \end{aligned} \quad (11)$$

where n is referred to as the “memory time”. Clearly, there are two differences, i.e. (a) TE is measured in the original state space while CME is in the delay embedding space; (b). TE is measured from cause y (n past points) to effect x while CME is from effect X (k neighbor points) to cause Y . However, for a 3-variable logistic system, it is found that CME is less effective than CMC, producing false negatives for the cascading case, that is, some direct causal relations cannot be detected (Supplementary Materials Fig. S21), which may be due to the misesimation of strong association problem [63–65]. The new concept of IC concavity introduced by the CMC method effectively solves this problem. Moreover, testing on a large number of simulated and real-world datasets demonstrates that our method not only inherits the advantages of the existing cross-mapping methods, such as solving the non-separability problem and less scale bias, but also is able to detect nonlinear causality, remove indirect causality in a robust and accurate manner.

Notably, in this work, the concept of causality is discussed from the perspective of dynamics, requiring time series data, different from the current statistical framework. For example, in the dynamical causality framework, the simplest way to define causality is the coupling in dynamical equations, while in POM or SCM, intervention is usually necessary, which requires intervention data rather than time-series data. But this does not mean that the two perspectives are antagonistic, on the contrary, the two frameworks are complementary to each other. The dynamic causality perspective can provide the missing parts of the statistical one, such as non-separability for a general nonlinear dynamical network. How to combine these two perspectives in practice is a topic worthy of further study.

Our method is data-driven by time series with similar requirements to other cross mapping-based methods. (1) It assumes that there is an unknown dynamical model [66] behind the observation data, rather than a system dominated by noise. (2) In a nonlinear dynamical system, strongly coupled (or synchronized) variables without any noise can make the complete system collapses into a sub-manifold. For such a case, the direction of the causation is indeterminate since the effect variable becomes an observation function on the cause system [2]. (3) Because the cross mapping-based methods need to find neighbor points in the delay embedding space, the samples of time series should not be too small (e.g., > 10). Too less of samples may greatly reduce the prediction performance. In addition to data, there are also some important problems in post-processing, one of which is how to set the threshold values to reconstruct the causal network after obtaining the causal strength matrix. There are some methods [67] for setting the threshold values automatically, but for real-world datasets whose causal intensities change in a wide range, they cannot always provide satisfactory results conforming to reality. Thus, to extract the correct causality, prior knowledge is needed for many real systems [68]. Further improvements regarding the above weaknesses will be included in our future work.

Data and code availability

Datasets and code used in this study have been integrated into a Python package, which is available at <https://github.com/PengTao-HUST/crossmapy>. The intact time-series data of the real neural network are available from the corresponding author on reasonable request.

Declaration of competing interest

The authors declare that they have no conflicts of interest in this work.

CRediT authorship contribution statement

Peng Tao: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Visualization. **Qifan Wang:** Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Visualization. **Jifan Shi:** Formal analysis, Methodology, Software, Investigation, Writing – original draft. **Xiaohu Hao:** Software, Investigation, Formal analysis. **Xiaoping Liu:** Investigation, Formal analysis. **Bin Min:** Investigation, Formal analysis. **Yiheng Zhang:** Resources, Data curation. **Chenyang Li:** Resources, Data curation. **He Cui:** Writing – review & editing, Project administration, Funding acquisition. **Luonan Chen:** Conceptualization, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Acknowledgments

We thank T. Wang and R. Zheng for their discussions and surgical assistance. This work was supported by National Key R&D Program ([2022YFA1004800](#), [2022YFC2704600](#), [2022YFC2704604](#)), National Science Foundation of China ([12131020](#), [31930022](#), [12026608](#)), CAS Strategic Priority Research Program ([XDB38040400](#), [XDB32040103](#)), Lingang Lab ([LG202105-01-02](#)), Shanghai Municipal Science and Technology Major Project ([2021SHZDZX](#)), Special Fund for Science and Technology Innovation Strategy of Guangdong Province ([2021B0909050004](#), [2021B0909060002](#)), Major Key Project of Peng Cheng Laboratory ([PCL2021A12](#)), and JST Moonshot R&D ([JPMJMS2021](#)).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.fmre.2023.01.007](https://doi.org/10.1016/j.fmre.2023.01.007).

References

- [1] J. Runge, S. Bathiany, E. Boltt, et al., Inferring causation from time series in Earth system sciences, *Nat. Commun.* **10** (2019) 2553.
- [2] G. Sugihara, R. May, H. Ye, et al., Detecting causality in complex ecosystems, *Science* **338** (2012) 496–500.
- [3] J.G. Richens, C.M. Lee, S. Johri, Improving the accuracy of medical diagnosis with causal machine learning, *Nat. Commun.* **11** (2020) 3923.
- [4] Y. Chen, S.L. Bressler, M. Ding, Frequency decomposition of conditional Granger causality and application to multivariate neural field potential data, *J. Neurosci. Meth.* **150** (2006) 228–237.
- [5] B. Schölkopf, F. Locatello, S. Bauer et al., Towards causal representation learning, arXiv preprint arXiv:2102.11107, (2021).
- [6] S.-N. Jerzy, D.M. Dabrowska, T.P. Speed, On the application of probability theory to agricultural experiments. Essay on principles. Section 9, *Stat. Sci.* **5** (1990) 465–472.
- [7] D.B. Rubin, Estimating causal effects of treatments in randomized and nonrandomized studies, *J. Educ. Psychol.* **66** (1974) 688.
- [8] P.R. Rosenbaum, D.B. Rubin, The central role of the propensity score in observational studies for causal effects, *Biometrika* **70** (1983) 41–55.
- [9] G.W. Imbens, D.B. Rubin, *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*, Cambridge University Press, Cambridge, 2015.
- [10] J. Pearl, Causal diagrams for empirical research, *Biometrika* **82** (1995) 669–688.
- [11] J. Pearl, Causal inference in statistics: an overview, *Stat. Surv.* **3** (2009) 96–146.
- [12] J. Pearl, D. Mackenzie, *The Book of Why: The New Science of Cause and Effect*, Basic books, 2018.
- [13] J. Pearl, *Causality*, 2nd ed., Cambridge University Press, Cambridge, 2009.
- [14] P. Spirtes, C. Glymour, An algorithm for fast recovery of sparse causal graphs, *Soc. Sci. Comput. Rev.* **9** (1991) 62–72.
- [15] S. Shimizu, P.O. Hoyer, A. Hyvärinen, et al., A linear non-Gaussian acyclic model for causal discovery, *J. Mach. Learn. Res.* **7** (2006).
- [16] D.M. Chickering, Optimal structure identification with greedy search, *J. Mach. Learn. Res.* **3** (2002) 507–554.
- [17] I. Tsamardinos, L.E. Brown, C.F. Aliferis, The max-min hill-climbing Bayesian network structure learning algorithm, *Mach. Learn.* **65** (2006) 31–78.
- [18] D. Bacciu, T.A. Etchells, P.J.G. Lisboa, et al., Efficient identification of independence networks using mutual information, *Comput. Stat.* **28** (2013) 621–646.
- [19] S. Shimizu, T. Inazumi, Y. Sogawa, et al., DirectLiNGAM: a direct method for learning a linear non-Gaussian structural equation model, *J. Mach. Learn. Res.* **12** (2011) 1225–1248.
- [20] J. Ramsey, M. Glymour, R. Sanchez-Romero, C. Glymour, et al., A million variables and more: the fast greedy equivalence search algorithm for learning high-dimensional graphical causal models, with an application to functional magnetic resonance images, *Int. J. Data Sci. Anal.* **3** (2017) 121–129.

- [21] R. Guo, L. Cheng, J. Li, et al., A survey of learning causality with data: problems and methods, *ACM Comput. Surv.* 53 (2020) 1–37.
- [22] J. Shi, L. Chen, K. Aihara, Embedding entropy: a nonlinear measure of dynamical causality, *J. R. Soc. Interface* 19 (2022) 20210766.
- [23] C.W.J. Granger, Investigating causal relations by econometric models and cross-spectral methods, (1969) 424–438.
- [24] T. Schreiber, Measuring information transfer, *Phys. Rev. Lett.* 85 (2000) 461.
- [25] M.C. McGraw, E.A. Barnes, Memory matters: a case for Granger causality in climate variability studies, *J. Climate* 31 (2018) 3289–3300.
- [26] C. Papagiannopoulou, D.G. Miralles, S. Decubber, et al., A non-linear Granger-causal framework to investigate climate–vegetation dynamics, *Geosci. Model. Dev.* 10 (2017) 1945–1960.
- [27] P. Spirtes, K. Zhang, in: *Causal Discovery and Inference: Concepts and Recent Methodological Advances*, SpringerOpen, 2016, pp. 1–28. in.
- [28] T. Bossomaier, L. Barnett, M. Harré, et al., in: *Transfer entropy*, in: *An introduction to Transfer Entropy*, Springer, 2016, pp. 65–95.
- [29] Y. Hirata, K. Aihara, Identifying hidden common causes from bivariate time series: a method using recurrence plots, *Phys. Rev. E* 81 (2010) 016203.
- [30] H. Ye, E.R. Deyle, L.J. Gilarranz, et al., Distinguishing time-delayed causal interactions using convergent cross mapping, *Sci. Rep.* 5 (2015) 1–9.
- [31] A.T. Clark, H. Ye, F. Isbell, et al., Spatial convergent cross mapping to detect causal relationships from short time series, *Ecology* 96 (2015) 1174–1181.
- [32] L. Breton, E.J. Leonardi, L.K. Quinn, et al., Convergent cross sorting for estimating dynamic coupling, *Sci. Rep.* 11 (2021) 1–10.
- [33] X. Ge, A. Lin, Dynamic causality analysis using overlapped sliding windows based on the extended convergent cross-mapping, *Nonlinear Dyn.* 104 (2021) 1753–1765.
- [34] H. Ma, K. Aihara, L. Chen, Detecting causality from nonlinear dynamics with short-term time series, *Sci. Rep.* 4 (2014) 1–10.
- [35] D. Harnack, E. Laminski, M. Schünemann, et al., Topological causality in dynamical systems, *Phys. Rev. Lett.* 119 (2017) 098301.
- [36] H. Ma, S. Leng, C. Tao, et al., Detection of time delays and directional interactions based on time series from complex dynamical systems, *Phys. Rev. E* 96 (2017) 012221.
- [37] S. Leng, H. Ma, J. Kurths, et al., Partial cross mapping eliminates indirect causal influences, *Nat. Commun.* 11 (2020) 2632.
- [38] F. Takens, Detecting strange attractors in turbulence, in: *Dynamical Systems and turbulence*, Warwick 1980, Springer, 1981, pp. 366–381.
- [39] X. Ying, S.-Y. Leng, H.-F. Ma, et al., Continuity scaling: a rigorous framework for detecting and quantifying causality accurately, *Research* 2022 (2022).
- [40] J. Runge, Causal network reconstruction from time series: from theoretical assumptions to practical estimation, *Chaos: Interdiscip. J. Nonlinear Sci.* 28 (2018) 075310.
- [41] M. Vejmelka, M. Palus, Inferring the directionality of coupling with conditional mutual information, *Phys. Rev. E* 77 (2008) 026214.
- [42] N. Xiao, A. Zhou, M.L. Kempher, et al., Disentangling direct from indirect relationships in association networks, *Proc. Natl Acad. Sci.* 119 (2022) e2109995119.
- [43] B. Dann, J.A. Michaels, S. Schaffelhofer, et al., Uniting functional network topology and oscillations in the fronto-parietal single unit network of behaving primates, *Elife* 5 (2016) e15719.
- [44] P. Erdős, A. Rényi, On the evolution of random graphs, *Publ. Math. Inst. Hung. Acad. Sci.* 5 (1960) 17–60.
- [45] H. Kantz, T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge university press, 2004.
- [46] T. Schaffter, D. Marbach, D. Floreano, GeneNetWeaver: in silico benchmark generation and performance profiling of network inference methods, *Bioinformatics* 27 (2011) 2263–2270.
- [47] D. Marbach, T. Schaffter, C. Mattiussi, et al., Generating realistic in silico gene networks for performance assessment of reverse engineering methods, *J. Comput. Biol.* 16 (2009) 229–239.
- [48] T.W. Wong, T.S. Lau, T.S. Yu, et al., Air pollution and hospital admissions for respiratory and cardiovascular diseases in Hong Kong, *Occup. Environ. Med.* 56 (1999) 679–683.
- [49] A. Milojevic, P. Wilkinson, B. Armstrong, et al., Short-term effects of air pollution on a range of cardiovascular events in England and Wales: case-crossover analysis of the MINAP database, hospital admissions and mortality, *Heart* 100 (2014) 1093–1098.
- [50] H. Mustafić, P. Jabre, C. Caussin, et al., Main air pollutants and myocardial infarction: a systematic review and meta-analysis, *JAMA* 307 (2012) 713–721.
- [51] A.S.V. Shah, J.P. Langrish, H. Nair, et al., Global association of air pollution and heart failure: a systematic review and meta-analysis, *Lancet* 382 (2013) 1039–1048.
- [52] E. Benincà, J. Huismans, R. Heerkloss, et al., Chaos in a long-term experiment with a plankton community, *Nature* 451 (2008) 822–825.
- [53] M.M. Churchland, J.P. Cunningham, M.T. Kaufman, et al., Neural population dynamics during reaching, *Nature* 487 (2012) 51–56.
- [54] T. Wang, Y. Chen, H. Cui, From parametric representation to dynamical system: shifting views of the motor cortex in motor control, *Neurosci. Bull.* (2022) 1–13.
- [55] E.M. Izhikevich, *Dynamical Systems in Neuroscience*, MIT press, 2007.
- [56] W. Gerstner, W.M. Kistler, R. Naud, et al., *Neuronal dynamics: From Single Neurons to Networks and Models of Cognition*, Cambridge University Press, 2014.
- [57] B.M. Yu, J.P. Cunningham, G. Santhanam, et al., Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity, *Adv. Neural Inf. Process. Syst.* 21 (2008).
- [58] K.V. Shenoy, M. Sahani, M.M. Churchland, Cortical control of arm movements: a dynamical systems perspective, *Annu. Rev. Neurosci.* 36 (2013) 337–359.
- [59] Y. Li, Y. Wang, H. Cui, Eye-hand coordination during flexible manual interception of an abruptly appearing, moving target, *J. Neurophysiol.* 119 (2018) 221–234.
- [60] Y. Li, Y. Wang, H. Cui, Posterior parietal cortex predicts upcoming movement in dynamic sensorimotor control, *P. Natl. Acad. Sci. USA* 119 (2022) e2118903119.
- [61] A.P. Georgopoulos, A.B. Schwartz, R.E. Kettner, Neuronal population coding of movement direction, *Science* 233 (1986) 1416–1419.
- [62] Y. Xie, P. Hu, J. Li, et al., Geometry of sequence working memory in macaque prefrontal cortex, *Science* 375 (2022) 632–639.
- [63] J. Zhao, Y. Zhou, X. Zhang, et al., Part mutual information for quantifying direct associations in networks, *P. Natl. Acad. Sci. USA* 113 (2016) 5130–5135.
- [64] J. Shi, J. Zhao, T. Li, et al., Detecting direct associations in a network by information theoretic approaches, *Sci. China Math.* 62 (2019) 823–838.
- [65] J. Shi, J. Zhao, X. Liu, et al., Quantifying direct dependencies in biological networks by multiscale association analysis, *IEEE ACM T. Comput. Bi.* 17 (2018) 449–458.
- [66] P.T. Clemson, A. Stefanovska, Discerning non-autonomous dynamics, *Phys. Rep.* 542 (2014) 297–368.
- [67] E. Siggiridou, C. Koutlis, A. Tsimpiris, et al., Evaluation of Granger causality measures for constructing networks from multivariate time series, *Entropy* 21 (2019) 1080.
- [68] J. Zhang, H. Wang, J. Xia, et al., Coevolutionary insights between promoters and transcription factors in the plant and animal kingdoms, *Zool. Res.* 43 (2022) 805.



Luonan Chen received the B.S. degree from the Huazhong University of Science and Technology, Wuhan, China, in 1984, and the M.S. and Ph.D. degrees from Tohoku University, Sendai, Japan, in 1988 and 1991, respectively. Since 1997, he has been an Associate Professor with Osaka Sangyo University, Osaka, Japan, where he became a Full Professor. Since 2009, he has been a Professor and the Executive Director with the Key Laboratory of Systems Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China. In recent years, he published over 400 journal papers and two monographs in the area of systems biology. His current research interests include computational systems biology and nonlinear dynamics.