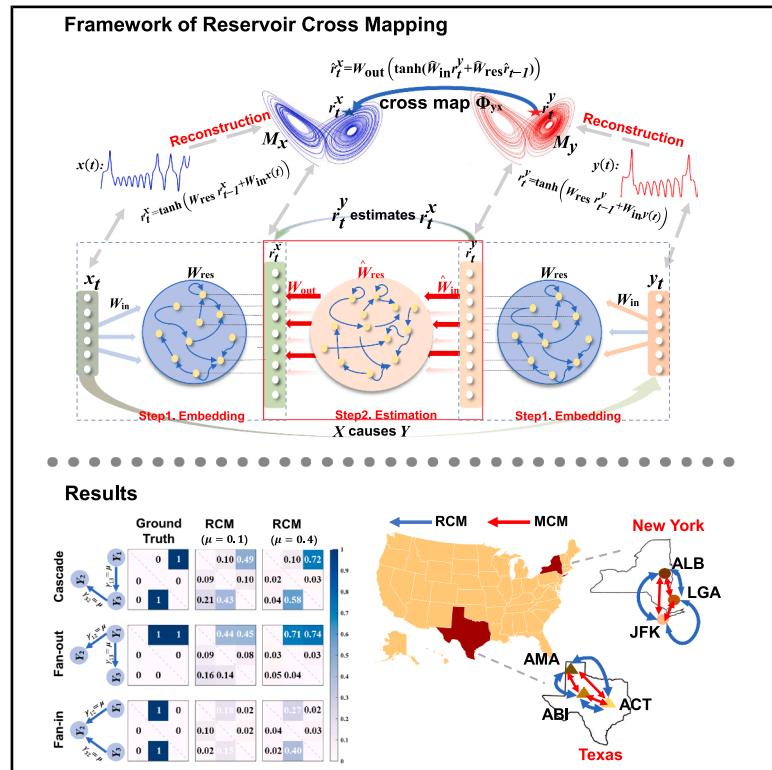


Reservoir cross mapping as a nonlinear framework for detecting dynamical causality

Graphical abstract



Authors

Ren Cao, Jintong Zhao, Chun Guan,
Huanfei Ma, Jifan Shi, Siyang Leng

Correspondence

hfma@suda.edu.cn (H.M.),
jfshi@fudan.edu.cn (J.S.),
syleng@fudan.edu.cn (S.L.)

In brief

Cao et al. present reservoir cross mapping, a causality detection framework that integrates reservoir computing with mutual cross mapping to detect dynamical causality. It reconstructs system dynamics and evaluates the smoothness of cross mapping between reservoirs to identify causal relations and their strengths.

Highlights

- Detect causality by integrating reservoir computing with mutual cross mapping
- Eliminate reliance on embedding parameters and achieve nonlinear estimation
- Reconstruct system dynamics and identify causal relations by reservoirs
- Demonstrate effectiveness and robustness using simulated and real-world datasets



Article

Reservoir cross mapping as a nonlinear framework for detecting dynamical causality

Ren Cao,¹ Jintong Zhao,¹ Chun Guan,¹ Huanfei Ma,^{2,*} Jifan Shi,^{3,4,*} and Siyang Leng^{1,3,5,*}

¹Institute of AI and Robotics, College of Intelligent Robotics and Advanced Manufacturing, Fudan University, Shanghai 200433, China

²School of Mathematical Sciences, Soochow University, Suzhou 215006, China

³Research Institute of Intelligent Complex Systems, Fudan University, Shanghai 200433, China

⁴Center for Integrative Spatial-Omics Research, Fudan University, Shanghai 200433, China

⁵Lead contact

*Correspondence: hfma@suda.edu.cn (H.M.), jfshi@fudan.edu.cn (J.S.), syleng@fudan.edu.cn (S.L.)

<https://doi.org/10.1016/j.xcrp.2025.102683>

SUMMARY

Identifying causal relations from observational time series is essential for understanding complex systems but remains a challenging problem. Mutual cross mapping (MCM), a widely used causality detection method, has addressed the nonseparability problem in nonlinear systems via state space reconstruction and approximates cross mapping between reconstructed manifolds. However, MCM relies on accurately estimated parameters and a locally linear estimation of the cross mapping, limiting its general applicability. This work presents reservoir cross mapping (RCM), a causality detection framework that integrates reservoir computing (RC) with MCM to accurately detect dynamical causality. RCM provides a nonlinear approach based on RC's embedding capability by reconstructing the dynamics of different variables and identifies causal relations and their strengths by evaluating the smoothness of the cross mapping between reservoirs. Evaluations of various simulated and real-world datasets demonstrate the effectiveness and robustness of RCM in detecting causality.

INTRODUCTION

Causality, unlike correlation, is directional and aims to discern the cause and the effect in a system, thereby addressing the problem of "why."¹ It plays an essential role in revealing the relations among variables and explaining the underlying mechanisms of complex systems. Additionally, identifying causality is an important step for reconstructing complex networks, such as in climate systems,² gene-regulatory networks,³ and social networks.⁴ Despite its importance, identifying causal and interactive relations remains a challenging task in understanding complex systems. Numerous methods have been developed to detect the potential causal links among variables, which can be primarily classified into two categories: statistical modeling and dynamical modeling.

From the perspective of statistical modeling, causality resides in the interactions among random variables, which can be detected through statistical randomization or intervention. The potential outcome framework (Neyman-Rubin causal model)⁵ and the structural causal model (Pearl's causal calculus)⁶ are two prominent statistical approaches, and these have been proven essentially equivalent.¹ However, the statistical approach, which defines causality under the weakest constraints, primarily focuses on identifying edges in a directed acyclic graph.¹ Theoretically, this can lead to multiple indistinguishable situations, all belonging to the same Markov equivalence class.⁷ Moreover, these methods merely ascertain the existence of a causal link

and struggle to quantify its strength. In practice, intervention experiments are often unfeasible, leaving accessibility only to observational time series.

Dynamical modeling employs time-series data to reconstruct causal relations from a dynamical perspective, and Shi et al.⁸ further unified causations in dynamical systems in a dynamical causality (DC) framework. Dynamical approaches include many numerical algorithms designed to measure the strength of causal influence from one part of a dynamical system to another. These approaches can be conveniently used to analyze time-series data recorded from the real world, where Granger causality (GC),⁹ transfer entropy (TE),¹⁰ and mutual cross mapping (MCM)¹¹ stand out as three prominent algorithms. GC is grounded in prediction improvement within regression models, but it is linear and requires to meet the separability assumption, which may not be apt for detecting causality in nonlinear systems, especially with weak to moderate couplings.^{12,13} While TE uses information entropy to measure causality and extends GC to nonlinear cases, it still cannot solve the nonseparability problem and is prone to false-negative problems in practical detections.^{11,14} As a complement, MCM, which utilizes state space reconstruction^{15,16} based on Takens' theorem,¹⁷ has been proposed. Convergent cross mapping (CCM) builds upon MCM by further evaluating changes in divergence and estimation accuracy as the time-series length increases.^{11,18} Recent efforts have been devoted to improving MCM and focusing on eliminating the effects of indirect causality.^{14,19} MCM employs a delay-embedding technique but relies



on local linear estimation with the Pearson correlation coefficient (PCC) as the measure.⁸ Therefore, selecting proper phase space parameters such as the embedding dimension and time lag in MCM poses several challenges. These include inconsistent results from different algorithms, high computational demands, and the susceptibility of detection accuracy to noise, window length, and data dimensionality.^{20–22}

To alleviate these problems, insights from reservoir computing (RC),²³ a prominent machine-learning technique that originated from recurrent neural networks (NNs), inform our nonlinear framework for detecting DC. Recent research underscores RC's capability to reconstruct the attractor of nonlinear dynamical systems without pre-selecting phase space parameters,²⁴ and Duan et al.²⁵ verified a rigorous theory for RC to embed the dynamics of real complex systems. Leveraging reservoirs to record historical information, RC maintains short-term memory for effective state space reconstruction and nonlinear estimation.²⁶ Some recent works have succeeded in exploiting different kinds of NNs to detect causality, most of which integrate the NN with the GC framework,^{27,28} addressing GC's nonlinearity but neglecting the nonseparability problem. Hence, leveraging RC's predictive and embedding capabilities inspires a method akin to the MCM framework, accommodating nonseparability while being nonlinear.

In this work, we propose a causality detection framework called reservoir cross mapping (RCM), which integrates RC with MCM and detects causality based on observed time series. Specifically, RCM incorporates two RC schemes in, respectively, the reconstruction and estimation steps of MCM. In the reconstruction step, we leverage the reservoir's embedding capability to reconstruct the state space, founding on the generalized embedding theorem and thus obviating the need to estimate embedding dimensions and time lags as required in MCM. In the estimation step, we employ another RC scheme and establish a connection between RC's estimation error and causal strength, utilizing information from the effect variable to estimate the cause variable. Lower estimation error signifies stronger causality, avoiding the need for nearest-neighbors search and the use of correlation coefficient as causal indicator, as in MCM. With these settings, RCM can inherit the advantages of traditional approaches while addressing challenges such as nonlinearity and nonseparability and eliminating the need for pre-selecting phase space parameters. Numerical simulation results demonstrate the validity of RCM, accurately identifying the direction and strength of causal relations. Compared to MCM, RCM achieves more precise identification, is less prone to false-positive issues, and can mitigate the effects of indirect causality, leading to more accurate and effective causal network reconstruction. Furthermore, causal inference on various real-world datasets also validates its effectiveness. Thus, RCM demonstrates potential as a universal framework for detecting DC in complex systems.

RESULTS

Detecting DC with RCM

In practice, MCM relies on the delay-embedding technique, which requires proper selection of phase space parameters,

whose procedures are reviewed in “methods.” The proposed RCM framework alleviates such shortcomings through exploiting a key idea to leverage RC for replacing delay embeddings with generalized embeddings,^{29,30} thereby achieving a nonlinear cross-map between the reservoirs. To be concrete, consider a simple case of two variables X and Y where X causes Y unidirectionally. Let $X = \{x(t)\}_{t=1}^L$ and $Y = \{y(t)\}_{t=1}^L$ represent the corresponding observational time series of length L , with the underlying evolving systems denoted by x and y . RCM applies two different RC schemes to identify this unidirectional causal link from X to Y , as shown in Figure 1.

The first step is to utilize the embedding capability of RC to generate the generalized embeddings r_t^X and r_t^Y for both time series X and Y , respectively. Such capability has been demonstrated in previous works²⁵ and thus enables the reconstruction of the underlying attractors M_X and M_Y without the need for estimation of time lags and dimensions, as required in MCM, which is built upon delay embeddings. Next, to identify the causation from X to Y , r_t^X is estimated using r_t^Y , defining the cross-map $\Phi_{yx} : M_Y \rightarrow M_X$. In MCM, the estimation, denoted as $\tilde{\Phi}_t^X$, involves searching for nearest neighbors and computing their weighted average. However, this approach is susceptible to issues such as false nearest neighbors and local nonlinearity. RCM addresses these challenges by also employing RC to approximate Φ_{yx} . Specifically, r_t^Y serves as the input to the RC, while r_t^X is the target output. The training loss, measured by the normalized mean squared error (NMSE), reflects the smoothness of Φ_{yx} and can be used as a measure of causal strength, as indicated in Ma et al.³¹ Since X causally interacts with Y unidirectionally, r_t^Y embeds the dynamics of r_t^X , making the cross-map Φ_{yx} smooth. Conversely, training an RC to approximate the unsmooth cross-map Φ_{xy} would result in a higher training loss. By setting an appropriate threshold $1 \gg T > 0$, the unidirectional causal link from X to Y can be identified.

More details about RCM are provided in methods. To validate and further demonstrate the applicability of our RCM method, we present extensive numerical experiments on various simulated and real-world datasets.

Effectiveness of RCM on simulated datasets

The first example is a two-dimensional logistic system with unidirectional/bidirectional couplings:

$$\begin{aligned} x_{t+1} &= x_t [y_x - (\gamma_x - \beta_{yx})x_t - \beta_{yx}y_t] + \varepsilon_{x,t}, \\ y_{t+1} &= y_t [y_y - (\gamma_y - \beta_{xy})y_t - \beta_{xy}x_t] + \varepsilon_{y,t}, \end{aligned} \quad (\text{Equation 1})$$

where $\gamma_x = 3.8$, $\gamma_y = 3.6$, and β_{yx} and β_{xy} denote the coupling strengths from Y to X and from X to Y , respectively. $\varepsilon_{.,t}$ represents independent Gaussian random noise with a mean of 0 and a standard deviation of 0.015. In both unidirectional and bidirectional cases, we keep β_{yx} constant while varying the coupling strength β_{xy} from 0 to 1 with a step of 0.1 to show ascending causal strength from X to Y . In the unidirectional scenario, where β_{yx} is fixed at 0, a drive-response relation emerges from X to Y , with no reciprocal influence

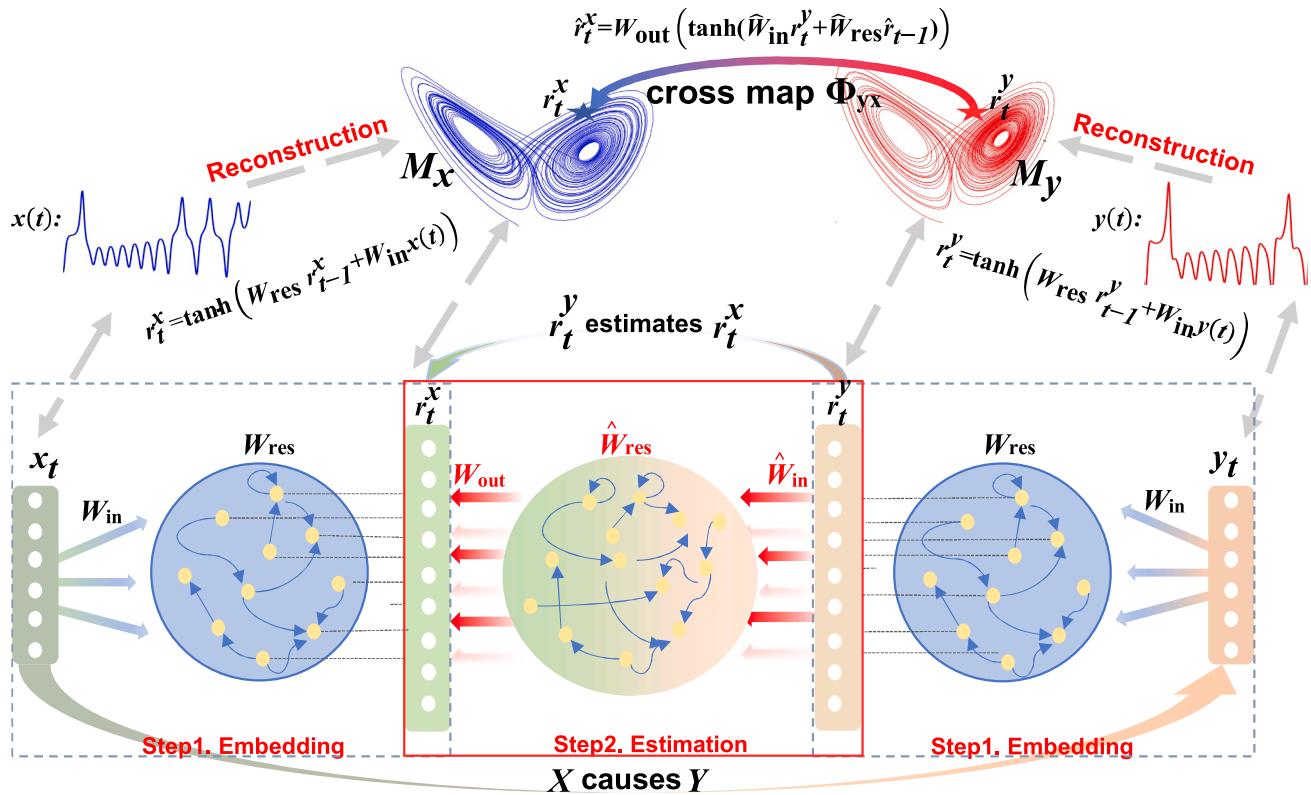


Figure 1. The framework of RCM

In step 1, time series x_t and y_t are input into the reservoir, producing outputs r_t^x and r_t^y , which correspond to points in attractors M_x and M_y , respectively. In step 2, r_t^y is used as input to another reservoir to estimate r_t^x . If variable X causes variable Y , the training loss in step 2 will be sufficiently low, which means the estimated \hat{r}_t^x will closely match r_t^x and the cross-map Φ_{xy} is smooth.

(Figure 2A). In the bidirectional case with β_{xy} set to 0.1, a reciprocal causative relation emerges between X and Y (Figure 2B). For each coupling strength, initial values are randomly selected from the interval $[0, 1]$, and 6,000 data points are generated using an iterative method, discarding the first 1,000. RCM and MCM are applied to detect causality in these scenarios.

In these scenarios, as β_{xy} increases, both $\text{RCM}[X \rightarrow Y]$ and $\text{MCM}[X \rightarrow Y]$ initially increase and subsequently saturate, while $\text{MCM}[Y \rightarrow X]$ and $\text{RCM}[Y \rightarrow X]$ are influenced by reverse causality when $\beta_{yx} \neq 0$. It is observed that, when $\beta_{xy} \in [0.1, 0.3]$, causal and noncausal relations can be distinguished by selecting an appropriate threshold T for RCM (e.g., $T = 0.05$). Moreover, $\text{RCM}[X \rightarrow Y]$ clearly indicates changes in causal strength, especially when $\beta_{xy} \in [0, 0.5]$. In contrast, $\text{MCM}[X \rightarrow Y]$ increases rapidly and saturates at approximately 0.97 when $\beta_{xy} \in [0.3, 1]$, indicating that MCM may fail to capture variations in causal strength at higher levels of coupling. Since these scenarios involve only two variables, we further evaluate and compare the causal detection capabilities of RCM and MCM in more complex multivariate settings in the following experiments.

Next, we consider a three-dimensional discrete logistic system defined by the following equation:

$$Y_j(t+1) = Y_j(t) \left(\gamma_{jj} - \sum_{i=1,2,3} \gamma_{ij} Y_i(t) \right) + \varepsilon_{j,t}, j = 1, 2, 3, \quad (\text{Equation 2})$$

where $\gamma_{11} = 3.6$, $\gamma_{22} = 3.72$, and $\gamma_{33} = 3.68$, with $\varepsilon_{j,t}$ representing white noise with a mean of 0 and a standard deviation of 0.005. The coupling strength is denoted by μ , with two values considered in this study: 0.1 and 0.4. By setting γ_{ij} to either 0 or μ , the causal relations among the three variables show three representative network motifs—fan-in, fan-out, and cascading structures—as depicted in Figure 3.

In general, causal and noncausal relations are distinguishable using RCM. By setting an appropriate threshold $T \in [0.21, 0.27]$, RCM can successfully reconstruct all three motifs, except for the fan-in motif when $\mu = 0.1$. In the fan-in motif (i.e., $\gamma_{12} = \gamma_{32} = \mu$), Y_2 is driven by both Y_1 and Y_3 simultaneously, and the detected causal strengths are weaker compared to other cases. In fact, the fan-in motif is generally considered difficult to infer due to the simultaneous impact of multiple inputs, which diminishes the influence of each individual driver.³² In contrast, spurious causal links may be identified using MCM. For example, in the cascading motif (i.e., $\gamma_{13} = \gamma_{32} = \mu$), MCM erroneously

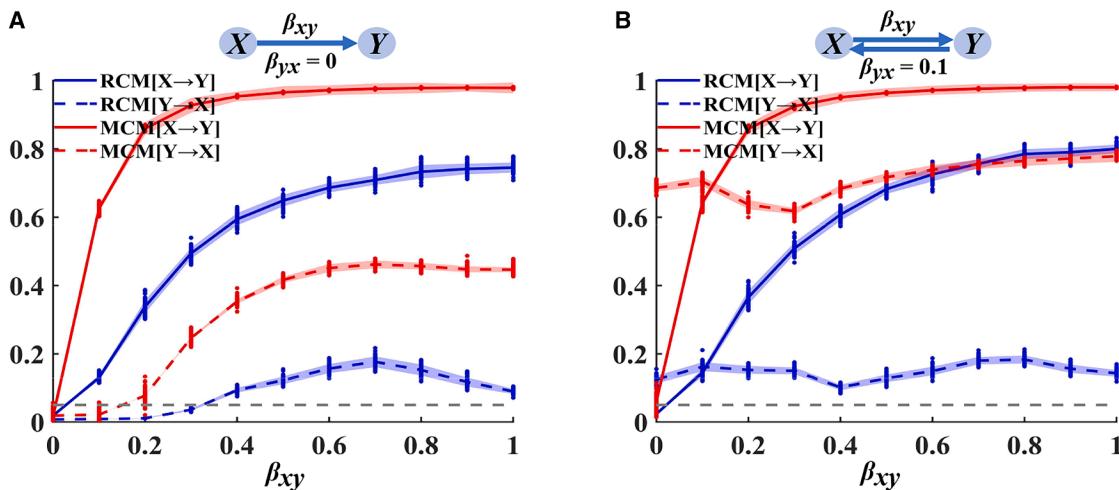


Figure 2. Comparison of RCM and MCM indices in detecting dynamical causality under unidirectional and bidirectional coupling scenarios

(A) The unidirectional case with $\beta_{yx} = 0$ and β_{xy} varying from 0 to 1.
(B) The bidirectional case with $\beta_{yx} = 0.1$ and β_{xy} varying from 0 to 1. The lines represent the mean values, and the shaded areas represent the standard deviations, both computed from 50 independently generated time series. The gray lines in each panel are the reference of 0.05. For MCM, the time lag $\tau = 1$ and embedding dimension $E = 4$ are used. For RCM, the leakage rates are set to $\alpha = \hat{\alpha} = 1$, and the regularization parameter β is selected from the set $\{10^{-1}, 10^{-5}\}$.

identifies an indirect causal link from Y_1 to Y_2 due to the influence of causation transitivity.¹⁴ In the fan-out motif (i.e., $\gamma_{12} = \gamma_{13} = \mu$), both Y_2 and Y_3 are influenced by a common source, Y_1 , leading to correlated dynamics. This makes it difficult for MCM, which relies on the correlation coefficient as its index, to correctly distinguish the true causal links between them. To accurately identify both the cascading and fan-out structures using MCM, the threshold needs to be set above 0.94—an impractically high value in real-world applications.

Reconstructing multi-node causal networks is challenging due to the complex interconnections among nodes, which often limits the reliability of traditional methods. To validate the effectiveness of RCM, we first consider the multi-variable coupled Hénon map, defined as

$$\begin{aligned} x_{i,t} &= 1 - 1.4x_{i,t-1}^2 + 0.2x_{i,t-2} + \sigma\epsilon_{i,t}, \text{ for } i = 1, 8, \\ x_{i,t} &= 1 - 1.4(0.5\mu(x_{i-1,t-1} + x_{i+1,t-1}) + (1 - \mu)x_{i,t-1})^2 \\ &\quad + 0.2x_{i,t-2} + \sigma\epsilon_{i,t}, \text{ for } i = 2, \dots, 7 \end{aligned} \quad (\text{Equation 3})$$

where $\mu = 0.1$ is the coupling strength between neighboring nodes, and $\epsilon_{i,t}$ is standard Gaussian noise with amplitude $\sigma = 0.02$.

To compare RCM and MCM, receiver-operating characteristic (ROC) curves are plotted, as shown in Figures 4C and 4E. For RCM and MCM, the thresholds corresponding to the maximum Youden indices are 0.6051 and 0.8682, respectively. Based on these thresholds, the reconstructed networks are plotted in Figures 4B and 4D, respectively. Compared to the ground truth (Figure 4A), the numbers of misidentified edges for RCM and MCM are 2 and 10, respectively, suggesting that RCM achieves superior accuracy in this scenario. Details about the ROC analysis and the Youden index are provided in [methods](#).

To further simulate more complex network structures, we consider a randomly generated causal network consisting of ten coupled Lorenz systems:

$$\begin{aligned} \dot{x}_n &= a(y_n - x_n) + \mu \sum_{m \in \{1, 2, \dots, 10\}} c_{mn}x_m, \\ \dot{y}_n &= bx_n - y_n - x_n z_n, \\ \dot{z}_n &= x_n y_n - cz_n, \end{aligned} \quad (\text{Equation 4})$$

where $n = 1, 2, \dots, 10$ denotes the ten interacting nodes coupled through variable x by the term $\sum_{m \in \{1, 2, \dots, 10\}} c_{mn}x_m$ without self-loops (i.e., $c_{nn} = 0$), and μ denotes the coupling strength. The coupling matrix $\mathbf{C} = [c_{mn}]$ is a randomly generated adjacency matrix with a connection probability ρ . The causal networks are generated under various connection probabilities ρ and coupling strength μ , with an example shown in Figure 5A. Four different combinations of ρ and μ are considered. For each combination, ten independent trials are conducted, each with a data length of 7,000. We compare the performance of the RCM method with GC, TE, and MCM in reconstructing these causal networks. For each edge between pairs of nodes in the generated network, we compute the values of four indices and record the corresponding true network structure. The ROC curves are then plotted according to Equations 21 and 22, as shown in Figure 5B.

As shown in Figure 5B, all four indices perform better when the network is sparse ($\rho = 0.1$). Among them, TE consistently struggles to reconstruct causal networks, with all area-under-the-curve (AUC) values below 0.65. MCM and GC show improved performance at a higher coupling strength ($\mu = 0.5$). In contrast, RCM consistently and effectively reconstructs Lorenz networks across various levels of sparsity and coupling strength, achieving the best performance among all methods, with AUC values exceeding 0.9 across all conditions.

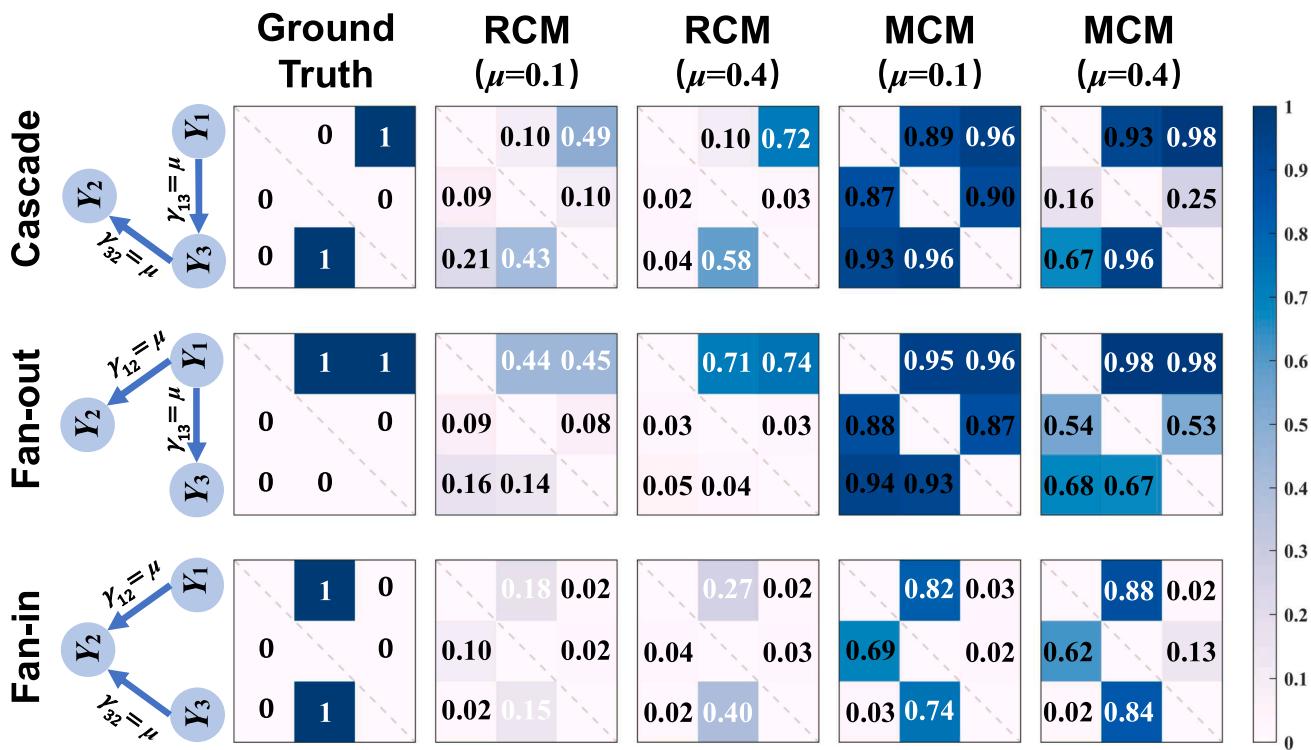


Figure 3. Illustration of basic network motifs and their reconstruction results using RCM and MCM

All results represent the average of ten trials, with each trial involving the generation of 5,000 data points. For MCM, the time lag $\tau = 1$ and embedding dimension $E = 4$ are used. For RCM, the regularization parameter is set to $\beta = 10^{-8}$.

Detecting causality in real-world datasets

The El Niño-Southern Oscillation (ENSO) is a recurring climate phenomenon characterized by anomalous increase in sea surface temperature (SST) in the tropical Pacific, which can disrupt atmospheric circulation, alter precipitation patterns, and impact the global climate system. Understanding the temporal dynamics of ENSO is crucial for improving climate prediction and developing strategies to mitigate its effects on global weather patterns. To identify ENSO events, four regional SST indicators (Niño 1 + 2, Niño 3, Niño 3.4, and Niño 4) are commonly used. The occurrence of El Niño is inferred by analyzing the SST deviations recorded by these indicators over time. The spatial distribution of these indices, shown in Figure 6A, is available through the National Center for Atmospheric Research (NCAR) at <https://climatedataguide.ucar.edu/climate-data>. The SST data, sourced from the National Oceanic and Atmospheric Administration (NOAA) (https://psl.noaa.gov/gcos_wgsp/Timeseries), includes monthly area-averaged SST records from January 1870 to December 2022, totaling $1,836 \times 4$ samples. An interpolation method is applied to improve the smoothness of the data.³³

By applying RCM and MCM, respectively, we aim to reconstruct the causal relations among these regions. Since the ground truth is unavailable, we employ the unsupervised k -means clustering method³⁴ to determine the threshold values for RCM and MCM by classifying the detection results into two groups (i.e., with or without direct causation), as shown in Figure 6C (right). The reconstructed causal networks are shown

in Figure 6A, based on the heatmaps (Figure 6C, left), where cells with index values exceeding the thresholds are boxed.

In accordance with the Walker circulation theory,³⁵ the interactions detected both by RCM and MCM among the four variables form a closed loop. This loop is due to the oceanic behavior of ENSO, which results from the alternating El Niño and La Niña cycles. During El Niño, SSTs in the eastern Pacific rise abnormally, strengthening upward air currents and weakening the trade winds. This shift moves warm water eastward, further increasing SSTs and creating a self-reinforcing positive feedback loop. Conversely, during La Niña, warmer sea temperatures in the western Pacific enhance the trade winds, pushing cold water to the eastern Pacific, leading to descending air currents and cooling effects, forming a negative feedback loop. The results of RCM and MCM both are consistent with previous studies³⁶ and align well with established conclusions regarding El Niño.

Next, we evaluate a dataset of hourly temperature values compiled from the ASOS Weather Network (<https://www.weather.gov/asos>). The dataset spans from 1949 to 2021, offering comprehensive and stationary temperature time series. Many of the longest-operating stations are airport weather stations. We select three stations in central Texas and three additional stations located several hundred miles away in eastern New York (refer to Figure 6B). The specific airport codes include ACT (Waco, TX), ABI (Abilene, TX), AMA (Amarillo, TX), and ALB (Albany, NY), as well as LaGuardia (LGA) and JFK, the two largest

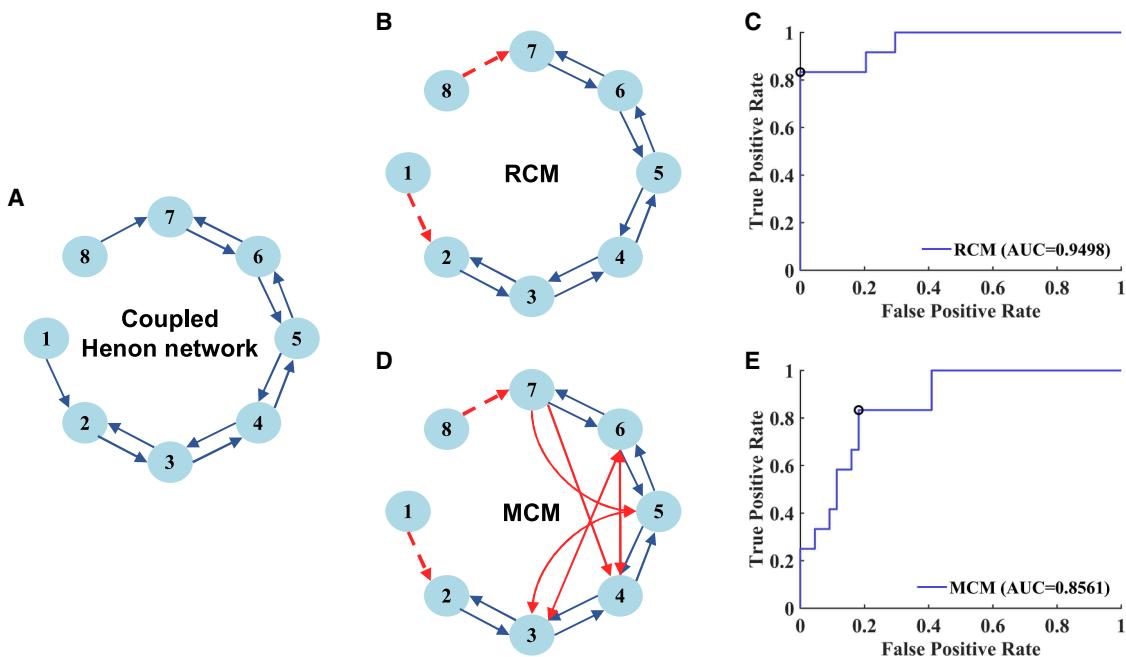


Figure 4. Ground-truth eight-node Hénon map network and its reconstructions by RCM and MCM

(A) Visualization of the coupled Hénon map network with eight nodes.

(B and D) Numerical reconstruction results of the coupled Hénon map using RCM (B) and MCM (D).

(C and E) ROC curves for RCM (C) and MCM (D) when inferring the eight-node coupled Hénon map causality network.

Blue solid arrows indicate true causal relations, while red solid and dashed lines represent false and missing causal relations. The black circles in (C) and (E) correspond to the maximum Youden indices. All results represent the average of ten trials, each involving the generation of 1,500 data points. For MCM, the time lag $\tau = 1$ and embedding dimension $E = 6$ are used. For RCM, the regularization parameter is set to $\beta = 10^{-5}$.

airports in the New York City metropolitan area. In total, we consider $569,443 \times 6$ samples.

Based on the heatmaps and k -means clustering results presented in Figure 6D, we reconstruct the causal networks for the selected stations via RCM and MCM, as shown in Figure 6B. Typically, airport weather stations in close proximity exhibit stronger connections, suggesting potential causal relations. In contrast, stations that are geographically distant generally show minimal or no correlation, suggesting an absence of causal relations. The results obtained from RCM and MCM both support this hypothesis and are consistent with the findings of Grigsby et al.³⁷ Although both methods yield consistent causal structures in this scenario, MCM is significantly more time consuming than RCM, as shown in Figure 7H.

Robustness

In this section, we evaluate the robustness of RCM and MCM under various settings, including intrinsic noise σ^{in} (innoise), extrinsic noise σ^{ex} (exnoise), training data length L , and the number of neurons m in the reservoirs (the embedding step). Specifically, we consider a five-node coupled Lorenz system arranged in a ring configuration, with each node governed by Equation 4. The coupling strength μ is set to 0.1, and the adjacency matrix \mathbf{C} , along with its visualization, is shown in Figure 7A.

We begin by evaluating the robustness of the RCM and MCM indices to varying levels of noise. First, we set the extrinsic

noise to 0 and vary the intrinsic noise from 0 to 0.2, with a step size of 0.02. As shown in Figures 7B and 7C, both RCM and MCM demonstrate strong robustness to intrinsic noise. Next, we set the intrinsic noise to 0 and increase the extrinsic noise from 0 to 1, with a step size of 0.1. As shown in Figures 7D and 7E, for both RCM and MCM, the indices for true causal relations decrease as extrinsic noise levels increase. However, in this scenario, RCM continues to accurately identify causal relations and reconstruct the ring-structured topology until the extrinsic noise exceeds 0.7, while MCM may fail to differentiate causal relations when the extrinsic noise exceeds 0.1.

Subsequently, we evaluate the robustness of RCM and MCM with respect to data length. When the data length is incrementally increased from 1,000 to 10,000, as shown in Figures 7F and 7G, both MCM and RCM indices initially decrease but then stabilize as the data length exceeds 7,000. This indicates that, for accurate causal relation reconstruction, a sufficient data length is essential for both RCM and MCM. Moreover, as shown in Figure 7H, when the data length exceeds 6,000, the time required to compute a single MCM index increases significantly. This is because MCM relies on searching for nearest neighbors in high-dimensional space to perform cross mapping, a process that becomes more time consuming as the embedding dimension and data length increase. Therefore, in the era of big data, RCM demonstrates greater practical potential due to its computational efficiency.

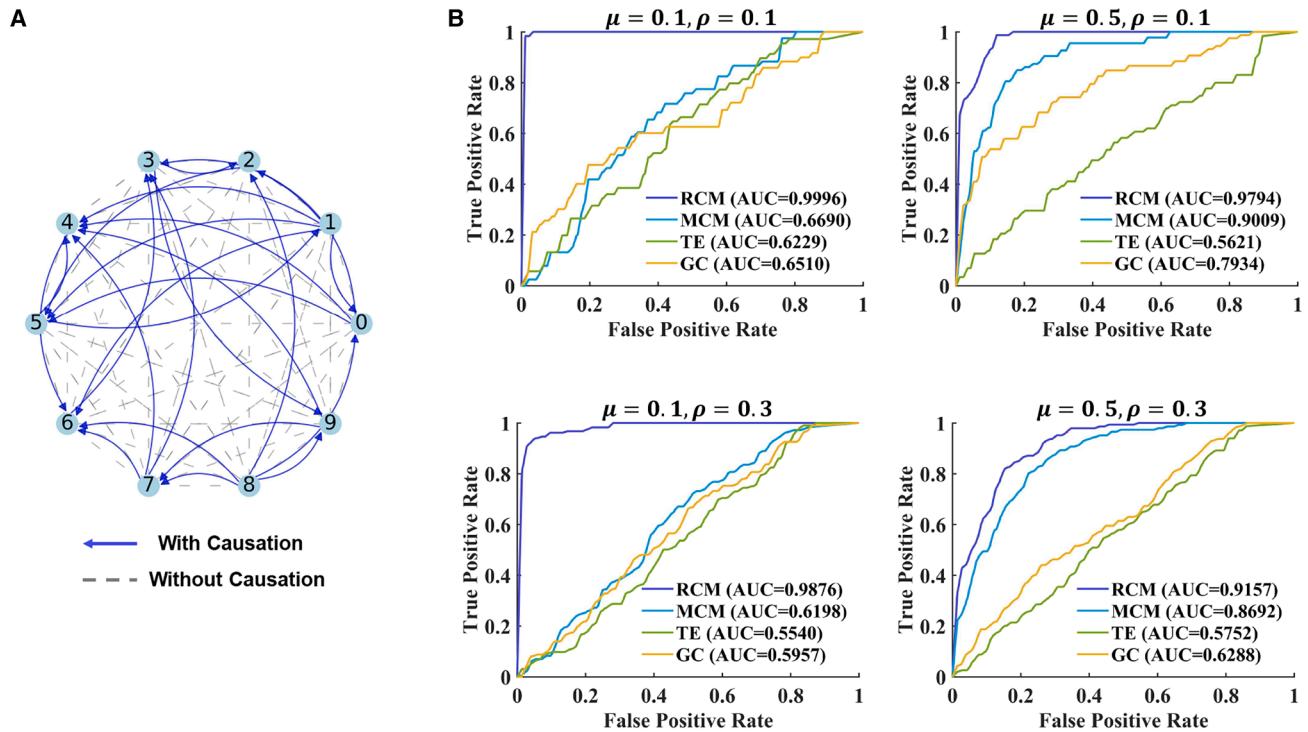


Figure 5. The comparison of reconstruction results for Lorenz networks using RCM, MCM, TE, and GC

(A) Visualization of a Lorenz network with ten nodes, randomly generated with a connection probability ρ . Blue solid arrows represent the presence of true causal relations, while gray dashed lines indicate their absence.

(B) ROC curves illustrating the performance of four methods under varying network parameters. ρ indicates the connection probability, defining the network's sparsity ($\rho = 0.1$ corresponds to a sparse network, $\rho = 0.3$ to a dense network). μ represents the coupling strength, with values of 0.1 and 0.5 corresponding to weak and strong coupling conditions, respectively. For MCM, the time lag $\tau = 0.1$ and embedding dimension $E = 7$ are used. For RCM, the regularization parameter is set to $\beta = 10^{-5}$. For GC and TE, the model order parameter is set to $\rho = 6$.

Lastly, we analyze the impact of the number of neurons in the reservoir during the embedding step. Let m_c and m_e denote the number of neurons in the reservoir for causal and effect nodes, respectively. Accordingly, in Equation 14, the random matrices \mathbf{W}_{in} and \mathbf{W}_{res} for causal nodes are of size $m_c \times n$ and $m_c \times m_c$, respectively, while those for effect nodes are of size $m_e \times n$ and $m_e \times m_e$. We consider three cases: (1) $m_c = m_e = m$, with m ranging from 10 to 1,000 (see Figure 7I); (2) $m_c = 200$, with m_e ranging from 10 to 1,000 (see Figure 7J); and (3) $m_e = 200$, with m_c ranging from 10 to 1,000 (see Figure 7K). The results show that the RCM indices stabilize when the number of neurons exceeds 200. Therefore, the random matrices \mathbf{W}_{in} and \mathbf{W}_{res} for the causal and effect nodes do not necessarily need to be identical, and differences between them do not affect the detection accuracy when the number of neurons is sufficiently large. In general, a larger number of neurons (e.g., $m_c = m_e = 200$) is recommended for robust causality detection with RCM.

DISCUSSION

In conclusion, we have developed a method called RCM, which integrates RC with MCM to accurately detect causality among variables in complex dynamical systems. The proposed RCM re-

tains advantages of MCM over various GC-based approaches. Similar to MCM, RCM measures causality within the embedding space using cross mapping, effectively addressing the nonseparability issue. Additionally, RCM overcomes MCM's limitations, such as the need for precise pre-selection of phase space parameters and challenges associated with nonlinear estimation.

The RCM causality detection process consists of two stages: embedding and estimation. Compared to MCM, RCM leverages the embedding capability of RC, providing a more efficient and flexible embedding construction method. In the estimation step, RCM constructs mappings between reservoir states and assesses their smoothness to detect causality, compensating for MCM's limitations of locally linear estimations. Extensive comparative experiments using synthetic and real-world datasets validate the performance and robustness of RCM in practical applications. The results show that RCM outperforms other methods, such as GC, TE, and MCM, in accurately identifying causal relations and reconstructing causal networks. Despite its advantages, RCM still has limitations; for example, its effectiveness may be reduced when the available time-series data are limited. Future research will focus on extending RCM's capability to detect higher-order causal relations, offering a more comprehensive tool for understanding complex dynamical interactions.

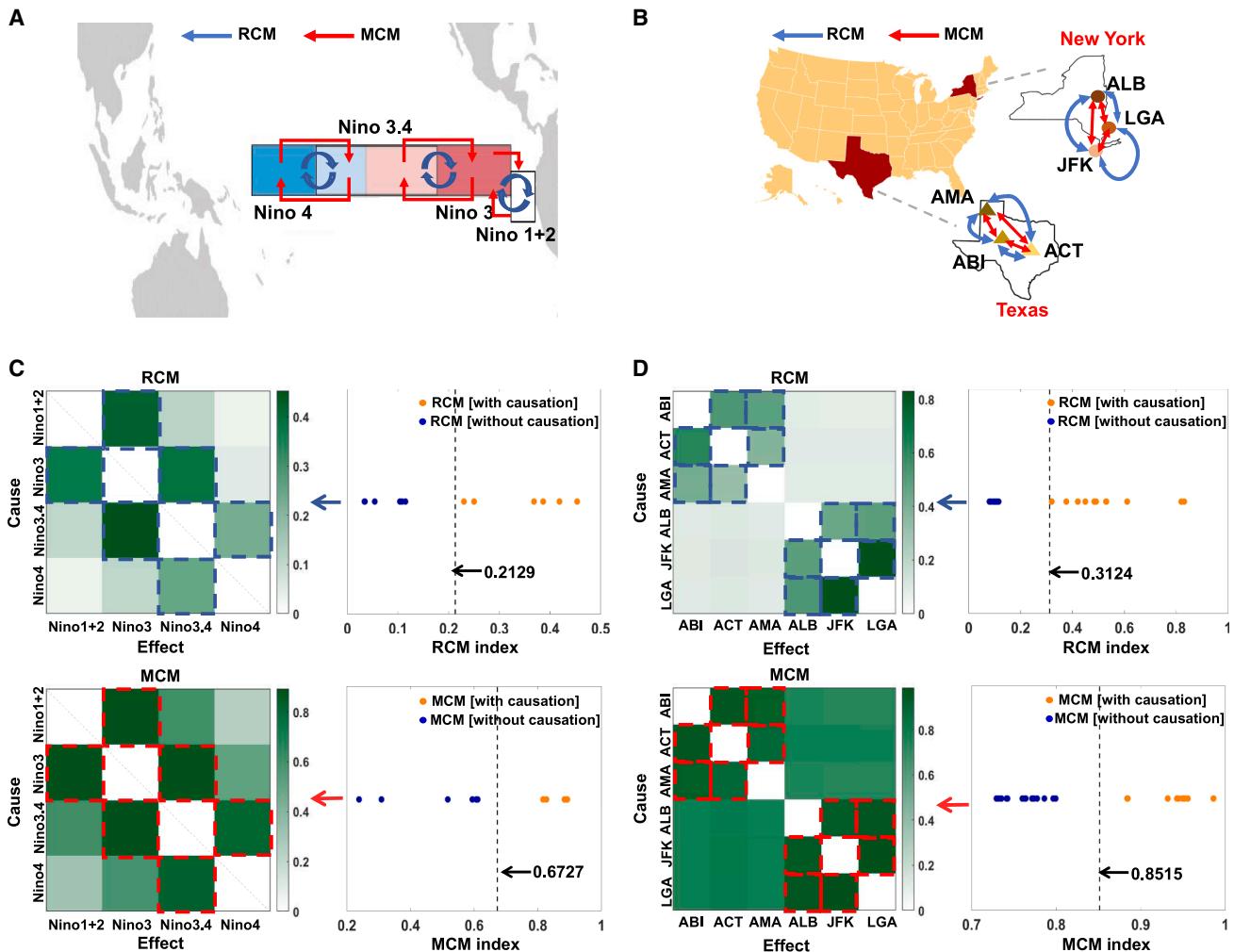


Figure 6. Reconstruction results obtained by applying RCM and MCM to real-world datasets

(A) Map of the regions corresponding to the Niño indices.
(B) Spatial distribution of three weather stations in Texas (ACT, ABI, and AMA) and three stations in New York (ALB, LGA, and JFK).
(C and D) The causal links (solid arrows) in (A) and (B) correspond to the cells boxed by dashed rectangles in the heatmaps shown in (C) and (D), where the RCM and MCM indices exceed the respective thresholds. The dashed lines in right-hand panels represent the thresholds determined by the k -means clustering method. For MCM, the time lag $\tau = 1$ and embedding dimension $E = 6$ are used. For RCM, the regularization parameter is set to $\beta = 10^{-5}$.

METHODS

We begin with a brief introduction to DC followed by an overview of classical causality detection methods, including GC, TE, and MCM. Subsequently, we provide a comprehensive description of how RCM is utilized for causality detection. Specifically, considering two time series, $X = \{x(t)\}_{t=1}^L$ and $Y = \{y(t)\}_{t=1}^L$ measured from variables X and Y in a dynamical system, we aim to detect whether a causal link (e.g., from X to Y) exists.

Dynamical causality

Consider a general system governed by delayed autonomous differential equations:

$$\dot{x}(t) = \mathbf{f}(x(t), x(t - \theta_1), \dots, x(t - \theta_K)) + \sigma_t^{\text{in}}, \quad (\text{Equation 5})$$

where $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ denotes the state vector, $\mathbf{f} = (f_1, f_2, \dots, f_n)^T$ represents smooth functions, $\theta_K > \dots > \theta_1 > 0$ are the time delays, and σ_t^{in} indicates intrinsic noise with a mean of zero, which represents internal disturbances within the system. In addition, independent extrinsic noise (σ_t^{ex}) with a constant mean can also be added to the dynamics, which refers to the noise influencing the observed or measured signal of the system. In practice, time-series data are generated using a discretized form of Equation 5, expressed as

$$\mathbf{x}(t) = \hat{\mathbf{f}}(\mathbf{x}(t-1), \dots, \mathbf{x}(t-p)) + \varepsilon_t. \quad (\text{Equation 6})$$

Dynamical causality from variable x_i to x_j in Equation 6 is established if there exists $k \in \{1, \dots, p\}$ such that $\partial \hat{f}_i / \partial x_j(t-k) \neq 0$ for almost any sampled t . Within this framework, conventional causality methods such as GC, TE, and

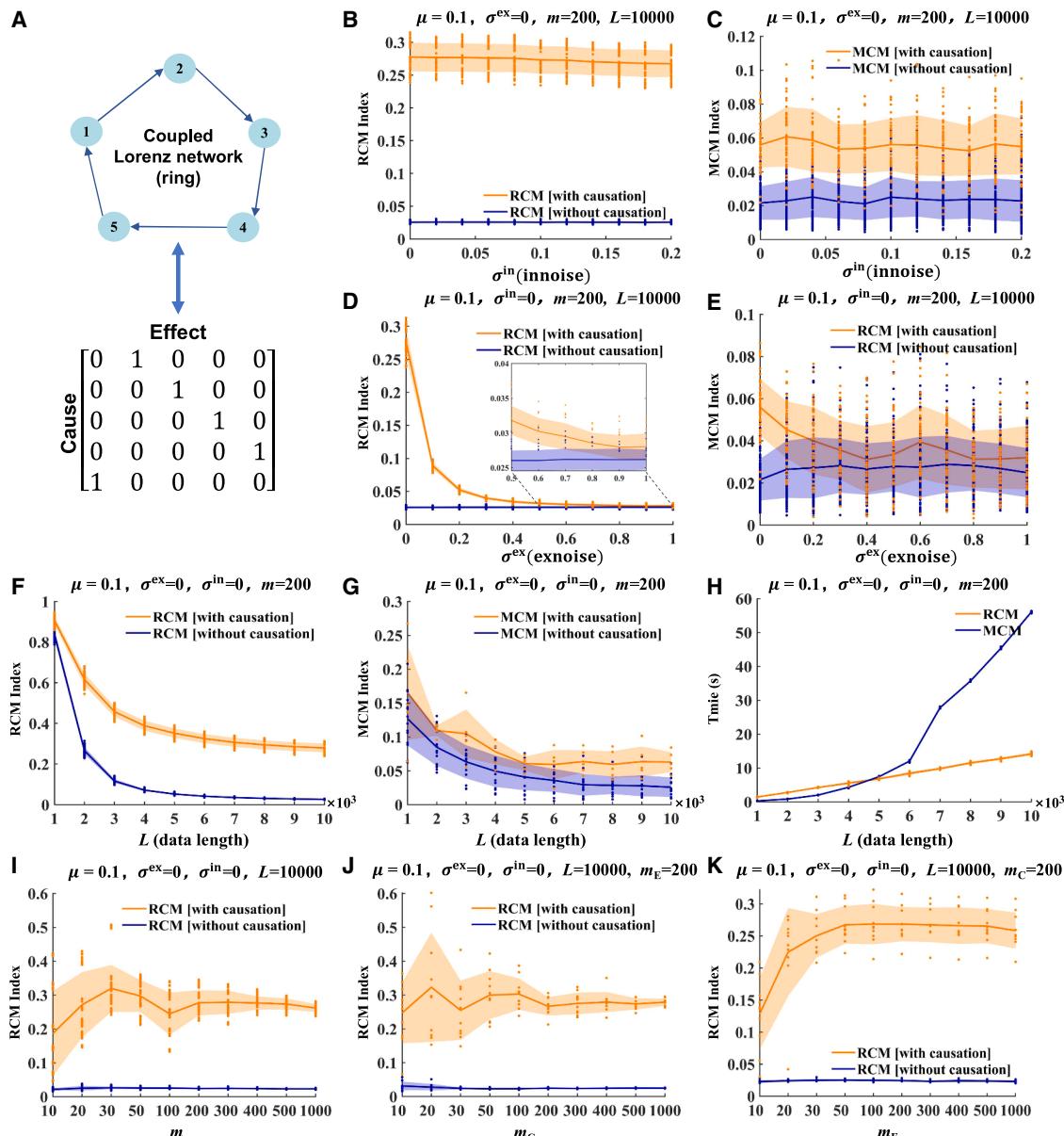


Figure 7. Robustness of RCM and MCM in a coupled Lorenz system with a ring structure

(A) Visualization of the Lorenz network (ring structure) and the corresponding adjacency matrix \mathbf{C} .

(B–E) Robustness of the RCM (B and D) and MCM (C and E) indices to varying levels of noise.

(F–H) Robustness of the RCM and MCM indices to the data length L .

(I–K) Robustness of the RCM indices to the number of neurons m in the reservoir during the embedding step.

Each panel presents the results from ten independent trials, each involving 10,000 generated data points. The points represent the detected RCM and MCM indices for all directions, with and without causation, in the ring structure. The solid line denotes their average, and the shaded areas indicate the variance across trials. For MCM, the time lag $\tau = 0.1$ and embedding dimension $E = 7$ are used. For RCM, the regularization parameter is set to $\beta = 10^{-1}$.

MCM are consistent.⁸ Furthermore, direct and indirect DC correspond to the explicit and implicit dependencies of \hat{f}_i on x_j , respectively.

Granger causality

GC is based on linear regression, namely the vector autoregressive (VAR) modeling. In GC, if the prediction of future values of

variable y is significantly improved by incorporating information from variable x , then x is said to cause y , and vice versa. To formalize this, consider the following models. On the one hand, $y(t)$ can be predicted using only its own history:

$$y(t) = \sum_{j=1}^p a_j y(t-j) + \varepsilon_y(t). \quad (\text{Equation 7})$$

On the other hand, $y(t)$ can be predicted using the history of both $x(t)$ and $y(t)$:

$$y(t) = \sum_{j=1}^p b_j x(t-j) + \sum_{j=1}^p a_j y(t-j) + \varepsilon_{y/x}(t), \quad (\text{Equation 8})$$

where a_j and b_j are the regression coefficients, p is the model order determined by the Akaike information criterion,³⁸ and $\varepsilon_{y/x}(t)$ and $\varepsilon_y(t)$ are the respective prediction errors. By comparing the error terms of the two models' approximations in [Equation 7](#) and [Equation 8](#), the GC index is defined as

$$\text{GC}[x \rightarrow y] = -\ln\left(\frac{\text{Var}(\varepsilon_{y/x}(t))}{\text{Var}(\varepsilon_y(t))}\right). \quad (\text{Equation 9})$$

Transfer entropy

TE is a generalization of GC to nonlinear cases, measuring the amount of information transferred from one variable to another. It evaluates the reduction in uncertainty (entropy) when predicting y using past information from both x and y , compared to using only the history of y . The TE index is defined as

$$\begin{aligned} \text{TE}[x \rightarrow y] &= H(y_t | y_{t-1}, \dots, y_{t-p}) \\ &\quad - H(y_t | x_{t-1}, \dots, x_{t-p}, y_{t-1}, \dots, y_{t-p}), \end{aligned} \quad (\text{Equation 10})$$

where the Shannon entropy $H(x)$ is defined as

$$H(x) = -\sum_i p(x_i) \ln p(x_i), \quad (\text{Equation 11})$$

which measures the uncertainty of a random variable x .

Mutual cross mapping

The MCM technique relies on simplex projection, a nearest-neighbor algorithm employing exponentially weighted distances from nearby points on a reconstructed manifold to estimate correlations.

Step 1: based on Takens' embedding theory,¹⁷ MCM creates two corresponding shadow manifolds, M_X and M_Y , by forming lagged-coordinate vectors $\mathbf{x}_t = [x(t), x(t-\tau), x(t-2\tau), \dots, x(t-(E-1)\tau)]$ and $\mathbf{y}_t = [y(t), y(t-\tau), y(t-2\tau), \dots, y(t-(E-1)\tau)]$ for $t = 1 + (E-1)\tau$ to $t = L$. Here, the embedding dimension E satisfies $E > \max(d_x, d_y)$, where d_x and d_y denote the dimensions of the compact manifolds on which X and Y evolve, respectively, and τ represents the time lag. In this paper, the parameters E and τ are determined by the methods of false nearest neighbors (FNNs) and delayed mutual information (DMI), respectively.^{39,40}

Step 2: MCM uses the local linear cross mapping technique to obtain estimation of \mathbf{x}_t , denoted as $\hat{\mathbf{x}}_t$. Specifically, for each point \mathbf{y}_t in M_Y , its $E+1$ nearest neighboring points can be identified, denoted by $\mathcal{N}(\mathbf{y}_t)$, and its mutual neighbors in M_X can be found, i.e., $\widehat{\mathcal{N}}^Y(\mathbf{x}_t) \triangleq \{\mathbf{x}_t' | \mathbf{y}_t' \in \mathcal{N}(\mathbf{y}_t)\}$. Note that the points in $\mathcal{N}(\mathbf{y}_t)$ and $\widehat{\mathcal{N}}^Y(\mathbf{x}_t)$ have the same indices. The map from nearest neighbors

of \mathbf{y}_t to mutual neighbors of \mathbf{x}_t is defined as cross-map $\Phi_{yx} : M_Y \rightarrow M_X$. If X causally influences Y , the mutual neighbors in $\widehat{\mathcal{N}}^Y(\mathbf{x}_t)$ should be close to \mathbf{x}_t in M_X , which actually implies that Φ_{yx} is locally smooth around \mathbf{y}_t . Specifically, $\hat{\mathbf{x}}_t$ is estimated by a linear combination with points in $\widehat{\mathcal{N}}^Y(\mathbf{x}_t)$:

$$\hat{\mathbf{x}}_t | M_Y = \sum_{i=1}^{E+1} w_i \mathbf{x}_{t,i}, \quad \mathbf{x}_{t,i} \in \widehat{\mathcal{N}}^Y(\mathbf{x}_t), \quad (\text{Equation 12})$$

$$\text{MCM}[x \rightarrow y] = \text{Corr}(\{\mathbf{x}_t\}, \{\hat{\mathbf{x}}_t\}), \quad (\text{Equation 13})$$

where w_i is a weight determined by the distance between \mathbf{y}_t and its i^{th} nearest neighbor on M_Y . If the correlation coefficient ([Equation 13](#)) exceeds an empirical threshold T , the MCM method concludes the presence of a causal influence from X to Y . We notice the key idea behind finding mutual neighbors is to measure the smoothness of the cross-map, and the smoothness of Φ_{yx} indicates the strength of causality from X to Y .³¹

While MCM addresses nonseparability by considering causality in the delay embedding space and utilizing the embedding mapping, it remains locally linear⁸ and sensitive to phase space parameter selection.^{22,41} The causal links it detects may be direct or indirect due to causality transitivity.^{14,19} Moreover, accurate results in MCM require the nearest neighbors for both \mathbf{x} and \mathbf{y} to be close to their true neighborhoods; otherwise, contradictory outcomes may occur.⁴²

Reservoir cross mapping

RCM utilizes RC to explore the causal relations between variables X and Y , and we can similarly summarize RCM into two steps. The framework of RCM is shown in [Figure 1](#).

Step 1: similar to MCM, RCM reconstructs the manifolds M_X and M_Y using time series $x(t)$ and $y(t)$. However, unlike MCM, RCM leverages the embedding ability of the reservoirs to reconstruct the manifolds without the need for accurate estimation of time lags and embedding dimensions. Specifically, as illustrated in [Figure 1](#) (the embedding step), we take time series $x(t)$ and $y(t)$ as input of the reservoir separately and update the states of the reservoir neurons, denoted as r_t^x and r_t^y . The evolution of neurons is governed by the following equations:

$$r_t^x = f(W_{\text{res}} r_{t-1}^x + W_{\text{in}} x(t)), \quad (\text{Equation 14})$$

$$r_t^y = f(W_{\text{res}} r_{t-1}^y + W_{\text{in}} y(t)), \quad t = 1, 2, \dots,$$

where the input weight matrix W_{in} and the reservoir network matrix W_{res} are random matrices of size $m \times n$ and $m \times m$, respectively, with elements independently drawn from the uniform distribution $\mathcal{U}(-1, 1)$. The spectral radius ρ is defined as the largest eigenvalue (in magnitude) of W_{res} . The component-wise activation function $f \in C^2(\mathbb{R}, (-1, 1))$ is set as a tanh function in this study. Additionally, we emphasize that the initial value r_0 does not affect the reconstruction process, owing to the echo state property of the reservoir. This property, crucial for constructing RC, ensures that the dynamics of the reservoir neurons, starting

from any different initial values $\mathbf{r}_0^{(1)}$ and $\mathbf{r}_0^{(2)}$, converge to the same trajectory, i.e., $\lim_{t \rightarrow \infty} \|\mathbf{r}_t^{(1)} - \mathbf{r}_t^{(2)}\| = 0$.⁴³ Based on the reservoir embedding theorem,²⁵ \mathbf{r}_t^x and \mathbf{r}_t^y correspond to points on the manifolds M_X and M_Y , with the number of neurons m in the reservoir representing the embedding dimension. In RC, m is typically large, ensuring $m > 2 \max(d_x, d_y)$.

Step 2: RCM employs RC to approximate the mapping Φ_{yx} , where the training error reflects the cross-map smoothness from \mathbf{r}_t^y to \mathbf{r}_t^x , which also corresponds to the causal strength. Notably, other forms of NNs, such as MLP, can also approximate the smoothness of the cross-map; however, RC performs comparably to other machine-learning methods on dynamical system tasks while requiring significantly smaller datasets and offering faster training speed, due to its minimal parameter requirements and the absence of gradient back-propagation.⁴⁴

As illustrated in Figure 1 (the estimation step), \mathbf{r}_t^y is used as the input to another reservoir, while \mathbf{r}_t^x serves as the estimation target during the training of the RC. Unlike MCM, which relies on local neighbors of \mathbf{y}_t in M_Y to estimate corresponding \mathbf{x}_t in M_X , which suffers from reduced accuracy with limited time-series data, RCM directly estimates \mathbf{r}_t^x from \mathbf{r}_t^y through RC, enabling nonlinear estimation. The estimation of \mathbf{r}_t^x (i.e., $\hat{\mathbf{r}}_t^x$) is generated as follows:

$$\hat{\mathbf{r}}_t = \tanh(\widehat{\mathbf{W}}_{\text{in}} \mathbf{r}_t^y + \widehat{\mathbf{W}}_{\text{res}} \hat{\mathbf{r}}_{t-1}), \quad (\text{Equation 15})$$

$$\hat{\mathbf{r}}_t^x = \mathbf{W}_{\text{out}} \hat{\mathbf{r}}_t, \quad (\text{Equation 16})$$

where the input weight matrix $\widehat{\mathbf{W}}_{\text{in}}$ and the reservoir network matrix $\widehat{\mathbf{W}}_{\text{res}}$ are random matrices of size $m \times N$ and $N \times N$, respectively. The spectral radius $\hat{\rho}$ is defined as the largest eigenvalue (in magnitude) of $\widehat{\mathbf{W}}_{\text{res}}$. \mathbf{W}_{out} is the output weight matrix and RC match $\hat{\mathbf{r}}_t^x$ to the desired output \mathbf{r}_t^x in a least-square sense using Tikhonov regularization, so that \mathbf{W}_{out} is given by

$$\mathbf{W}_{\text{out}} = \mathbf{R}_x \mathbf{R}_{\text{total}}^\top (\mathbf{R}_{\text{total}} \mathbf{R}_{\text{total}}^\top + \beta \mathbf{I})^{-1}, \quad (\text{Equation 17})$$

where the matrix \mathbf{R}_x and $\mathbf{R}_{\text{total}}$ are the horizontal concatenation of \mathbf{r}_t^x and $\hat{\mathbf{r}}_t$. The regularization parameter β , also referred to as the ridge parameter, is introduced to prevent overfitting during training, and \mathbf{I} denotes the identity matrix.

The training loss, quantified by NMSE, serves as an indicator of the smoothness of the cross-map Φ :

$$\text{NMSE} = \frac{1}{m \times L} \sum_{t=1}^L \frac{\|\hat{\mathbf{r}}_t^x - \mathbf{r}_t^x\|^2}{\sigma^2}, \quad (\text{Equation 18})$$

$$\text{RCM}[x \rightarrow y] = \frac{1}{\exp(\gamma \cdot \text{NMSE})}, \quad (\text{Equation 19})$$

where $\|\cdot\|$ represents the L_2 norm, L denotes the length of the training data, and σ^2 represents the variance of \mathbf{r}^x . Note that the factor γ is a positive constant introduced to normalize the in-

dex and suppress the long-tail effect,³¹ and it is empirically set to $\gamma = 5$ throughout this study.

Moreover, a more general form of reservoir dynamics can be considered in practice,⁴⁵ whereby Equations 14 and 15 are reformulated as

$$\mathbf{r}_t^x = \phi((1 - \alpha)\mathbf{r}_{t-1}^x + \alpha \cdot \tanh(\mathbf{W}_{\text{res}} \mathbf{r}_{t-1}^x + \mathbf{W}_{\text{in}} \mathbf{x}(t))),$$

$$\mathbf{r}_t^y = \phi((1 - \alpha)\mathbf{r}_{t-1}^y + \alpha \cdot \tanh(\mathbf{W}_{\text{res}} \mathbf{r}_{t-1}^y + \mathbf{W}_{\text{in}} \mathbf{y}(t))),$$

$$\begin{aligned} \hat{\mathbf{r}}_t = \phi((1 - \hat{\alpha})\hat{\mathbf{r}}_{t-1} + \hat{\alpha} \cdot \tanh(\widehat{\mathbf{W}}_{\text{res}} \hat{\mathbf{r}}_{t-1} \\ + \widehat{\mathbf{W}}_{\text{in}} \mathbf{r}_t^y)), t = 1, 2, \dots, \end{aligned} \quad (\text{Equation 20})$$

where α and $\hat{\alpha}$ denote the leakage rates that control the update speed of the reservoir dynamics. The operator ϕ acts component-wise, squaring the elements at even-numbered indices of the vector. In this study, unless otherwise stated, we set the parameters for RCM with $\alpha = 0.44$, $\hat{\alpha} = 0.9$, $m = 200$, $N = 1000$, $\rho = 0.67$, and $\hat{\rho} = 0.9$. The regularization parameter β is optimized for specific systems through a simple grid search.⁴⁶

In summary, if X causes Y , the dynamics of X is encoded in the manifold M_Y , and the embedding \mathbf{r}_t^y of Y contains information about \mathbf{r}_t^x . We can use RC to fit the mapping from \mathbf{r}_t^y to \mathbf{r}_t^x , which is grounded in the principle that a NN can approximate any smooth map, whereas attempting to approximate an unsmooth map will fail with larger training errors. A stronger causal relation from X to Y implies that more information about \mathbf{r}_t^x is encoded in \mathbf{r}_t^y . As a result, the estimation $\hat{\mathbf{r}}_t^x$ obtained from \mathbf{r}_t^y using RC will be closer to \mathbf{r}_t^x . In this work, we use the training error (NMSE) of RC to evaluate the smoothness of the mapping, which actually measures the distance between $\hat{\mathbf{r}}_t^x$ and \mathbf{r}_t^x . A large NMSE indicates an unsmooth map, suggesting that Y lacks information about X , meaning X does not cause Y . Conversely, a smaller NMSE indicates a smoother mapping, demonstrating that RC can extract more information about X from Y , which indicates a stronger causality from X to Y . We design a metric (Equation 19), which is proportional (\propto) to the strength of causality.

Receiver-operating characteristic curves

The ROC curves are plotted using the following notations.

- P : total positive edges; N : total negative edges;
- TP: true positive; FP: false positive; TN: true negative; FN: false negative;
- TPR: true positive rate; FPR: false positive rate; and
- AUC: area under the ROC curve.

At a fixed threshold value, the numbers of TP and FP in the estimated network (assuming the true network is known) can be counted. The TPR and FPR can then be computed as

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (\text{Equation 21})$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}. \quad (\text{Equation 22})$$

By varying the threshold from 0 to 1, the ROC curve is generated. Furthermore, the Youden Index⁴⁷ is defined as

$$J = \text{TPR} - \text{FPR}. \quad (\text{Equation 23})$$

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Siyang Leng (syleng@fudan.edu.cn).

Materials availability

This study did not generate new materials.

Data and code availability

The code and data can be found online at <https://doi.org/10.5281/zenodo.15516899>. Any additional information needed to reanalyze the data in the paper will be addressed by the [lead contact](#) upon reasonable request.

ACKNOWLEDGMENTS

This work is sponsored by the National Key R&D Program of China (no. 2022YFC2704604) and the National Natural Science Foundation of China (nos. 12471457, 12301620, 12171350, and 42450192). This work is also sponsored by the Science and Technology Commission of Shanghai Municipality (no. 2021SHZDXZ0103), the Shanghai Municipal Data Bureau special fund for urban digital transformation (no. 202401065), and the AI for Science Foundation of Fudan University (no. FudanX24AI041).

AUTHOR CONTRIBUTIONS

R.C. and S.L. conceived the idea. R.C., J.Z., and S.L. designed the research. R.C. performed the research. R.C., J.Z., C.G., H.M., J.S., and S.L. analyzed the data and wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 5, 2025

Revised: April 29, 2025

Accepted: June 10, 2025

Published: July 4, 2025

REFERENCES

- Neuberg, L.G. (2003). Causality: Models, Reasoning, and Inference, by Judea Pearl, Cambridge University Press, 2000. *Econom. Theory* 19, 675–685. <https://doi.org/10.1017/S0266466603004109>.
- Wiedermann, M., Radebach, A., Donges, J.F., Kurths, J., and Donner, R.V. (2016). A climate network-based index to discriminate different types of el niño and la niña. *Geophys. Res. Lett.* 43, 7176–7185. <https://doi.org/10.48550/arXiv.1604.04432>.
- Wang, Y.X.R., and Huang, H. (2014). Review on statistical methods for gene network reconstruction using expression data. *J. Theor. Biol.* 362, 53–61. <https://doi.org/10.1016/j.jtbi.2014.03.040>.
- Graham, B.S. (2015). Methods of identification in social networks. *Annu. Rev. Econom.* 7, 465–485. <https://doi.org/10.1146/annurev-economics-080614-115611.2>.
- Splawa-Neyman, J., Dabrowska, D.M., and Speed, T.P. (1990). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Stat. Sci.* 5, 465–472. <https://doi.org/10.1214/ss/1177012031>.
- Pearl, J. (2009). Causality: Models, Reasoning, and Inference, 2nd ed., ed. (Cambridge University Press). <https://doi.org/10.1017/CBO9780511803161>.
- He, Y.B., Geng, Z., and Liang, X. (2005). Learning causal structures based on markov equivalence class. In *Algorithmic Learning Theory: 16th International Conference*, Singapore, October 8–11, 2005. Proceedings 16 (Springer), pp. 92–106. https://doi.org/10.1007/11564089_9.
- Shi, J., Chen, L., and Aihara, K. (2022). Embedding entropy: a nonlinear measure of dynamical causality. *J. R. Soc. Interface* 19, 20210766. <https://doi.org/10.1098/rsif.2021.0766>.
- Granger, C.W.J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 424–438. <https://doi.org/10.2307/1912791>.
- Schreiber, T. (2000). Measuring information transfer. *Phys. Rev. Lett.* 85, 461–464. <https://doi.org/10.1103/PhysRevLett.85.461>.
- Sugihara, G., May, R., Ye, H., Hsieh, C.H., Deyle, E., Fogarty, M., and Munch, S. (2012). Detecting causality in complex ecosystems. *Science* 338, 496–500. <https://doi.org/10.1126/science.1227079>.
- Arnhold, J., Grassberger, P., Lehnertz, K., and Elger, C.E. (1999). A robust method for detecting interdependences: application to intracranially recorded eeg. *Phys. Nonlinear Phenom.* 134, 419–430. [https://doi.org/10.1016/S0167-2789\(99\)00140-2](https://doi.org/10.1016/S0167-2789(99)00140-2).
- Quiroga, R.Q., Arnhold, J., and Grassberger, P. (2000). Learning driver-response relationships from synchronization patterns. *Phys. Rev.* 61, 5142–5148. <https://doi.org/10.1103/PhysRevE.61.5142>.
- Leng, S., Ma, H., Kurths, J., Lai, Y.C., Lin, W., Aihara, K., and Chen, L. (2020). Partial cross mapping eliminates indirect causal influences. *Nat. Commun.* 11, 2632. <https://doi.org/10.1038/s41467-020-16238-0>.
- Casdagli, M., Eubank, S., Farmer, J.D., and Gibson, J. (1991). State space reconstruction in the presence of noise. *Phys. Nonlinear Phenom.* 51, 52–98. https://doi.org/10.1007/978-1-4899-2305-9_5.
- Kugiumtzis, D., Lillekjendlie, B., and Christoffersen, N. (1994). Chaotic time series: Part 1: estimation of some invariant properties in state space. *Modeling, Identification and Control* 15, 205–224. <https://doi.org/10.48550/arXiv.chao-dyn/9401004>.
- Takens, F. (2006). Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence, Warwick 1980: proceedings of a symposium held at the University of Warwick 1979/80* (Springer), pp. 366–381. <https://doi.org/10.1007/BFb0091924>.
- Ying, X., Leng, S.Y., Ma, H.F., Nie, Q., Lai, Y.C., and Lin, W. (2022). Continuity scaling: A rigorous framework for detecting and quantifying causality accurately. *Research* 2022, 9870149. <https://doi.org/10.34133/2022/9870149>.
- Yang, L., Lin, W., and Leng, S. (2023). Conditional cross-map-based technique: From pairwise dynamical causality to causal network reconstruction. *Chaos* 33, 063101. <https://doi.org/10.1063/5.0144310>.
- Hegger, R., and Kantz, H. (1999). Improved false nearest neighbor method to detect determinism in time series data. *Phys. Rev.* 60, 4970–4973. <https://doi.org/10.1103/PhysRevE.60.4970>.
- Sitz, A., Schwarz, U., Kurths, J., and Voss, H.U. (2002). Estimation of parameters and unobserved components for nonlinear systems from noisy time series. *Phys. Rev.* 66, 016210. <https://doi.org/10.1103/PhysRevE.66.016210>.
- Tsonis, A.A. (2007). Reconstructing dynamics from observables: the issue of the delay parameter revisited. *Int. J. Bifurcation Chaos* 17, 4229–4243. <https://doi.org/10.1142/S0218127407019913>.
- Lukoševičius, M., and Jaeger, H. (2009). Reservoir computing approaches to recurrent neural network training. *Comp. Sci. Rev.* 3, 127–149. <https://doi.org/10.1016/j.cosrev.2009.03.005>.
- Lu, Z., Hunt, B.R., and Ott, E. (2018). Attractor reconstruction by machine learning. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 28, 061104. <https://doi.org/10.48550/arXiv.1805.03362>.
- Duan, X.Y., Ying, X., Leng, S.Y., Kurths, J., Lin, W., and Ma, H.F. (2023). Embedding theory of reservoir computing and reducing reservoir network using time delays. *Phys. Rev. Res.* 5, L022041. <https://doi.org/10.1103/PhysRevResearch.5.L022041>.

26. Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat, f. (2019). Deep learning and process understanding for data-driven earth system science. *Nature* 566, 195–204. <https://doi.org/10.1038/s41586-019-0912-1>.
27. Duggento, A., Guerrisi, M., and Toschi, N. (2021). Echo state network models for nonlinear granger causality. *Philos. Trans. A Math. Phys. Eng. Sci.* 379, 20200256. <https://doi.org/10.1101/651679>.
28. Wang, M., and Fu, Z. (2022). A new method of nonlinear causality detection: Reservoir computing granger causality. *Chaos Solitons Fractals* 154, 111675. <https://doi.org/10.1016/j.chaos.2021.111675>.
29. Ma, H., Leng, S., Aihara, K., Lin, W., and Chen, L. (2018). Randomly distributed embedding making short-term high-dimensional data predictable. *Proc. Natl. Acad. Sci. USA* 115, E9994–E10002. <https://doi.org/10.1073/pnas.1802987115>.
30. Deyle, E.R., and Sugihara, G. (2011). Generalized theorems for nonlinear state space reconstruction. *PLoS One* 6, e18295. <https://doi.org/10.1371/journal.pone.0018295>.
31. Ma, H., Aihara, K., and Chen, L. (2014). Detecting causality from nonlinear dynamics with short-term time series. *Sci. Rep.* 4, 7464. <https://doi.org/10.1038/srep07464>.
32. Marbach, D., Prill, R.J., Schaffter, T., Mattiussi, C., Floreano, D., and Stolovitzky, G. (2010). Revealing strengths and weaknesses of methods for gene network inference. *Proc. Natl. Acad. Sci. USA* 107, 6286–6291. <https://doi.org/10.1073/pnas.0913357107>.
33. Wahba, G. (1981). Spline interpolation and smoothing on the sphere. *SIAM J. Sci. Stat. Comput.* 2, 5–16. <https://doi.org/10.1137/0902002>.
34. Hartigan, J.A., and Wong, M.A. (1979). Algorithm as 136: A k-means clustering algorithm. *Appl. Stat.* 28, 100–108. <https://doi.org/10.2307/2346830>.
35. Walker, G.T. (1925). Correlation in seasonal variations of weather—a further study of world weather. *Mon. Weather Rev.* 53, 252–254. [https://doi.org/10.1175/1520-0493\(1925\)53%3C252:CISVOW%3E2.0.CO;2](https://doi.org/10.1175/1520-0493(1925)53%3C252:CISVOW%3E2.0.CO;2).
36. Zhao, J., Gan, Z., Huang, R., Guan, C., Shi, J., and Leng, S. (2024). Detecting dynamical causality via intervened reservoir computing. *Commun. Phys.* 7, 232. <https://doi.org/10.1038/s42005-024-01730-6>.
37. Grigsby, J., Wang, Z., Nguyen, N., and Qi, Y. (2021). Long-range transformers for dynamic spatiotemporal forecasting. Preprint at arXiv. <https://doi.org/10.48550/arXiv.2109.12218>.
38. Akaike, H. (1974). Stochastic theory of minimal realization. *IEEE Trans. Automat. Contr.* 19, 667–674. <https://doi.org/10.1109/TAC.1974.1100707>.
39. Ma, H., Leng, S., Tao, C., Ying, X., Kurths, J., Lai, Y.C., and Lin, W. (2017). Detection of time delays and directional interactions based on time series from complex dynamical systems. *Phys. Rev. E* 96, 012221. <https://doi.org/10.1103/PhysRevE.96.012221>.
40. Fraser, A.M., and Swinney, H.L. (1986). Independent coordinates for strange attractors from mutual information. *Physical review A* 33, 1134–1140. <https://doi.org/10.1103/PhysRevA.33.1134>.
41. Casaleggio, A., and Braiotta, S. (1997). Estimation of Lyapunov exponents of ECG time series—the influence of parameters. *Chaos Solitons Fractals* 8, 1591–1599. [https://doi.org/10.1016/S0960-0779\(97\)00040-4](https://doi.org/10.1016/S0960-0779(97)00040-4).
42. Quyen, M.L.V., Martinerie, J., Adam, C., and Varela, F.J. (1999). Nonlinear analyses of interictal EEG map the brain interdependences in human focal epilepsy. *Phys. Nonlinear Phenom.* 127, 250–266. [https://doi.org/10.1016/S0167-2789\(98\)00258-9](https://doi.org/10.1016/S0167-2789(98)00258-9).
43. Yildiz, I.B., Jaeger, H., and Kiebel, S.J. (2012). Re-visiting the echo state property. *Neural Netw.* 35, 1–9. <https://doi.org/10.1016/j.neunet.2012.07.005>.
44. Vlachas, P.R., Pathak, J., Hunt, B.R., Sapsis, T.P., Girvan, M., Ott, E., and Koumoutsakos, P. (2020). Backpropagation algorithms and reservoir computing in recurrent neural networks for the forecasting of complex spatiotemporal dynamics. *Neural Netw.* 126, 191–217. <https://doi.org/10.1016/j.neunet.2020.02.016>.
45. Hertaux, J., and Räth, C. (2020). Breaking symmetries of the reservoir equations in echo state networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 30, 123142. <https://doi.org/10.1063/5.0028993>.
46. Racca, A., and Magri, L. (2021). Robust optimization and validation of echo state networks for learning chaotic dynamics. *Neural Netw.* 142, 252–268. <https://doi.org/10.1016/j.neunet.2021.05.004>.
47. Fluss, R., Faraggi, D., and Reiser, B. (2005). Estimation of the Youden index and its associated cutoff point. *Biom. J.* 47, 458–472. <https://doi.org/10.1002/bimj.200410135>.