Research Article

Continuity Scaling: A Rigorous Framework for Detecting and Quantifying Causality Accurately

Xiong Ying,[1,2,3] **Si-Yang Leng** ⬤,[2,4] **Huan-Fei Ma** ⬤,[5] **Qing Nie** ⬤,[6] **Ying-Cheng Lai** ⬤,[7] **and Wei Lin** ⬤[1,2,3,8]

[1]*School of Mathematical Sciences, SCMS, and SCAM, Fudan University, Shanghai 200433, China*
[2]*Research Institute for Intelligent Complex Systems, CCSB, and LCNBI, Fudan University, Shanghai 200433, China*
[3]*State Key Laboratory of Medical Neurobiology and MOE Frontiers Center for Brain Science, Institutes of Brain Science, Fudan University, Shanghai 200032, China*
[4]*Institute of AI and Robotics, Academy for Engineering and Technology, Fudan University, Shanghai 200433, China*
[5]*School of Mathematical Sciences, Soochow University, Suzhou 215006, China*
[6]*Department of Mathematics, Department of Developmental and Cell Biology, And NSF-Simons Center for Multiscale Cell Fate Research, University of California, Irvine, CA 92697-3875, USA*
[7]*School of Electrical, Computer, And Energy Engineering, Arizona State University, Tempe, Arizona 85287-5706, USA*
[8]*Shanghai Artificial Intelligence Laboratory, Shanghai 200232, China*

Correspondence should be addressed to Wei Lin; wlin@fudan.edu.cn

Data-based detection and quantification of causation in complex, nonlinear dynamical systems is of paramount importance to science, engineering, and beyond. Inspired by the widely used methodology in recent years, the cross-map-based techniques, we develop a general framework to advance towards a comprehensive understanding of dynamical causal mechanisms, which is consistent with the natural interpretation of causality. In particular, instead of measuring the smoothness of the cross-map as conventionally implemented, we define causation through measuring the *scaling law* for the continuity of the investigated dynamical system directly. The uncovered scaling law enables accurate, reliable, and efficient detection of causation and assessment of its strength in general complex dynamical systems, outperforming those existing representative methods. The continuity scaling-based framework is rigorously established and demonstrated using datasets from model complex systems and the real world.

1. Introduction

Identifying and ascertaining causal relations are a problem of paramount importance to science and engineering with broad applications [1–3]. For example, accurate detection of causation is the key to identifying the origin of diseases in precision medicine [4] and is important to fields such as psychiatry [5]. Traditionally, associational concepts are often misinterpreted as causation [6, 7], while causal analysis in fact goes one step further beyond association in a sense that, instead of using static conditions, causation is induced under changing conditions [8]. The principle of Granger causality formalizes a paradigmatic framework [9–11], quantifying causality in terms of prediction improvements, but, because of its linear, multivariate, and statistical regression nature, the various derived methods require extensive data [12]. Entropy-based methods [13–20] face a similar difficulty. Another issue with the Granger causality is the fundamental requirement of separability of the underlying dynamical variables, which usually cannot be met in the real world systems. To overcome these difficulties, the cross-map-based techniques, paradigms tailored to dynamical systems, have been developed and have gained widespread attention in the past decade [21–36].

The cross-map is originated from nonlinear time series analysis [37–42]. A brief understanding of such a map is as follows. Consider two subsystems: X and Y. In the reconstructed phase space of X, if for any state vector at a time a set of neighboring vectors can be found, the set of the cross-mapped vectors, which are the partners with equal time of X, could be available in Y. The cross-map underlying the reconstructed spaces can be written as $Y_t = \Phi(X_t)$ (where X_t and Y_t are delay coordinates with sufficiently large dimensions) for the case of Y *unidirectionally* causing X while, mathematically, its inverse map does not exist [34]. In practice, using the prior knowledge on the true causality in toy models or/and the assumption on the expanding property of Φ (representing by its Jacobian's singular value larger than one in the topological causality framework [24]), scientists developed many practically useful techniques based on the cross-map for causality detection. For instance, the "activity" method, originally designed to measure the continuity of the inverse of the cross-map, compares the divergence of the cross-mapped vectors to the state vector in X with the divergence of the independently-selected neighboring vectors to the same state vector [22, 23]. The topological causality measures the divergence rate of the cross-mapped vectors from the state vectors in Y [24], and the convergent cross-mapping (CCM), increasing the length of time series, compares the true state vector Y with the average of the cross-mapped vectors, as the estimation of Y [21, 25–36]. Then, the change of the divergence or the accuracy of the estimation is statistically evaluated for determining the causation from Y to X. Inversely, the causation from X to Y can be evaluated in an analogous manner. The above evaluations [21, 24, 26–36] can be understood at a conceptional and qualitative level and perform well in many demonstrations.

In this work, striving for a comprehensive understanding of causal mechanisms and inspired by the cross-map-based techniques, we develop a mathematically rigorous framework for detecting causality in nonlinear dynamical systems, turning eyes towards investigating the original systems from their cross-maps, which is also logically consistent with the natural interpretation of causality as functional dependences [2, 8]. The skills used in cross-map-based methods are assimilated in our framework, while we directly study the original dynamical systems or the reconstructed systems instead of the cross-maps. The foundation of our framework is the *scaling law* for the changing relation of ε with δ arising from the continuity for the investigated system, henceforth the term "continuity scaling". In addition to providing a theory, we demonstrate, using synthetic and real-world data, that our continuity scaling framework is accurate, computationally efficient, widely applicable, showing advantages over the existing methods.

2. Continuity Scaling Framework

To explain the mathematical idea behind the development of our framework, we use the following class of discrete time dynamical systems: $\mathbf{x}_{t+1} = \mathbf{f}(\mathbf{x}_t, \mathbf{y}_t)$ and $\mathbf{y}_{t+1} = \mathbf{g}(\mathbf{x}_t, \mathbf{y}_t)$ for $t \in \mathbb{N}$, where the state variables $\{\mathbf{x}_t\}_{t \in \mathbb{N}}$, $\{\mathbf{y}_t\}_{t \in \mathbb{N}}$ evolve in

the compact manifolds $\mathcal{M}$, $\mathcal{N}$ of dimension $D_{\mathcal{M}}$, $D_{\mathcal{N}}$ under sufficiently smooth map $\mathbf{f}$, $\mathbf{g}$, respectively. We adopt the common recognition of causality in dynamical systems.

Definition 1. If the dependence of $\mathbf{f}(\mathbf{x}, \mathbf{y})$ on $\mathbf{y}$ is nontrivial (i. e., a directional coupling exists), a variation in $\mathbf{y}$ results in a change in the value of $\mathbf{f}(\mathbf{x}, \mathbf{y})$ for any given $\mathbf{x}$, which, according to the natural interpretation of causality [2, 43], admits that $\mathbf{y} : \{\mathbf{y}_t\}_{t \in \mathbb{N}}$ has a direct causal effect on $\mathbf{x} : \{\mathbf{x}_t\}_{t \in \mathbb{N}}$, denoted by $\mathbf{y} \hookrightarrow \mathbf{x}$, as shown in the upper panel of Figure 1(a).

We now interpret the causal relationship stipulated by the continuity of a function. Let $\mathbf{f}_{\mathbf{x}_g}(\cdot) \triangleq \mathbf{f}(\mathbf{x}_g, \cdot)$ for a given point $\mathbf{x}_g \in \mathcal{M}$. For any $\mathbf{y}_P \in \mathcal{N}$, we denote its image under the given function by $\mathbf{x}_I \triangleq \mathbf{f}_{\mathbf{x}_g}(\mathbf{y}_P)$. Applying the logic statement of a continuous function to $\mathbf{f}_{\mathbf{x}_g}(\cdot)$, we have that, for any neighborhood $\mathcal{O}(\mathbf{x}_I, \varepsilon)$ centered at $\mathbf{x}_I$ and of radius $\varepsilon > 0$, there exists a neighborhood $\mathcal{O}(\mathbf{y}_P, \delta)$ centered at $\mathbf{y}_P$ of radius $\delta > 0$, such that $\mathbf{f}_{\mathbf{x}_g}(\mathcal{O}(\mathbf{y}_P, \delta)) \subset \mathcal{O}(\mathbf{x}_I, \varepsilon)$. The neighborhood and its radius are defined by

$$\mathcal{O}(\mathbf{p}, h) = \{\mathbf{s} \in \mathcal{S} | \text{dist}_{\mathcal{S}}(\mathbf{s}, \mathbf{p}) < h\}, \mathbf{p} \in \mathcal{S}, h > 0, \quad (1)$$

where $\text{dist}_{\mathcal{S}}(\cdot, \cdot)$ represents an appropriate metric describing the distance between two given points in a specified manifold $\mathcal{S}$ with $\mathcal{S} = \mathcal{M}$ or $\mathcal{N}$. The meaning of this mathematical statement is that, if we have a neighborhood of the resulting variable $\mathbf{x}_I$ first, we can then find a neighborhood for the causal variable $\mathbf{y}_P$ to satisfy the above mapping and inclusion relation. This operation of "first-ε-then-δ" provides a rigorous base for the principle that the information about the resulting variable can be used to estimate the information of the causal variable and therefore to ascertain causation, as indicated by the long arrow in the middle panels of Figure 1(a). Note that, the existence of the $\delta > 0$ neighborhood is always guaranteed for a continuous map $\mathbf{f}_{\mathbf{x}_g}$. In fact, due to the compactness of the manifold $\mathcal{N}$, a largest value of δ exists. However, if $\mathbf{y}_P$ does not have an explicit causal effect on the variable $\mathbf{x}_I$, i.e., $\mathbf{f}_{\mathbf{x}_g}$ is independent of $\mathbf{y}_P$, the existence of δ is still assured but it is independent of the value of ε, as shown in the upper panel of Figure 1(b). This means that merely determining the existence of a δ-neighborhood is not enough for inferring causation - it is necessary to vary ε systematically and to examine the scaling relation between δ and ε. In the following we discuss a number of scenarios.

Case I. Dynamical variables $\{(\mathbf{x}_t, \mathbf{y}_t)\}_{t \in \mathbb{N}}$ are fully measurable. For any given constant $\varepsilon_{\mathbf{x}} > 0$, the set $\{\mathbf{x}_\tau \in \mathcal{M} | \tau \in I_{\mathbf{x}}^t(\varepsilon_{\mathbf{x}})\}$ can be used to approximate the neighborhood $\mathcal{O}(\mathbf{x}_{t+1}, \varepsilon_{\mathbf{x}})$, where the time index set is

$$I_{\mathbf{x}}^t(\varepsilon_{\mathbf{x}}) \triangleq \{\tau \in \mathbb{N} | \text{dist}_{\mathcal{M}}(\mathbf{x}_{t+1}, \mathbf{x}_\tau) < \varepsilon_{\mathbf{x}}\}. \quad (2)$$

The radius $\delta_{\mathbf{y}}^t = \delta_{\mathbf{y}}^t(\varepsilon_{\mathbf{x}})$ of the neighborhood $\mathcal{O}(\mathbf{y}_t, \delta_{\mathbf{y}}^t)$
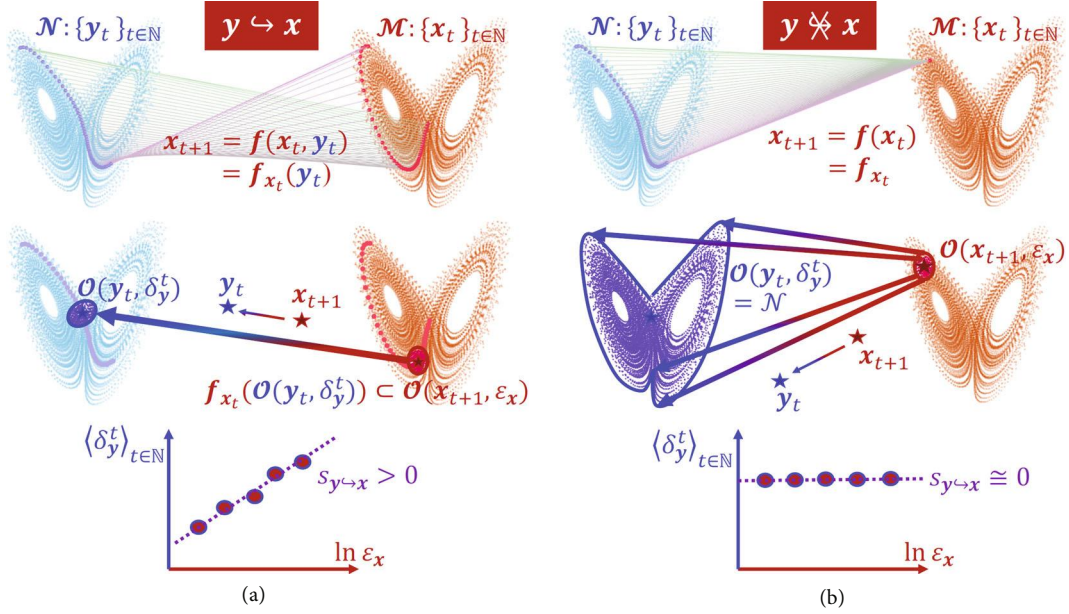
FIGURE 1: Illustration of causal relation between two sets of dynamical variables. (a) Existence of causation from $\mathbf{y}$ in $\mathcal{N}$ to $\mathbf{x}$ in $\mathcal{M}$, where each correspondence from $\mathbf{x}_{t+1}$ to $\mathbf{y}_t$ is one-to-one, represented by the line or the arrow, respectively, in the upper and the middle panels. In this case, a change in $\ln \varepsilon_\mathbf{x}$ results in a direct change in $\delta_\mathbf{y}$ (the lower panel) with $\varepsilon_\mathbf{x}$ and $\delta_\mathbf{y}$ denoting the neighborhood size of the resulting variable $\mathbf{x}$ and of the causal variable $\mathbf{y}$, respectively. (b) Absence of causation from $\mathbf{y}$ to $\mathbf{x}$, where every point on each trajectory, $\{\mathbf{y}_t\}$, in $\mathcal{N}$ could be the correspondent point from $\mathbf{x}_{t+1}$ in $\mathcal{M}$ (the upper panel) and thus every point in $\mathcal{N}$ belongs to the largest δ-neighborhood of $\mathbf{y}_t$ (the middle panel). In this case, $\delta_\mathbf{y}$ does not depend on $\varepsilon_\mathbf{x}$ (the lower panel). Also refer to the supplemental animation for illustration.

satisfying $\mathbf{f}_{\mathbf{x}_g = \mathbf{x}_t}(\mathcal{O}(\mathbf{y}_t, \delta_\mathbf{y}^t)) \subset \mathcal{O}(\mathbf{x}_{t+1}, \varepsilon_\mathbf{x})$ can be estimated as

$$\delta_\mathbf{y}^t(\varepsilon_\mathbf{x}) \triangleq \left\{ \# \left[\bar{I}_\mathbf{x}^t(\varepsilon_\mathbf{x}) \right] \right\}^{-1} \sum_{\tau \in \bar{I}_\mathbf{x}^t(\varepsilon_\mathbf{x})} \mathrm{dist}_\mathcal{N}(\mathbf{y}_t, \mathbf{y}_{\tau-1}), \qquad (3)$$

where $\#[\cdot]$ is the cardinality of the given set and the index set is $\bar{I}_\mathbf{x}^t(\varepsilon_\mathbf{x}) \triangleq \{\tau \in I_\mathbf{x}^t(\varepsilon_\mathbf{x}) | \mathrm{dist}_\mathcal{M}(\mathbf{x}_t, \mathbf{x}_{\tau-1}) < \varepsilon_\mathbf{x}\}$.

The strict mathematical steps for estimating $\delta_\mathbf{y}^t$ are given in Section II of Supplementary Information (SI). We emphasize that here correspondence between $\mathbf{x}_{t+1}$ and $\mathbf{y}_t$ is investigated, differing from the cross-map-based methods, with one-step time difference naturally arising. This consideration yields a key condition [DD], which is only need when considering the original iteration/flow and whose detailed description and universality are demonstrated in SI. We reveal a linear scaling law between $\langle \delta_\mathbf{y}^t \rangle_{t\in\mathbb{N}}$ and $\ln \varepsilon_\mathbf{x}$, as shown in the lower panels of Figure 1, whose slope $s_{\mathbf{y}\hookrightarrow\mathbf{x}}$ is an indicator of the correspondent relation between ε and δ and hence the causal relation $\mathbf{y}\hookrightarrow\mathbf{x}$. Here, $\langle\cdot\rangle_{t\in\mathbb{N}}$ denotes the average over time. In particular, a larger slope value implies a stronger causation in the direction from $\mathbf{y}$ to $\mathbf{x}$ as represented by the map functions $\mathbf{f}(\mathbf{x}_t, \mathbf{y}_t)$ (Figure 1(a)), while a near zero slope indicates null causation in this direction (Figure 1(b)). Likewise, possible causation in the reversed direction, $\mathbf{x}\hookrightarrow\mathbf{y}$, as represented by the function $\mathbf{g}(\mathbf{x}_t, \mathbf{y}_t)$, can be assessed analogously. And the unidirectional case when $\mathbf{f}(\mathbf{x}, \mathbf{y}) = \mathbf{f}_0(\mathbf{x})$ independent of $\mathbf{y}$ is uniformly considered in Case II. We summarize the consideration below

and an argument for the generic existence of the scaling law is provided in Section II of SI.

Theorem 2. *For dynamical variables $\{(\mathbf{x}_t, \mathbf{y}_t)\}_{t\in\mathbb{N}}$ measured directly from the dynamical systems, if the slope $s_{\mathbf{y}\hookrightarrow\mathbf{x}}$ defined above is zero, no causation exists from $\mathbf{y}$ to $\mathbf{x}$. Otherwise, a directional coupling can be confirmed from $\mathbf{y}$ to $\mathbf{x}$ and the slope $s_{\mathbf{y}\hookrightarrow\mathbf{x}}$ increases monotonically with the coupling strength.*

Case II. The dynamical variables $\{(\mathbf{x}_t, \mathbf{y}_t)\}_{t\in\mathbb{N}}$ are not directly accessible but measurable time series $\{u_t\}_{t\in\mathbb{N}}$ and $\{v_t\}_{t\in\mathbb{N}}$ are available, where $u_t = u(\mathbf{x}_t)$ and $v_t = v(\mathbf{y}_t)$ with $u: \mathcal{M} \longrightarrow \mathbb{R}^{r_u}$ and $v: \mathcal{N} \longrightarrow \mathbb{R}^{r_v}$ being smooth observational functions. To assess causation from $\mathbf{y}$ to $\mathbf{x}$, we assume one-dimensional observational time series (for simplicity): $r_u = r_v = 1$, and use the classical delay-coordinate embedding method [37–42, 44] to reconstruct the phase space: $\mathbf{u}_t = (u_t, u_{t+\tau_u}, \cdots, u_{t+(d_u-1)\tau_u})^T$ and $\mathbf{v}_t = (v_t, v_{t+\tau_v}, \cdots, v_{t+(d_v-1)\tau_v})^T$, where $\tau_{u,v}$ is the delay time and $d_{u,v} > 2(D_\mathcal{M} + D_\mathcal{N})$ is the embedding dimension that can be determined using some standard criteria [45]. As illustrated in Figure 2, the dynamical evolution of the reconstructed states $\{(\mathbf{u}_t, \mathbf{v}_t)\}_{t\in\mathbb{N}}$ is governed by

$$\mathbf{u}_{t+1} = \tilde{\mathbf{f}}(\mathbf{u}_t, \mathbf{v}_t), \mathbf{v}_{t+1} = \tilde{\mathbf{g}}(\mathbf{u}_t, \mathbf{v}_t). \qquad (4)$$

The map functions can be calculated as $\tilde{\mathbf{f}}(\mathbf{u}, \mathbf{v}) \triangleq \mathbf{E}_u \circ [\mathbf{f}, \mathbf{g}](\mathbf{\Pi}_1 \circ \mathbf{E}_u^{-1}(\mathbf{u}), \mathbf{\Pi}_2 \circ \mathbf{E}_v^{-1}(\mathbf{v}))$, $\tilde{\mathbf{g}}(\mathbf{u}, \mathbf{v}) \triangleq \mathbf{E}_v \circ [\mathbf{f}, \mathbf{g}](\mathbf{\Pi}_1 \circ \mathbf{E}_u^{-1}(\mathbf{u}), \mathbf{\Pi}_2 \circ \mathbf{E}_v^{-1}(\mathbf{v}))$, where the embedding (diffeomorphism)
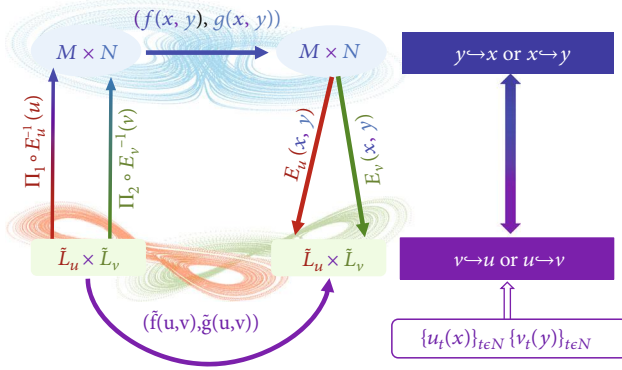
FIGURE 2: Illustration of system dynamics before and after embedding for Case II. In the left panel, the arrows describe how the original systems $(\mathbf{f}, \mathbf{g})$ is equivalent to the system $(\tilde{\mathbf{f}}, \tilde{\mathbf{g}})$ after embedding. In the right panel, causation between the internal variables $\mathbf{x}$ and $\mathbf{y}$ can be ascertained by detecting the causation between the variables $\mathbf{u}$ and $\mathbf{v}$ reconstructed from measured time series.

$\mathbf{E}_s: \mathcal{M} \times \mathcal{N} \longrightarrow \tilde{\mathcal{L}}_s \subset \mathbb{R}^{d_s}$ with $\tilde{\mathcal{L}}_s \triangleq \mathbf{E}_s(\mathcal{M} \times \mathcal{N})$, $s = u$ or v, is given by

$$
\begin{aligned}
\mathbf{E}_u(\mathbf{x}, \mathbf{y}) &\triangleq (u(\mathbf{x}), u \circ \mathbf{\Pi}_1 \circ [\mathbf{f}, \mathbf{g}]^{\tau_u}(\mathbf{x}, \mathbf{y}), u \circ \mathbf{\Pi}_1 \circ \\
&\quad [\mathbf{f}, \mathbf{g}]^{2\tau_u}(\mathbf{x}, \mathbf{y}), \cdots, u \circ \mathbf{\Pi}_1 \circ [\mathbf{f}, \mathbf{g}]^{(d_u-1)\tau_u}(\mathbf{x}, \mathbf{y})), \\
\mathbf{E}_v(\mathbf{x}, \mathbf{y}) &\triangleq (v(\mathbf{y}), v \circ \mathbf{\Pi}_2 \circ [\mathbf{f}, \mathbf{g}]^{\tau_v}(\mathbf{x}, \mathbf{y}), v \circ \mathbf{\Pi}_2 \circ \\
&\quad [\mathbf{f}, \mathbf{g}]^{2\tau_v}(\mathbf{x}, \mathbf{y}), \cdots, v \circ \mathbf{\Pi}_2 \circ [\mathbf{f}, \mathbf{g}]^{(d_v-1)\tau_v}(\mathbf{x}, \mathbf{y})),
\end{aligned}
\tag{5}
$$

with the inverse function $\mathbf{E}_s^{-1}$ defined on $\tilde{\mathcal{L}}_s$, $[\mathbf{f}, \mathbf{g}]^k$ representing the kth iteration of the map and the projection mappings as $\mathbf{\Pi}_1(\mathbf{x}, \mathbf{y}) = \mathbf{x}$ and $\mathbf{\Pi}_2(\mathbf{x}, \mathbf{y}) = \mathbf{y}$. Case II has now been reduced to Case I, and our continuity scaling framework can be used to ascertain the causation from $\mathbf{v}$ to $\mathbf{u}$ based on the measured time series with the indices $I_{\mathbf{u}}^t(\varepsilon_{\mathbf{u}})$, $\delta_{\mathbf{v}}^t(\varepsilon_{\mathbf{u}})$ and $s_{\mathbf{v} \hookrightarrow \mathbf{u}}$ (equations (2) and (3)).

Does the causation from $\mathbf{v}$ to $\mathbf{u}$ imply causation from $\mathbf{y}$ to $\mathbf{x}$? The answer is affirmative, which can be argued, as follows. If the original map function $\mathbf{f}$ is independent of $\mathbf{y}$: $\mathbf{f}(\mathbf{x}, \mathbf{y}) = \mathbf{f}_0(\mathbf{x})$, there is no causation from $\mathbf{y}$ to $\mathbf{x}$. In this case, the embedding $\mathbf{E}_u(\mathbf{x}, \mathbf{y})$ becomes independent of $\mathbf{y}$, degenerating into the form of $\mathbf{E}_u(\mathbf{x}, \mathbf{y}) = \mathbf{E}_{u0}(\mathbf{x})$, a diffeomorphism from $\mathcal{M}$ to $\tilde{\mathcal{L}}_{u0} = \mathbf{E}_{u0}(\mathcal{M})$ only. As a result, equation (4) becomes $\mathbf{u}_{t+1} = \tilde{\mathbf{f}}_0(\mathbf{u}_t)$ and $\mathbf{v}_{t+1} = \tilde{\mathbf{g}}(\mathbf{u}_t, \mathbf{v}_t)$, where $\tilde{\mathbf{f}}_0(\mathbf{u}) = \mathbf{E}_{u0} \circ \mathbf{f} \circ \mathbf{E}_{u0}^{-1}(\mathbf{u})$ and the resulting mapping $\tilde{\mathbf{f}}_0$ is independent of $\mathbf{v}$. The independence can be validated by computing the slope $s_{\mathbf{v} \hookrightarrow \mathbf{u}}$ associated with the scaling relation between $\langle \delta_{\mathbf{v}}^t \rangle_{t \in \mathbb{N}}$ and $\ln \varepsilon_{\mathbf{u}}$, where a zero slope indicates null causation from $\mathbf{v}$ to $\mathbf{u}$ and hence null causation from $\mathbf{y}$ to $\mathbf{x}$. Conversely, a finite slope signifies causation between the variables. Thus, any type of causal relation (unidirectional or bi-directional) detected between the reconstructed state variables $\{(\mathbf{u}_t, \mathbf{v}_t)\}_{t \in \mathbb{N}}$ implies the same type of causal relation between the internal but inaccessible variables $\mathbf{x}$ and $\mathbf{y}$ of the original system.

Case III. The structure of the internal variables is completely unknown. Given the observational functions $\tilde{u}, \tilde{v}: \mathcal{M} \times \mathcal{N} \longrightarrow \mathbb{R}$ with $\tilde{u}_t = \tilde{u}(\mathbf{x}_t, \mathbf{y}_t)$ and $\tilde{v}_t = \tilde{v}(\mathbf{x}_t, \mathbf{y}_t)$, we first reconstruct the state space: $\tilde{\mathbf{u}}_t = (\tilde{u}_t, \tilde{u}_{t+\tau}, \cdots, \tilde{u}_{t+(d-1)\tau})^T$ and $\tilde{\mathbf{v}}_t = (\tilde{v}_t, \tilde{v}_{t+\tau}, \cdots, \tilde{v}_{t+(d-1)\tau})^T$. To detect and quantify causation from $\tilde{\mathbf{v}}$ to $\tilde{\mathbf{u}}$ (or vice versa), we carry out a continuity scaling analysis with the modified indices $I_{\tilde{\mathbf{u}}}^t(\varepsilon_{\tilde{\mathbf{u}}})$, $\delta_{\tilde{\mathbf{v}}}^t(\varepsilon_{\tilde{\mathbf{u}}})$ and $s_{\tilde{\mathbf{v}} \hookrightarrow \tilde{\mathbf{u}}}$. Differing from Case II, here, due to the lack of knowledge about the correspondence structure between the internal and observational variables, a causal relation for the latter does not definitely imply the same for the former.

Case IV. Continuous-time dynamical systems possessing a sufficiently smooth flow $\{\mathbf{S}_t; t \in \mathbb{R}\}$ on a compact manifold $\mathcal{H}$: $d\mathbf{S}_t(\mathbf{u}_0)/dt = \chi(\mathbf{S}_t(\mathbf{u}_0))$, where χ is the vector field. Let $\{\hat{u}_{t=\omega n+\nu}\}_{n \in \mathbb{Z}}$ and $\{\hat{v}_{t=\omega n+\nu}\}_{n \in \mathbb{Z}}$ be two respective time series from the smooth observational functions $\hat{u}, \hat{v}: \mathcal{H} \longrightarrow \mathbb{R}$ with $\hat{u}_t = \hat{u}(\mathbf{S}_t)$ and $\hat{v}_t = \hat{v}(\mathbf{S}_t)$, where $1/\omega$ is the sampling rate and ν is the time shift. Defining $\mathbf{\Xi} \triangleq \mathbf{S}_\omega: \mathcal{H} \longrightarrow \mathcal{H}$ and $\hat{\mathbf{S}}_n \triangleq \mathbf{S}_{\omega n+\nu}(\mathbf{u}_0)$, we obtain a discrete-time system as $\hat{\mathbf{S}}_{n+1} = \mathbf{\Xi}(\hat{\mathbf{S}}_n)$ with the observational functions as $\hat{u}_n = \hat{u}(\hat{\mathbf{S}}_n)$ and $\hat{v}_n = \hat{v}(\hat{\mathbf{S}}_n)$, reducing the case to Case III and rendering applicable our continuity scaling analysis to unveil and quantify the causal relation between $\{\hat{u}_{t=\omega n+\nu}\}_{n \in \mathbb{Z}}$ and $\{\hat{v}_{t=\omega n+\nu}\}_{n \in \mathbb{Z}}$. If the domains of $\hat{u}$ and $\hat{v}$ have their own restrictions on some particular subspaces, e.g., $\hat{u}: \mathcal{H}_u \longrightarrow \mathbb{R}$ and $\hat{v}: \mathcal{H}_v \longrightarrow \mathbb{R}$ with $\mathcal{H} = \mathcal{H}_u \oplus \mathcal{H}_v$, the case is further reduced to Case II, so the detected causal relation between the observational variables imply causation between the internal variables belonging to their respective subspaces.

3. Demonstrations: From Complex Dynamical Models to Real-World Networks

To demonstrate the efficacy of our continuity scaling framework and its superior performance, we have carried out extensive numerical tests with a large number of synthetic and empirical datasets, including those from gene regulatory networks as well as those of air pollution and hospital admission. The practical steps of the continuity scaling framework together with the significance test procedures are described in Methods. We present three representative examples here, while leaving others of significance to SI.

The first example is an ecological model of two unidirectionally interacting species: $x_{1,t+1} = x_{1,t}(3.8 - 3.8x_{1,t} - \mu_{12} x_{2,t})$ and $x_{2,t+1} = x_{2,t}(3.7 - 3.7x_{2,t} - \mu_{21}x_{1,t})$. With time series $\{(x_{1,t}, x_{2,t})\}_{t \in \mathbb{N}}$ obtained from different values of the coupling parameters, our continuity scaling framework yields correct results of different degree of unidirectional causation, as shown in Figures 3(a) and 3(b). In all cases, there exists a reasonable range of $\ln \varepsilon_{x_2}$ (neither too small nor too large) from which the slope $s_{x_1 \hookrightarrow x_2}$ of the linear scaling can be extracted. The statistical significance of the estimated slope values and consequently the strength of causation can be assessed with the standard p-value test [46] (Methods and SI). An ecological model with bidirectional coupling has also been tested (see Section III of SI). Figures 3(c) and 3(d)
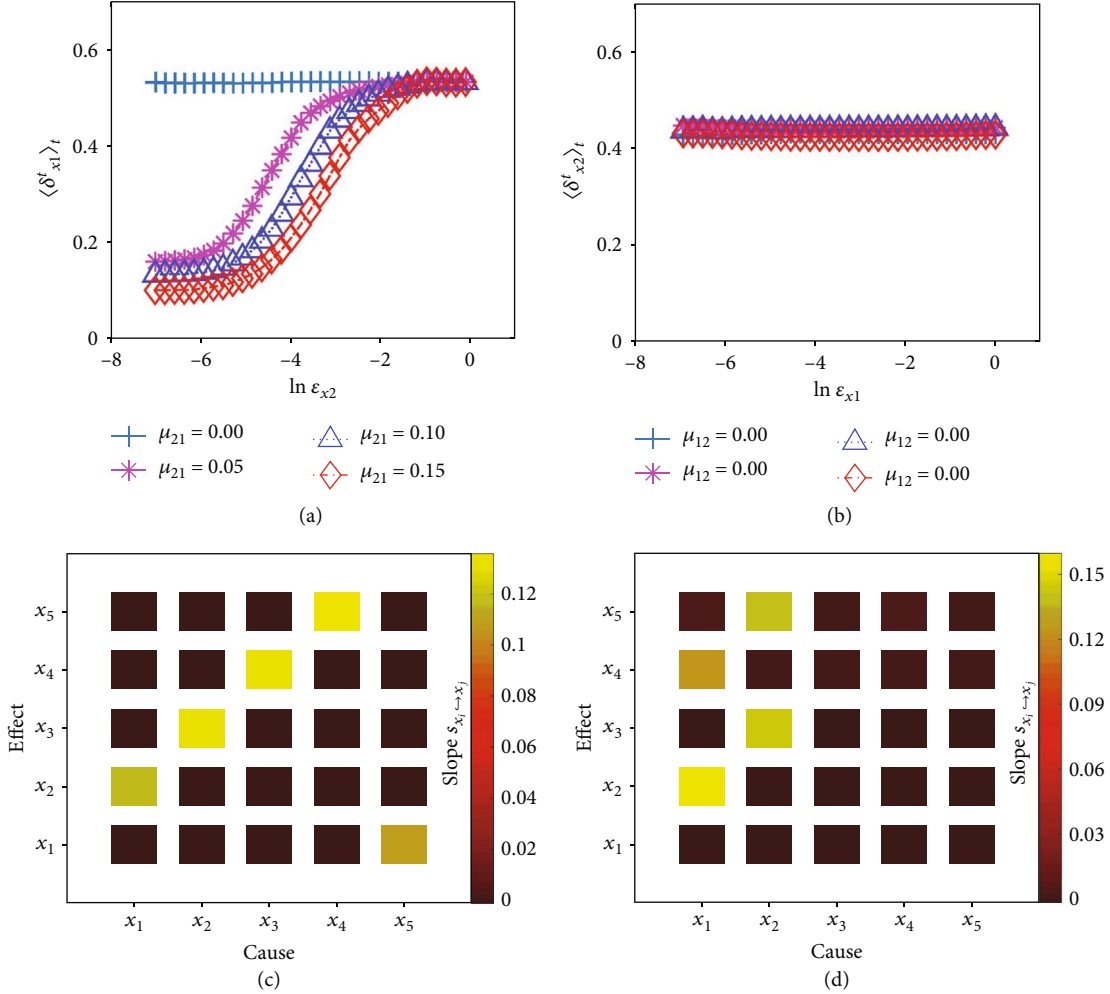
FIGURE 3: Ascertaining and characterizing causation in various ecological systems of interacting species. (a, b) Unidirectional causation of two coupled species. In (a), the values of the slope $s_{x_1 \hookrightarrow x_2}$ associated with the causal relation $x_1 \hookrightarrow x_2$ are approximately 0.0004, 0.1167, 0.1203, and 0.1238 for four different values of the coupling parameter μ_{21}. (b) Near zero slope values $s_{x_2 \hookrightarrow x_1}$ for $x_2 \hookrightarrow x_1$, indicating its nonexistence. (c, d) Inferred causal network of five species whose interacting structure is, respectively, that of a ring: $x_i \hookrightarrow x_{i+1 (\mathrm{mod}\ 5)}$ ($i = 1, \cdots, 5$) and of a tree: $x_j \hookrightarrow x_{j+1, j+3}$ ($j = 1, 2$), where the estimated slope values are color-coded. Results of a statistical analysis of the accuracy and reliability of the determined causal interactions are presented in SI Section III. Time series of length 5000 are used in all these simulations. The embedding parameters are $\tau_s = 1$ and $d_s = 3$ with $s = x_1, \cdots, x_5$.

show the results from ecological networks of five mutually interacting species on a ring and on a tree structure, respectively, where the color-coded slope values reflect accurately the interaction patterns in both cases.

The second example is the coupled Lorenz system: $\dot{x}_i = \sigma_i(y_i - x_i) + \mu_{ij}x_j$, $\dot{y}_i = x_i(\rho_i - z_i) - y_i$, $\dot{z}_i = x_iy_i - \beta_iz_i$ with i, $j = 1, 2$ and $i \neq j$. We use time series $\{y_{1,t}, y_{2,t}\}_{t=n\omega}$ for detecting different configurations of causation (see Section III of SI). Figure 4 presents the overall result, where the color-coded estimated values of the slope from the continuity scaling are shown for different combinations of the sampling rate $1/\omega$ and coupling strength. Even with relatively low sampling rate, our continuity scaling framework can successfully detect and quantify the strength of causation. Note that the accuracy does not vary monotonously with the sampling rate, indicating the potential of our framework

to ascertain and quantify causation even with rare data. Moreover, the proposed index can accurately reflect the true causal strength (denoting by the coupling parameter), which is also evidenced by numerical tests in Sections III and IV of SI. Robustness tests against different noise perturbations are provided in Section III of SI demonstrating the practical usefulness of our framework. Additionally, analogous to the first example, we present in SI several examples on causation detection in the coupled Lorenz system with nonlinear couplings, and the Rössler-Lorenz system, etc., which further demonstrates the generic efficacy of our framework.

In addition, we present study on several real-world dataset, which brings new insights to the evolutionary mechanism of underlying systems. We study gene expression data from DREAM4 in silico Network Challenge [47, 48], whose intrinsic gene regulatory networks (GRNs) are known for verification (Figure 5(a) and Figure S17 of SI). Applying
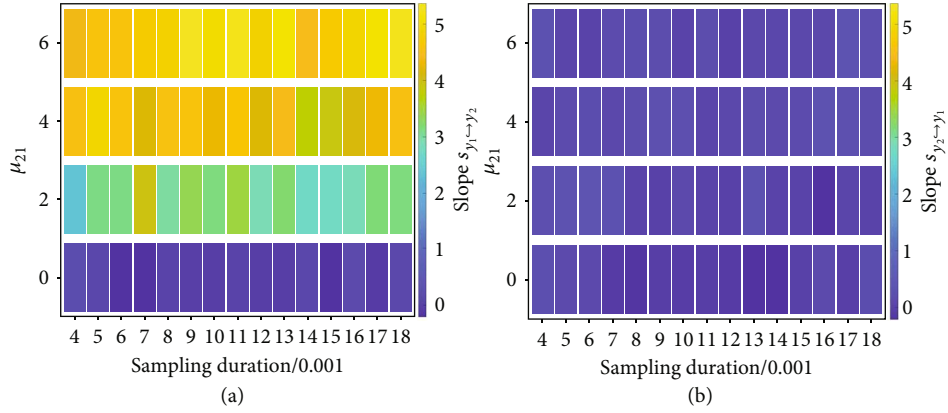
FIGURE 4: Detecting causation in the unidirectionally coupled Lorenz system. The results are for different values of μ_{21} ($\mu_{12} = 0$) and sampling rate $1/\omega$. (a, b) Color-coded values of the slopes $s_{y_1 \hookrightarrow y_2}$ and $s_{y_2 \hookrightarrow y_1}$, respectively. The integration time step is 10^{-3} and the embedding parameters are $d_s = 7$, $\tau_s \approx 0.05$ with $\omega | \tau_s$ ($s = y_1$ or y_2). See Section III and Table S9 of SI for all the other parameters including the time series lengths used in the simulations.
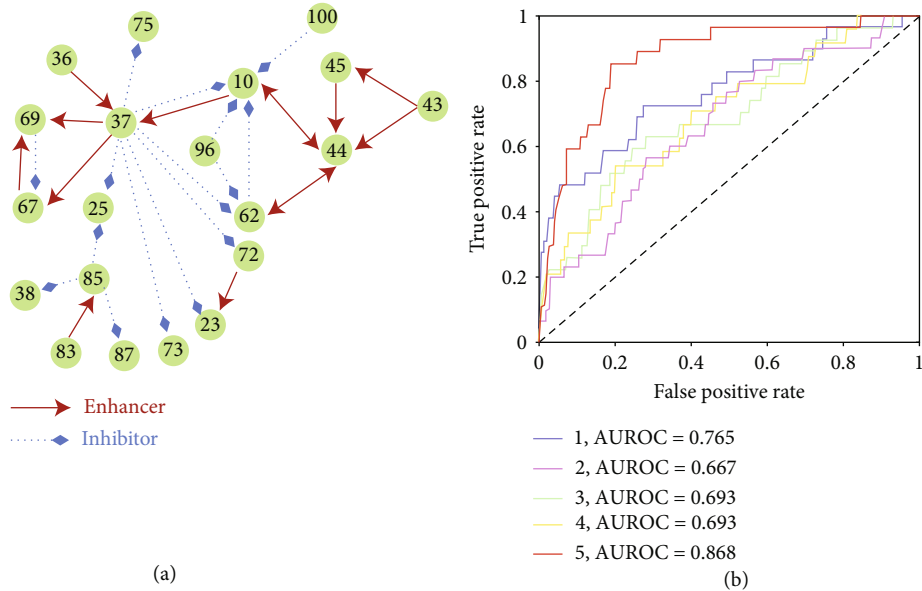
FIGURE 5: Detecting causal interactions in five GRNs. (a) One representative GRN containing 20 randomly selected genes. Other four structures can be found in Figure S17 of SI. (b) The ROC curves as well as their AUROC values demonstrate the efficacy of our framework.

our framework to these data, we ascertain the causations between each pair of genes by using the continuity scaling framework. The corresponding ROC curves for five different networks as well as their AUROC values are shown in Figure 5(b), which indicates a high detection accuracy in dealing with real-world data.

We then test the causal relationship in a marine ecosystem consisting of Pacific sardine landings, northern anchovy landings and sea surface temperature (SST). We reveal new findings to support the competing relationship hypothesis stated in [49] which cannot be detected by CCM [25]. As pointed out in Figure 6, while common influence from SST to both species is verified with both methods, our continuity scaling additionally illuminates notable influence from anchovy to sardine with its reverse direction being less significant. While competing relationship plays an important

role in ecosystems, continuity scaling can reveal more essential interaction mechanism. See Section III.E of SI for more details.

Moreover, we study the transmission mechanism of the recent COVID-19 pandemic. Particularly, we analyze the daily new cases of COVID-19 of representative countries for two stages: day 1 (January 22nd 2020) to day 100 (April 30th 2020) and day 101 (May 1st 2020) to day 391 (February 15th 2021). Our continuity scaling is pairwisely applied to reconstruct the transmission causal network. As shown in Figure 7, China shows a significant effect on a few countries at the first stage and this effect disappears at the second stage. However, other countries show a different situation with China, whose external effect lasts as shown in Section III.E and Figure S18 of SI. Our results accord with that China holds stringent epidemic control strategies with
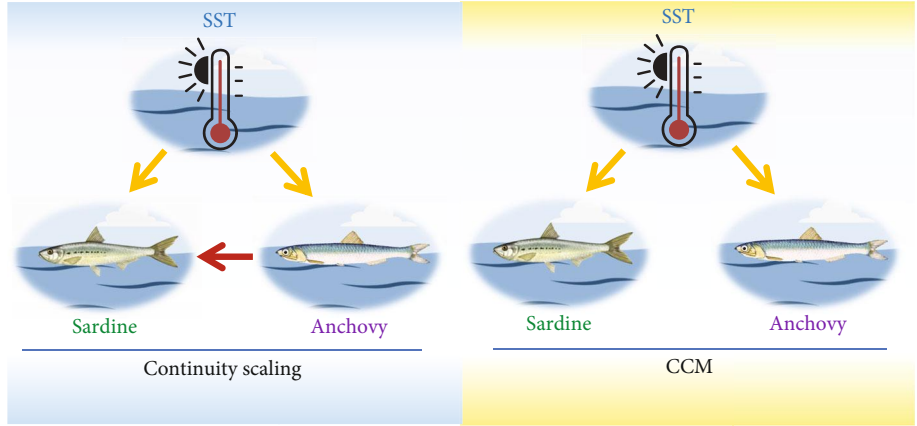
FIGURE 6: The comparison of causal network structure detected by continuity scaling and CCM among sea surface temperature, sardine, and anchovy.
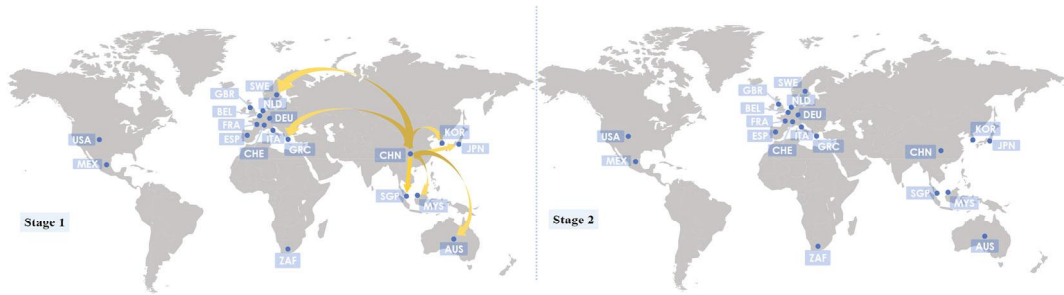

FIGURE 7: The causal effect from China to other countries of the COVID-19 pandemic detected by continuity scaling between stages 1 and 2. Here, stage 1 is from January 22nd 2020 to April 30th 2020, and stage 2 is from May 1st 2020 to February 15th 2021. For those detected causal links between all countries, refer to Section III.E and Figure S18 of SI. These maps are for illustration only.

sporadic domestic infections, as evidenced by official daily briefings, demonstrating the potential of continuity scaling in detecting causal networks for ongoing complex systems. Additionally, We emphasize that day 100 is a suitable critical day to distinguish the early severe stage and the late well-under-control stage of the pandemic (see Figure S18(a) of SI), while slight change of the critical day will not nullify our result. As shown in Figure S18(b) of SI, when the critical day varies from day 94 to day 106, no significant change (less than 5%) of the detected causal links occurs at both stages, and the number of countries under influence of China at Stage 2 remains zero. See more details in Section III.E of SI.

Additional real world examples including air pollutants and hospital admission record from Hong Kong are also shown in Section III of SI.

4. Discussion

To summarize, we have developed a novel framework for data-based detection and quantification of causation in complex dynamical systems. On the basis of the widely used cross-map-based techniques, our framework enjoys a rigorous foundation, focusing on the continuity scaling law of the concerned system directly instead of only investigating the continuity of its cross-map. Therefore, our framework

is consistent with the standard interpretation of causality, and works even in the typical cases where several existing typical methods do not perform that well or even they fail (see the comparison results in Section IV of SI). In addition, the mathematical reasoning leading to the core of our framework, the continuity scaling, helps resolve the long-standing issue associated with techniques directly using cross-map that information about the resulting variables is required to project the dynamical behavior of the causal variables, whereas several works in the literature [50], which directly studied the continuity or the smoothness of the cross-map, likely yielded confused detected results on causal directions.

Computational complexity. The computational complexity of the algorithm is $O(T^2 N_\varepsilon)$, which is relatively smaller than the CCM method, whose computational complexity is $O(T^2 \log T)$.

Limitations and future works. Nevertheless, there are still some spaces for improving the presently proposed framework. First, currently, only bivariate detection algorithm is designed, so generalization to multivariate network inference requires further considerations, as analogous to those works presented in Refs. [51–53]. Second, the causal time delay has not been taken into account in the current framework, so it also could be further investigated, similar to the work reported in Ref. [33]. Also, more advanced algorithms, such as the one developed in Ref. [54], could be

integrated into this framework for detecting those temporal causal structures. Definitely, we will settle these questions in our future work.

Detecting causality in complex dynamical systems has broad applications not only in science and engineering, but also in many aspects of the modern society, demanding accurate, efficient, and rigorously justified and hence trustworthy methodologies. Our present work provides a vehicle along this feat and indeed resolves the puzzles arising in the use of those influential methods.

5. Methods

Continuity scaling framework: a detailed description of algorithms. Let $\{u_t\}_{t=1,2,\cdots,T}$ and $\{v_t\}_{t=1,2,\cdots,T}$ be two experimentally measured time series of internal variables $\{(x_t, y_t)\}_{t\in\mathbb{N}}$. Typically, if the dynamical variables $\{(x_t, y_t)\}_{t\in\mathbb{N}}$ are accessible, $\{(u_t, v_t)\}$ reduce to one-dimensional coordinate of the internal system. The key computational steps of our continuity scaling framework are described, as follows.

We reconstruct the phase space using the classical method of delay coordinate embedding [37] with the optimal embedding dimension d_z and time lag τ_z determined by the methods in Refs. [55, 56] (i.e., the false nearest neighbors and the delayed mutual information, respectively):

$$\mathscr{L}_z \triangleq \left\{ \mathbf{z}(t) = \left(z_t, z_{t+\tau_z}, \cdots, z_{t+(d_z-1)\tau_z} \right) \middle| t=1, \cdots, T_0 \right\}, \quad (6)$$

where $z = u, v$, $T_0 = \min\{T - (d_z - 1)\tau_z | z = u, v\}$, and Euclidean distance is used for both $\mathscr{L}_{u,v}$.

We present the steps for causation detection using the case of $\mathbf{v} \hookrightarrow \mathbf{u}$ as an example.

We calculate the respective diameters for $\mathscr{L}_{u,v}$ as

$$D_z \triangleq \max \left\{ \text{dist}_{\mathscr{L}_z}(\mathbf{z}(t), \mathbf{z}(\tau)) \middle| 1 \le t, \tau \le T_0 \right\}, \quad (7)$$

where $z = u, v$, and $\mathbf{z} = \mathbf{u}, \mathbf{v}$. We set up a group of numbers, $\{\varepsilon_{\mathbf{u},j}\}_{j=1,\cdots,N_\varepsilon}$, as $\varepsilon_{\mathbf{u},1} = e \cdot D_u$, $\varepsilon_{\mathbf{u},N_\varepsilon} = D_u$, with the other elements satisfying

$$\frac{\ln \varepsilon_{\mathbf{u},j} - \ln \varepsilon_{\mathbf{u},1}}{j-1} = \frac{\ln \varepsilon_{\mathbf{u},N_\varepsilon} - \ln \varepsilon_{\mathbf{u},1}}{N_\varepsilon - 1}, \quad (8)$$

for $j = 2, \cdots, N_\varepsilon - 1$. Then, in light of (2) with (3), we have

$$\delta_{\mathbf{v}}^t(\varepsilon_{\mathbf{u}}) = \# \left[I_{\mathbf{u}}^t(\varepsilon_{\mathbf{u}}) \right]^{-1} \sum_{\tau \in I_{\mathbf{u}}^t(\varepsilon_{\mathbf{u}})} \text{dist}_{\mathscr{L}_v}(\mathbf{v}(t), \mathbf{v}(\tau - 1)), \quad (9)$$

with

$$I_{\mathbf{u}}^t(\varepsilon_{\mathbf{u}}) = \left\{ \tau \in \mathbb{N} \middle| \text{dist}_{\mathscr{L}_u}(\mathbf{u}(t+1), \mathbf{u}(\tau)) < \varepsilon_{\mathbf{u}}, |t+1-\tau| > E \right\} \quad (10)$$

where numerically, $\varepsilon_{\mathbf{u}}$ alters its value successively from the set $\{\varepsilon_{\mathbf{u},j}\}_{j=1,\cdots,N_\varepsilon}$, and the threshold E is a positive number

chosen to avoid the situation where the nearest neighboring points are induced by the consecutive time order only.

As defined, $\langle \delta_{\mathbf{v}}^t(\varepsilon_{\mathbf{u}}) \rangle_{t\in\mathbb{N}}$ is the average of $\delta_{\mathbf{v}}^t(\varepsilon_{\mathbf{u}})$ over all possible time t. We use a finite number of pairs $\left\{ \left(\langle \delta_{\mathbf{v}}^t(\varepsilon_{\mathbf{u},j}) \rangle_{t\in\mathbb{N}_{T_0}}, \ln \varepsilon_{\mathbf{u},j} \right) \right\}_{j=1,\cdots,N_\varepsilon}$ to approximate the scaling relation between $\langle \delta_{\mathbf{v}}^t(\varepsilon_{\mathbf{u}}) \rangle_{t\in\mathbb{N}}$ and $\ln \varepsilon_{\mathbf{u}}$, where $\mathbb{N}_{T_0} = \{1, 2, \cdots, T_0\}$. Theoretically, a larger value of N_ε and a smaller value of e will result in a more accurate approximation of the scaling relation. In practice, the accuracy is determined by the length of the observational time series, the sampling duration, and different types of noise perturbations. In numerical simulations, we set $e = 0.001$ and $N_\varepsilon = 33$. In addition, a too large or a too small value of $\varepsilon_{\mathbf{u}}$ can induce insufficient data to restore the neighborhood and/or the entire manifold. We thus set $\delta_{\mathbf{v}}^t(\varepsilon_{\mathbf{u},j}) = \delta_{\mathbf{v}}^t(\varepsilon_{\mathbf{u},j+1})$ as a practical technique as the number of points is limited practically in a small neighborhood. As a result, near zero slope values would appear on both sides of the scaling curve $\langle \delta_{\mathbf{v}}^t(\varepsilon_{\mathbf{u}}) \rangle_{t\in\mathbb{N}}$-$\ln \varepsilon_{\mathbf{u}}$, as demonstrated in Figure 3 and in SI. In such a case, to estimate the slope of the scaling relation, we take the following approach.

Define a group of numbers by

$$S_j \triangleq \frac{\langle \delta_{\mathbf{v}}^t(\varepsilon_{\mathbf{u},j+1}) \rangle_{t\in\mathbb{N}_{T_0}} - \langle \delta_{\mathbf{v}}^t(\varepsilon_{\mathbf{u},j}) \rangle_{t\in\mathbb{N}_{T_0}}}{\ln \varepsilon_{\mathbf{u},j+1} - \ln \varepsilon_{\mathbf{u},j}}, \quad (11)$$

where $j = 1, \cdots, N_\varepsilon - 1$, sort them in a descending order, from which we determine the $[N_\varepsilon + 1/2]$ largest numbers, collect their subscripts - j's together as an index set $\widehat{J}$, and set $H \triangleq \{j, j+1 | j \in \widehat{J}\}$. Applying the least squares method to the linear regression model:

$$\langle \delta_{\mathbf{v}}^t(\varepsilon_{\mathbf{u}}) \rangle_{t\in\mathbb{N}} = S \cdot \ln \varepsilon_{\mathbf{u}} + b \quad (12)$$

with the dataset $\left\{ \left(\langle \delta_{\mathbf{v}}^t(\varepsilon_{\mathbf{u},j}) \rangle_{t\in\mathbb{N}_{T_0}}, \ln \varepsilon_{\mathbf{u},j} \right) \right\}_{j\in H}$, we get the optimal values $(\widehat{S}, \widehat{b})$ for the parameters (S, b) in (12) and finally obtain the slope of the scaling relation as $s_{\mathbf{v}\hookrightarrow\mathbf{u}} \triangleq \widehat{S}$.

For the other causal direction from $\mathbf{u}$ to $\mathbf{v}$, these steps are equally applicable to estimating the slope $s_{\mathbf{u}\hookrightarrow\mathbf{v}}$.

To assess the statistical significance of the numerically determined causation, we devise the following surrogate test using the case of $\mathbf{v}$ causing $\mathbf{u}$ as an illustrative example.

Divide the time series $\{\mathbf{u}(t)\}_{t\in\mathbb{N}_{T_0}}$ into N_G consecutive segments of equal length (except for the last segment - the shortest segment). Randomly shuffle these segments and then regroup them into a surrogate sequence $\{\widehat{\mathbf{u}}(t)\}_{t\in\mathbb{N}_{T_0}}$. Applying such a random permutation method to $\{\mathbf{v}(t)\}_{t\in\mathbb{N}_{T_0}}$ generates another surrogate sequence $\{\widehat{\mathbf{v}}(t)\}_{t\in\mathbb{N}_{T_0}}$. Carrying out the slope computation yields $s_{\widehat{\mathbf{v}}\hookrightarrow\widehat{\mathbf{u}}}$. The procedure can be repeated for a sufficient number of times, say Q, which consequently yields a group of estimated slopes, denoted as $\{s_{\widehat{\mathbf{v}}\hookrightarrow\widehat{\mathbf{u}}}^q\}_{q=0,1\cdots,Q}$, where $s_{\widehat{\mathbf{v}}\hookrightarrow\widehat{\mathbf{u}}}^0$ is set as $s_{\mathbf{v}\hookrightarrow\mathbf{u}}$ obtained from the original time series. For all the estimated slopes, we calculate

their mean $\widehat{\mu}_{\mathbf{v} \to \mathbf{u}}$ and the standard deviation $\widehat{\sigma}_{\mathbf{v} \to \mathbf{u}}$. The p-value for $s_{\mathbf{v} \to \mathbf{u}}$ is calculated as

$$p_{s_{\mathbf{v} \to \mathbf{u}}} \triangleq 1 - \text{normcdf} \left[\frac{s_{\mathbf{v} \to \mathbf{u}} - \widehat{\mu}_{\mathbf{v} \to \mathbf{u}}}{\widehat{\sigma}_{\mathbf{v} \to \mathbf{u}}} \right], \qquad (13)$$

where $\text{normcdf}[\cdot]$ is the cumulative Gaussian distribution function. The principle of statistical hypothesis testing guarantees the existence of causation from $\mathbf{v}$ to $\mathbf{u}$ if $p_{s_{\mathbf{v} \to \mathbf{u}}} < 0.05$.

In simulations, we set the number of segments to be $N_G = 25$ and the number of times for random permutations to be $Q \geq 20$.

Additional Points

Code Availability. The source codes for our CS framework are available at https://github.com/bianzhiyu/ContinuityScaling.

Conflicts of Interest

The authors declare no competing interests.

Authors' Contributions

W.L. conceived idea. X.Y., S.-Y.L., and W.L. designed and performed the research. X.Y., S.-Y.L., H.-F.M., and W.L. analyzed the data. H.-F.M., Y.-C.L., and Q.N. contributed data and analysis tools, and all the authors wrote the paper. X.Y. and S.-Y.L. equally contributed to this work.

Acknowledgments

Supplementary Materials

Supplementary materials: SI.pdf (where we include analytic and computational details of the results in the main text. This SI is helpful but not essential for understanding the main results of the paper.) *(Supplementary Materials)*

References

[1] M. Bunge, *Causality and Modern Science*, Routledge, 2017.

[2] J. Pearl, *Causality*, Cambridge university press, 2013.

[3] J. Runge, S. Bathiany, E. Bollt et al., "Inferring causation from time series in earth system sciences," *Nature Communications*, vol. 10, no. 1, p. 2553, 2019.

[4] F. S. Collins and H. Varmus, "A new initiative on precision medicine," *New England Journal of Medicine*, vol. 372, no. 9, pp. 793–795, 2015.

[5] G. N. Saxe, A. Statnikov, D. Fenyo et al., "A complex systems approach to causal discovery in psychiatry," *PloS One*, vol. 11, no. 3, article e0151174, 2016.

[6] D. R. Cox and D. V. Hinkley, *Theoretical Statistics*, CRC Press, 1979.

[7] T. M. Cover, *Elements of Information Theory*, John Wiley & Sons, 1999.

[8] J. Pearl, "Causal inference in statistics: an overview," *Statistics Surveys*, vol. 3, pp. 96–146, 2009.

[9] N. Wiener, *The Theory of Prediction*, Modern mathematics for engineers, 1956.

[10] C. W. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica: Journal of the Econometric Society*, vol. 37, no. 3, pp. 424–438, 1969.

[11] S. Haufe, V. V. Nikulin, K. R. Müller, and G. Nolte, "A critical assessment of connectivity measures for EEG data: a simulation study," *NeuroImage*, vol. 64, pp. 120–133, 2013.

[12] M. Ding, Y. Chen, and S. L. Bressler, "Granger causality: basic theory and application to neuroscience," *Handbook of Time Series Analysis: recent theoretical developments and applications*, vol. 437, 2006.

[13] T. Schreiber, "Measuring information transfer," *Physical Review Letters*, vol. 85, no. 2, pp. 461–464, 2000.

[14] S. Frenzel and B. Pompe, "Partial mutual information for coupling analysis of multivariate time series," *Physical Review Letters*, vol. 99, no. 20, article 204101, 2007.

[15] R. Vicente, M. Wibral, M. Lindner, and G. Pipa, "Transfer entropy–a model-free measure of effective connectivity for the neurosciences," *Journal of Computational Neuroscience*, vol. 30, no. 1, pp. 45–67, 2011.

[16] J. Runge, J. Heitzig, V. Petoukhov, and J. Kurths, "Escaping the curse of dimensionality in estimating multivariate transfer entropy," *Physical Review Letters*, vol. 108, no. 25, article 258701, 2012.

[17] J. Sun, C. Cafaro, and E. M. Bollt, "Identifying the coupling structure in complex systems through the optimal causation entropy principle," *Entropy*, vol. 16, no. 6, pp. 3416–3433, 2014.

[18] C. Cafaro, W. M. Lord, J. Sun, and E. M. Bollt, "Causation entropy from symbolic representations of dynamical systems," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 25, article 043106, 2015.

[19] J. Sun, D. Taylor, and E. M. Bollt, "Causal network inference by optimal causation entropy," *SIAM Journal on Applied Dynamical Systems*, vol. 14, no. 1, pp. 73–106, 2015.

[20] M. Solyanik-Gorgone, J. Ye, M. Miscuglio, A. Afanasev, A. E. Willner, and V. J. Sorger, "Quantifying information via Shannon entropy in spatially structured optical beams," *Research*, vol. 2021, article 9780760, 2021.

[21] Y. Hirata and K. Aihara, "Identifying hidden common causes from bivariate time series: a method using recurrence plots," *Physical Review E*, vol. 81, no. 1, article 016203, 2010.

[22] R. Q. Quiroga, J. Arnhold, and P. Grassberger, "Learning driver-response relationships from synchronization patterns," *Physical Review E*, vol. 61, no. 5, pp. 5142–5148, 2000.

[23] J. Arnhold, P. Grassberger, K. Lehnertz, and C. E. Elger, "A robust method for detecting interdependences: application to intracranially recorded eeg," *Physica D: Nonlinear Phenomena*, vol. 134, no. 4, pp. 419–430, 1999.

[24] D. Harnack, E. Laminski, M. Schünemann, and K. R. Pawelzik, "Topological causality in dynamical systems," *Physical Review Letters*, vol. 119, no. 9, article 098301, 2017.

[25] G. Sugihara, R. May, H. Ye et al., "Detecting causality in complex ecosystems," *Science*, vol. 338, no. 6106, pp. 496–500, 2012.

[26] E. R. Deyle, M. Fogarty, C. H. Hsieh et al., "Predicting climate effects on pacific sardine," *Proceedings of the National Academy of Sciences*, vol. 110, no. 16, pp. 6430–6435, 2013.

[27] X. Wang, S. Piao, P. Ciais et al., "A two-fold increase of carbon cycle sensitivity to tropical temperature variations," *Nature*, vol. 506, no. 7487, pp. 212–215, 2014.

[28] H. Ma, K. Aihara, and L. Chen, "Detecting causality from nonlinear dynamics with short-term time series," *Scientific Reports*, vol. 4, pp. 1–10, 2014.

[29] J. M. McCracken and R. S. Weigel, "Convergent cross-mapping and pairwise asymmetric inference," *Physical Review E*, vol. 90, no. 6, article 062903, 2014.

[30] H. Ye, E. R. Deyle, L. J. Gilarranz, and G. Sugihara, "Distinguishing time-delayed causal interactions using convergent cross mapping," *Scientific Reports*, vol. 5, no. 1, article 14750, 2015.

[31] A. T. Clark, H. Ye, F. Isbell et al., "Spatial convergent cross mapping to detect causal relationships from short time series," *Ecology*, vol. 96, no. 5, pp. 1174–1181, 2015.

[32] J.-J. Jiang, Z.-G. Huang, L. Huang, H. Liu, and Y.-C. Lai, "Directed dynamical influence is more detectable with noise," *Scientific Reports*, vol. 6, no. 1, article 24088, 2016.

[33] H. Ma, S. Leng, C. Tao et al., "Detection of time delays and directional interactions based on time series from complex dynamical systems," *Physical Review E*, vol. 96, no. 1, article 012221, 2017.

[34] J. M. Amigó and Y. Hirata, "Detecting directional couplings from multivariate flows by the joint distance distribution," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 28, article 075302, 2018.

[35] Y. Wang, J. Yang, Y. Chen, P. De Maeyer, Z. Li, and W. Duan, "Detecting the causal effect of soil moisture on precipitation using convergent cross mapping," *Scientific Reports*, vol. 8, no. 1, pp. 1–8, 2018.

[36] S. Leng, H. Ma, J. Kurths et al., "Partial cross mapping eliminates indirect causal influences," *Nature Communications*, vol. 11, no. 1, pp. 1–9, 2020.

[37] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980*, Springer, 1981.

[38] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, "Geometry from a time series," *Physical Review Letters*, vol. 45, no. 9, pp. 712–716, 1980.

[39] T. Sauer, J. A. Yorke, and M. Casdagli, "Embedology," *Journal of Statistical Physics*, vol. 65, no. 3-4, pp. 579–616, 1991.

[40] J. Stark, "Delay embeddings for forced systems. I. Deterministic forcing," *Journal of Nonlinear Science*, vol. 9, no. 3, pp. 255–332, 1999.

[41] J. Stark, D. S. Broomhead, M. E. Davies, and J. Huke, "Delay embeddings for forced systems. II. Stochastic forcing," *Journal of Nonlinear Science*, vol. 13, no. 6, pp. 519–577, 2003.

[42] M. R. Muldoon, D. S. Broomhead, J. P. Huke, and R. Hegger, "Delay embedding in the presence of dynamical noise," *Dynamics and Stability of Systems*, vol. 13, no. 2, pp. 175–186, 1998.

[43] P. Spirtes, C. N. Glymour, R. Scheines, and D. Heckerman, *Causation, Prediction, and Search*, MIT press, 2001.

[44] B. Cummins, T. Gedeon, and K. Spendlove, "On the efficacy of state space reconstruction methods in determining causality," *SIAM Journal on Applied Dynamical Systems*, vol. 14, no. 1, pp. 335–381, 2015.

[45] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, vol. 7, Cambridge university press, 2004.

[46] G. Lancaster, D. Iatsenko, A. Pidde, V. Ticcinelli, and A. Stefanovska, "Surrogate data for hypothesis testing of physical systems," *Physics Reports*, vol. 748, pp. 1–60, 2018.

[47] D. Marbach, T. Schaffter, C. Mattiussi, and D. Floreano, "Generating realistic in silico gene networks for performance assessment of reverse engineering methods," *Journal of Computational Biology*, vol. 16, no. 2, pp. 229–239, 2009.

[48] D. Marbach, R. J. Prill, T. Schaffter, C. Mattiussi, D. Floreano, and G. Stolovitzky, "Revealing strengths and weaknesses of methods for gene network inference," *Proceedings of the National Academy of Sciences*, vol. 107, no. 14, pp. 6286–6291, 2010.

[49] R. Lasker and A. Mac Call, "New ideas on the fluctuations of the clupeoid stocks off California," *Proceedings of the Joint Oceanographic Assembly, Halifax, August 1982: General Symposia*, Department of Fisheries and Oceans, Ontario, 1983.

[50] M. L. V. Quyen, J. Martinerie, C. Adam, and F. J. Varela, "Nonlinear analyses of interictal EEG map the brain interdependences in human focal epilepsy," *Physica D: Nonlinear Phenomena*, vol. 127, no. 3-4, pp. 250–266, 1999.

[51] J. Peters, D. Janzing, and B. Schölkopf, *Elements of Causal Inference: Foundations and Learning Algorithms*, MIT Press, 2017.

[52] J. Runge, "Causal network reconstruction from time series: from theoretical assumptions to practical estimation," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 28, article 075310, 2018.

[53] Y. Lou, L. Wang, and G. Chen, "Enhancing controllability robustness of q-snapback networks through redirecting edges," *Research*, vol. 2019, article 7857534, 23 pages, 2019.

[54] J.-W. Hou, H.-F. Ma, D. He, J. Sun, Q. Nie, and W. Lin, "Harvesting random embedding for high-frequency change-point detection in temporal complex," *National Science Review*, vol. 9, article nwab228, 2022.

[55] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Physical Review A*, vol. 33, no. 2, pp. 1134–1140, 1986.

[56] M. B. Kennel, R. Brown, and H. D. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Physical Review A*, vol. 45, no. 6, pp. 3403–3411, 1992.