Unified understanding of nonparametric causality detection in time series

Yutaka Osada[1], Masayuki Ushio[2], and Michio Kondoh[1]
[1] *Graduate School of Life Sciences, Tohoku University, Sendai, Japan.*
[2] *Department of Ocean Science, The Hong Kong University of Science and Technology, Hong Kong SAR, China.*
(Dated: July 14, 2023)

Most complex systems in the real world are driven by multiple interactions among components. Identifying these interactions is critical for understanding, forecasting, and controlling system-level phenomena. Transfer entropy (TE) and convergent cross mapping (CCM) are two widely-used approaches for nonparametric causality detection based on time-series data. However, the theoretical relationship between TE and CCM has not been formally explained. Here, we provide a theoretical formulation that links TE and CCM in an information-theoretic framework, showing that they have different definitions of causal influence. Furthermore, from the formulation, we propose a novel nonparametric causality test, named unified information-theoretic causality (UIC), which has lower data requirements than those of TE (due to robustness to noise) and lower false-positive rates than those of CCM. Numerical experiments confirmed that UIC outperforms TE and CCM in terms of the performance of causality detection for both linear and nonlinear dynamical systems. Finally, we demonstrate the practical importance of the conditional test based on UIC using numerical simulations with indirect causality and empirical data for microbial communities in irrigated water samples from rice fields. Our results contribute to a unified understanding of nonparametric causality tests in time series and the accurate reconstruction of complex real-world systems.

I. INTRODUCTION

Detection of cause–effect relationships among variables, events or objects is a fundamental topic in both natural and social sciences. Most attempts at causality detection are easily disturbed by confounding factors [1] and nonlinear dynamics [2]. These issues are particularly challenging in large systems where intervention is technically or ethically difficult or even impossible [3]. In recent years, nonparametric tests for time series have attracted attention as promising methods for detecting potential causal relations with the minimal need to intervene in the target system [2, 4]. Time-series data support Granger's statistical concepts of causality: (1) the cause necessarily occurs before the effect, and (2) the cause contains information about the effect when accounting for confounding variables [5, 6].

Nonparametric causality tests based on time-series data have been proposed from the perspectives of information theory and dynamical systems theory. In this study, we focus on two widely-used methods: transfer entropy (TE) [7] and convergent cross mapping (CCM) [2]. TE is an information-theoretic causality test, known as a nonparametric generalization of Granger causality [8]. TE has a clear mathematical definition of causal influence, which is advantageous for theoretical advances [9, 10] and practical applications [3, 11]. CCM is a nonparametric causality test developed mainly for nonlinear dynamical systems [2, 12]. The most important feature of CCM is the use of time-delay embedding theorems [13], which allow for the detection of causal relations without addressing confounding variables embedded by state space reconstruction [2]. As the performance of these methods can be reversed depending on the system [2, 14], it is essential to clarify the mathematical connection between TE and CCM in order to understand their advantages and disadvantages as causal detection in various systems.

Here, we provide a theoretical formulation that links TE and CCM in an information-theoretic framework. Furthermore, based on the formulation, we propose a novel nonparametric causality test, named unified information-theoretic causality (UIC). UIC incorporates the advantages of both TE and CCM, including (1) a clear mathematical definition of causal influence, (2) noise robustness for causal variables in nonlinear dynamical systems, and (3) efficient computation of statistical significance. Based on the previously developed extension of TE [9], we also introduce a conditional UIC test and apply it to an artificial model system and an empirical microbial system. In the numerical experiments, UIC outperformed TE and CCM for both linear and nonlinear dynamical systems, consistent with the theoretical results.

II. UNIFYING CAUSALITY TESTS

A. Setting

We consider a multivariate time series with $\boldsymbol{x} = \{x_t\}$ as the effect variable and $\boldsymbol{y} = \{y_t\}$ as the cause variable and test whether y_{t-p} has a causal influence on x_t with time lag $p > 0$. Under generic conditions, x_t can be embedded by y_{t-p} and $\boldsymbol{x}_t^{(E,\tau)} = \{x_{t-\tau}, x_{t-2\tau}, ..., x_{t-E\tau}\}$ based on the time-delay embedding theorem [13, 15]. Note that, in most real-world cases, the embedding dimensions ($E > 0$) and time interval ($\tau > 0$) are not known *a priori* and need to be estimated [16].

B. TE and CCM

In the above setting, we give the definitions of TE and CCM following their original formulations. Using the joint probability $p(x, y)$ and the conditional probability $p(x|y)$, TE measures causal influence as the flow of information from a cause variable to an effect variable as follows [7]:

$$\begin{aligned}&\mathrm{TE}(y_{t-p} \to x_t | \boldsymbol{x}_t^{(E,\tau)}) \\ =&\sum_t p(y_{t-p}, x_t, \boldsymbol{x}_t^{(E,\tau)}) \log \frac{p(x_t | y_{t-p}, \boldsymbol{x}_t^{(E,\tau)})}{p(x_t | \boldsymbol{x}_t^{(E,\tau)})}. \quad (1)\end{aligned}$$

Thus, the causal influence from y_{t-p} to x_t is tested by rejecting the null hypothesis, $\mathrm{TE}(y_{t-p} \to x_t | \boldsymbol{x}_t^{(E,\tau)}) = 0$. CCM evaluates the predictive performance of cross mapping (i.e., a nonparametric regression model based on the nearest neighbor method) to test for causal influence. Under the assumption of Gaussian noise, the predictive performance $X(\boldsymbol{L}, \boldsymbol{P})$ is defined as follows [2]:

$$\begin{aligned}X(\boldsymbol{L}, \boldsymbol{P}) =& \sum_{t \in \boldsymbol{P}} p(y_{t-p}, x_t, \boldsymbol{x}_t^{(E,\tau)}) \log p(\boldsymbol{x}_{\boldsymbol{L}}, \boldsymbol{y}_{\boldsymbol{L}}) \\ &\times \int d\boldsymbol{w} p(y_{t-p} | \boldsymbol{y}_{\boldsymbol{L}}, \boldsymbol{w}) p(\boldsymbol{w} | x_t, \boldsymbol{x}_t^{(E,\tau)}, \boldsymbol{x}_{\boldsymbol{L}}),\end{aligned}$$

where $\boldsymbol{L}$ and $\boldsymbol{P}$ are the sets of time indices for state space reconstruction (i.e., model training) and prediction (i.e., model cross-validation), respectively. $\boldsymbol{x}_{\boldsymbol{L}} = \{x_t | t \in \boldsymbol{L}\}$ and $\boldsymbol{y}_{\boldsymbol{L}} = \{y_t | t \in \boldsymbol{L}\}$. $\boldsymbol{w}$ denotes the model parameters (e.g., projection weights). Note that $p(\boldsymbol{w} | x_t, x_t^{(E,\tau)}, \boldsymbol{x}_{\boldsymbol{L}})$ and $p(y_{t-p} | \boldsymbol{y}_{\boldsymbol{L}}, \boldsymbol{w})$ describe the process where nearest neighbors are searched from the effect variable and the process where the predictive performance of the cause variable is evaluated based on the searched nearest neighbors, respectively. Sugihara *et al.* [2] measured causal influence by the improvement in predictive performance with increasing training data (so-called "convergence") as follows:

$$\mathrm{CCM}(y_{t-p} \to x_t | \boldsymbol{x}_t^{(E,\tau)}) = X(\boldsymbol{L}, \boldsymbol{P}) - X(\boldsymbol{L}_{\min}, \boldsymbol{P}), \quad (2)$$

where $\boldsymbol{L}_{\min}$ is the set of time indices of training data at the minimum size (minimum library length) [16]. As the training data size approaches zero, we can prove that the expectation of causal influence defined by Eq. (2) is equivalent to that of the following mutual information (see Supplemental Material S1):

$$\begin{aligned}&\mathrm{CCM}^{\mathrm{IT}}(y_{t-p} \to x_t | \boldsymbol{x}_t^{(E,\tau)}) \\ =&\sum_t p(y_{t-p}, x_t, \boldsymbol{x}_t^{(E,\tau)}) \log \frac{p(y_{t-p} | x_t, \boldsymbol{x}_t^{(E,\tau)})}{p(y_{t-p})}, \quad (3)\end{aligned}$$

which is considered the information-theoretic definition of the CCM causal influence.

C. Unified Information-theoretic Causality

Using the information-theoretic definitions, here we clarify the theoretical connection between TE and CCM. We define the following novel measure of causal influence, hereafter referred to as unified information-theoretic causality (UIC):

$$\begin{aligned}&\mathrm{UIC}(y_{t-p} \to x_t | \boldsymbol{x}_t^{(E,\tau)}) \\ =&\sum_t p(y_{t-p}, x_t, \boldsymbol{x}_t^{(E,\tau)}) \log \frac{p(y_{t-p} | x_t, \boldsymbol{x}_t^{(E,\tau)})}{p(y_{t-p} | \boldsymbol{x}_t^{(E,\tau)})}. \quad (4)\end{aligned}$$

By the Bayes' rule, $p(y|x, z)p(x|z) = p(x|y, z)p(y|z)$, we can immediately prove the equivalence between TE and UIC from Eqs. (1) and (4). If there are enough data to accurately estimate the conditional probabilities, we will have the same value of causal influence for TE and UIC. It is also obvious from Eqs. (3) and (4) that CCM and UIC (or TE) are equivalent only if y_{t-p} and $\boldsymbol{x}_t^{(E,\tau)}$ are independent. In general, CCM tends to perform better for nonlinear systems because this independence can be more satisfied due to long-term unpredictability. CCM is more likely to detect false causality than TE [3, 14] because y_{t-p} may have causal influence on x_t via $\boldsymbol{x}_t^{(E,\tau)}$ (i.e., the previous states of x_t).

In CCM and UIC, $\boldsymbol{x}$ and $\boldsymbol{y}$ show complete separation as conditional and predicted variables for conditional probabilities. This variable separation is not trivial for nonparametric causality tests for two reasons. First, it relates to noise robustness because noise of conditional variables has a significant effect on the estimation of causal influence, particularly in nonlinear dynamical systems. As CCM and UIC do not include any cause variables (i.e., y_{t-p}) as conditional variables, they show greater robustness to noise than that of TE. Second, variable separation allows us to efficiently generate surrogate data (see Supplemental Material S3 for details), which are required to evaluate the statistical significance of causal influences [17]. These theoretical expectations are indeed confirmed in the following numerical experiments.

D. Numerical Experiments

Using artificial model systems, we compared the detection performance of three causality tests. Synthetic time-series were generated by numerical simulations from two dynamical systems with varying noise: a nonlinear logistic model and linear vector autoregression (VAR) model (Supplemental Material S5). The performance of causality tests was evaluated by the area under the receiver-operator characteristic curve (AUC). We found that UIC outperforms TE and CCM for both nonlinear and linear systems (Fig. 1). The performance of CCM was similar to that of UIC for the nonlinear system (Fig. 1a). However, for the linear system, with an increasing number of time points, the improvement in performance was

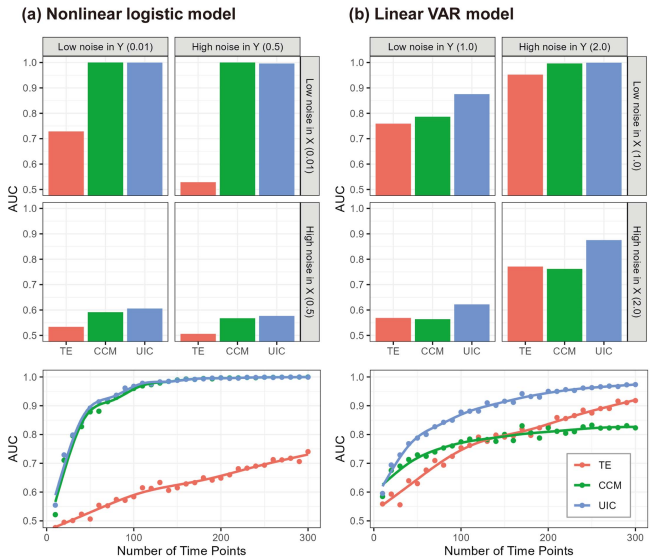

FIG. 1. Simulation experiments to compare the detection performance of transfer entropy (TE, red), convergent cross mapping (CCM, green), and unified information-theoretic causality (UIC, blue) in (a) nonlinear logistic model systems and (b) linear VAR model systems. In the upper panels, areas under the receiver-operator characteristic curve (AUC) are compared for different noise levels of effect (X) and cause (Y) variables. In the lower panels, we examine the relation between data requirements (i.e., the number of time points in time-series data) and AUC at the same noise level. Lines are obtained by fitting simulation results (points) by generalized additive models.

slower for CCM than for TE and UIC (Fig. 1b) because false positives do not decrease in CCM; even with an infinite number of observations, CCM did not necessarily detect true causality (i.e., a lack of statistical consistency). Consistent with its theoretical expectation, TE was quite sensitive to the noise levels of causal variables for the nonlinear dynamical system and thus requires a considerable number of time points for robust causality detection (Fig. 1a).

III. DEVELOPMENT OF CONDITIONAL CAUSALITY TESTS

In most large real-world systems, the causal effect between two variables can be detected by causality tests even when there are no direct relations between them. There are two possible scenarios for the misidentification of causal effects. In the first scenario, the causal effect is indirectly mediated by third variables; for example, when y_{t-2} affects z_{t-1}, which affects x_t, we find the causal effect of y_{t-2} on x_t. In the second scenario, target variables are strongly driven by external forces, which are never embedded by time-delayed variables. In this case, an external force z_{t-2} affects both y_{t-1} and x_t, resulting in the spurious detection of a causal effect of y_{t-1} on x_t. The misidentifications of causality are ameliorated by conducting the causality test conditional on third variables. In the following section, we introduce the conditional UIC test based on previous studies of TE [9].

A. Conditional UIC

The measure of causal influence conditional on third variables, $\boldsymbol{z} = \{z_t\}$, is defined as follows:

$$\begin{aligned}&\mathrm{UIC}(y_{t-p} \to x_t | \boldsymbol{x}_t^{(E,\tau)}, \boldsymbol{z}_t) \\ &= \sum_t p(y_{t-p}, x_t, \boldsymbol{x}_t^{(E,\tau)}, \boldsymbol{z}_t) \log \frac{p(y_{t-p}|x_t, \boldsymbol{x}_t^{(E,\tau)}, \boldsymbol{z}_t)}{p(y_{t-p}|\boldsymbol{x}_t^{(E,\tau)}, z_t)}.\end{aligned} \tag{5}$$

The difference from unconditional UIC defined by Eq. 4 is that the conditional probabilities include z_t as conditional variables. By conducting two causal tests:

$$\begin{aligned}&\mathrm{UIC}(y_{t-p} \to x_t | \boldsymbol{x}_t^{(E,\tau)}) > 0 \text{ and} \\ &\mathrm{UIC}(y_{t-p} \to x_t | \boldsymbol{x}_t^{(E,\tau)}, z_t) = 0,\end{aligned} \tag{6}$$

we can examine whether the causal influence of y_{t-p} on x_t is indirectly mediated by z_t or not.

B. Applications

First, we applied the conditional UIC test to a four-species food chain model to demonstrate the identification of direct and indirect effects [12]. In the food chain model, one species directly affects another species at time-lag 1 along a transitive causal chain (Fig. 2; Supplemental Material S5). Using conditional UIC tests expressed by conditions (6), we successfully detected the direct causal effects at time lag 1 (red circle in Fig. 2; $P < 0.05$ using a surrogate-based significance test). However, when using only unconditional UIC tests, we detected many indirect effects (blue circle in Fig. 2) in addition to direct effects.

Next, we applied the conditional UIC test to bacterial DNA time-series data collected from an experimental rice field [18]. These time-series data were obtained by quantifying DNA copy numbers for over 1000 taxa from irrigated water in experimental rice plots. In this application, we chose 20 abundant bacterial taxa with sufficient temporal fluctuations (see Supplemental Material S6 for more details). Conditional UIC detected 16 direct interactions and 28 indirect interactions among these bacterial taxa. To investigate the effects of identifying direct and indirect interactions on dynamical properties, we further estimated the signs and strengths of these interactions by S-map [19, 20]. Then, we evaluated the dynamic stability of the system [21] in two situations: (1) the system

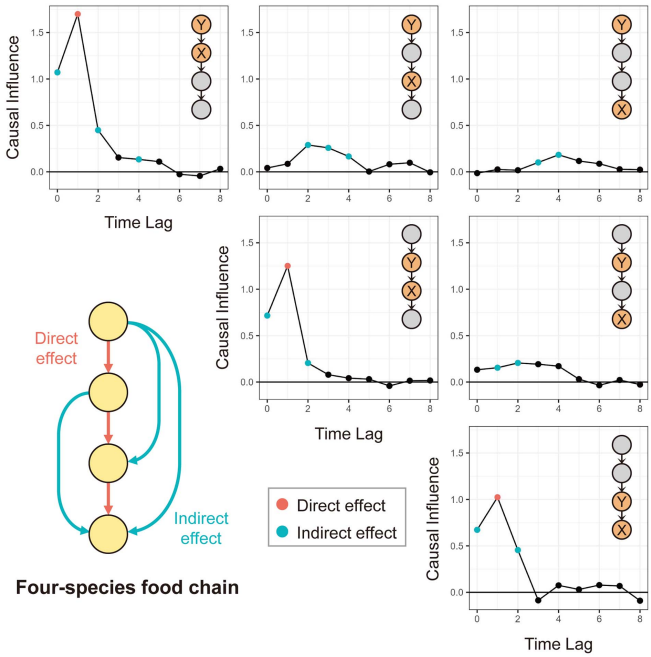

FIG. 2. Identification of direct and indirect effects based on conditional UIC in a four-species food chain model. In this model, species directly affect their prey at time lag 1. The causality tests were performed for species pairs with direct effects (diagonal panels) and with indirect effects (off-diagonal panels). For each species pair, species X and Y were used as effect and cause variables, respectively. Red and blue circles indicate direct and indirect effects detected by significance tests, while black circles indicate no causal effects. Causal influences (vertical axes) are estimated by unconditional UIC (also see Supplemental Material Fig. S1).

is incorrectly assumed to be driven by both direct and indirect interactions (Fig. 3a) and (2) the system is correctly assumed to be driven by only direct interactions (Fig. 3b). In this analysis, the sign was reversed for 4 of 16 direct interactions due to a failure to distinguish between indirect and direct interactions. These interactions had weak (i.e., near-zero) strengths, and their estimates may be influenced by other strong interactions. Using the conditional UIC, the system was identified as more stable because the reconstructed network eliminated suspicious indirect interactions (lower panels in Fig. 3), suggesting that the time-series analyses that do not distinguish between direct and indirect effects will misestimate dynamical properties such as system stability.

IV. DISCUSSION

In this paper, we develop a unified framework for nonparametric causal tests based on information-theoretic formulations and propose a novel causality test (i.e., UIC), establishing the theoretical connection between TE and CCM. Numerical experiments reveal that UIC shows higher robustness to noise in time-series data and have lower data requirements than TE and CCM (Fig. 1). Importantly, UIC outperforms TE and CCM in both linear and nonlinear dynamical systems. These results support the real-world applications of UIC because we do not know *a priori* whether TE or CCM is an appropriate causality test in weakly nonlinear systems.

In addition to its high performance, our method has two key advantages over conventional approaches. First, UIC has access to previous theoretical advances in both information theory [6, 10] and dynamical systems theory [22, 23]. As such, our framework facilitates the development of causality tests in an integrative way, with the help of both theories. Second, our method enables efficient computations of statistical significance. Specifically, we can reduce the computational requirements for searching nearest neighbors for state space reconstruction, which is the greatest computational challenge in nonparametric causality tests (Supplemental Material S3). Because recent studies have focused on large systems with thousands of components or more [18], efficient computation is important for real-world applications.

Interestingly, despite the mathematical equivalence of their definitions, TE and UIC show distinct statistical behavior. This arises from the following three characteristics of the nonparametric causality tests we studied: (1) the use of time-delayed variables for state space re-

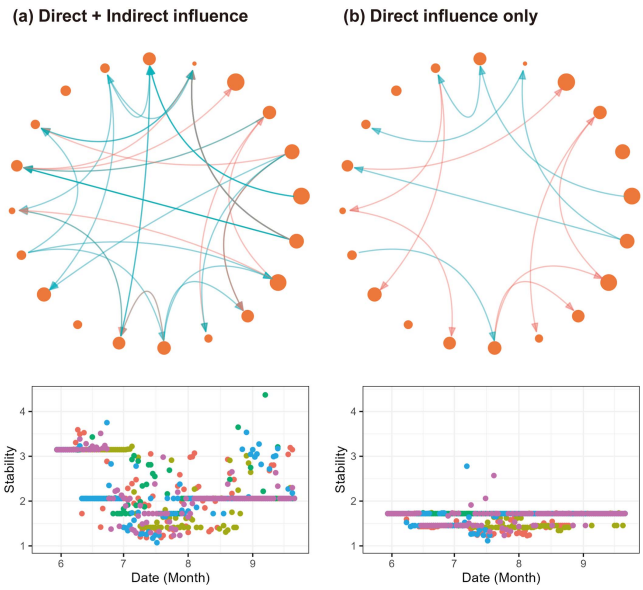

FIG. 3. Application of conditional UIC to empirical bacterial community data. (a) Unconditional UIC detects 44 interactions between 20 Proteobacteria taxa. Each node size indicates the abundance of each bacterial DNA. Blue and red arrows represent positive and negative interactions estimated by S-map, respectively. The dynamic stability estimated from these interactions exhibits large temporal fluctuation. Colors in the scatter plots indicate different experimental plots. (b) Conditional UIC finds only 16 direct interactions among 44 interactions. The dynamic stability indicates that the system is more stable when estimated from only direct interactions.

construction, (2) the need to search nearest neighbors to obtain nonparametric estimators, and (3) the need to separately calculate two (conditional) probabilities and their ratio to obtain causal influence. In contrast to (1) and (2), the last characteristic may depend on the algorithm used in nonparametric causality tests. For example, the Frenzel-Pompe algorithm [9] gives an identical value for causal influence for both TE and UIC. Although comparing computational algorithms is beyond the scope of our study, we stress the caveat that TE and UIC might have similar statistical behavior when implemented using other computational algorithms.

To understand complex systems in the real world, it is critical to identify interactions among components. For example, under anthropogenic stress, ecosystems sometimes exhibit unexpected regime shifts, including decreases in net primary production [24, 25] and mass bleaching of corals [26], due to transitions in dynamical structure emergent from interacting components [27]. With improved ecosystem monitoring techniques [28], nonparametric causality detection allows us to capture such critical transitions, providing a basis for preventing unexpected regime shifts and supporting ecosystem management to maintain biodiversity and ecosystem functions. Our method can improve the application of causality detection to real-world complex systems, thereby contributing to human well-being in a changing world.

ACKNOWLEDGMENTS

This study was supported financially by JSPS KAKENHI grants (nos. 18K14801, 21K15171, 19H05641 and 21H05312).

[1] P. Spirtes, C. N. Glymour, R. Scheines, and D. Heckerman, *Causation, prediction, and search* (MIT press, 2000).

[2] G. Sugihara, R. May, H. Ye, C.-h. Hsieh, E. Deyle, M. Fogarty, and S. Munch, Detecting causality in complex ecosystems, science **338**, 496 (2012).

[3] J. Runge, S. Bathiany, E. Bollt, G. Camps-Valls, D. Coumou, E. Deyle, C. Glymour, M. Kretschmer, M. D. Mahecha, J. Muñoz-Marí, *et al.*, Inferring causation from time series in earth system sciences, Nature communications **10**, 2553 (2019).

[4] J. Runge, P. Nowack, M. Kretschmer, S. Flaxman, and D. Sejdinovic, Detecting and quantifying causal associations in large nonlinear time series datasets, Science advances **5**, eaau4996 (2019).

[5] C. W. Granger, Investigating causal relations by econometric models and cross-spectral methods, Econometrica **37**, 424 (1969).

[6] K. Hlaváčková-Schindler, M. Paluš, M. Vejmelka, and J. Bhattacharya, Causality detection based on information-theoretic approaches in time series analysis, Physics Reports **441**, 1 (2007).

[7] T. Schreiber, Measuring information transfer, Physical review letters **85**, 461 (2000).

[8] L. Barnett, A. B. Barrett, and A. K. Seth, Granger causality and transfer entropy are equivalent for gaussian variables, Physical review letters **103**, 238701 (2009).

[9] S. Frenzel and B. Pompe, Partial mutual information for coupling analysis of multivariate time series, Physical review letters **99**, 204101 (2007).

[10] M. Oizumi, N. Tsuchiya, and S.-i. Amari, Unified framework for information integration based on information geometry, Proceedings of the National Academy of Sciences **113**, 14817 (2016).

[11] J. Runge, J. Heitzig, V. Petoukhov, and J. Kurths, Escaping the curse of dimensionality in estimating multivariate transfer entropy, Physical review letters **108**, 258701 (2012).

[12] H. Ye, E. R. Deyle, L. J. Gilarranz, and G. Sugihara, Distinguishing time-delayed causal interactions using convergent cross mapping, Scientific reports **5**, 1 (2015).

[13] F. Takens, Detecting strange attractors in turbulence, in *Dynamical Systems and Turbulence, Warwick 1980*, Vol. 898 (1981).

[14] T. Edinburgh, S. J. Eglen, and A. Ercole, Causality indices for bivariate time series data: A comparative review of performance, Chaos: An Interdisciplinary Journal of Nonlinear Science **31**, 083111 (2021).

[15] E. R. Deyle and G. Sugihara, Generalized theorems for nonlinear state space reconstruction, Plos one **6**, e18295 (2011).

[16] C.-W. Chang, M. Ushio, and C.-h. Hsieh, Empirical dynamic modeling for beginners, Ecological research **32**, 785 (2017).

[17] J. Runge, Conditional independence testing based on a nearest-neighbor estimator of conditional mutual information, in *International Conference on Artificial Intelligence and Statistics* (PMLR, 2018) pp. 938–947.

[18] M. Ushio, Interaction capacity as a potential driver of community diversity, Proceedings of the Royal Society B **289**, 20212690 (2022).

[19] G. Sugihara, Nonlinear forecasting for the classification of natural time series, Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences **348**, 477 (1994).

[20] E. R. Deyle, R. M. May, S. B. Munch, and G. Sugihara, Tracking and forecasting ecosystem interactions in real time, Proceedings of the Royal Society B: Biological Sciences **283**, 20152258 (2016).

[21] M. Ushio, C.-h. Hsieh, R. Masuda, E. R. Deyle, H. Ye, C.-W. Chang, G. Sugihara, and M. Kondoh, Fluctuating interaction network and time-varying stability of a natural fish community, Nature **554**, 360 (2018).

[22] H. Ye and G. Sugihara, Information leverage in interconnected ecosystems: Overcoming the curse of dimensionality, Science **353**, 922 (2016).

[23] H. Ma, S. Leng, K. Aihara, W. Lin, and L. Chen, Randomly distributed embedding making short-term high-dimensional data predictable, Proceedings of the National Academy of Sciences **115**, E9994 (2018).

[24] L. A. Gherardi and O. E. Sala, Enhanced precipitation variability decreases grass-and increases shrub-

productivity, Proceedings of the National Academy of Sciences **112**, 12735 (2015).

[25] L. J. Gilarranz, A. Narwani, D. Odermatt, R. Siber, and V. Dakos, Regime shifts, trends, and variability of lake productivity at a global scale, Proceedings of the National Academy of Sciences **119**, e2116413119 (2022).

[26] T. P. Hughes, K. D. Anderson, S. R. Connolly, S. F. Heron, J. T. Kerry, J. M. Lough, A. H. Baird, J. K. Baum, M. L. Berumen, T. C. Bridge, *et al.*, Spatial and temporal patterns of mass bleaching of corals in the anthropocene, Science **359**, 80 (2018).

[27] M. Scheffer, J. Bascompte, W. A. Brock, V. Brovkin, S. R. Carpenter, V. Dakos, H. Held, E. H. Van Nes, M. Rietkerk, and G. Sugihara, Early-warning signals for critical transitions, Nature **461**, 53 (2009).

[28] M. Besson, J. Alison, K. Bjerge, T. E. Gorochowski, T. T. Høye, T. Jucker, H. M. Mann, and C. F. Clements, Towards the fully automated monitoring of ecological communities, Ecology Letters **25**, 2753 (2022).