Information transfer from causal history in complex system dynamics

Peishi Jiang and Praveen Kumar

Phys. Rev. E **99**, 012306 — Published 3 January 2019

DOI: 10.1103/PhysRevE.99.012306

Information Transfer from Causal History in Complex System Dynamics

Peishi Jiang and Praveen Kumar*
Ven Te Chow Hydrosystem Laboratory, Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA
(Dated: December 17, 2018)

In a multivariate evolutionary system, the present state of a variable is a resultant outcome of all interacting variables through the temporal history of the system. How can we quantify the information transfer from the history of all variables to the outcome of a specific variable at a specific time? We develop information theoretic metrics to quantify the information transfer from the entire history, called causal history. Further, we partition this causal history into immediate causal history, as a function of lag τ from the recent time, to capture the influence of recent dynamics, and the complementary distant causal history. Further, each of these influences are decomposed into self and cross feedbacks. By employing a Markov property for directed acyclic time series graph, we reduce the dimensions of the proposed information-theoretic measure to facilitate an efficient estimation algorithm. This approach further reveals an information aggregation property, that is, the information from historical dynamics are accumulated at the preceding time directly influencing the present state of variable(s) of interest. These formulations allow us to analyze complex inter-dependencies in unprecedented ways. We illustrate our approach for: (1) characterizing memory dependency by analyzing a synthetic system with short memory; (2) distinguishing from traditional methods such as lagged mutual information using the Lorenz chaotic model; (3) comparing the memory dependencies of two long-memory processes with and without the strange attractor using the Lorenz model and a linear Ornstein-Uhlenbeck process; and (4) illustrating how dynamics in a complex system is sustained through the interactive contribution of self and cross dependencies in both immediate and distant causal histories, using the Lorenz model and observed stream chemistry data known to exhibit $1/f$ long-memory.

I. INTRODUCTION

The dynamics of natural systems, such as ecosystems and climate, arise as a result of spontaneous self-organization. Their dynamical characteristics, such as existence of strange attractors or $1/f$ long-memory dependencies, arise as a result of feedback between all interacting variables. Information theory offers compelling approaches for characterizing the complex non-linear inter-dependencies present in such systems [1]. For example, a recent study has argued that the spontaneous formation of a self-organized structure is reflected as decrease of joint entropy of the system as well as increase of contemporaneous inter-dependencies among interacting components [2]. However, most of the existing information-theoretic approaches are anchored on characterizing either bivariate information transfer using transfer entropy or momentary information transfer [3–7], or the interactions among a specific set of variables by using methods based on partial information decomposition [8–12], which becomes difficult when more than three variables are involved. These approaches provide important and insightful views associated with specific interactions within a system, but do not allow us to assess the entire range of information transfer among all variables. For example, we may ask how the interactions of several or all variables in a system determine the state of an individual variable at a specific time. Alternatively, we may ask how a finite time history of interactions results in an observed outcome of a specific variable at a specific time. To answer these questions, we require metrics that allow us to characterize full range of causal dependency in the system (in the Granger sense [13]), which structures the transfer of information that progressively influences a target variable.

Consider a system composed of N variables, $\vec{X}_t = \{X_t, Y_t, Z_t, ...\}_N$, varying in time. The current state of a variable, say $Z_t \in \vec{X}_t$, is a result of the evolutionary history of the system $\vec{X}_t^- = \{\vec{X}_{t-1}, \vec{X}_{t-2}, \vec{X}_{t-3}, ...\}$, which we call *causal history*. We partition this history, based on a partitioning time lag τ with respect to the present, into recent or *immediate causal history* $\{\vec{X}_{t-1}, \vec{X}_{t-2}, ..., \vec{X}_{t-\tau}\}$ and the complementary *distant causal history* $\{\vec{X}_{t-\tau-1}, \vec{X}_{t-\tau-2}, ...\}$. Generally, while the information from the immediate causal history is expected to be nondecreasing with τ, the degree and convergence of information from the distant causal history informs the influence from the remaining historical dynamics beyond the lag τ. Thus, quantification of the information transfer to a target variable at time t, from both its immediate and distant causal histories, would delineate the dependency of the variable on the prior dynamics as well as the memory in the system, which are keys for understanding various complex systems [14–17]. Therefore, the objective of this study is to quantify and characterize the influence of a immediate, distant, and/or entire causal history on Z_t by using an information-theoretic framework.

We use a directed acyclic time series graph approach

* kumar1@illinois.edu

to characterize the temporal dependencies of the system as well as for simplifying the computation of the information transfer. Specifically, we demonstrate the features of our approach in terms of: (1) Information aggregation property in the causal history, achieved through simplification from Markovian assumption in directed acyclic time series graph; (2) Discerning system memory, and its advantage over traditional methods such as lagged mutual information; (3) Characterizing the changing interaction information jointly provided by a target variable's self and cross dependencies, as a function τ, from both immediate and distant causal histories; and (4) Quantifying the change in memory dependency in a system when the influence of any particular variable is isolated from the remaining variables.

This paper is organized as follows. First, in Section II, we provide the definitions and the properties of the information transfer in both immediate and distant causal histories based on directed acyclic time series graph representation of the system. Then, in Section III we implement this approach to delineate the temporal dynamics of three different systems by quantifying the information transfer from causal history. We first identify the memory dependency of a trivariate logistic model – a short-memory system, in Section III A. Next in Section III B, we analyze the chaotic and long-memory Lorenz model for comparing the proposed approach with lagged mutual information in delineating the memory dependency of the system. Then, we investigate the information transfer in a linear trivariate Ornstein-Uhlenbeck process, whose dynamics also shows long memory property but without the existence of a stranger attractor. While the model-generated synthetic data are used for analysis in the previous three example, in the third example in Section III D, we demonstrate an application using observed stream chemistry time series data, obtained in the Upper Hafren catchment in Wales, United Kingdom [17]. Last, summary and conclusions are given in Section IV.

II. INFORMATION TRANSFER FROM CAUSAL HISTORY

We represent the temporal dependency in the multivariate system $\vec{X}_t$ as a time series directed acyclic graph [18, 19] as illustrated in Fig. 1, where each node represents a variable at a specific time step t (e.g., Z_t) and the parents of a target node or a set of nodes are denoted as $P_\bullet$ (e.g., P_{Z_t}). The directed edge linking two lagged nodes (e.g., $X_{t-\tau_X}$ and Z_t with $\tau_X > 0$) in the graph indicates the direct influence from $X_{t-\tau_X}$ to Z_t. The causal influence, assumed here in a Granger sense [13], from a lagged node $X_{t-\tau_X}$ to a target Z_t can be either through a directed edge or indirectly via a causal path $C_{X_{t-\tau_X}\to Z_t}$, which is a set of nodes connected by a sequence of directed edges from $X_{t-\tau_X}$ to Z_t. That is, $C_{X_{t-\tau_X}\to Z_t} \equiv \{V_{t-\tau_V} : V_t \in \vec{X}_t, \tau_V > 0, X_{t-\tau_X} \to \cdot \to V_{t-\tau_V} \to \cdot \to Z_{t-\tau_Z}\} \cup \{X_{t-\tau_X}\}$. We consider the causal influence to a target node as arising only from a node earlier in time, which results in a directed acyclic graph (DAG) of time series. In this section, based on this DAG time series graph representation, we provide the mathematical definition of causal history, its simplification for computation, the associated properties, and further analyses of causal history in terms of self and cross dependencies.

A. Definitions of Causal History

The causal history of a target node Z_t includes all the nodes that influence Z_t through causal paths in the graph, and is represented by $\vec{X}_t^- = \{\vec{X}_{t-1}, \vec{X}_{t-2}, ...\}$. Therefore, the total information, $\mathcal{T}$, to Z_t given by the causal history, can be expressed as the mutual information (MI) [20] between the two, which is given by:

$$\mathcal{T} = I(Z_t; \vec{X}_t^-). \tag{1}$$

Further, an immediate causal history of Z_t is considered as a finite length causal history immediately preceding time t, $\vec{X}_{t-\tau} = \{X_{t-\tau}, Y_{t-\tau}, ...\}_N$ starting from all the contemporaneous source nodes at lag τ. It is represented by a multitude of causal paths, that is, $C_{\vec{X}_{t-\tau}\Rightarrow Z_t} = \cup_{X_{t-\tau}\in\vec{X}_{t-\tau}} C_{X_{t-\tau}\to Z_t}$ (the blue dashed box in Fig. 1a). To generalize the following theory, we define the immediate causal history as a subgraph preceding Z_t arising from a set of lagged sources $\vec{V} = \{X_{t-\tau_X}, Y_{t-\tau_Y}, ...\}$ to Z_t, $C_{\vec{V}\Rightarrow Z_t} = \cup_{V_{t-\tau_V}\in\vec{V}} C_{V_{t-\tau_V}\to Z_t}$. Then, the complementary distant causal history can be naturally expressed as the remaining part of the causal history, $\vec{X}_t^- \backslash C_{\vec{V}\Rightarrow Z_t}$, where $\backslash$ is the subtraction operator (the red dashed box in Fig. 1a). By using the chain rule of MI [20], the total information $\mathcal{T}$ can be decomposed into the information from (1) the immediate causal history, $\mathcal{J}$, and (2) the distant causal history, $\mathcal{D}$, such that:

$$\begin{aligned}\mathcal{T} &= I(Z_t; C_{\vec{V}\Rightarrow Z_t}, \vec{X}_t^- \backslash C_{\vec{V}\Rightarrow Z_t}) \\ &= \underbrace{I(Z_t; \vec{X}_t^- \backslash C_{\vec{V}\Rightarrow Z_t})}_{=\mathcal{D}} + \underbrace{I(Z_t; C_{\vec{V}\Rightarrow Z_t} | \vec{X}_t^- \backslash C_{\vec{V}\Rightarrow Z_t})}_{=\mathcal{J}} \\ &= \mathcal{D} + \mathcal{J}.\end{aligned} \tag{2}$$

Eq.(2) expresses that the information from the distant causal history, $\mathcal{D}$, is provided by all the lagged nodes not in the immediate history, i.e., $\vec{X}_t^- \backslash C_{\vec{V}\Rightarrow Z_t}$, through their mutual information with Z_t; while the information from the recent dynamics, $\mathcal{J}$, is accounted for by the conditional mutual information (CMI) between the target and the immediate causal history conditioned on the distant history.

B. Simplifications of $\mathcal{T}$, $\mathcal{J}$, and $\mathcal{D}$

It is noted that the empirical computations of $\mathcal{T}$, $\mathcal{J}$, and $\mathcal{D}$ in Eq.(2) are infeasible due to the potentially in-

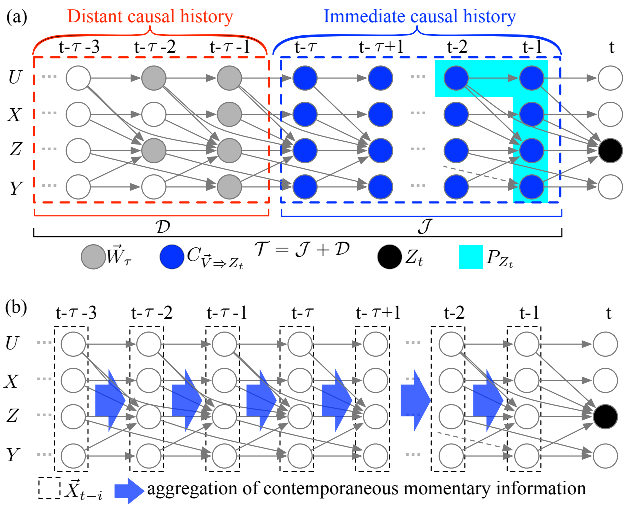

FIG. 1. (color online) Illustration of the causal history $\vec{X}_t^-$ of a target node Z_t. (a) The partition of $\vec{X}_t^-$ into an immediate causal history, $C_{\vec{V}\Rightarrow Z_t}$ (the dashed blue box), and the complementary distant causal history, $\vec{X}_t^-\backslash C_{\vec{V}\Rightarrow Z_t}$ (the dashed red box). The parents of the target Z_t [Eq.(3)], P_{Z_t}, are identified by the cyan colored box. (b) The aggregation of contemporaneous momentary information from each set of contemporaneous nodes $\vec{X}_{t-i}$ (the dashed hollow box) at an early time step $t-i$ in the causal history [Eq.(10)].

finite number of nodes in $\vec{X}_t^-$ and $\vec{X}_t^-\backslash C_{\vec{V}\Rightarrow Z_t}$. Therefore, to address this challenge and connect the time series graph with the underlying joint probability, we assume the Markov property for DAG ([21], theorem 1). This is consistent with prior work [5], which states that any node Z_t in the graph is independent of all its non-descendants given the knowledge of its parents P_{Z_t} [22]. For the graph in Fig. 1, for example, this implies that given its parents P_{Z_t} (the cyan colored box), the target node Z_t is conditionally independent of the rest of its non-descendants, $\vec{X}_t^-\backslash P_{Z_t}$.

Now, the main idea of reducing the dimensions in $\mathcal{T}$, $\mathcal{J}$ and $\mathcal{D}$ originates from the connection between conditional independence and the node separation in the graph based on the Markov property [5]. The simplification of $\mathcal{T}$ can be immediately achieved by using chain rule as follows (note that $P_{Z_t}\subset\vec{X}_t^-$):

$$\begin{aligned}\mathcal{T}&=I(Z_t;P_{Z_t},\vec{X}_t^-\backslash P_{Z_t})\\&=I(Z_t;P_{Z_t})+\underbrace{I(Z_t;\vec{X}_t^-\backslash P_{Z_t}\mid P_{Z_t})}_{=0}\\&=I(Z_t;P_{Z_t}),\end{aligned}\tag{3}$$

which is the mutual information between Z_t and its parents P_{Z_t} (see Fig. 1a). The zero value for $I(Z_t;\vec{X}_t^-\backslash P_{Z_t}\mid P_{Z_t})$ results from the Markov property that separates Z_t from the remaining historical nodes given its parents.

Furthermore, the distant causal history, $\vec{X}_t^-\backslash C_{\vec{V}\Rightarrow Z_t}$, which serves in Eq.(2) as the condition set and information contributor in $\mathcal{J}$ and $\mathcal{D}$, respectively, can be partitioned into two parts: (1) the parents of both Z_t and the immediate causal history $C_{\vec{V}\Rightarrow Z_t}$ excluding those in the immediate causal history, denoted as $\vec{W}_\tau=P_{C_{\vec{V}\Rightarrow Z_t}\cup Z_t}\backslash C_{\vec{V}\Rightarrow Z_t}$ (the grey nodes in Fig. 1a), and (2) the remaining nodes, $\vec{X}_t^-\backslash(C_{\vec{V}\Rightarrow Z_t}\cup\vec{W}_\tau)$. Then, in a similar manner as for $\mathcal{T}$, the Markov property and the chain rule also facilitate the simplifications for $\mathcal{D}$:

$$\begin{aligned}\mathcal{D}&=I(Z_t;\vec{W}_\tau,\vec{X}_t^-\backslash(C_{\vec{V}\Rightarrow Z_t}\cup\vec{W}_\tau))\\&=I(Z_t;\vec{W}_\tau)+\underbrace{I(Z_t;\vec{X}_t^-\backslash(C_{\vec{V}\Rightarrow Z_t}\cup\vec{W}_\tau)\mid\vec{W}_\tau)}_{=0}\\&=I(Z_t;\vec{W}_\tau),\end{aligned}\tag{4}$$

and for $\mathcal{J}$:

$$\begin{aligned}\mathcal{J}&=I(Z_t;C_{\vec{V}\Rightarrow Z_t}|\vec{X}_t^-\backslash C_{\vec{V}\Rightarrow Z_t})\\&=I(Z_t;C_{\vec{V}\Rightarrow Z_t}|\vec{W}_\tau).\end{aligned}\tag{5}$$

Both, the zero value for $I(Z_t;\vec{X}_t^-\backslash(C_{\vec{V}\Rightarrow Z_t}\cup\vec{W}_\tau)\mid\vec{W}_\tau)$ and the reduction of the condition set of $\mathcal{J}$ into $\vec{W}_\tau$ in Eqs.(4) and (5), respectively, are due to the conditional independence between Z_t and $\vec{X}_t^-\backslash(C_{\vec{V}\Rightarrow Z_t}\cup\vec{W}_\tau)$ given the knowledge of $\vec{W}_\tau$, which separates the immediate finite history associated with Z_t and Z_t itself from the remaining history. In fact, a decomposition of $C_{\vec{V}\Rightarrow Z_t}$, into (1) $P_{Z_t}^{C_{\vec{V}\Rightarrow Z_t}}\equiv P_{Z_t}\cap C_{\vec{V}\Rightarrow Z_t}$ – the direct causes of Z_t in the immediate causal history, and (2) $C_{\vec{V}\Rightarrow Z_t}\backslash P_{Z_t}^{C_{\vec{V}\Rightarrow Z_t}}$ – the remaining intermediate nodes in $C_{\vec{V}\Rightarrow Z_t}$, enables a further simplification of $\mathcal{J}$, that is (see Appendix A for derivations):

$$\mathcal{J}=I(Z_t;P_{Z_t}^{C_{\vec{V}\Rightarrow Z_t}}\mid\vec{W}_\tau),\tag{6}$$

which is achieved by taking the chain rule expansion based on $C_{\vec{V}\Rightarrow Z_t}$ and dropping off the other term because of the conditional independence of Z_t with the remaining history given its parents. Also, by substituting Eqs.(4) and (5) back to Eq.(2) and noticing $P_{Z_t}\subset P_{Z_t}^{C_{\vec{V}\Rightarrow Z_t}}\cup\vec{W}_\tau$, we can again utilize the Markov property to get:

$$\mathcal{T}=I(Z_t;P_{Z_t}^{C_{\vec{V}\Rightarrow Z_t}},\vec{W}_\tau)=I(Z_t;P_{Z_t}),$$

which reduces to Eq.(3) as we should expect and is constant in terms of the time lag τ. We also note that the quantities $\mathcal{J}$ and $\mathcal{D}$ are functions of τ, but this is not included in the notation for brevity as this does not cause any ambiguity.

C. Information Aggregation Property of $\mathcal{T}$ and $\mathcal{J}$

The simplifications in Eqs.(3)-(6) imply an important property of information aggregation from intermediate nodes to the direct causes of the node(s) of interest. For all the three information transfer measures, the information accumulate at the nodes that are either the parents of the target node Z_t [P_{Z_t} for $\mathcal{T}$ in Eq.(3) and $P_{Z_t}^{C_{\vec{V}\Rightarrow Z_t}}$ for $\mathcal{J}$ in Eq.(6)] or the parents of the union of Z_t and its immediate causal history [$\vec{W}_\tau$ for $\mathcal{D}$ in Eq.(4)]. This property, derived from the Markov property for DAG, illustrates that the latest observations actually contain all the information of the earlier dynamics in the system, transferred via the causal paths, and influence the states of the variables at the next stage.

Further insights associated with such information aggregation property can be obtained by a decomposition of both $\mathcal{T}$ and $\mathcal{J}$. We separate $C_{\vec{V}\Rightarrow Z_t}$ into τ set of nodes, where τ is the maximum time lag between the target Z_t and the earliest node in the source nodes $\vec{V}$, that is, $\tau = \arg\max_k\{X_{t-k} : X_{t-k} \in C_{\vec{V}\Rightarrow Z_t}\}$. Each set of nodes $\vec{V}_{t-i}$ represents all the contemporaneous nodes in $C_{\vec{V}\Rightarrow Z_t}$ at the time step $t-i$ ($1 \le i \le \tau$), that is, $\vec{V}_{t-i} = \{V_{t-\tau_V} : V_{t-\tau_V} \in C_{\vec{V}\Rightarrow Z_t} \mid \tau_V = i\}$. It is clear that $C_{\vec{V}\Rightarrow Z_t} = \cup_{i=1}^{\tau}\vec{V}_{t-i}$ and $\vec{V}_{t-i_1} \cap \vec{V}_{t-i_2} = \emptyset$ for $i_1 \neq i_2$. Therefore, we can express $\mathcal{J}$ in Eq.(5) as:

$$\mathcal{J} = I(Z_t; \vec{V}_{t-1}, ..., \vec{V}_{t-\tau} \mid \vec{W}_\tau),$$

and by using the chain rule for conditional mutual information [23], we get:

$$\mathcal{J} = \sum_{i=1}^{\tau} I(Z_t; \vec{V}_{t-i} \mid \vec{W}_\tau, \vec{V}_{t-i-1}, ..., \vec{V}_{t-\tau}). \tag{7}$$

Note that $\{\vec{V}_{t-i-1}, ..., \vec{V}_{t-\tau}\}$ are actually the remaining parents of both Z_t and the subgraph $C_{\vec{V}_{t-i}\Rightarrow Z_t}$ initiated by $\vec{V}_{t-i}$, which are not in $\vec{W}_\tau$. Therefore, the condition set in Eq.(7), $\{\vec{W}_\tau, \vec{V}_{t-i-1}, ..., \vec{V}_{t-\tau}\}$, in fact contains the parents of the union of Z_t and $C_{\vec{V}_{t-i}\Rightarrow Z_t}$, or $P_{C_{\vec{V}_{t-i}\Rightarrow Z_t}\cup Z_t}$. Also, due to the Markov property of the time series DAG, $P_{C_{\vec{V}_{t-i}\Rightarrow Z_t}\cup Z_t}$ separates $C_{\vec{V}_{t-i}\Rightarrow Z_t} \cup Z_t$ from their non-descendants, including the remaining nodes in the conditions in Eq.(7), and thus gives:

$$\begin{aligned}\mathcal{G}_i &\equiv I(Z_t; \vec{V}_{t-i} \mid \vec{W}_\tau, \vec{V}_{t-i-1}, ..., \vec{V}_{t-\tau}) \\ &= I(Z_t; \vec{V}_{t-i} \mid P_{C_{\vec{V}_{t-i}\Rightarrow Z_t}\cup Z_t} \backslash C_{\vec{V}_{t-i}\Rightarrow Z_t})\end{aligned} \tag{8}$$

where $\mathcal{G}_i$ is the generalized version of the momentary information transfer along causal paths [12, 18] from multiple source nodes $\vec{V}_{t-i}$ to Z_t along the multiple causal paths $C_{\vec{V}_{t-i}\Rightarrow Z_t}$. It quantifies the uncertainty reduction in Z_t due to $\vec{V}_{t-i}$ conditioned on the parents of both Z_t and $C_{\vec{V}_{t-i}\Rightarrow Z_t} \cup Z_t$,

Correspondingly, Eq.(7) can thus be simplified as:

$$\mathcal{J} = \sum_{i=1}^{\tau} \mathcal{G}_i = \sum_{i=1}^{\tau} I(Z_t; \vec{V}_{t-i} \mid P_{C_{\vec{V}_{t-i}\Rightarrow Z_t}\cup Z_t} \backslash C_{\vec{V}_{t-i}\Rightarrow Z_t}). \tag{9}$$

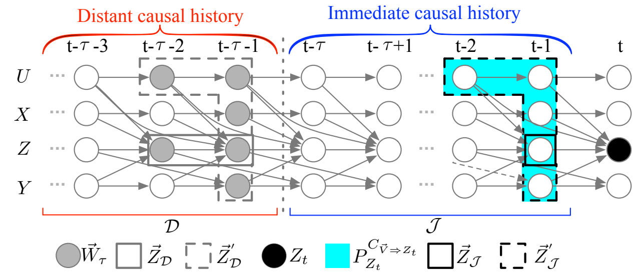

FIG. 2. (color online) Illustration of the self and cross dependencies in both simplified immediate and distant causal histories for a target Z_t (the black node). The self-dependencies, $\vec{Z}_{\mathcal{J}}$, and the complementary part, $\vec{Z}'_{\mathcal{J}}$, in the simplified immediate causal history, $P_{Z_t}^{C_{\vec{V}\Rightarrow Z_t}}$, are identified in solid and dashed black boxes, respectively. The self-dependencies, $\vec{Z}_{\mathcal{D}}$, and the complementary part, $\vec{Z}'_{\mathcal{D}}$, in the simplified distant causal history, $\vec{W}_\tau$, are identified in solid and dashed grey boxes, respectively.

This equation elucidates that the information given by a sequence of dynamics preceding Z_t, i.e., its immediate causal history, is an accumulation of the momentary information transfer from the contemporaneous dynamics at each time step involved in this finite history.

Such accumulation of momentary information can be generalized to the total information $\mathcal{T}$ if the source nodes $\vec{V}$ of the immediate causal history are taken as all the variables at an infinite past, $\vec{X}_{t-\tau} = \{V_{t-\tau}, X_{t-\tau}, Y_{t-\tau}, Z_{t-\tau}, ...\}$, with $\tau \to \infty$. In this case, the immediate causal history is naturally the whole causal history itself, and thus $\mathcal{J} = \mathcal{T}$, which based on Eq.(9) gives:

$$\mathcal{T} = \lim_{\tau\to\infty} \sum_{i=1}^{\tau} I(Z_t; \vec{X}_{t-i} \mid P_{C_{\vec{X}_{t-i}\Rightarrow Z_t}\cup Z_t} \backslash C_{\vec{V}_{t-i}\Rightarrow Z_t}). \tag{10}$$

By relating the above equation with Eq.(2), again we see that the momentary information from all the previous intermediate nodes in the causal history are accumulated at the nodes that directly influence the target Z_t, as shown in Fig. 1b. Note that, a measure similar to Eqs.(7)-(10) is proposed in [5], called the decomposed transfer entropy. It approximates the information coming from all the historical states of a source variable $\vec{X}_t^-$ as the summation of individual conditional mutual information from each lagged $X_{t-\tau}$ in a finite set of $\vec{X}_t^-$. This is different from the information aggregation of $\mathcal{J}$ and $\mathcal{T}$ proposed here in that Eqs.(9) and (10) approximate the information from the historical states of multiple source variables to the target.

D. Interactions from Self-Feedbacks in $\mathcal{J}$ and $\mathcal{D}$

To further dissect the information transfer we characterize the interaction information arising from self and cross dependencies of a target variable Z_t in both immediate and distant causal histories. Note that interaction information between two sets of source nodes $\vec{A}$ and $\vec{B}$ contributing information to Z_t is given as:

$$\begin{aligned}\mathcal{I} &= I(Z_t;\vec{A}|\vec{B}) - I(Z_t;\vec{A}) \\ &= I(Z_t;\vec{A},\vec{B}) - [I(Z_t;\vec{A}) + I(Z_t;\vec{B})].\end{aligned} \quad (11)$$

For distant causal history, represented by $\vec{W}_\tau$, the two decomposed parts include: (1) a self-feedback component of Z_t, $\vec{Z}_\mathcal{D} \equiv \{V_{t-\tau} \in \vec{W}_\tau \mid V = Z\}$ (the grey box in Fig. 2); and (2) the complementary component, $\vec{Z}'_\mathcal{D} \equiv \vec{W}_\tau \backslash \vec{Z}_\mathcal{D}$ (the dashed grey box in Fig. 2). The difference between $\mathcal{D}$ and the summation of the mutual information between Z_t and each of the two components in $\vec{W}_\tau$ then accounts for an interaction information, $\mathcal{I}_\mathcal{D}$, which is given by:

$$\mathcal{I}_\mathcal{D} = \mathcal{D} - [I(Z_t;\vec{Z}_\mathcal{D}) + I(Z_t;\vec{Z}'_\mathcal{D})]. \quad (12)$$

$\mathcal{I}_\mathcal{D}$ quantifies the interaction information in Eq.(11) transferred to the target Z_t from its self-dependency, $\vec{Z}_\mathcal{D}$, as well as the complementary component, $\vec{Z}'_\mathcal{D}$,in distant history. A negative $\mathcal{I}_\mathcal{D}$ [i.e., $\mathcal{D} < I(Z_t;\vec{Z}_\mathcal{D}) + I(Z_t;\vec{Z}'_\mathcal{D})$] shows a net redundancy in the interaction between the two components, while a positive $\mathcal{I}_\mathcal{D}$ [i.e., $\mathcal{D} > I(Z_t;\vec{Z}_\mathcal{D}) + I(Z_t;\vec{Z}'_\mathcal{D})$] illustrates a net synergistic influence on the target.

Similarly, the simplified immediate causal history of Z_t, represented by $P_{Z_t}^{C_{\vec{V}\Rightarrow Z_t}}$, can be partitioned into (1) a component containing the self-dependence of the target, $\vec{Z}_\mathcal{J} \equiv \{V_{t-\tau} \in P_{Z_t}^{C_{\vec{V}\Rightarrow Z_t}} \mid V = Z\}$ (the black box in Fig. 2); and (2) the complementary part, $\vec{Z}'_\mathcal{J} \equiv P_{Z_t}^{C_{\vec{V}\Rightarrow Z_t}} \backslash \vec{Z}_\mathcal{J}$ (the dashed black box in Fig. 2). The corresponding interaction information from the two parts of immediate causal history, $\mathcal{I}_\mathcal{J}$, can be computed as:

$$\mathcal{I}_\mathcal{J} = \mathcal{J} - [I(Z_t;\vec{Z}_\mathcal{J} \mid \vec{W}_\tau) + I(Z_t;\vec{Z}'_\mathcal{J} \mid \vec{W}_\tau)], \quad (13)$$

quantifying the conditional interaction information to Z_t from its self and cross dependencies in the immediate causal history.

We also note that in [18], the interaction information is used for investigating how the influence from a source node $X_{t-\tau}$ to Z_t is intervened by the immediate nodes in the causal path $C_{X_{t-\tau}\to Z_t}$. In this study, we evaluate the interaction effects on Z_t from immediate and distant causal histories in terms of: first, Z_t's own history, and second, historical states of the other variables.

III. APPLICATIONS

To illustrate the capability of the approach described above for delineating the temporal dependency of a system, we quantify the information transfer from the causal history in three different systems. We first characterize the temporal dependency of a short-memory system through a trivariate logistic model. Then, we illustrate how the proposed approach is different from lagged mutual information in addressing system's memory dependency using an example of a chaotic system – the Lorenz model. Further, we compare the Lorenz model with a trivariate Ornstein-Uhlenbeck process to investigate how the information transfer differs in processes with and without strange attractor. Finally, we quantify the memory dependency from time series observations, representing catchment chemistry, which is known to have long-term dependency. Especially, by decomposing the distant history into the self-feedback of the target and the complementary component characterizing information transfer from other interacting variables, we observe the redundancy-dominated $\mathcal{J}$, as well as consistent nonzero and synergy-dominated $\mathcal{D}$ in both the Lorenz model and the stream chemistry system, which we conjecture as sustaining chaotic and fractal features of the two systems.

A. Trivariate Logistic System: a Short-memory System

In the following, we empirically analyze the information transfer in the causal history of a nonlinear model-generated synthetic data. Consider a trivariate coupled logistic system, mathematically expressed as:

$$X_{i,t} = \frac{1-\epsilon}{3}\sum_{j=1}^{3} 4X_{j,t-1}(1-X_{j,t-1}) + \epsilon\eta_t^{X_i}, i \in \{1,2,3\} \quad (14)$$

where $\eta_t^{X_i} \in [0,1]$ is a uniform noise term and $0 < \epsilon < 1$ is its coupling strength. To investigate the total information and its two components to the target node $X_{3,t}$, we consider the immediate causal history as the causal subgraph $C_{\{X_{1,t-\tau},X_{2,t-\tau},X_{3,t-\tau}\}\Rightarrow X_{3,t}}$ starting at an earlier time step $t-\tau$ ($\tau \geq 1$) (see Fig. 3a). $\mathcal{J}$, $\mathcal{D}$ and $\mathcal{T}$ are calculated for τ ranging from 1 to 50 and $\epsilon \in [0.1, 0.2, 0.3, 0.5, 0.8]$. For each pair of τ and ϵ, 10,000 data points are generated to conduct the empirical estimations, with an ensemble of 10 runs for each to get an average behavior. To avoid the infinite dimensions in Eq.(2) in the computation, we compute $\mathcal{T}$, $\mathcal{D}$ and $\mathcal{J}$ based on Eqs.(3), (4), and (6), respectively. The k-nearest-neighbor (kNN) estimator [4, 24] is adopted for the estimation of $\mathcal{J}$, $\mathcal{T}$ and $\mathcal{D}$ with $k = 5$ (low k facilitates a low bias of the estimated MI and CMI [4]). In the next two applications, the computation of $\mathcal{T}$, $\mathcal{D}$, and $\mathcal{J}$ are also conducted in the same manner.

The contribution of immediate causal history $\mathcal{J}$, and the proportion of distant causal history, $\mathcal{D}$, in the total information transfer $\mathcal{T}$, $\mathcal{D}/\mathcal{T}$, are shown in Fig. 3b. We observe that for the range of noise coupling strengths ϵ,

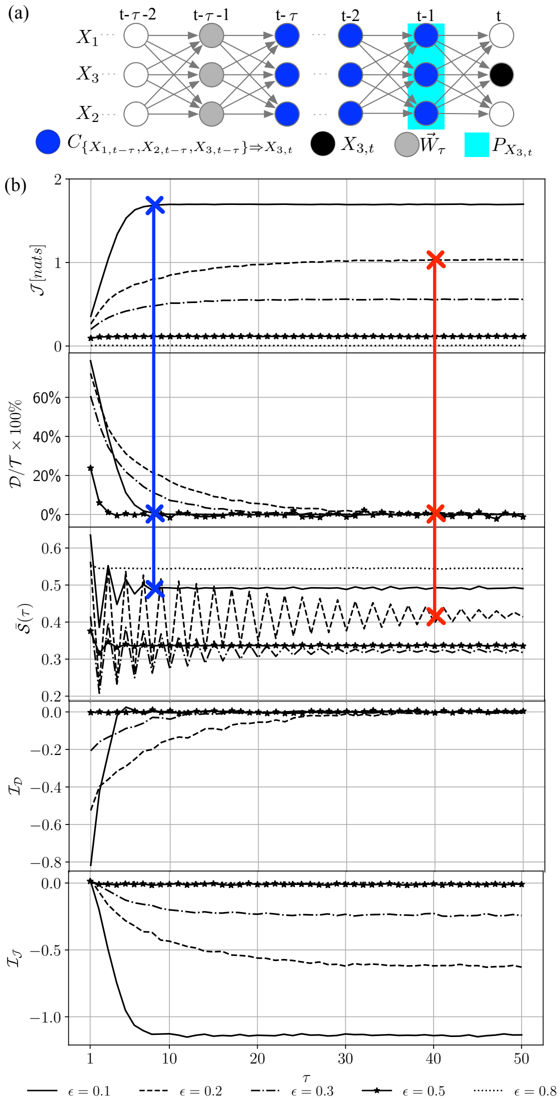

FIG. 3. (color online) Illustration of the trivariate coupled logistic model. (a) The times series graph of the system with the causal subgraph $C_{\{X_{1,t-\tau},X_{2,t-\tau},X_{3,t-\tau}\}\Rightarrow X_{3,t}}$ as the immediate causal history (the representations of the nodes are the same as in Fig. 1a). (b) Plots of $\mathcal{J}$, $\mathcal{D}/\mathcal{T}$, $\bar{\mathcal{S}}$, $\mathcal{I}_\mathcal{D}$ and $\mathcal{I}_\mathcal{J}$ for τ ranging from 1 to 50 with $\epsilon \in [0.1, 0.2, 0.3, 0.5, 0.8]$ (blue and red crosses, connected through a vertical line, represent the convergence points of $\mathcal{J}$, $\mathcal{D}/\mathcal{T}$, and $\bar{\mathcal{S}}$ for $\epsilon = 0.1$ and $\epsilon = 0.2$, respectively; note that results for $\epsilon = 0.8$ are not plotted (except $\mathcal{J}$) due to its high variability resulting from a near-zero total information $\mathcal{T}$).

$\mathcal{J}$ and $\mathcal{D}/\mathcal{T}$ increases and decreases, respectively, with lag τ, and $\mathcal{D}/\mathcal{T}$ achieves asymptotic convergence to zero when the lag is sufficiently large. In particular, the convergence to zero of $\mathcal{D}/\mathcal{T}$ illustrates that this trivariate coupled logistic model has a short memory for influencing the target. Further, the decrease of $\mathcal{J}$ with increasing coupling strength ϵ implies that a strong noise can reduce the information transfer from the preceding finite length period and, thus, also reduce the total information in this short-memory system.

Also, it is noted that the curves in $\mathcal{D}/\mathcal{T}$ decrease with increasing τ but intersect for different values of ϵ. This is because of different interactions and synchronization of coupled logistic maps as a function of ϵ [25–27]. Therefore, we compute the lag synchronization for each pair of lagged variables $X_{i,t-\tau}$ and $X_{j,t}$ $(i, j \in \{1, 2, 3\})$ with τ ranging from 1 to 50, which is given by:

$$\mathcal{S}_{ij}(\tau) = \left\{ \frac{E[(X_{i,t-\tau} - X_{j,t})^2]}{[E(X^2_{i,t-\tau})E(X^2_{j,t})]^{1/2}} \right\}^{0.5}, \tag{15}$$

where E is the expectation function. Since the dynamics is highly symmetric in terms of $\{X_1, X_2, X_3\}$ for this trivariate model, we compute the averaged lag synchronization $\bar{\mathcal{S}}(\tau)$ as:

$$\bar{\mathcal{S}}(\tau) = \frac{\sum_{i,j} \mathcal{S}_{ij}(\tau)}{9}, \tag{16}$$

which is sketched in the middle plot of Fig. 3b. It shows that for each noise coupling strength ϵ, $\bar{\mathcal{S}}$ oscillates for small τ, and then the amplitude decreases and $\bar{\mathcal{S}}$ eventually converges with increasing τ, implying a consistent similarity structure between each pair of lagged variables given an ϵ. The convergence of the averaged lag synchronization, $\bar{\mathcal{S}}$, implies that the similarity between a target $X_{j,t}$ and a lagged history node $X_{i,t-\tau}$ gradually becomes invariant with increasing τ. It is consistent with the convergences of both $\mathcal{J}$ and $\mathcal{D}/\mathcal{T}$ for each ϵ, which are illustrated for $\epsilon = 0.1$ and $\epsilon = 0.2$ in blue and red crosses, respectively.

Further, the interaction information $\mathcal{I}_\mathcal{D}$ and $\mathcal{I}_\mathcal{J}$ increases and decreases with time lag τ, and then converges to zero and a negative value, respectively. The rapid convergence to the asymptotic values suggests no synergy or redundancy for this short-memory model. Meanwhile, the drop of $\mathcal{I}_\mathcal{J}$ with increasing τ means the contributions from self and cross dependencies in the immediate causal history share a higher redundancy.

B. The Lorenz Model: a Comparison with Lagged Mutual Information

Now, we perform the analysis of the Lorenz model to investigate the difference between the proposed measures of causal history and traditional methods such as lagged mutual information in capturing the temporal dependency of a system, as well as to understand the potential interdependencies embedded in its chaotic behavior. The Lorenz model is prototypical of its chaotic behavior [28], that is, its dynamics are contained in a strange attractor with a fractal dimension between 2 and 3, and its governing equation is given by a system of three variables

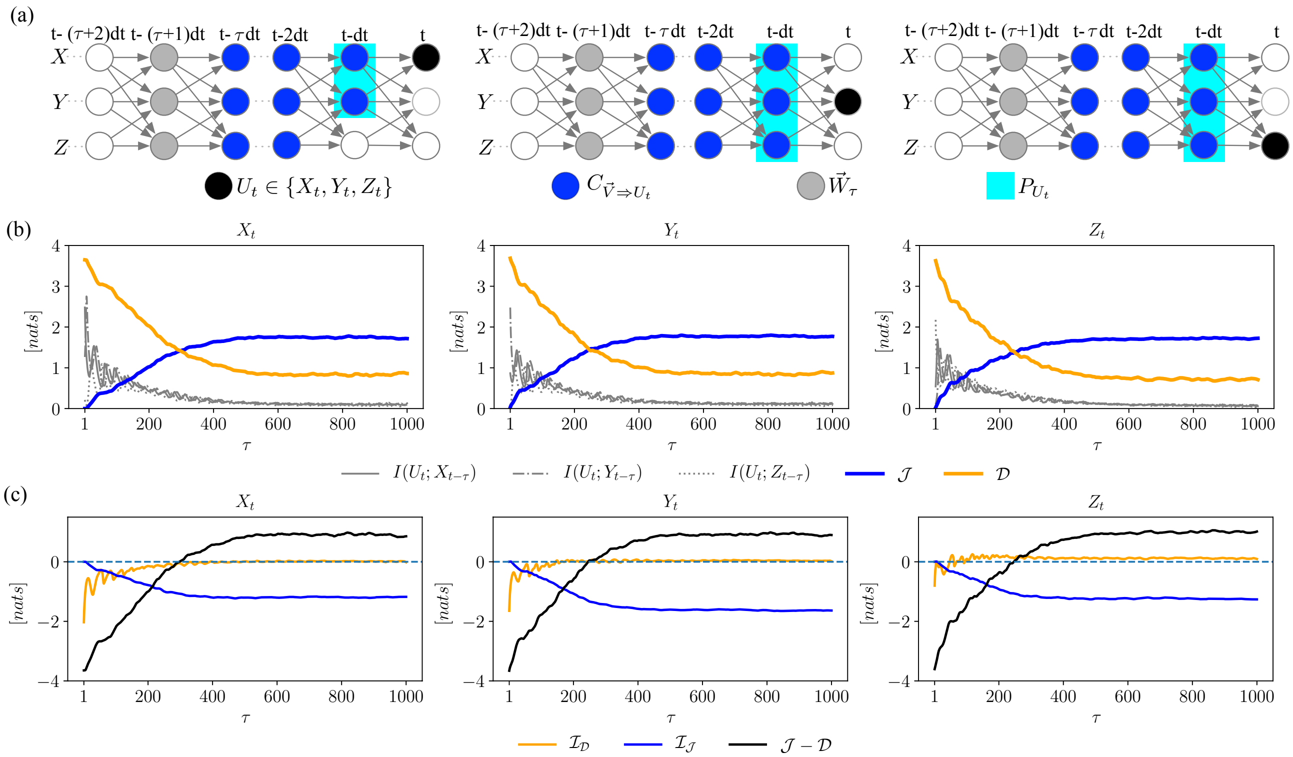

FIG. 4. (color online) Illustration of the Lorenz model with parameters $\sigma = 10$, $\rho = 28$ and $\beta = 8/3$. (a) The times series graph of the system with the causal subgraph $C_{\{X_{t-\tau}, Y_{t-\tau}, Z_{t-\tau}\} \Rightarrow U_t}$ ($U \in \{X, Y, Z\}$) as the immediate causal history. (b) The corresponding plots of the lagged mutual information, $\mathcal{J}$, and $\mathcal{D}$ for the time lag τ ranging from 1 to 1000. (c) The corresponding plots of $\mathcal{I}_\mathcal{D}$, $\mathcal{I}_\mathcal{J}$, and $\mathcal{J} - \mathcal{D}$ for the time lag τ ranging from 1 to 1000.

X_t, Y_t and Z_t as:

$$\frac{dX_t}{dt} = \sigma(Y_t - X_t) \tag{17a}$$

$$\frac{dY_t}{dt} = X_t(\rho - Z_t) - Y_t \tag{17b}$$

$$\frac{dZ_t}{dt} = X_t Y_t - \beta Z_t, \tag{17c}$$

where the parameters σ, ρ and β in this study are set as 10, 28, and 8/3, respectively.

To analyze the information dynamics in the system as well as the resulting long-term dependence, we empirically quantify the influence on a target $U_t \in \{X_t, Y_t, Z_t\}$ based on (1) the lagged mutual information between each pair of variables $I(U_t; V_{t-\tau dt})$, where $V_t \in \{X_t, Y_t, Z_t\}$, and τ and dt are the lag step and the time interval, respectively; (2) the information transfer from the immediate and distant causal histories for each variable, $\mathcal{J}$ and $\mathcal{D}$, respectively; and (3) the interaction information contributed by a self-feedback and the corresponding complementary components in both distant and immediate causal history, $\mathcal{I}_\mathcal{D}$ and $\mathcal{I}_\mathcal{J}$, as indicated in Eqs.(12) and (13), respectively. The immediate causal history is now the subgraph $C_{\{X_{t-\tau dt}, Y_{t-\tau dt}, Z_{t-\tau dt}\} \Rightarrow U_t}$ (see Fig. 4a), from which we can observe that given a time lag τdt the representative distant causal history $\vec{W}_\tau = \{X_{t-(\tau+1)dt}, Y_{t-(\tau+1)dt}, Z_{t-(\tau+1)dt}\}$. The measures are calculated for τ ranging from 1 to 1000 with the time interval $dt = 0.01$. 10,000 data points are generated to conduct the empirical estimations, with an ensemble of 10 runs to get an average behavior.

The results of the lagged mutual information, $\mathcal{D}$, and $\mathcal{J}$ are shown in Fig. 4b. The quantities $\mathcal{J}$ and $\mathcal{D}$ increases and decreases, respectively, with increasing τ, converging to some nonzero values when τ is around 500. The consistent nonzero $\mathcal{D}$ for large τ arises from the fact that the Lorenz system is a long-memory process such that information provided from the distant history informs the present dynamics. Meanwhile, the lagged mutual information, $I(U_t; V_{t-\tau dt})$, for all the three variables shows strong oscillations and gradually decays to zero. The oscillations are due to the chaotic behavior where the 'butterfly' trajectory of the strange attractor in this phase space determines the frequency of these oscillations, and the slow decay to zero reflects the long term dependency. However, the lagged mutual information does not show the consistent information contributed from the past as $\mathcal{D}$ does. Therefore, the proposed information transfer from the causal history provides a view for analyzing the memory dependency of the system that is complementary to traditional methods such as lagged mutual information.

Furthermore, the difference between $\mathcal{J}$ and $\mathcal{D}$ as well

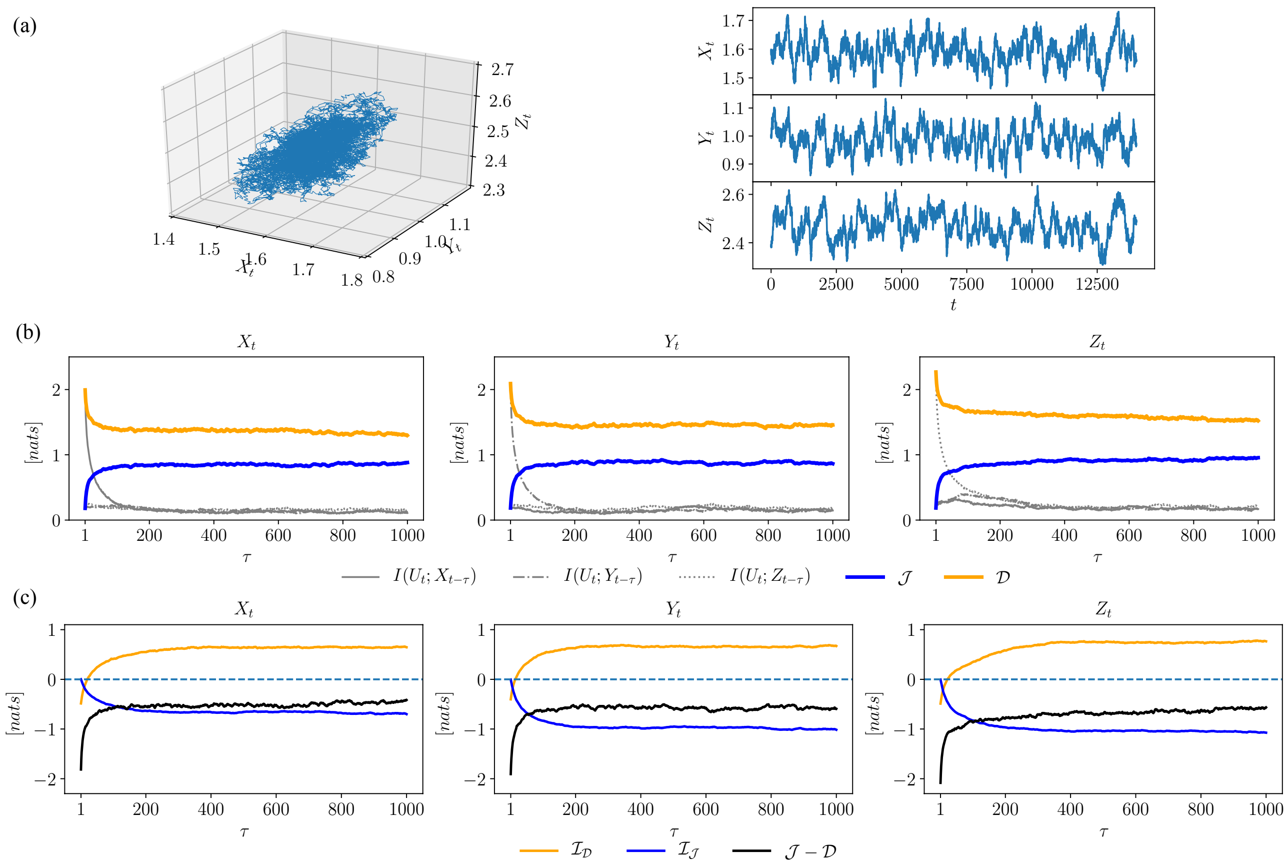

FIG. 5. (color online) Illustration of the Ornstein-Uhlenbeck process in Eq.(18). (a) The trajectories of the process (left) and the time series of each variable (right). (b) The corresponding plots of the lagged mutual information, $\mathcal{J}$, and $\mathcal{D}$ for the time lag τ ranging from 1 to 1000. (c) The corresponding plots of $\mathcal{I}_{\mathcal{D}}$, $\mathcal{I}_{\mathcal{J}}$, and $\mathcal{J} - \mathcal{D}$ for the time lag τ ranging from 1 to 1000.

as their interaction information $\mathcal{I}_{\mathcal{J}}$ and $\mathcal{I}_{\mathcal{D}}$, shown in Fig. 4(c), illustrate different roles of the immediate and distant causal histories in shaping the target. First, the recent dynamics of the Lorenz model has a stronger influence on the target than the remaining earlier dynamics as time lag τ becomes larger than around 200. This is evidenced by the convergence of $\mathcal{J} - \mathcal{D}$ to a positive value (the black thick line). Also, the convergence of $\mathcal{I}_{\mathcal{J}}$ to a negative value (the blue thick line) implies a higher redundancy effect from the interaction information of cross and self dependencies in the immediate causal history, as observed in the trivariate chaotic map. Meanwhile, the convergence of $\mathcal{I}_{\mathcal{D}}$ to zero (the orange thick line) suggests a balanced contribution from synergistic and redundant effects, each of which are not necessarily zero in the Lorenz model due to the nonzero convergence of $\mathcal{D}$ plotted in Fig.4(b). In short, the Lorenz model with a strange attractor shows each variable is affected by (1) a strong influence given by immediate causal history with dominant redundant effects from the self and cross dependencies, and (2) less influence from distant causal history with balanced redundancy and synergistic effects.

C. The Ornstein-Uhlenbeck process: a Long-Memory Process without Strange Attractor

To investigate the difference between processes with and without strange attractors in terms of the information transfer from causal history, we now conduct the analysis on a trivariate linear Ornstein-Uhlenbeck (OU) process with long-term dependency. The OU process is chosen such that the model has the same structure of the directed acyclic time series graph as the Lorenz model shown in Fig. 4(a) and it is stationary, which is given by:

$$\frac{dX_t}{dt} = -0.5X_t + 0.3Y_t + \zeta_X \tag{18a}$$

$$\frac{dY_t}{dt} = 0.4X_t - 0.4Y_t - 0.3Z_t + \zeta_Y \tag{18b}$$

$$\frac{dZ_t}{dt} = 0.4X_t + 0.6Y_t - 0.7Z_t + \zeta_Z, \tag{18c}$$

where ζ_X, ζ_Y and ζ_Z are independently and identically distributed noise terms following standard normal distribution. As in the analysis of the Lorenz model, we quantify the influence on each variable in the OU process in terms of lagged mutual information, the information from immediate and distant causal history $\mathcal{J}$ and $\mathcal{D}$, and their interaction information $\mathcal{I}_{\mathcal{J}}$ and $\mathcal{I}_{\mathcal{D}}$. The computation settings of the above information-theoretic measures are the same as the Lorenz model. The trajectory and the time series of each variable of the OU process are plotted in Fig. 5(a) with time interval $dt = 0.01$ and 10,000 simulated data points, visually showing that the dynamics are confined in a three-dimensional confined domain which is not a strange attractor.

The long-memory property of the OU process in Eq.(18) is evidenced in the non-zero convergence of $\mathcal{D}$ and a slow-decay of the auto mutual information of each variable in Fig. 5(b), as also observed in the Lorenz model (Fig. 4(b)). Nevertheless, different from the Lorenz model which shows a higher convergence value in $\mathcal{J}$, the convergence value of $\mathcal{D}$ in the OU process is larger. It indicates that, for the OU process, the distant causal history always provides more information to the target than the immediate causal history no matter how much of the finite recent dynamics are considered. Further, while the interaction information $\mathcal{I}_{\mathcal{J}}$ and $\mathcal{I}_{\mathcal{D}}$ still decreases and increases with the time lag τ, respectively, similar to the Lorenz model, $\mathcal{I}_{\mathcal{D}}$ in the OU process converges a value larger than zero. The convergence of $\mathcal{I}_{\mathcal{D}}$ to a positive value implies a net synergistic effect from the interaction contribution to the target. In summary, compared with the Lorenz model, the evolutionary dynamics of the OU process, which shows a similar long-term dependency but without a strange attractor, contains a more dominant influence from distant causal history with a net synergistic effect on each variable in the process.

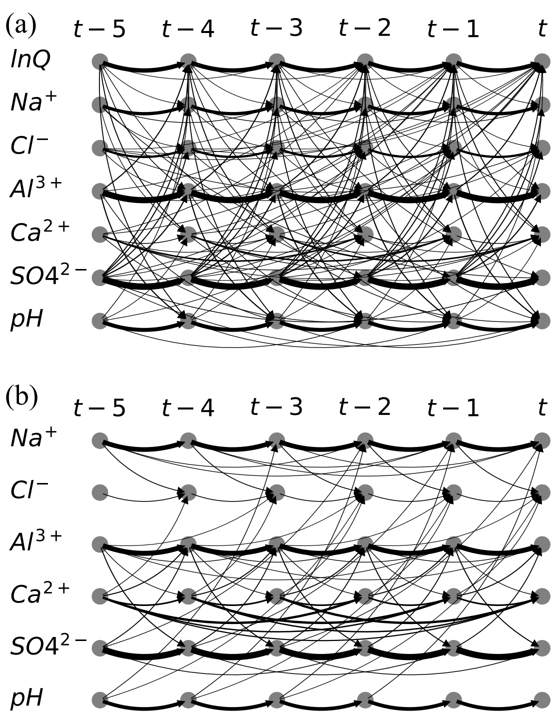

FIG. 6. Time series graph constructed by using the Tigramite algorithm from (a) observed logarithm of flow rate and six catchment chemistry time series data; and (b) the six catchment chemistry data with the variation of logarithmic flow rate corrected. The thickness of edges represents the coupling strength between two nodes computed by momentary information transfer shown in Fig. 9 (see the details of the graph construction in Appendix B).

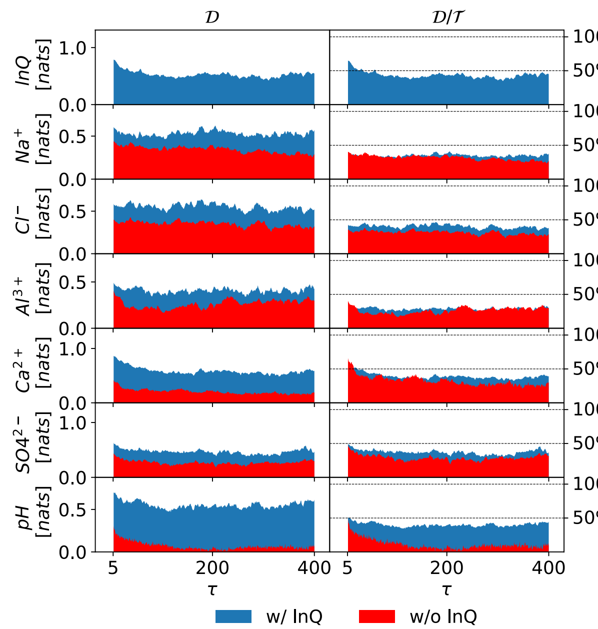

FIG. 7. (color online) Plots of the information transfers $\mathcal{D}$ (left) and the proportion $\mathcal{D}/\mathcal{T}$ (right) over the time lag τ for the raw data and the flow rate-corrected data taking the immediate causal history initiated from all the variables with a same lag τ based on the estimated time series graph in Fig. 6.

D. Catchment Chemistry Data: an Observed Long-Memory System

We now employ our approach to analyze the water solutes in the Upper Hafren in Wales, where the stream chemistry records are found to have $1/f$ fractal signatures reflecting long-term dependencies due to the complex interactions occurring in the catchment [17, 29]. In this application, the logarithm of flow rate, $\ln Q$, and six water chemistry variables, Na^+, Cl^-, Al^{3+}, Ca^{2+}, $SO4^{2-}$ and pH, are chosen for analysis, which are sampled every 7-h from March 2007 to Jan 2009. The 1/f fractal signatures are found in the corrected chemistry data, where the trend of the logarithm of stream flow is excluded [17]. Both the raw and the flow rate-corrected data are available from [17], which are used here. Here, we construct the time series graph for both the raw data and the flow rate-corrected data by using the Tigramite

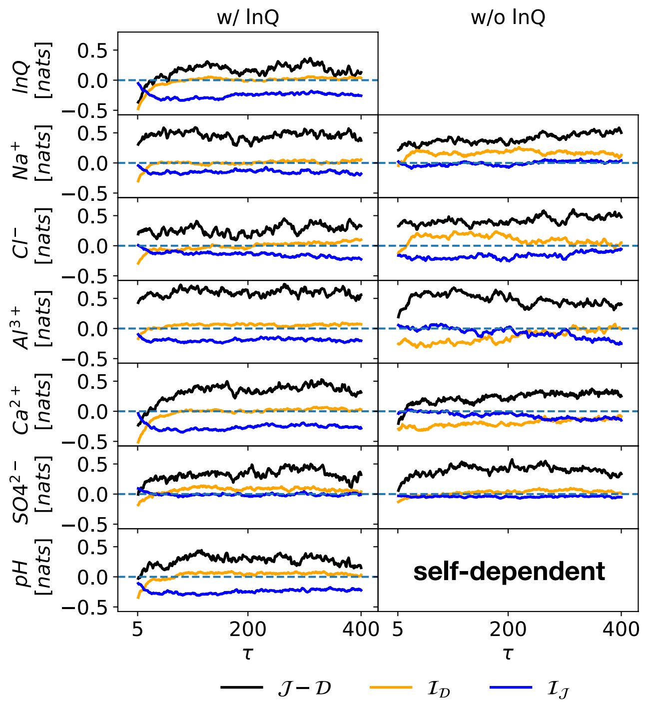

FIG. 8. (color online) Plots of the interaction information from distant causal history, $\mathcal{I}_{\mathcal{D}}$ in Eq.(12) (black line), and immediate causal history, $\mathcal{I}_{\mathcal{J}}$ in Eq.(13) (blue line), over the time lag τ for the raw data and the flow rate-corrected data taking the immediate causal history initiated from all the variables with a same lag τ based on the estimated time series graph in Fig. 6.

algorithm [5, 18, 30, 31] – a modified PC algorithm [22] anchored on the conditional independence test to remove any spurious relationship between each pair of nodes.

The two resulting time series graphs are shown in Fig. 6 (see the details of the graph construction in Appendix B), where coupling strengths in each directed edge, represented as the thickness of the edge, is computed as the momentary information transfer (MIT) [32] between the two nodes. We can observe strong self-feedback dependencies (shown as thick edges) for most variables in both graphs. Meanwhile, the remaining "hairy" causal influences, in a Granger sense, illustrate the relatively weaker lagged interdependencies (shown as thin edges) among the variables, which, along with the self-feedback dependency, contribute to the current state of each variable. Furthermore, the comparison between the two graphs shows that with the influence of flow rate excluded, the graph constructed from the flow rate-corrected data (Fig. 6b) contains fewer cross-dependencies (Fig. 6a). It reflects the fact that flow rate (based mixing) plays a key role in establishing the connectivities among the stream chemistry variables.

Based on the graphs, we now compute the information transfer measures, $\mathcal{T}$ and $\mathcal{D}$, and the interaction information $\mathcal{I}_{\mathcal{J}}$ and $\mathcal{I}_{\mathcal{D}}$ in Eqs.(12) and (13), respectively. The immediate causal history is initiated by all the five variables with a same time lag τ ranging from 1 to 400 (117 days for 7hr dataset). Again, $\mathcal{T}$ and $\mathcal{D}$ are first calculated based on Eqs.(3)-(4) with the number of nearest neighbors $k = 5$ (in kNN method).

The plots of $\mathcal{D}$ and the proportion $\mathcal{D}/\mathcal{T}$ as a function of τ shown in Fig. 7 are insightful. First, for all the variables in both graphs, the information from the distant causal history, $\mathcal{D}$ (the left column of Fig. 7), drops rapidly at small lags τ but starts to converge to a value far from zero for larger time lags (except for pH). Such persistent non-zero $\mathcal{D}$ reflects the long-term dependence present in the water chemistry data, and illustrates that the dynamics from a distant causal history in the stream plays an important role in shaping the current states of the solutes [29]. Further, the right column of Fig. 7 shows that, for each variable in both networks, the percentage of the convergence value of $\mathcal{D}$ in the total information $\mathcal{T}$ is less than 50%, illustrating a more dominant influence from the immediate causal history. Also, by comparing the dynamics with and without flow rate, both $\mathcal{D}$ and its percentage in the total information, $\mathcal{D}/\mathcal{T}$, decrease when the influence of flow rate is excluded. It illustrates that flow rate is an important driving variable that connects various water stream variables, and contributes to maintaining the long-memory dependence. However, this dependence varies for different variables. Specifically, for variables that are highly dependent on flow rate, such as Ca^{2+} and pH, $\mathcal{D}$ declines significantly when the influence of flow rate is excluded. For other variables, especially Na^{+} and Cl^{-} the majority of which originates from the oceanic sources through atmospheric pathways in this close-to-coast location [33], $\mathcal{D}$ drops to a lesser degree and thus still holds a relatively strong memory persistence due to their lower dependencies on flow rate.

Further, the interaction information $\mathcal{I}_{\mathcal{J}}$ and $\mathcal{I}_{\mathcal{D}}$ of the immediate and distant causal histories, respectively, as a function of lag τ are plotted in Fig. 8. First, we see that when the influence of the flow rate is included (the left column of Fig. 8), $\mathcal{I}_{\mathcal{J}}$ decreases with increasing τ and converges to a negative value, suggesting the prevalence of strong redundant influence in the immediate causal history. Meanwhile, $\mathcal{I}_{\mathcal{D}}$ flattens out to zero as τ becomes larger than around 20. The convergence of $\mathcal{I}_{\mathcal{D}}$ to zero implies a balanced synergistic and redundant effects from the self and cross dependencies in the distant causal history. Moreover, in the network without the influence of flow rate (the right column of Fig. 8), $\mathcal{I}_{\mathcal{J}}$ also converges to zero, indicating a balance of synergistic and redundant contribution.

Also, notice that there exist oscillations in different information-theoretic measures shown in both Figs. 7 and 8 even when the values converge for large τ. This is possibly due to the bias induced by the estimation of the proposed high-dimensional information-theoretic measures [12, 18, 32] with a limited amount of data points, which are around 1000~2000 for the estimation of $\mathcal{D}$ for different time lags. A shuffle test is also conducted for the computation of $\mathcal{D}$, to ensure that most of the values are statistically significant at $\alpha = 0.05$ significance level (see Appendix B for details).

IV. CONCLUSION

We have developed information-theoretic measures to partition the influence of total causal history ($\mathcal{T}$) into two components, immediate ($\mathcal{J}$) and distant ($\mathcal{D}$) causal history. While the information from the immediate causal history quantifies the impact on the state of a specific variable from trajectories of recent dynamics, its complement, the distant causal history, illustrates such impact stemming from the remaining older history.

By employing the Markov property for directed acyclic graph, we reduce the dimensions of $\mathcal{T}$, $\mathcal{D}$ and $\mathcal{J}$ to make the computations of the three measures feasible. The Markov property based simplification further results in the information aggregation property of the time series directed acyclic graph, that is, the information transferred from earlier dynamics in the causal history accumulate at the nodes directly influencing the target node(s). Moreover, the dimension reduction also enables further partitions of both the immediate and distant causal histories into self and cross dependencies, and allows us to quantify their interaction information contribution to a target.

It is noted that while the dimension of $\mathcal{T}$ is now reduced to only the parents of the target, the cardinalities of $\mathcal{D}$ and $\mathcal{J}$ can still be high due to the inclusion of the parents of the immediate causal history. For instance, in the stream chemistry example, the dimensions of $\mathcal{D}$ and $\mathcal{J}$ are around 30 and 40, respectively, as shown in Fig. 11. Such high dimensions might result in biased information-theoretic estimation based on limited datasets. Future research is required to further reduce the dimensionality.

We take the opportunity to distinguish the causal history formulation presented here with some relevant prior work. These include transfer entropy [3], momentary information transfer [5], causation entropy [7], and directed information [6, 34]. These existing information-theoretic measures quantify the coupling strength between two (lagged) variables with or without the knowledge of other variable(s), while the proposed causal history analysis investigates how the entire evolutionary dynamics involving all variables in a system influences a target variable. This uniqueness of considering contribution from multiple variables enables analyses that are not possible otherwise. The followings is a brief summary of the differences with these different information-theoretic approaches.

Transfer entropy (TE) [3] quantifies the information transfered to a target, Z_t, from a sequence of previous states of another variable, $X_{t-1:t-\tau} = \{X_{t-1}, X_{t-2}, ..., X_{t-\tau}\}$, given the knowledge of the past states of itself, $Z_{t-1:t-\tau} = \{Z_{t-1}, Z_{t-2}, ..., Z_{t-\tau}\}$. It is computed through a conditional mutual information, and is given by:

$$I^{TE}_{X\to Z}(\tau) = I(Z_t; X_{t-1:t-\tau} \mid Z_{t-1:t-\tau}). \tag{19}$$

Momentary information transfer (MIT) [5], on the other hand, considers the information transfer to Z_t from a specific lagged variable $X_{t-\tau}$ given the knowledge of the entire historical states, and is obtained as the conditional mutual information given as:

$$I^{MIT}_{X\to Z}(\tau) = I(Z_t; X_{t-\tau} \mid P_{C_{X_{t-\tau}\to Z_t}} \backslash P_{Z_t}). \tag{20}$$

The condition set $P_{C_{X_{t-\tau}\to Z_t}} \backslash P_{Z_t}$, anchored on the Markov property, is a simplified set of all the dynamics preceding the time t, $\vec{X}^-_t = \{\vec{X}_{t-1}, \vec{X}_{t-2}, ...\}$.

The idea of conditioning, which prevents the influence from the nodes in the condition set in influencing the quantification of coupling strength, is also used in causation entropy (CE) [7]. CE from a source variable with lag 1, X_{t-1}, to the a target, Z_t, conditioned on a third variable, Y_t, with lag 1, and is given by:

$$I^{CE}_{X\to Z|Y} = I(Z_t; X_{t-1} \mid Y_{t-1}). \tag{21}$$

Notice that causation entropy is a generalization of transfer entropy in Eq.(19) with $\tau = 1$, that is $I^{CE}_{X\to Z|Z} = I^{TE}_{X\to Z}(1)$.

Further, another measure called Directed Information (DI) [6] quantifies how a limited historical dynamics of a source variable, $X_{t-\tau:t}$, affects the dynamical trajectory of the target variables, $Z_{t-\tau:t}$. This is given as:

$$I^{DI}_{X\to Z}(\tau) = \sum_{i=1}^{\tau} I(Z_{t-i}; X_{t-1:t-i} \mid Z_{t-1:t-i+1}). \tag{22}$$

When the knowledge of the dynamical trajectory of the third variables, $Y_{t-\tau:t}$ is given, it is converted into a conditional directed information (CDI) [6], given by:

$$I^{CDI}_{X\to Z|Y}(\tau) = \sum_{i=1}^{\tau} I(Z_{t-i}; X_{t-1:t-i} \mid Z_{t-1:t-i+1}, Y_{t-1:t-i}). \tag{23}$$

Different from I^{TE}, I^{MIT} and I^{CE}, which quantify the influence to a target from a lagged source variable, I^{DI} and I^{CDI} consider the influence from the past dynamics preceding time t as well as the instantaneous dynamics at time t.

In addition to pairwise interactions, a variation of Eq.(21), temporal causation entropy (TCE) [35] is used for inferring the Markov order of a process, which is given by:

$$I^{TCE}(\tau) = I(Z_t; \vec{Z}^-_t \backslash Z_{t-1:t-\tau} \mid Z_{t-1:t-\tau}). \tag{24}$$

which is the conditional mutual information between Z_t and its earlier dynamics, $\vec{Z}^-_t \backslash Z_{t-1:t-\tau}$, given the immediate dynamics $Z_{t-1:t-\tau}$. The calculation of I^{TCE} in Eq.(24) involves the division of the entire history of a process into two parts based on a time lag τ, which looks similar to the partition of immediate and distant causal histories at a first glance. However, they differ in both the purposes and the technical details. While I^{TCE} is used to infer the Markov order of a process based on the smallest τ when I^{TCE} equals to zero in Eq.(24), the causal history analysis investigates the contribution from both immediate and distant causal histories. The different orientation in the causal history analysis, along with its multivariate nature of the analysis, indicate that this work adds significantly to the discourse associated with such studies.

All these existing information-theoretic measures (i.e., I^{TE}, I^{MIT}, I^{CE}, I^{DI} and I^{CDI}), except I^{TCE}, quantify the coupling strengths between two (lagged) variables from different perspectives. On the other hand, the proposed approach for causal history analysis presented in our work is initiated from a different perspective. It aims at analyzing how the target is driven by the entire evolutionary dynamics, which involves multivariate interactions in a complex system. By analyzing the whole history of the system, it allows the partition of the causal history into an immediate and distant components as well as quantification of these quantities. Furthermore, the instantaneous influence, which is explored in I^{DI} and I^{CDI}, is not considered as cause-effect relationship in this study. This is because the directionality of such causal influence between two contemporaneous nodes is unclear and the contemporaneous dynamics is not considered as causal 'history'.

The quantification of the information from the immediate and distant causal histories sketches the memory dependency of the system, which are illustrated with four examples with varying memories. Further, in addition to characterizing the memory dependency of a complex system, the proposed approach also delineates some key features of the complexity associated with its dynamics, which are not captured by other traditional method such as lagged mutual information. First, for the Lorenz model and the OU process, while lagged mutual information slowly goes to zero with increasing time lag τ, the information from distant causal history $\mathcal{D}$ converges to a nonzero value with large lags. It implies a persistent information influence over long time scale in the system's evolutionary dynamics. Second, we observe that the analyzed models have different characteristics of information transfer. For instance, while the interaction information of distant causal history, $\mathcal{I}_{\mathcal{D}}$, flattens out in both the Lorenz model and the logistic map, the convergence of $\mathcal{I}_{\mathcal{D}}$ to zero in the Lorenz model suggests that there is a balanced synergy and redundancy jointly contributed by the self and cross dependencies. However, in the OU process, which also has long memory but no strange attractor, there turns out to be a net synergy effect in the distant causal history as $\mathcal{I}_{\mathcal{D}}$ converges to a positive value. Further, the differences in the interaction information of the immediate causal history, $\mathcal{I}_{\mathcal{J}}$, also illustrate the various dynamics in different systems. The comparison between the stream chemistry system with and without the influence of flow rate shows that the existence of the flow rate is able to enhance the redundnant effect from self and cross dependencies in immediate causal history.

By involving multiple components as well as the causal influences among them, the proposed measures address an unresolved problem, that of quantifying the causal influence on the current state of a variable from the evolutionary dynamics of the entire system. It is different from what has been addressed so far by existing information-theoretic measures, which is usually anchored on pairwise interactions or multivariate analysis associated with specific parts of the system [3, 5, 7, 12]. This uniqueness, therefore, facilitates addressing the questions of how the complexity of a system is sustained over time, which is reflected in varying memory dependency. With the increasing availability of observations in various domains, this work can open up avenues for new data-driven approaches for the study of complex system dynamics.

APPENDIX

Appendix A: Derivations for Information from Immdeidate Causal History, $\mathcal{J}$

This section provides the derivations of Eqs.(6). We separate the immediate causal history $C_{\vec{V}\Rightarrow Z_t}$ into two sets: (1) those belonging to the parents of Z_t, $P_{Z_t}^{C_{\vec{V}\Rightarrow Z_t}} = P_{Z_t} \cap C_{\vec{V}\Rightarrow Z_t}$, and (2) the remaining nodes, $C_{\vec{V}\Rightarrow Z_t} \backslash P_{Z_t}^{C_{\vec{V}\Rightarrow Z_t}}$. Then, using the chain rule, $\mathcal{J}$ defined in Eq.(5) can be written as:

$$\mathcal{J} = I(Z_t; P_{Z_t}^{C_{\vec{V}\Rightarrow Z_t}}, C_{\vec{V}\Rightarrow Z_t} \backslash P_{Z_t}^{C_{\vec{V}\Rightarrow Z_t}} \mid \vec{W}_\tau) \tag{A1}$$

$$= I(Z_t; P_{Z_t}^{C_{\vec{V}\Rightarrow Z_t}} \mid \vec{W}_\tau) \tag{A2}$$

$$+ \underbrace{I(Z_t; C_{\vec{V}\Rightarrow Z_t} \backslash P_{Z_t}^{C_{\vec{V}\Rightarrow Z_t}} \mid \vec{W}_\tau, P_{Z_t}^{C_{\vec{V}\Rightarrow Z_t}})}_{=0} \tag{A3}$$

$$= I(Z_t; P_{Z_t}^{C_{\vec{V}\Rightarrow Z_t}} \mid \vec{W}_\tau), \tag{A4}$$

yielding Eq.(6). The chain rule of the conditional mutual information (CMI) facilitates the transition from Eq.(A1) to Eq.(A2). Moreover, in the 2^{nd} term of Eq.(A2), $I(Z_t; C_{\vec{V}\Rightarrow Z_t} \backslash P_{Z_t}^{C_{\vec{V}\Rightarrow Z_t}} \mid \vec{W}_\tau, P_{Z_t}^{C_{\vec{V}\Rightarrow Z_t}})$, the parents of Z_t are contained in the condition set, which is the union of $P_{Z_t}^{C_{\vec{V}\Rightarrow Z_t}}$ and $\vec{W}_\tau$, including the parents of Z_t in $C_{\vec{V}\Rightarrow Z_t}$ and the remaining parents not in the subgraph, respectively. Therefore, due to the Markov property, given P_{Z_t} (included in the union of $\vec{W}_\tau$ and $P_{Z_t}^{C_{\vec{V}\Rightarrow Z_t}}$), Z_t is independent of its non-descendants, which contains both $C_{\vec{V}\Rightarrow Z_t} \backslash P_{Z_t}^{C_{\vec{V}\Rightarrow Z_t}}$ and the remaining nodes in the condition set $\{\vec{W}_\tau, P_{Z_t}^{C_{\vec{V}\Rightarrow Z_t}}\}$, thus leading to $I(Z_t; C_{\vec{V}\Rightarrow Z_t} \backslash P_{Z_t}^{C_{\vec{V}\Rightarrow Z_t}} \mid \vec{W}_\tau, P_{Z_t}^{C_{\vec{V}\Rightarrow Z_t}}) = 0$.

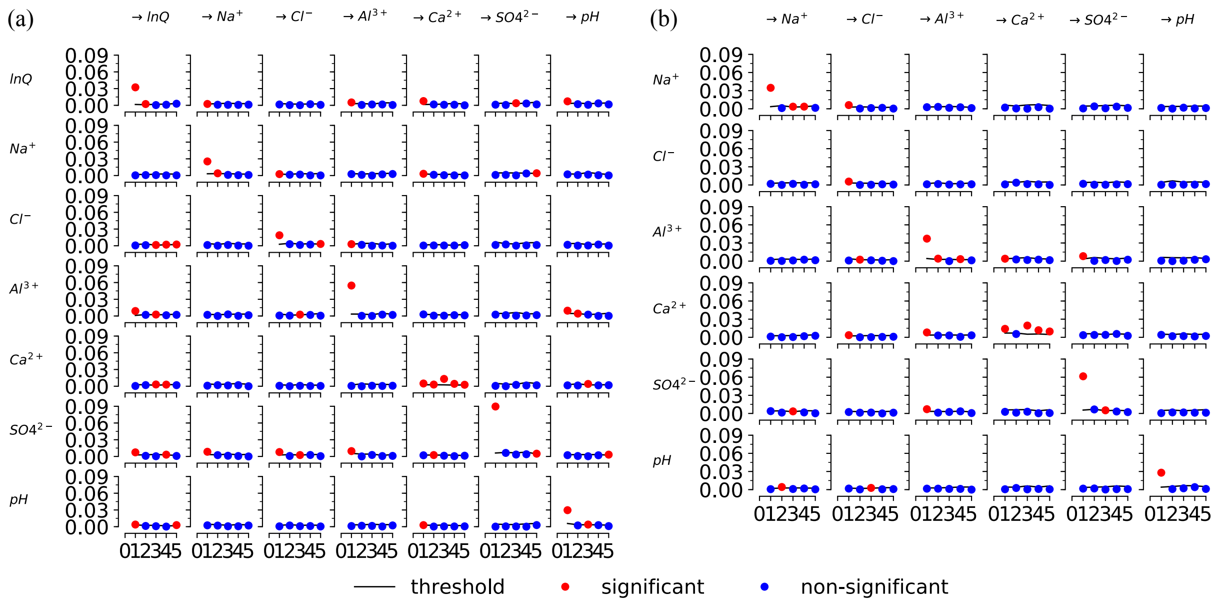

FIG. 9. (color online) Illustration of the estimated lag functions (y-axis: the coupling strength [nats] computed based on momentary information transfer (MIT) [32]; x-axis: the time lag τ) of the catchment chemistry data by using Tigramite algorithm for: (a) the logarithm of flow rate and six chemistry variable; and (b) the six chemistry variables with the variation of the logarithm of flow rate excluded.

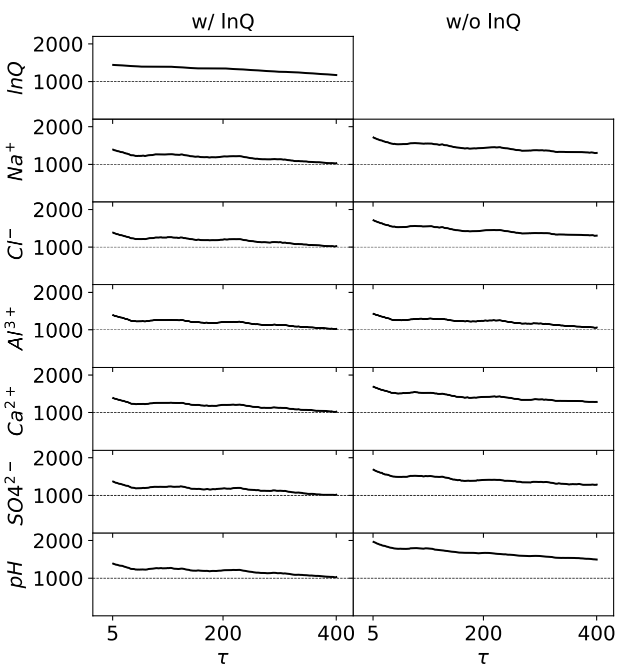

FIG. 10. Number of data points for computing $\mathcal{D}$ in Eq.(4) in terms of the time lag τ for each variable in the two time series graphs constructed in Fig. 9.

Appendix B: Construction of the Time Series Graph for Water Chemistry Data

The catchment chemistry data in the Upper Hafren in Wales, sampled and analyzed every 7-h from March 2007 to Jan 2009, are available as the supporting information of [17]. In this study we use, the logarithmic flow rate ($\ln Q$) and six water quality variables (i.e., Na^+, Cl^-, Al^{3+}, Ca^{2+}, $SO4^{2-}$ and pH), as well as the data with flow-dependent variations corrected [17], are used. We construct two time series graphs for the raw data and the flow rate-corrected one, separately, with the total number 2375 data points including gaps for each graph. The existence of the gaps in the data would reduce the lengths of samples in computing conditional mutual information (CMI) or mutual information (MI), thus potentially worsening the estimation. To minimize this effect, we use the whole dataset to get the sample data points for estimating MI or CMI and then remove the data points containing gaps in the samples [9].

The time series graph is constructed by using Tigramite algorithm [5, 18, 30, 31], which is a modified PC algorithm [22] and anchored on the conditional independence test to remove any spurious relationship between two nodes. We employ the k-nearest-neighbor (kNN) CMI-based conditional independence test, with the number of nearest-neighbor $k = 100$ (high k facilitates a low variance of the estimated CMI [4]). Each test is conducted based on 100 samples with a significance level $\alpha = 95\%$. The graphs are constructed with a maximum time lag $\tau_{max} = 5$. The resulting dependencies for

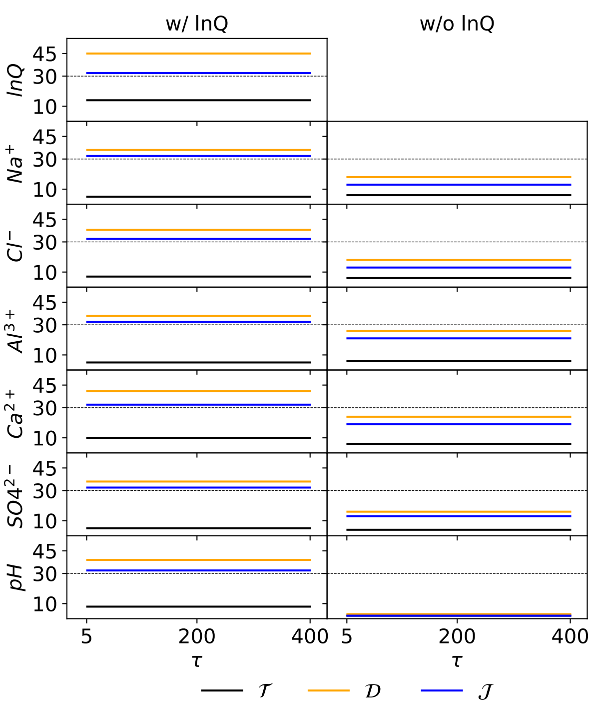

FIG. 11. The cardinality of the estimated $\mathcal{T}$, $\mathcal{D}$ and $\mathcal{J}$ in Eq.(3), Eq.(4) and Eq.(6), respectively, in terms of the time lag τ for each variable in the two time series graphs constructed in Fig. 9.

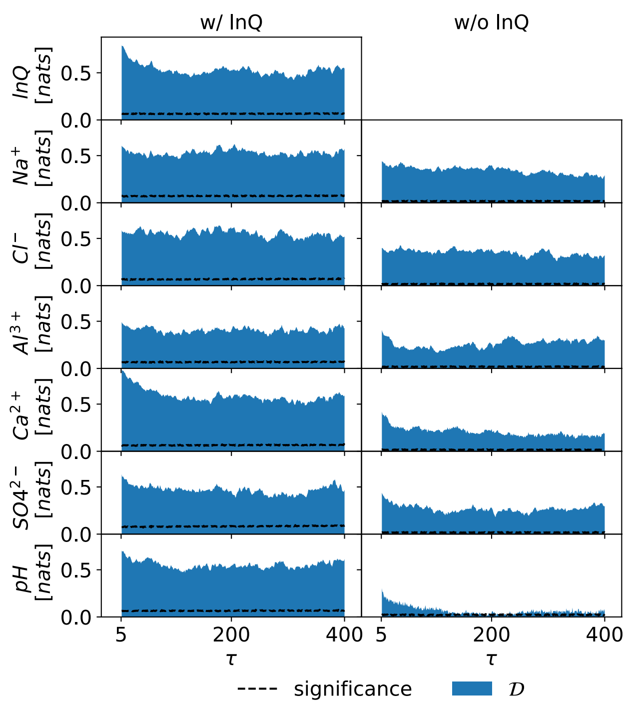

FIG. 12. The estimated $\mathcal{D}$ in Eq.(4) from the two networks constructed in Fig. 9 as well as the corresponding threshold for shuffle test with significance level $\alpha = 0.05$.

the two networks are shown in Fig. 9, sketching the lag function in terms of the momentary information transfer [32] between each pair of lagged components. Based on the two time series graphs, $\mathcal{D}$ and $\mathcal{T}$ for each variable are computed based on Eqs.(4) and (3), respectively, by using kNN approach with $k = 5$. The dimensions of $\mathcal{T}$, $\mathcal{D}$, and $\mathcal{J}$ are shown in Fig. 11. As the computation of $\mathcal{D}$ requires higher dimensions, the numbers of data points used for computing $\mathcal{D}$ are shown in Fig. 10, where in each case more than 1000 are used. Further, to check the significance of $\mathcal{D}$, shuffle test is conducted for $\mathcal{D}$ with a significance level of 95% based on 100 shuffles. The result of shuffle tests in Fig. 12 shows most $\mathcal{D}$ are statistically significant.

ACKNOWLEDGMENTS

Funding support from the following NSF grants are acknowledged: ICER 1440315, EAR 1331906, ACI 1261582, and EAR 1417444.

[1] H. Haken, *Information and Self-Organization* (Springer-Verlag Berlin Heidelberg, 2006).
[2] F. Rosas, P. A. Mediano, M. Ugarte, and H. J. Jensen, arXiv preprint arXiv:1808.05602 (2018).
[3] T. Schreiber, Phys. Rev. Lett. **85**, 461 (2000).
[4] S. Frenzel and B. Pompe, Phys. Rev. Lett. **99**, 204101 (2007).
[5] J. Runge, J. Heitzig, V. Petoukhov, and J. Kurths, Phys. Rev. Lett. **108**, 258701 (2012).
[6] P.-O. Amblard and O. J. J. Michel, Entropy **15**, 113 (2013).
[7] J. Sun and E. M. Bollt, Physica D: Nonlinear Phenomena **267**, 49 (2014).
[8] P. L. Williams and R. D. Beer, CoRR **abs/1004.2515** (2010).
[9] A. E. Goodwell and P. Kumar, Water Resources Research **53**, 5920 (2017).
[10] A. E. Goodwell and P. Kumar, Water Resources Research **53**, 5899 (2017).
[11] A. E. Goodwell, P. Kumar, A. W. Fellows, and G. N. Flerchinger, Proceedings of the National Academy of Sciences (2018), 10.1073/pnas.1800236115.
[12] P. Jiang and P. Kumar, Phys. Rev. E **97**, 042310 (2018).
[13] C. W. J. Granger, Econometrica **37**, 424 (1969).
[14] P. Kumar and B. L. Ruddell, Entropy **12**, 2085 (2010).
[15] J. T. Lizier, J. Heinzle, A. Horstmann, J.-D. Haynes, and M. Prokopenko, Journal of Computational Neuroscience **30**, 85 (2011).
[16] P. Jizba, H. Kleinert, and M. Shefaat, Physica A: Statistical Mechanics and its Applications **391**, 2971 (2012).
[17] J. W. Kirchner and C. Neal, Proceedings of the National Academy of Sciences **110**, 12213 (2013).
[18] J. Runge, Phys. Rev. E **92**, 062829 (2015).
[19] M. Eichler, Probability Theory and Related Fields **153**, 233 (2012).
[20] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)* (Wiley-Interscience, 2006).
[21] S. L. Lauritzen, A. P. Dawid, B. N. Larsen, and H. Leimer, Networks **20**, 491 (1990).
[22] P. Spirtes, C. Glymour, and R. Scheines, *Causation, Prediction, and Search.*, Vol. 81 (MIT Press, Cambridge, 2000).
[23] C. E. Shannon and W. Weaver, *A Mathematical Theory of Communication* (University of Illinois Press, Urbana, IL, 1949).
[24] A. Kraskov, H. Stögbauer, and P. Grassberger, Phys. Rev. E **69**, 066138 (2004).
[25] M. G. Rosenblum, A. S. Pikovsky, and J. Kurths, Phys. Rev. Lett. **78**, 4193 (1997).
[26] F. M. Atay, J. Jost, and A. Wende, Phys. Rev. Lett. **92**, 144101 (2004).
[27] G. Paredes, O. Alvarez-Llamoza, and M. G. Cosenza, Phys. Rev. E **88**, 042920 (2013).
[28] E. N. Lorenz, Journal of the Atmospheric Sciences **20**, 130 (1963).
[29] J. W. Kirchner, X. Feng, and C. Neal, Nature **403**, 524 (2000).
[30] J. Runge, V. Petoukhov, J. F. Donges, J. Hlinka, N. Jajcay, M. Vejmelka, D. Hartman, N. Marwan, M. Paluš, and J. Kurths, Nature communications **6**, 8502 (2015).
[31] J. Runge, D. Sejdinovic, and S. Flaxman, arXiv preprint arXiv:1702.07007 (2017).
[32] J. Runge, J. Heitzig, N. Marwan, and J. Kurths, Phys. Rev. E **86**, 061121 (2012).
[33] X. Feng, J. W. Kirchner, and C. Neal, Journal of Hydrology **292**, 296 (2004).
[34] G. Kramer, *Directed information for channels with feedback*, Ph.D. thesis, Swiss Federal Institute of Technology Zurich (1998).
[35] C. Cafaro, W. M. Lord, J. Sun, and E. M. Bollt, Chaos: An Interdisciplinary Journal of Nonlinear Science **25**, 043106 (2015).