

Unraveling the cause-effect relation between time series

X. San Liang*

School of Marine Sciences, Nanjing University of Information Science and Technology (Nanjing Institute of Meteorology), Nanjing 210044 and China Institute for Advanced Study, Central University of Finance and Economics, Beijing 100081, China
(Received 8 June 2014; revised manuscript received 14 October 2014; published 24 November 2014)

Given two time series, can one faithfully tell, in a rigorous and quantitative way, the cause and effect between them? Based on a recently rigorized physical notion, namely, information flow, we solve an inverse problem and give this important and challenging question, which is of interest in a wide variety of disciplines, a positive answer. Here causality is measured by the time rate of information flowing from one series to the other. The resulting formula is tight in form, involving only commonly used statistics, namely, sample covariances; an immediate corollary is that causation implies correlation, but correlation does not imply causation. It has been validated with touchstone linear and nonlinear series, purportedly generated with one-way causality that evades the traditional approaches. It has also been applied successfully to the investigation of real-world problems; an example presented here is the cause-and-effect relation between the two climate modes, El Niño and the Indian Ocean Dipole (IOD), which have been linked to hazards in far-flung regions of the globe. In general, the two modes are mutually causal, but the causality is asymmetric: El Niño tends to stabilize IOD, while IOD functions to make El Niño more uncertain. To El Niño, the information flowing from IOD manifests itself as a propagation of uncertainty from the Indian Ocean.

DOI: [10.1103/PhysRevE.90.052150](https://doi.org/10.1103/PhysRevE.90.052150)

PACS number(s): 02.50.-r, 89.70.Cf, 05.45.Tp, 89.75.-k

I. INTRODUCTION

In a philosophical discussion, Dempster says, “Causal analyses are guides to higher understanding” [1]. Indeed, identification of the causality between dynamical events is a subject of enormous interest in a wide variety of disciplines. Examples are seen everywhere in neuroscience [2], biology [3], network dynamics [4], financial economics [5], earth sciences [6], and physics [7], to name a few. Toward the end of this study, we will see a climate science example that has been linked to the severe weather and natural disasters on a global scale.

Given that, more often than not, what is known about the two events in question is in the form of time series, causality analysis between time series is therefore of particular importance. This is, however, a very difficult problem; in fact, it is said to be “one of the biggest challenges” in data science (Ref. [8], p. 274). Presently a common practice, particularly a practice in climate science, is computing time-lagged correlation. However, it is well known that correlation does not carry the needed directedness or asymmetry and hence does not necessarily imply causality. As stated by Barnard (Ref. [9], p. 387): “That correlation is not causation is perhaps the first thing that must be said.” Besides, one may argue that, for recurrent processes, there is actually no way to distinguish a lag from an advance, unless one has enough *a priori* knowledge of the processes of concern. This is particularly a problem when the processes are nonsequential, with circular cause and consequence that easily lead one to the chicken-egg dilemma.

Another practice is performing a Granger causality test, which is a statistical test of the usefulness of one time series in forecasting another. This kind of test, as the name implies, provides only a yes-or-no judgment, lacking the quantitative information that may be required in many circumstances. In attempt to remedy the deficiency, recently people have begun to

turn their eyes to an empirical information-theoretic measure, namely, transfer entropy [10]. Unfortunately, this kind of measure is notoriously difficult to evaluate, requiring long time series and imposing challenges to computation [11]; moreover, recent studies have shown that it is biased as its value depends on the autodependency coefficient in a dynamical system [12]. Perhaps this is the reason why in many applied sciences, climate science in particular, time-delayed correlation analysis is still the major tool.

In this study, we will show that causality analysis can be rigorously formulated and quantitatively realized, and that the resulting formula turns out to be remarkably concise and very easy to compute. Considering the large body of related research in climate science, among other sciences, this study is expected to provide timely help.

We use information flow, or information transfer as it may be referred to in the literature, to measure the causation between dynamical events. It has long been recognized as an appropriate causality measure, as the amount of information exchange between two events tells not only the magnitude but also the direction of the cause-effect relation (cf. Refs. [2–6]). Due to its importance, the past decades have seen a surge of interest in formulating this notion. Formalisms have been established empirically or half-empirically, among which is the aforementioned transfer entropy. Realizing that information flow is a real physical notion, and that a real physical notion should be rigorously, rather than empirically, built on a solid foundation so as to be universally applicable for problems in different disciplines, Liang and Kleeman [13] take the initiative to establish, followed by a series of research afterwards, a rigorous formalism for the information flow within given dynamical systems, both deterministic and stochastic (see Ref. [14] for a review). The problem is that it relies on a given dynamical system and hence all the information in the sample space (one can produce arbitrarily many realizations from the given system) or somewhat partially prescribed causality (but dynamics or laws alone do not have causal efficacy; see, e.g.,

*sanliang@courant.nyu.edu

Ref. [15]). What we have here, however, are just two time series, i.e., one single realization for a two-dimensional (2D) system. If we refer to it as a “forward problem,” then what we need for the present study is an “inverse problem,” i.e., whether and how the same notion can be realized with no dynamics but two time series previously given; that way, the quantitative causal relation between the series will be obtained. This could be challenging since only one realization is available, while the resulting formula must verify, to an acceptable extent, the touchstone properties established based on the knowledge of essentially infinite many realizations. This study shows that such a formula, which is expected to have far-reaching implications, indeed can be obtained.

In the following we first briefly introduce the “forward problem.” The major derivation is presented in Sec. III, where a concise formula for causality analysis is obtained. This formula is validated with touchstone time series purportedly generated with only one-way causality (Sec. IV); it is then applied to the study of the causal relation between El Niño and the Indian Ocean Dipole (IOD), the two major modes of climate variation, with important results which have been evidenced for years but missed in previous observational data analyses with existing tools (Sec. V). This study is discussed in Sec. VI and summarized in Sec. VII.

II. THE “FORWARD PROBLEM”

Before getting into the “inverse problem,” let us give a brief review of the “forward problem,” i.e., the recently established formalism of information flow with dynamics given [16]. Consider a d -dimensional stochastic system

$$d\mathbf{X} = \mathbf{F}(\mathbf{X}; \boldsymbol{\theta})dt + \mathbf{B}(\mathbf{X}; \boldsymbol{\theta})d\mathbf{W}, \quad (1)$$

where \mathbf{F} is the vector of drift coefficients, $\boldsymbol{\theta}$ a vector of parameters, $\mathbf{B} = (b_{ij})$ a matrix of diffusion coefficients (or volatility in finance), and \mathbf{W} a vector of standard Wiener process (\mathbf{W} is the so-called white noise). As usual, the vector field \mathbf{F} is assumed to be differentiable. The rate of information flowing from a component, say, X_2 , to another, say, X_1 , is the change rate of the marginal entropy of X_1 , minus the same change rate but with the effect from X_2 instantaneously excluded from the system. These rates of information flow or transfer have been derived analytically in a closed form for any given dynamical systems. The derivation is lengthy and rather technically involved, requiring, for example, the evaluation of some Frobenius-Perron operator [17] and the establishment of a kind of Fokker-Planck equation for the X_2 -excluded system [16]. But the results appear to be tractable. Particularly, for a system of dimensionality 2 (2D), which we will be considering in this study, we have the following theorem.

Theorem II.1: For the system (1), if $d = 2$, then the rate of flow from X_2 to X_1 is

$$T_{2 \rightarrow 1} = -E \left[\frac{1}{\rho_1} \frac{\partial(F_1 \rho_1)}{\partial x_1} \right] + \frac{1}{2} E \left[\frac{1}{\rho_1} \frac{\partial^2 (b_{11}^2 + b_{12}^2) \rho_1}{\partial x_1^2} \right], \quad (2)$$

where ρ_1 is the marginal probability density of X_1 , and E the mathematical expectation.

The proof is referred to Ref. [16], and the same for the theorems in this section. $T_{2 \rightarrow 1}$ may be either zero or nonzero.

A nonzero $T_{2 \rightarrow 1}$ means that X_2 is causal to X_1 : a positive value means that X_2 makes X_1 more uncertain, and vice versa. This measure of information flow possesses an important property, namely, the property of causality or property of flow or transfer asymmetry: a one-way information flow implies nothing about the flow in the opposite direction. In particular, we have the following proved fact:

Theorem II.2: For the system (1), if the evolution of X_1 does not depend on X_2 , then the information flow from X_2 to X_1 vanishes, i.e., $T_{2 \rightarrow 1} = 0$. Specifically, if the vector field component F_1 is independent of x_2 ,

$$E \left[\frac{1}{\rho_1} \frac{\partial(F_1 \rho_1)}{\partial x_1} \right] = 0;$$

if $b_{11}^2 + b_{12}^2$ is independent of x_2 , then

$$E \left[\frac{1}{\rho_1} \frac{\partial^2 (b_{11}^2 + b_{12}^2) \rho_1}{\partial x_1^2} \right] = 0.$$

This property, which one may expect on physical grounds, is very important and will form the touchstone for causality studies.

III. CAUSALITY ANALYSIS

The above result on information flow is rigorous. It is, however, not about causality analysis, as the causal relation is somewhat partially prescribed in the given dynamical system, though dynamical rules alone do not mean causality [15]. Since from a dynamical system one can generate as many realizations as one likes, this formalism is essentially based on infinitely many realizations. In this study, however, the only knowns are two time series, i.e., one realization, with which we need to find the corresponding information flow. Generally it could be very challenging, considering the above touchstone property of one-way causality, among others, to be verified.

Since the dynamics is unknown, we first need to choose a model. As always, a linear model is the natural choice, at least at the initial stage of development. In this case, (2) can be greatly simplified. In Eq. (1), assume $\mathbf{F} = \mathbf{f} + \mathbf{A}\mathbf{X}$, with $\mathbf{f} = (f_1, f_2)^T$, $\mathbf{A} = (a_{ij})$, and $\mathbf{B} = \begin{bmatrix} b_{11} & 0 \\ 0 & b_{22} \end{bmatrix}$ being constant vectors and matrices.

In the equation, if initially (X_1, X_2) has a bivariate Gaussian distribution, as the system is linear, the density is always Gaussian. Denote the mean by $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$, and the covariance matrix by $\boldsymbol{\Sigma} = (\sigma_{ij})$, where $\sigma_{ii} = \sigma_i^2$. They evolve following the differential equations

$$\frac{d\boldsymbol{\mu}}{dt} = \mathbf{f} + \mathbf{A}\boldsymbol{\mu}, \quad (3a)$$

$$\frac{d\boldsymbol{\Sigma}}{dt} = \mathbf{A}\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T. \quad (3b)$$

The information flows between X_1 and X_2 are now easy to evaluate. We only need to consider $T_{2 \rightarrow 1}$. Denote the two terms on the r.h.s. of (2) as (I) and (II), respectively, and write $b_{11}^2 + b_{12}^2$ as β for short. Substitution of

$$F_1 = f_1 + a_{11}X_1 + a_{12}X_2, \\ \rho_1 = \frac{1}{\sqrt{2\pi}\sigma_1} \exp \left[-\frac{(x_1 - \mu_1)^2}{2\sigma_1^2} \right]$$

into (2) for F_1 and ρ_1 yields

$$\begin{aligned} \text{(I)} &= -E \left[F_1 \frac{\partial \log \rho_1}{\partial x_1} + \frac{\partial F_1}{\partial x_1} \right] \\ &= -E \left[-\frac{(X_1 - \mu_1)}{\sigma_1^2} (f_1 + a_{11}X_1 + a_{12}X_2) + a_{11} \right] \\ &= a_{12} \frac{\sigma_{12}}{\sigma_1^2}, \end{aligned}$$

where σ_{ij} is obtained by solving (3), and

$$\text{(II)} = \frac{\beta}{2} E \left[\frac{1}{\rho_1} \frac{\partial^2 \rho_1}{\partial x_1^2} \right] = \frac{\beta}{2} E \left[\frac{(X_1 - \mu_1)^2}{\sigma_1^4} - \frac{1}{\sigma_1^2} \right] = 0.$$

The former is the theorem in Ref. [13] restated in the linear limit. The vanishing (II) is actually a particular case of the property of causality proved in Ref. [16]: when β is independent of x_2 , (II) = 0. Causality is reflected in the asymmetry of the flow. In (I), it goes with a_{12}/σ_{11} . To understand how asymmetry arises in (I), let us assume that β is a linear function of x_2 : $\beta = \beta_0 + \kappa x_2$. As shown above, β_0 does not contribute to $T_{2 \rightarrow 1}$; we need only look at the contribution from κx_2 . Usually this cannot be found explicitly; one has to solve a density evolution equation and then evaluate (2). For the sake of interpretation, let us assume that the density is also a Gaussian (this is generally not true, of course). With this,

$$\begin{aligned} \text{(II)} &= \frac{1}{2} E \left[\frac{\beta}{\rho_1} \frac{\partial^2 \rho_1}{\partial x_1^2} \right] \\ &= \frac{1}{2} E \left\{ (\beta_0 + \kappa X_2) \left[\frac{(X_1 - \mu_1)^2}{\sigma_1^4} - \frac{1}{\sigma_1^2} \right] \right\} \quad (4) \end{aligned}$$

$$= \frac{\kappa}{2\sigma_1^4} E(X_2 - \mu_2)(X_1 - \mu_1)^2, \quad (5)$$

which is clearly asymmetric between X_1 and X_2 .

Our strategy to approach the causality problem is first to estimate the linear model, i.e., to estimate the parameters $(\mathbf{f}, \mathbf{A}, \mathbf{B})$, and then compute the information flow. Without loss of much generality, assume the time series are equal-distanced.

We use maximum likelihood estimation (MLE) to fulfill the purpose. As the population covariances σ_{ij} ($i, j = 1, 2$) can be fairly accurately estimated by the corresponding sample covariances, we focus on the estimation of a_{12} . Consider an interval $[n\Delta t, (n+1)\Delta t]$, Δt being the time stepsize. If the transition probability density function $\rho(\mathbf{X}_{n+1}|\mathbf{X}_n; \boldsymbol{\theta})$ can be obtained, since $\{X_n\}$ is a Markov process, the likelihood

$$L_N(\boldsymbol{\theta}) = \rho(\mathbf{X}_1)\rho(\mathbf{X}_2|\mathbf{X}_1; \boldsymbol{\theta}) \cdots \rho(\mathbf{X}_N|\mathbf{X}_{N-1}; \boldsymbol{\theta})$$

is then given (N the sample size), or, alternatively, the log likelihood

$$\ell_N(\boldsymbol{\theta}) = \sum_{n=1}^N \log \rho(\mathbf{X}_{n+1}|\mathbf{X}_n; \boldsymbol{\theta}) + \log \rho(\mathbf{X}_1)$$

is obtained. When N is large, $\rho(\mathbf{X}_1)$ can be dropped without causing much error. In this linear case, analytical solution of ρ can be found; the result, however, is complicated. For the sake of practical applicability, we turn to the discretized version of the SDE. Using the Euler-Bernstein scheme,

$$\mathbf{X}_{n+1} = \mathbf{X}_n + \mathbf{F}(\mathbf{X}_n; \boldsymbol{\theta})\Delta t + \mathbf{B}(\mathbf{X}_n; \boldsymbol{\theta})\Delta \mathbf{W},$$

where $F = \mathbf{f} + \mathbf{A}\mathbf{x}$. Suppose Δt is small, we know that $\Delta \mathbf{W}$ is a normal $\mathcal{N}(\mathbf{0}, \Delta t \mathbf{I})$ and hence $\mathbf{X}_{n+1}|\mathbf{X}_n = \mathbf{x}_n$ is a normal $\mathcal{N}(\mathbf{x}_n + \mathbf{F}\Delta t, \mathbf{B}\mathbf{B}^T \Delta t)$. So

$$\begin{aligned} \rho(\mathbf{X}_{n+1} = \mathbf{x}_{n+1}|\mathbf{X}_n = \mathbf{x}_n) &= \frac{1}{[(2\pi)^2 \det(\mathbf{B}\mathbf{B}^T \Delta t)]^{1/2}} \\ &\times e^{-\frac{1}{2}(\mathbf{x}_{n+1} - \mathbf{x}_n - \mathbf{F}\Delta t)^T (\mathbf{B}\mathbf{B}^T \Delta t)^{-1} (\mathbf{x}_{n+1} - \mathbf{x}_n - \mathbf{F}\Delta t)} \end{aligned}$$

and

$$\begin{aligned} \ell_N(\boldsymbol{\theta}) &= \sum_{n=1}^N \log \rho(\mathbf{X}_{n+1}|\mathbf{X}_n; \boldsymbol{\theta}) \\ &= \text{const} - \frac{1}{2} \sum_{n=1}^N \log[\det(\mathbf{B}\mathbf{B}^T)] \\ &\quad - \frac{1}{2\Delta t} \sum_{n=1}^N (\mathbf{x}_{n+1} - \mathbf{x}_n - \mathbf{F}\Delta t)^T (\mathbf{B}\mathbf{B}^T)^{-1} \\ &\quad \times (\mathbf{x}_{n+1} - \mathbf{x}_n - \mathbf{F}\Delta t). \quad (6) \end{aligned}$$

Note here we have assumed that N is large enough such that $\rho(X_1)$ can be dropped without causing much error. The MLE of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}$, is therefore the $\boldsymbol{\theta}$ that maximizes $\ell_N(\boldsymbol{\theta})$.

For convenience, denote, for $i = 1, 2$,

$$\dot{X}_{i,n} := \frac{X_{i,n+1} - X_{i,n}}{\Delta t}, \quad (7)$$

$$R_{i,n} := \dot{X}_{i,n} - (f_i + a_{i1}X_{1,n} + a_{i2}X_{2,n}). \quad (8)$$

Also notice that $\mathbf{B}\mathbf{B}^T = \begin{bmatrix} b_1^2 & 0 \\ 0 & b_2^2 \end{bmatrix}$. Substituting into Eq. (6), the log likelihood $\ell_N(\boldsymbol{\theta}) = \ell_N(\mathbf{f}, \mathbf{A}, \mathbf{B})$ is

$$\begin{aligned} \ell_N(\mathbf{f}, \mathbf{A}, \mathbf{B}) &= \text{const} - \frac{N}{2} \log b_1^2 b_2^2 \\ &\quad - \frac{\Delta t}{2} \left[\frac{1}{b_1^2} \sum_{n=1}^N R_{1,n}^2 + \frac{1}{b_2^2} \sum_{n=1}^N R_{2,n}^2 \right]. \end{aligned}$$

The estimators of \mathbf{f} , \mathbf{A} , and \mathbf{B} can be found by maximizing ℓ_N . It is interesting to note that the equations governing the estimators $(\hat{f}_1, \hat{a}_{11}, \hat{a}_{12})$, $(\hat{f}_2, \hat{a}_{21}, \hat{a}_{22})$, and (\hat{b}_1, \hat{b}_2) are actually decoupled. This makes the estimation much easier. Additionally, notice that maximizing ℓ_N over (f_i, a_{i1}, a_{i2}) is equivalent to minimizing $\sum_n R_{i,n}^2 := Q_{N,i}$ over the same group of parameters, for $i = 1, 2$. So $(\hat{f}_1, \hat{a}_{11}, \hat{a}_{12})$ are precisely the least square estimators, which satisfies

$$\begin{aligned} \begin{bmatrix} N & \sum X_{1,n} & \sum X_{2,n} \\ \sum X_{1,n} & \sum X_{1,n}^2 & \sum X_{1,n}X_{2,n} \\ \sum X_{2,n} & \sum X_{2,n}X_{1,n} & \sum X_{2,n}^2 \end{bmatrix} \begin{bmatrix} \hat{f}_1 \\ \hat{a}_{11} \\ \hat{a}_{12} \end{bmatrix} \\ = \begin{bmatrix} \sum \dot{X}_{1,n} \\ \sum X_{1,n} \dot{X}_{1,n} \\ \sum X_{2,n} \dot{X}_{1,n} \end{bmatrix}. \end{aligned}$$

The above equation set can be written in a more familiar and succinct form:

$$\begin{bmatrix} 1 & \bar{X}_1 & \bar{X}_2 \\ \bar{X}_1 & \bar{X}_1^2 & \bar{X}_1\bar{X}_2 \\ \bar{X}_2 & \bar{X}_1\bar{X}_2 & \bar{X}_2^2 \end{bmatrix} \begin{bmatrix} \hat{f}_1 \\ \hat{a}_{11} \\ \hat{a}_{12} \end{bmatrix} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_1\bar{X}_1 \\ \bar{X}_2\bar{X}_1 \end{bmatrix},$$

where, as usual, the overline signifies sample mean. Let

$$C_{ij} := \overline{(X_i - \bar{X}_i)(X_j - \bar{X}_j)},$$

$$C_{i,dj} := \overline{(X_i - \bar{X}_i)(\dot{X}_j - \bar{X}_j)},$$

$\mathbf{C} = (C_{ij})$ being the sample covariance matrix. It is easy to show that

$$\begin{bmatrix} 1 & \bar{X}_1 & \bar{X}_2 \\ 0 & C_{11} & C_{12} \\ 0 & C_{12} & C_{22} \end{bmatrix} \begin{bmatrix} \hat{f}_1 \\ \hat{a}_{11} \\ \hat{a}_{12} \end{bmatrix} = \begin{bmatrix} \bar{X}_1 \\ C_{1,d1} \\ C_{2,d1} \end{bmatrix},$$

which yields

$$\hat{a}_{11} = \frac{C_{22}C_{1,d1} - C_{12}C_{2,d1}}{\det \mathbf{C}},$$

$$\hat{a}_{12} = \frac{-C_{12}C_{1,d1} + C_{11}C_{2,d1}}{\det \mathbf{C}},$$

$$\hat{f}_1 = \bar{X}_1 - \hat{a}_{11}\bar{X}_1 - \hat{a}_{12}\bar{X}_2.$$

To find the remaining parameter, let $\frac{\partial \ell_N}{\partial b_1} = 0$. This gives

$$\hat{b}_1 = \sqrt{\frac{Q_{N,1}\Delta t}{N}}, \quad (9)$$

where

$$Q_{N,1} = \sum_{n=1}^N R_{1,n}^2 = \sum_{n=1}^N [\dot{X}_{1,n} - (\hat{f}_1 + \hat{a}_{11}X_{1,n} + \hat{a}_{12}X_{2,n})]^2.$$

On the other hand, as the population covariance matrix can be estimated by the sample covariance matrix, we estimate σ_{12}/σ_{11} to be C_{12}/C_{11} .

Substituting these results to the linear version of (2), $T_{2 \rightarrow 1} = a_{12}\sigma_{12}/\sigma_{11}$, we finally obtain the rate of information flowing from X_2 to X_1 :

$$T_{2 \rightarrow 1} = \frac{C_{11}C_{12}C_{2,d1} - C_{12}^2C_{1,d1}}{C_{11}^2C_{22} - C_{11}C_{12}^2}, \quad (10)$$

where C_{ij} is the sample covariance between X_i and X_j , and $C_{i,dj}$ the covariance between X_i and \dot{X}_j . Strictly speaking, here $T_{2 \rightarrow 1}$ should bear a caret since it is an estimator of the true information flow. But we will abuse the notation a little bit for the sake of terseness. The flow in the opposite direction, $T_{1 \rightarrow 2}$, can be directly written out by switching the indices 1 and 2. The units are in nats per unit time.

Given a significance level, we may estimate the confidence interval for (4). This actually can always be achieved with bootstrap. But here things can be simplified. When N is large, $T_{2 \rightarrow 1}$ is approximately normally distributed around its true value with a variance $(\frac{C_{12}}{C_{11}})^2 \hat{\sigma}_{a_{12}}^2$, thanks to the MLE property. Here $\hat{\sigma}_{a_{12}}^2$ is determined as follows (e.g., Ref. [18]).

Denote $\theta = (f_1, a_{11}, a_{12}, b_1)$. Compute

$$NI_{ij} = - \sum_{n=1}^N \frac{\partial^2 \log \rho(\mathbf{X}_{n+1} | \mathbf{X}_n; \hat{\theta})}{\partial \theta_i \partial \theta_j}$$

to form a matrix $N\mathbf{I}$, \mathbf{I} being the Fisher information matrix. The inverse $(N\mathbf{I})^{-1}$ is the covariance matrix of $\hat{\theta}$, within which is $\hat{\sigma}_{a_{12}}^2$. Given a significance level, the confidence interval can be found accordingly.

As there has been a great deal of debate over correlation versus causation, it might be of interest to write (10) in terms of correlation and/or correlation-like quantities. Let $r = C_{12}/\sqrt{C_{11}C_{22}}$ be the sample correlation coefficient and $r'_{i,dj} = C_{i,dj}/\sqrt{C_{ii}C_{jj}}$ ($i, j = 1, 2$) be the ‘‘correlation’’ between X_i and \dot{X}_j but normalized with the variances of X_i and X_j . We then have

$$T_{2 \rightarrow 1} = \frac{r}{1 - r^2} (r'_{2,d1} - r r'_{1,d1}). \quad (11)$$

Obviously, two uncorrelated events ($r = 0$) must be noncausal ($T_{2 \rightarrow 1} = 0$); in other words, causation implies correlation. The converse, however, does not hold; i.e., correlation does not imply causation.

IV. VALIDATION

A. A linear problem

To validate (10), consider a 2D stochastic differential equation (SDE) set

$$dX_1 = (-X_1 + 0.5X_2)dt + 0.1dW_1, \quad (12a)$$

$$dX_2 = -X_2dt + 0.1dW_2. \quad (12b)$$

Clearly, X_2 drives X_1 , but not vice versa. This kind of problem is very typical in causality analysis: One component causes another, but the latter has no feedback to the former. We now generate a sample path $(X_{1,n}, X_{2,n})$ and expect to recover this touchstone one-way causality using Eq. (10) from the only realization. Using a time step $\Delta t = 0.001$, we generate 100 000 steps, corresponding to a time span $t = 0 - 100$. For later use, we purportedly initialize the series far from the equilibrium to allow for a period of spin-down, as shown in Fig. 1. The series reach a stationary state after approximately $t = 4$.

As here the dynamics is given, the true information flow between X_1 and X_2 can be evaluated by solving (3). The computed $T_{2 \rightarrow 1}$ and $T_{1 \rightarrow 2}$ are plotted in Fig. 1 (dashed lines). As expected, $T_{1 \rightarrow 2} \equiv 0$, since the evolution of X_2 does not depend on X_1 . That is to say, to X_2 , X_1 is not causal. On the other hand, X_2 drives X_1 and hence is causal to X_1 ; correspondingly $T_{2 \rightarrow 1} \neq 0$. In this example, $T_{2 \rightarrow 1}$ approaches a constant 0.1111 no matter how the covariances are initialized, although the result may be different during the spin-up period ($t < 4$).

Now let us see whether the concise formula (10) can help to recover the one-way causality. The data we have are just one realization, i.e., the above sample path. Our objective is, of course, not to estimate the whole evolution course of $T_{2 \rightarrow 1}$ and $T_{1 \rightarrow 2}$ as shown in Fig. 1; what we expect is to see whether the stationary values ($T_{2 \rightarrow 1} = 0.1111$, $T_{1 \rightarrow 2} = 0$) can be estimated with acceptable confidence. For this purpose, we form different series from the sample path with different

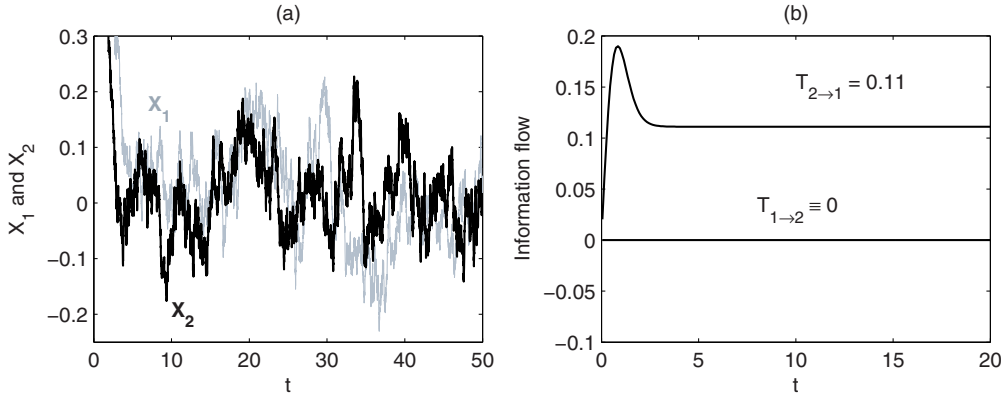


FIG. 1. (Color online) (a) A sample path (X_1 and X_2) of the linear SDE (12) generated using the Euler-Bernstein scheme with $\Delta t = 0.001$ and initialized with (1,2). (b) The true information flow rates $T_{2 \rightarrow 1}$ and $T_{1 \rightarrow 2}$ (in nats per unit time) within the system through solving (3) initialized with $\sigma_{11} = \sigma_{22} = 0.1, \sigma_{12} = 0$ and computing (2).

resolutions and different time intervals and lengths, and then test the performance of estimation. The testing results are listed in Table I.

Clearly, so long as the time span of the series is long enough, the estimation can be made rather accurate. With stationary data ($t = 5-100$), even when one samples the path every 100 time steps (corresponding to a time resolution of 0.1), which yields a time series of only 1000 data points, the result is still acceptable to some extent. To see how the formula may perform if the time span is short, we make the series spanning only 10 time units: $t = 10-20$. The result, $T_{2 \rightarrow 1} = 0.60, T_{1 \rightarrow 2} = 0.17$, unfortunately, is not as good as one would like to expect, even though the data points amount to 10 000. Nonetheless, if we estimate the standard error at a 95% significance level, which gives a value of 0.54 and 0.47, respectively, the result still sounds informative. It tells that $T_{2 \rightarrow 1}$ is significantly (at a 95% level) greater than zero, while $T_{1 \rightarrow 2}$ is not significantly different from zero, consistent with the accurate result.

We have also tested the formula (10) or (11) for systems with different drift and diffusion coefficients, particularly for systems where noises dominate. Again, so long as the series have a span long enough, the causal relation can be faithfully recovered. As an example, for the system

$$dX_1 = -0.5X_1 + X_2 + 20 dW_1,$$

$$dX_2 = -0.7X_2 + 10 dW_2,$$

TABLE I. The information flow computed with different series formed by sampling from the path shown in Fig. 1. Δn is the number of sampling steps (i.e., sampling every Δn time steps from the original series). The accurate result is $T_{2 \rightarrow 1} = 0.1111, T_{1 \rightarrow 2} = 0$ (in nats per unit time).

| Time span | Δn | $T_{2 \rightarrow 1}$ | $T_{1 \rightarrow 2}$ |
|-------------|------------|-----------------------|-----------------------|
| $t = 5-100$ | 1 | 0.11 | -0.002 |
| | 20 | 0.10 | -0.004 |
| | 100 | 0.09 | -0.01 |
| $t = 10-20$ | 1 | 0.60 | 0.17 |
| | 10 | 0.57 | 0.20 |

if the series span 1000 time units or longer, the estimation of $T_{2 \rightarrow 1}$ and $T_{1 \rightarrow 2}$ can be made fairly accurate [$(T_{2 \rightarrow 1}, T_{1 \rightarrow 2}) = (0.116, -0.0169)$ versus the true result (0.115, 0)].

It would be of interest to check how the popular approaches may work with the present series. Since the time-delayed correlation analysis is widely used (especially in climate science), though it has long been recognized inappropriate for causality inference (e.g., Refs. [8,9,19]), we compute the time delayed correlation r with the above time series X_1 and X_2 (the stationary segment $t = 5-100$) and plot it as a function of the time lag Δt (positive if X_1 lags X_2). We know X_2 causes X_1 , while X_1 does not cause X_2 . So, by the logic of correlation advocators, maximal correlation should take place between X_1 and some past X_2 ; i.e., X_2 should lead X_1 . For this specific example, unfortunately, this is not true at all. In Fig. 2 the computed $r(\Delta t)$ as a function of Δt is very small for $\Delta t > 0$ and takes much larger (still small) values at negative Δt . That is to say, a correlation advocator would have concluded that X_1 causes X_2 , not the opposite!

B. A nonlinear problem

We now look at how the formula (10), which stems from a linear assumption, may work for nonlinear systems. The system we will be considering is a one-way coupled map from Ref. [20]:

$$X_{1,n+1} = f(X_{1,n}), \tag{13a}$$

$$X_{2,n+1} = (1 - \xi)f(X_{2,n}) + \xi g_\alpha(X_{1,n}), \tag{13b}$$

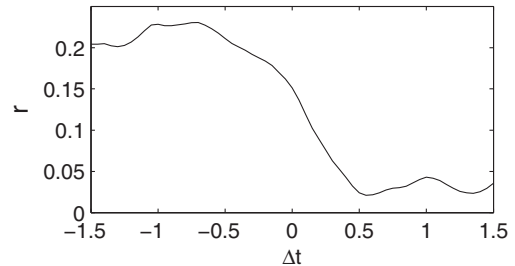


FIG. 2. Time delayed correlation coefficient r as a function of time delay Δt (in unit time; positive if X_1 lags X_2). The time series is from the sample path in Fig. 1 (the stationary segment $t = 5-100$).

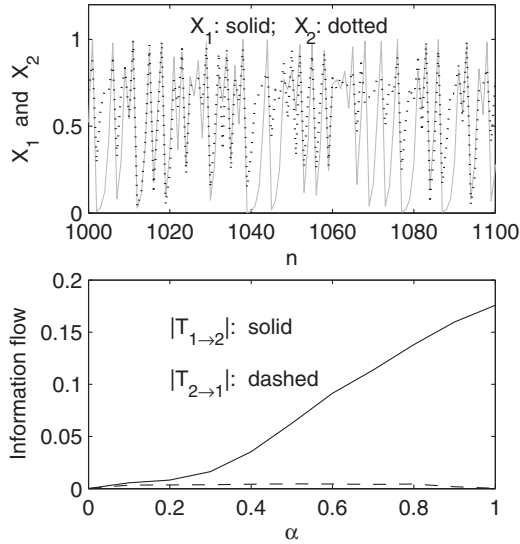


FIG. 3. Information flow within the system (13) with $g_\alpha(x) = \alpha f(x)$. The series are generated with initial condition $X_{1,1} = 0.4$, $X_{2,1} = 0.1$. Upper: Segments of $\{X_{1,n}\}$ (light solid) and $\{X_{2,n}\}$ (dotted) for $\xi = 0.3$ and $\alpha = 1.0$. Lower: Absolute rates of information flow (in nats per unit time) versus anticipation α .

where x

$$f(x) = 4x(1-x) \quad (14)$$

is the chaotic logistic map, and g_α a functional of f . In Ref. [20], Hahs and Pethel examine the “anticipatory system” with $g_\alpha(x) = (1-\alpha)f(x) + \alpha f^2(x)$, where f^2 means that the logistic map f applies twice. This is a very special and chaotic case, which needs a careful discussion; we defer it to Sec. VI. Here we consider

$$g_\alpha(x) = \alpha f(x), \quad (15)$$

a normal case where f applies only once. Obviously, when $\xi \neq 0$ and $\alpha \neq 0$, X_1 is the drive and X_2 the response; in other words, X_1 causes X_2 but not vice versa.

Pick the typical case $\xi = 0.3$. An example series pair is shown in the upper panel of Fig. 3 for $\alpha = 1.0$ [initialized with $X_1(1) = 0.4$, $X_2(1) = 0.1$]. We then use Eq. (10) to compute $T_{2 \rightarrow 1}$ and $T_{1 \rightarrow 2}$ and draw them as functions of α (see the lower panel of the figure). Note what we plot are $|T_{2 \rightarrow 1}|$ and $|T_{1 \rightarrow 2}|$, as it is the absolute values that tell the causality. Besides, the flow rates need not be continuous functions of α , and interpolation between a positive point and a negative point may lead to a spurious zero information flow, i.e., nil causality. (Here it is not a problem, though: all the values are negative.) The figure shows that, compared with $T_{1 \rightarrow 1}$, $T_{2 \rightarrow 1}$ is generally negligible when $\alpha \geq 0.3$. When α is small, the two may be comparable, but still $T_{1 \rightarrow 2}$ is much larger. Besides, $T_{1 \rightarrow 2}$ increases with α , consistent with the observation that in the equation of $X_{2,n+1}$, α controls the effect from X_1 . Particularly, when $\alpha = 0$, both $T_{1 \rightarrow 2}$ and $T_{2 \rightarrow 1}$ vanish, just as one would expect based on the independence between X_1 and X_2 . It is, therefore, safe to say that the one-way causality within the unidirectionally coupled map, albeit highly nonlinear, has been fairly well recovered with (10).

V. A REAL-WORLD APPLICATION

Finally let us look at an application to a real-world problem: the causal relation between the two famous climate modes, El Niño and the Indian Ocean Dipole (IOD). El Niño is the strongest interannual climate variation in the tropical Pacific air-sea coupled system, occurring at irregular intervals of two to seven years and lasting nine months to two years [21]; it is well known through its linkage to natural disasters in far-flung regions of the globe, such as floods in Ecuador, droughts in Southeast Asia and southern Africa, and an increased number of storms over the Pacific Ocean. There are several indices measuring the strength of El Niño, the popular ones being Niño3 and Niño4 [22]. IOD is another air-sea coupled climate mode, characterized by an aperiodic oscillation of sea surface temperature (SST) in the Indian Ocean [23]. It has been related to, among others, the floods in East Africa and droughts in Indonesia and parts of Australia. IOD is measured by an index called the Dipole Mode Index (DMI).

As the dominant modes in, respectively, the Pacific and Indian Oceans, the relationship between El Niño and IOD is of enormous interest; see Ref. [24] for a review. It has long been recognized that the tropical Indian and Pacific Oceans are interrelated on their SST anomalies [25]. But the relation between the two modes is still to be clarified. In general, it is believed that El Niño may induce IOD, which usually peaks in the fall; see Ref. [26]. The causality of El Niño is easily understood, considering the alteration of the Walker circulation during the El Niño years, which causes the subsidence over the Indian Ocean (e.g., Refs. [27,28]).

On the other hand, the impact from IOD has been recognized just recently; in fact, in early efforts, the Indian Ocean used to be treated as a slab of mixed layer responding passively to El Niño [29]. The recognition is mainly through climate predictions, such as the El Niño predictions (e.g., Refs. [30,31]) and the coupled atmosphere-ocean general circulation model experiments (e.g., Ref. [32]). Due to the correlation between the two modes, it has been suggested that an Indo-Pacific perspective should be adopted in research on the El Niño and IOD problems (e.g., Ref. [33]).

The linkage between IOD and El Niño, though evidenced from different aspects, is still an issue in debate. This is partly due to the failure to recover the IOD pattern in the Indian Ocean in many correlation or time-lagged correlation analyses with the indices and the SST observations, albeit the correlation could be significant. Such an example is shown in Fig. 4 of Ref. [24]. This has led people to conclude with caution that IOD might be partially independent of El Niño, though El Niño tends to induce IOD.

This problem, which is obviously a problem on cause and effect, could be due to the inadequacy of using correlation analysis for causality purposes. Now that we have arrived at the formula (10), let us see how things will come out if it is applied to the IOD-El Niño causality analysis.

First, look at the causal relation between the indices of the two modes. The DMI used is the monthly series in terms of SST gradient by the Japan Agency for Marine-Earth Science and Technology (JAMSTEC); the El Niño indices are from the NOAA ESRL Physical Sciences Division [34]. Since this DMI series spans from January 1958 to September 2010, the

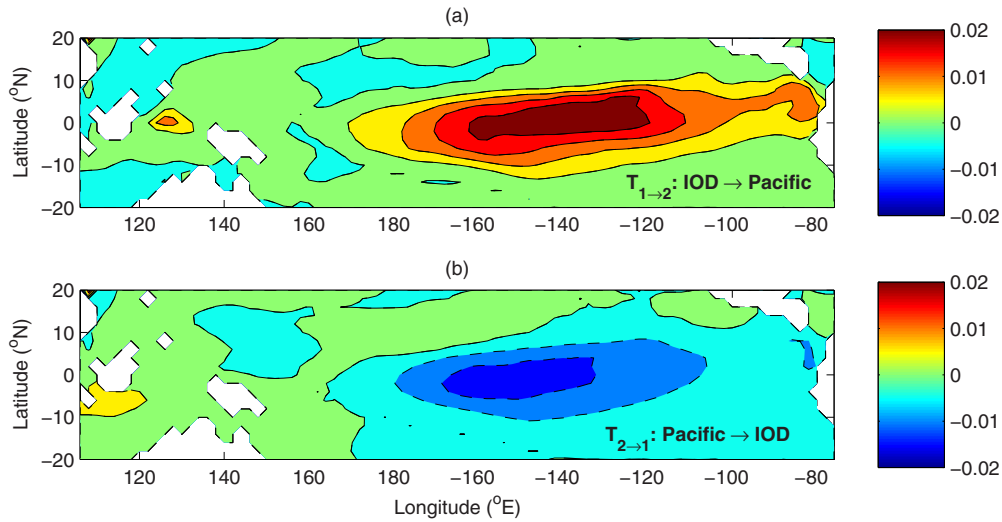


FIG. 4. (Color online) Information flow between the IOD index (symbolized by subscript 1) and the tropical Pacific SST (symbolized by subscript 2): (a) $T_{1 \to 2}$; (b) $T_{2 \to 1}$ (in nats/month). Negative values are in dashed lines.

El Niño indices, which have a much longer time span (1870 to present), are also tailored to the same period. These series then have a total of 633 time points. Application of (10) to the DMI and Niño4 yields a flow of information from El Niño to IOD: $T_{E \to I} = -6$, and a flow from IOD to El Niño : $T_{I \to E} = 13$ (units: 10^{-3} nats/month). Using Niño3 one arrives at a similar result: $T_{E \to I} = -6$ and $T_{I \to E} = 16$. That is to say, El Niño and IOD are mutually causal, and the causality is asymmetric, with the one from the latter to the former larger than its counterpart. Moreover, the different signs indicate that El Niño tends to stabilize IOD, while IOD tends to make El Niño more uncertain.

By our previous experience, series length is crucial. To see whether the length here is adequate, we have tried shortened series: 1963–2010 and 1968–2010. The former yields essentially the same result as the one in its full length (1958–2010); the result of the latter is also similar. In fact, even series with a span as short as 1970–2010 can give a qualitatively similar result. So the series may be long enough to form an ensemble with sufficient statistics. The resulting information flow rates hence may be acceptable.

With this assertion we proceed to extract the causality patterns out of the SST in the Pacific and Indian Oceans. The SST data are from the above NOAA site. As before, we use only the data for the period 1958–2010. First, compute the causality between DMI and the tropical Pacific SST. The results are shown in Fig. 4. Clearly DMI and the Pacific SST are mutually causal, and the information flow in either direction has a El Niño-like pattern. The signs of the flows are the same as computed above with two index series.

If the above El Niño-like pattern in the Pacific is common, the following pattern in the Indian Ocean is unique to our causality analysis. We compute the information flow between Niño4 and the Indian Ocean SST and plot the result in Fig. 5. Figure 5(a) shows the flow from El Niño to the SST. In the tropical region, there are indeed two poles, though the eastern one covers a rather limited region over Indonesia. On the map of the feedback from the Indian Ocean SST [Fig. 5(b)], this structure is much clearer, with a pattern in between the IOD and Indian Ocean Basin Mode (IOBM). This is remarkable, as in previous correlation analyses or time-lagged correlation analyses, the IOD pattern is not seen. Note in Fig. 5(b) that both

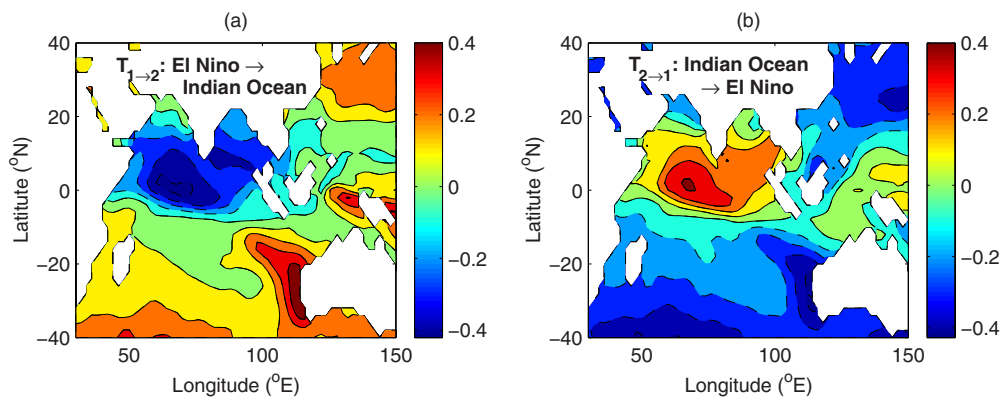


FIG. 5. (Color online) Information flow between Niño4 (symbolized by subscript 1) and the Indian Ocean SST (symbolized by subscript 2): (a) $T_{1 \to 2}$; (b) $T_{2 \to 1}$ (in nats/month). Negative values are in dashed lines.

centers are positive, indicating that the Indian Ocean influences the El Niño through IOD, which functions to amplify the El Niño oscillations. In contrast, the impact of El Niño on the Indian Ocean SST divides over the two centers, which have opposite signs of information flow. The causality analysis with Niño3 shows similar patterns, though the features may not be as pronounced as that with Niño4.

The relation between El Niño and IOD is an extensively investigated subject; further discussion, however, is beyond the scope of this study. Our purpose here is to use it as one real-world example to demonstrate the utility of the newly derived formula (10). The result is encouraging: the IOD-like pattern, which is anticipated as evidence builds up, but missed in previous data analyses, shows up in the index-SST causality patterns. Moreover, it is found that, on the whole, the El Niño influences IOD by making the latter more certain, while the causality from IOD to El Niño is generally the opposite. This is perhaps the reason why it has long been observed in a definite way that El Niño may induce IOD, while the influence from IOD to El Niño, albeit stronger, has been just recognized recently through predictability studies. The latter, which carries a positive information of flow, means that, to El Niño, the Indian Ocean is a source of uncertainty, and the causality from IOD to El Niño is manifested as a propagation of uncertainty from the former to the latter. Clearly an efficient way to observe and forecast an event is to put extensive observation at its uncertainty source: a way in which adaptive observing systems are designed. This is why knowledge of the Indian Ocean facilitates the El Niño forecast (e.g., Refs. [30,31]), i.e., increases the predictability of El Niño.

VI. DISCUSSION

While for linear systems Eq. (10) is satisfactory, and it also succeeds with nonlinear problems such as that in Sec. IV B, caution should be exercised in applying it to highly chaotic cases. Look again at the one-way causal map system (13), with g now defined as

$$g_\alpha(x) = (1 - \alpha)f(x) + \alpha f^2(x), \quad (16)$$

where f^2 means that the logistic map f applies twice. This is the anticipatory system introduced by Hahs and Pethel [20], with α being the anticipation parameter. It is very unusual in causality, as reflected in the computed transfer entropy rates, written as $T_{2 \rightarrow 1}^S$ and $T_{1 \rightarrow 2}^S$ here, versus α . Transfer entropy depends on the choices of delay length m , and the sample space resolution when coarse graining, which is presented with d , namely, the number of partitions in each dimension of the sample space [10]. Let $m = 2, d = 2$; Hahs and Pethel find that $T_{1 \rightarrow 2}^S$ decreases with $\alpha \in [0, 1]$, while $T_{2 \rightarrow 1}^S$ increases with α . The two intersect around $\alpha = 0.5$, and $T_{2 \rightarrow 1}$ finally dominates the computed information flow (see Fig. 1 of Ref. [20]). This is, of course, not what one would expect, as ideally the information flow from X_2 to X_1 should be around zero due to the one-way causality in Eq. (13).

The above anticipatory system seems to be a nightmare for causality analysis tools; our formula (10) together with (7) also encounters the similar difficulty. Of course, it could be argued that the formula is developed for linear systems, while this

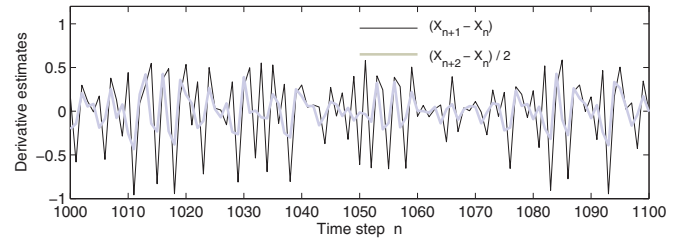


FIG. 6. (Color online) Estimates of dX/dt using different Euler forward schemes. The series is generated from the anticipatory system (13). Shown here is $\{X_{2,n}\}$ for $\alpha = 1$. Thin line: $(X_{2,n+1} - X_{2,n})$; light thick line: $(X_{2,n+2} - X_{2,n})/2$.

current system is highly nonlinear. However, as linearization is a common practice in nonlinear research, one expects that (10) should give qualitatively correct result. A possible reason why there is a difficulty here could be the inadequacy of the differencing scheme for highly chaotic series. It is well known that, for a field, its derivative is usually much more wildly distributed; a good example is velocity versus vorticity in turbulence (e.g., Ref. [35]). In fact, for a system with weak solutions, the shocklike structure may yield infinite derivatives which, when represented using differencing schemes, are very sensitive in time, resulting in spuriously large local differences, and hence disguise the differential field on larger scales. That is to say, the spuriously large local differences may overwhelm the whole series and make the resulting differences unable to characterize the large-scale (and hence global) trend of the differential field. The above highly chaotic anticipatory system is similar to this. In Fig. 6 a plot of $(X_{2,n+1} - X_{2,n})$ (thin line) and $(X_{2,n+2} - X_{2,n})/2$ (light thick line) versus n is drawn ($\alpha = 1$, initialized with $X_{1,1} = 0.4, X_{2,1} = 0.1$). Clearly, the former is much wilder and larger in amplitude. In this case, the latter represents better the general characteristics of the derivative series, although it is only accurate up to $O(2\Delta t)$.

Based on the above observation and argument, we modify (7) to

$$\dot{X}_{i,n} = \frac{X_{i,n+2} - X_{i,n}}{2\Delta t} \quad (17)$$

to compute the $C_{1,d1}$ and $C_{2,d1}$ in Eq. (10) for the anticipatory system problem. The resulting rates of information flow are plotted in Fig. 7(a). For clarity and for the sake of comparison, in Fig. 7(b) we replot their absolute values $|T_{2 \rightarrow 1}|$ and $|T_{1 \rightarrow 2}|$, which measure the strength of the underlying causality. The figure shows that, compared with $T_{1 \rightarrow 2}$, $T_{2 \rightarrow 1}$ is negligible; even at $\alpha = 0.5$ where the two stay closest, $|T_{1 \rightarrow 2}| > 5 \times |T_{2 \rightarrow 1}|$. The one-way causality within the anticipatory system has been fairly faithfully recovered with only modest computational effect, although this is a highly nonlinear system! We have also tried the differencing scheme $\dot{X}_{i,n} = \frac{X_{i,n+3} - X_{i,n}}{3}$, and the resulting structure of information flow is similar, except for a left translation of the extrema (minimum and maximum) by about $\Delta\alpha = 0.1$.

In general, time series are formed from measurements. The sampling cannot be as dense as all time steps are resolved. Actually, more often than not, it is coarse. We may then ask whether a resampled pair of the anticipatory series can give the right result, with $\dot{X}_{i,n} = (X_{i,n+1} - X_{i,n})/\Delta t$ as defined before

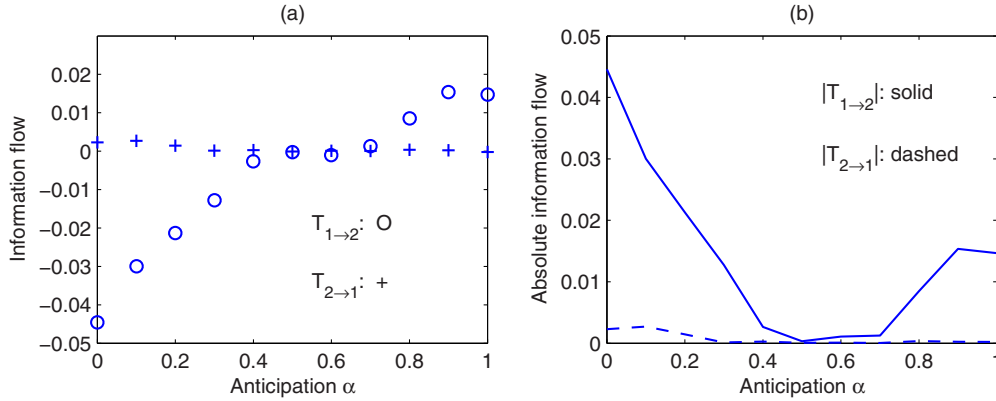


FIG. 7. (Color online) Information flow within the anticipatory system (13) and (16). The series is formed on MATLAB with initial condition $X_{1,1} = 0.4$, $X_{2,1} = 0.1$. (a) $T_{1 \rightarrow 2}$ and $T_{2 \rightarrow 1}$ (in nats per unit time) vs anticipation α estimated with $\dot{X}_{i,n} = \frac{X_{i,n+2} - X_{i,n}}{2}$. (b) Same as (a), but for absolute values. Note that $|T_{1 \rightarrow 2}| \gg |T_{2 \rightarrow 1}|$; even at the point $\alpha = 0.5$, where they seem to stay close in this figure, the former is actually five times larger ($3.26e-4$ vs $6.28e-5$ nats/unit time).

in Eq. (7). Sample the original series every two time steps and redo the computation. Indeed, the computed $T_{1 \rightarrow 2}$ and $T_{2 \rightarrow 1}$ are almost the same as that in Fig. 7; the tiny difference lies at $\alpha = 0.9$ and 1, where $T_{1 \rightarrow 2}$ takes the same value, but in Fig. 7 $\alpha = 0.9$ is a local maximum. The success of this “series coarsening” implies that, for realistic problems such as that in Sec. V, $\dot{X}_{i,n} = \frac{X_{i,n+2} - X_{i,n}}{\Delta t}$ can be used in Eq. (10) safely, for measurements essentially cannot be taken at extremely fine resolution.

A closer look at Fig. 7 shows, besides the fact that $T_{2 \rightarrow 1}$ is negligible compared with $T_{1 \rightarrow 2}$, that $T_{1 \rightarrow 2}$ actually changes sign as α varies on $[0, 1]$. The change of causal structure makes the anticipatory system of Ref. [20] particularly interesting. Physically, this means that, when α is small, X_1 functions to reduce the marginal entropy, and hence the uncertainty, of X_2 ; in other words, X_1 tends to stabilize X_2 . As α exceeds 0.5, however, the function is reversed; X_1 now makes X_2 more uncertain or more unpredictable. It is still unclear whether this fundamental change in causal structure is related to the result of Hahs and Pethel, but the change itself may deserve further study.

In a word, with highly chaotic and densely sampled time series as that generated with the anticipatory system, the formula (10) still works fairly well, provided that \dot{X}_i is differenced every two or more time steps to avoid spuriously large differences that prevent from characterizing the large scale trend of the derivative field. We remark that, in real-world applications, this should not be a problem, as measurements generally cannot be taken at extremely fine resolution.

VII. CONCLUDING REMARKS

We have obtained, in a rigorous and quantitative way, a concise formula for causal analysis between time series. The causality is measured by information flow, a physical notion rigorized just recently. For series X_1 and X_2 , the rate of information flowing (units: nats per unit time) from the latter to the former is

$$T_{2 \rightarrow 1} = \frac{C_{11}C_{12}C_{2,d1} - C_{12}^2C_{1,d1}}{C_{11}^2C_{22} - C_{11}C_{12}^2},$$

where C_{ij} is the sample covariance between X_i and X_j , $C_{i,dj}$ the covariance between X_i and \dot{X}_j , and \dot{X}_j the difference approximation of $\frac{dX_j}{dt}$ using the Euler forward scheme:

$$\dot{X}_{j,n} = \frac{X_{j,n+k} - X_{j,n}}{k\Delta t},$$

with $k \geq 1$ but should not be large to ensure precision. For general time series including those from realistic problems, $k = 1$ should be fine; for highly chaotic and densely sampled series, one may pick $k = 2$. Practically, one may first make a comparison between the results with $k = 1$ and $k = 2$. If they are qualitatively different, then $k = 1$ should be discarded.

The rate of information flow computed from the above formula, $T_{2 \rightarrow 1}$, could be zero or nonzero. If $T_{2 \rightarrow 1} = 0$, X_2 does not cause X_1 ; if not, it is causal. In the presence of causality, two cases can be distinguished according to the sign of the flow: a positive $T_{2 \rightarrow 1}$ means that X_2 functions to make X_1 more uncertain, while a negative value indicates that X_2 tends to stabilize X_1 . This formula is tight in form, involving only common statistics, namely, the sample covariances, and is hence very easy to compute. In addition, it clearly shows that causation implies correlation, but correlation does not necessarily lead to causation, resolving the continuing debate over causation versus correlation. We have validated it with a couple of touchstone series. In the linear case, the stationary preset one-way causality can be rather accurately recovered, provided that the series is long enough (time span, not number of data points) to contain sufficient statistics; this same causality cannot be inferred with the traditional time-delayed correlation analysis. When the series length is guaranteed, the formula works even with series with rather coarse resolution. In the nonlinear case, it also shows remarkable success with a chaotic logistic map, and an unusually chaotic anticipatory system with a very special causal structure.

The above formula has been applied to investigating the relation between El Niño and IOD, the two climate modes which have been linked to many natural hazards on earth. In general, the two modes are mutually causal, though the causality is asymmetric. Particularly, El Niño tends to stabilize IOD, while IOD functions to make El Niño more uncertain. In other words, to El Niño El Niño, the Indian Ocean is

a source of uncertainty, and the causality from IOD to El Niño is manifested as a propagation of uncertainty from the former to the latter. This assertion is also substantiated by the causality analyses between the El Niño index and the Indian Ocean sea surface temperature. The resulting information flow rates, both $T_{\text{ElNiño} \rightarrow \text{Indian}}$ and $T_{\text{Indian} \rightarrow \text{ElNiño}}$, clearly show a structure reminiscent of the IOD and IOBM, which has been missing in previous observational data analyses with traditional approaches.

It would be of interest to compare (10) to Schreiber's transfer entropy. This is, however, a task far beyond the scope of the present study. Previously in Ref. [36], a brief comparison has been made, which demonstrates that our formalism qualitatively agrees with transfer entropy, though quantitatively they are different. Recently Runge *et al.* [12] found that transfer entropy is biased since, within the framework of a dynamical system, it depends on the autodependency coefficient; as a remedy, they propose their own measure, namely, momentary information transfer based on graphical models. Of course, our formalism does not have this kind of bias, as implied by Theorem II.2. Nonetheless, the computational results with the linear problem in Sec. IV A seem to be similar, if the transfer entropy is computed using the analytical result such as that in Ref. [37]. (Note, if one does the computation through bin counting, the one-way causality is not seen in the resulting transfer entropy.) A careful and detailed comparison is yet to be made.

It should be pointed out that the derivation of the formula (10) is based on a formalism with respect to Shannon entropy, or absolute entropy, while it has been argued [38], in predictability research, relative entropy is a more advantageous choice due to its nice properties such as invariance under nonlinear transformation. Fortunately, as we have proved in Ref. [17], for 2D systems, the information flow thus obtained is the same with both absolute and relative entropies. But,

of course, the sign has to be changed, as the former is for uncertainty, while the latter is for predictability.

It should also be pointed out that originally our formalism is developed purportedly for physical systems from nature, such as the climate modes examined in Sec. V. The vector field of concern is differentiable. That is to say, in the governing equation (1), \mathbf{F} should be a differentiable function of \mathbf{X} . The resulting general formula (2) further asserts this, or the derivative cannot be taken. While in practice the formula (10) may actually go beyond the assumption and work with some nondifferentiable \mathbf{F} , the problems with systems such as those formed with the logical exclusive *or* operation should be excluded.

Some issues remain. First, in arriving at (10), a linear model has been adopted, while in reality, nonlinearity is ubiquitous. Although a linear model could be a good approximation in many cases, and, indeed, it appears to be remarkably successful for the benchmark nonlinear systems examined, we would just take it as an approximation, and do expect some more sophisticated model in order to arrive at a universally applicable formula. Second, what we have worked with is a 2D system. For general multidimensional systems, in principle this should be straightforward, though a formula like (2) is yet to be obtained for stochastic systems. These issues, among others, will be considered in the forthcoming studies.

ACKNOWLEDGMENTS

The comment from an anonymous referee is appreciated. This study was partially supported by the Jiangsu Provincial Government through the "Specially Appointed Professor Program" (Jiangsu Chair Professorship) to X.S.L., and by the National Science Foundation of China (NSFC) under Grant No. 41276032.

-
- [1] A. P. Dempster, *J. Stat. Plan. Infer.* **25**, 261 (1990).
 - [2] E. Pereda, R. Q. Quiroga, and J. Bhattacharya, *Prog. Neurobiol.* **77**, 1 (2005); A. Korzeniewska, M. Manczak, M. Kaminski, K. J. Blinowska, and S. Kasicki, *J. Neurosci. Meth.* **125**, 195 (2003); J. Wu, X. Liu, and J. Feng, *ibid.* **167**, 367 (2008).
 - [3] K. Sachs, O. Perez, D. Pe'er, D. A. Lauffenburger, and G. P. Nolan, *Science* **308**, 523 (2005); S. Okasha, *J. R. Soc. Interf.* **2**, 49 (2012); C. Jost and R. Arditi, *Theor. Pop. Biol.* **57**, 325 (2000).
 - [4] N. Ay and D. Polani, *Adv. Complex Syst.* **11**, 17 (2008); L. Sommerlade, M. Eichler, M. Jachan, K. Henschel, J. Timmer, and B. Schelter, *Phys. Rev. E* **80**, 051128 (2009).
 - [5] C. R. Chen, P. P. Lung, and N. S. P. Tay, *Rev. Financ. Econ.* **14**, 1 (2005); S. S. Lee, *Rev. Financ. Stud.* **25**, 439 (2011).
 - [6] R. Dontextner, S. Barbosa, J. Kurths, and N. Marwan, *Eur. Phys. J.* **74**, 1 (2009).
 - [7] J. P. Crutchfield and C. R. Shalizi, *Phys. Rev. E* **59**, 275 (1999); M. S. Baptista, S. P. Garcia, S. K. Dana, and J. Kurths, *Eur. Phys. J.* **165**, 119 (2008).
 - [8] C. O'Neil and R. Schutt, *Doing Data Science: Straight Talk from the Frontline* (O'Reilly, Cambridge, MA, 2013).
 - [9] G. A. Barnard, in *Encyclopedia of Statistical Sciences*, edited by S. Kotz, N. Johnson, and C. Read (Wiley, New York, 1982), Vol. 1, p. 387.
 - [10] T. Schreiber, *Phys. Rev. Lett.* **85**, 461 (2000).
 - [11] K. Hlaváčková-Schindler, M. Paluš, M. Vejmelka, and J. Bhattacharya, *Phys. Rep.* **441**, 1 (2007).
 - [12] J. Runge, J. Heitzig, N. Marwan, and J. Kurths, *Phys. Rev. E* **86**, 061121 (2012).
 - [13] X. S. Liang and R. Kleeman, *Phys. Rev. Lett.* **95**, 244101 (2005).
 - [14] X. S. Liang, *Entropy* **15**, 327 (2013).
 - [15] P. A. Corning, *Complexity* **7**, 18 (2002).
 - [16] X. S. Liang, *Phys. Rev. E* **78**, 031113 (2008).
 - [17] X. S. Liang, *Physica D* **248**, 1 (2013).
 - [18] P. H. Garthwaite, I. T. Jolliffe, and B. Jones, *Statistical Inference* (Prentice-Hall, Hertfordshire, UK, 1995).
 - [19] J. Runge, V. Petoukhov, and J. Kurths, *J. Climate* **27**, 720 (2014).
 - [20] D. W. Hahs and S. D. Pethel, *Phys. Rev. Lett.* **107**, 128701 (2011).
 - [21] See, e.g., M. A. Cane, *Science* **222**, 1189 (1983); S. G. Philander, *El Niño, La Niña, and the Southern Oscillation* (Academic Press,

- London, 1990); M. J. Suarez and P. S. Schopf, *J. Atmos. Sci.* **45**, 3283 (1988); F.-F. Jin, *Science* **274**, 76 (1996); J. M. Wallace, E. M. Rasmusson, P. P. Mitchell, V. E. Kousky, E. S. Sarachik, and H. von Storch, *J. Geophys. Res.* **103**, 14241 (1998).
- [22] K. E. Trenberth, *Bull. Am. Meteor. Soc.* **78**, 2771 (1997).
- [23] H. H. Saji, B. N. Goswami, P. N. Vinayachandran, and T. Yamagata, *Nature (London)* **401**, 360 (1999); P. J. Webster, A. M. Moore, J. P. Loschnigg, and R. R. Leben, *ibid.* **401**, 356 (1999).
- [24] C. Z. Wang, S.-P. Xie, and J. A. Carton, in *Earth's Climate: The Ocean-Atmosphere Interaction*, edited by C. Wang, S.-P. Xie, and J. A. Carton (AGU, Washington, DC, 2004), pp. 1–19.
- [25] G. N. Kiladis and H. F. Diaz, *J. Climate* **2**, 1069 (1989).
- [26] X. Wang and C. Wang, *Climate Dyn.* **42**, 991 (2014).
- [27] S. A. Klein, B. J. Soden, and N. C. Lau, *J. Climate* **12**, 917 (1999).
- [28] K. Ashok, Z. Guan, and T. Yamagata, *J. Meteor. Soc. Jpn.* **81**, 41 (2003).
- [29] N.-C. Lau and M. J. Nath, *J. Climate* **13**, 4287 (2000).
- [30] D. Chen and M. A. Cane, *J. Comput. Phys.* **227**, 3625 (2008).
- [31] T. Izumo, M. Lengaigne, J. Vialard, J.-J. Luo, T. Yamagata, and G. Madec, *Climate Dyn.* **42**, 291 (2013).
- [32] H. Annamalai, S. Kida, and J. Hafner, *J. Climate* **23**, 3933 (2010).
- [33] P. J. Webster and C. D. Hoyos, *Nature Geosci.* **3**, 152 (2010).
- [34] <http://www.jamstec.go.jp/>; <http://www.esrl.noaa.gov/psd>
- [35] A. J. Chorin, *Vorticity and Turbulence* (Springer, New York, 1994).
- [36] X. S. Liang and R. Kleeman, *Physica D* **231**, 1 (2007).
- [37] L. Barnett, A. B. Barrett, and A. K. Seth, *Phys. Rev. Lett.* **103**, 238701 (2009).
- [38] R. Kleeman, *J. Atmos. Sci.* **59**, 2057 (2002).