PHYSICAL REVIEW E **90**, 062921 (2014)

Quantification of causal couplings via dynamical effects: A unifying perspective

Dmitry A. Smirnov[*]

Saratov Branch of V.A. Kotel'nikov Institute of RadioEngineering and Electronics of the Russian Academy of Sciences, 38 Zelyonaya St., Saratov 410019, Russia

(Received 2 May 2014; published 29 December 2014)

Quantitative characterization of causal couplings from time series is crucial in studies of complex systems of different origin. Various statistical tools for that exist and new ones are still being developed with a tendency to creating a single, universal, model-free quantifier of coupling strength. However, a clear and generally applicable way of interpreting such universal characteristics is lacking. This work suggests a general conceptual framework for causal coupling quantification, which is based on state space models and extends the concepts of virtual interventions and dynamical causal effects. Namely, two basic kinds of interventions (state space and parametric) and effects (orbital or transient and stationary or limit) are introduced, giving four families of coupling characteristics. The framework provides a unifying view of apparently different well-established measures and allows us to introduce new characteristics, always with a definite "intervention-effect" interpretation. It is shown that diverse characteristics cannot be reduced to any single coupling strength quantifier and their interpretation is inevitably model based. The proposed set of dynamical causal effect measures quantifies different aspects of "how the coupling manifests itself in the dynamics," reformulating the very question about the "causal coupling strength."

DOI: 10.1103/PhysRevE.90.062921 PACS number(s): 05.45.Tp, 05.45.Xt, 02.50.Ga, 02.50.Tt

I. INTRODUCTION

Necessity to understand and characterize causal couplings among complex systems from time series arises in diverse research fields ranging from engineering [1,2] and physics [3,4] to ecology [5], cardiology [6–8], neuroscience [10–15], and climate science [16–24]. As a basic tool for inferring the very presence of couplings, one exploits Granger causality [25], which is reflected by nonzero values of autoregressive prediction improvement [26–28] or transfer entropy (TE) [29]. The coupling detection is most valuable as a step towards assessing the coupling "strength." To quantify the latter, one often uses the prediction improvement [16,17,26–28] and TE [2,6,15] with their extensions and modifications [13,24,30], including partial correlation [23], spectral Granger causality [31,32], decomposed TE [21], and phase dynamics modelling [4,7,33], with a natural desire [22,34] to have a universally applicable, clearly interpretable, and model-free measure of the coupling strength.

However, even the coupling detection is problematic in case of a "model-free" inference under quite usual conditions [35–38]. The problem of coupling quantification is more difficult. Thus, the concept of Granger causality does not suffice to assess long-term effects [20]. Furthermore, the above-mentioned coupling characteristics turn out to vary considerably under variations of individual systems' properties [22,39] so a relevant question has recently been highlighted: "Why should a measure of coupling strength between X and Y depend on internal dynamics of X?" [22]. Based on the TE idea, the authors of Ref. [22] have managed to develop a model-free measure called "momentary information transfer" (MIT), which is independent of internal dynamics properties for a certain class of systems. The MIT was further suggested as a universal tool to compare strengths of coupling mechanisms across pairs of systems. Still, one is often interested in a broader characterization of the coupling role in the dynamics and asks (i) "To what extent is the global warming over the last decades (i.e., concrete values of the surface temperature) determined by anthropogenic and natural factors [17,19,20,40]?," (ii) "Does an epileptic seizure or Parkinsonian tremor (i.e., a qualitative dynamical regime with a strong periodic component) result primarily from enhanced regularity within a certain brain area or from increased couplings between participating subsystems [11]?," and so on. Any single characteristic of coupling could hardly be appropriate to answer all such diverse questions.

It seems now relevant to broaden the question about the "coupling strength" and characterize the "coupling role in the dynamics." Then, focusing on model-free approaches seems to limit one's possibilities to introduce appropriate and meaningful characteristics. In contrast, it is shown below that a model-based perspective is fruitful in solving the latter task. The purpose of this work is to develop a general conceptual framework for causal coupling quantification from that perspective. It is accomplished by using the formalism of state space models and extending the concepts of virtual interventions [41,42] and dynamical causal effects [20]. The framework unites various coupling analysis techniques into a single consistent picture and clarifies interpretations of some well-known measures in vivid terms of "intervention-effect." It reveals irreducibility of multiple coupling characteristics to any single one and allows to introduce a consistent set of coupling measures, describing "how the coupling manifests itself in the dynamics" instead of assessing "the coupling strength."

Well-known coupling characteristics and their insufficiency are outlined in Sec. II. The unifying framework is introduced in Sec. III, along with concrete novel measures. Section IV presents numerical examples showing irreducibility of diverse coupling characteristics to a single one and nontrivial relationships among them. Estimation issues and implications of this study for a further research are discussed in Sec. V.

[*] smirnovda@yandex.ru

Conclusions are given in Sec. VI. Technical derivations and supplementary illustrations are presented in the Appendices.

II. INCOMPLETENESS OF EXISTING COUPLING CHARACTERISTICS

Consider two systems X and Y whose dynamics are described by finite-dimensional vectors $\mathbf{x}(t)$ and $\mathbf{y}(t)$, respectively, and t is time. The task is to quantify the "strengths" of influences (causal couplings) in the directions $X \to Y$ and $Y \to X$ from observations of the systems' dynamics. Throughout this paper, the combined process $\mathbf{z}(t) = [\mathbf{x}(t),\mathbf{y}(t)]$ is assumed to be Markovian, i.e., the conditional probability density $\rho_t(\mathbf{z}|\mathbf{z}_0)$ of $\mathbf{z}(t)$ at $t > 0$, given an initial state $\mathbf{z}(0) = \mathbf{z}_0$, does not depend on states $\mathbf{z}(t)$ at $t < 0$. This rather general property is typically implied when one speaks of states of stochastic systems, e.g., Ref. [43].

This section discusses insufficiency of existing model-free approaches. Section II A considers the case of a time series of complete state vectors $\{\mathbf{x}_n,\mathbf{y}_n\}_{n=1}^{N}$, where $\mathbf{x}_n = \mathbf{x}(t_n)$, $\mathbf{y}_n = \mathbf{y}(t_n)$, $t_n = n\Delta t$, and Δt is a sampling interval, and discusses restricted applicability and a model-based interpretation of the most advanced coupling characteristics. Section II B further stresses inevitably model-based causal inference in case of partially observed states. Section II C outlines the necessity of a broader coupling characterization in comparison to the approaches of Sec. II A.

A. Transfer entropy and momentary information transfer

To characterize an influence $X \to Y$, one often uses TE, which is conditional mutual information between $\mathbf{y}_n$ and $\mathbf{x}_{n-1}$, given $\mathbf{y}_{n-1}$ [29]. It is conveniently defined as a difference between the conditional distributions $\rho(\mathbf{y}_n|\mathbf{x}_{n-1},\mathbf{y}_{n-1})$ and $\rho(\mathbf{y}_n|\mathbf{y}_{n-1})$ in terms of the Kullback-Leibler (KL) divergence. The latter reads $D(p||q) = \int p(\mathbf{y})\ln[p(\mathbf{y})/q(\mathbf{y})]d\mathbf{y}$ for two arbitrary probability densities $p(\mathbf{y})$ and $q(\mathbf{y})$ [44]. Then TE in the direction $X \to Y$ is $T_{X\to Y} = \langle D(\rho(\mathbf{y}_n|\mathbf{x}_{n-1},\mathbf{y}_{n-1})||\rho(\mathbf{y}_n|\mathbf{y}_{n-1}))\rangle_{\mathbf{x}_{n-1},\mathbf{y}_{n-1}}$, where the angle brackets denote averaging over the stationary probability density of $\mathbf{z}$. $T_{X\to Y}$ equals zero if and only if $\mathbf{y}_n$ is conditionally independent of the previous $\mathbf{x}$, i.e., $\mathbf{y}(t)$ is a state vector of the system Y *per se* and, hence, the influence $X \to Y$ is absent. Positiveness of $T_{X\to Y}$ implies a causal coupling $X \to Y$ and $T_{X\to Y}$ is often used as a measure of "coupling strength." It is also called complete TE [42] (since complete states are observed) and information transfer to Y (ITY) [22]. Everything is similar for the direction $Y \to X$.

TE is model-free and sensitive to any statistical dependencies. However, it depends on internal dynamical properties of X and Y [22] rather than only on "the strength of the coupling mechanism." Indeed, consider an example of univariate autoregressive processes,

$$x_n = \alpha x_{n-1} + \xi_n, \quad y_n = \beta y_{n-1} + c x_{n-1} + \eta_n, \tag{1}$$

where n is discrete time, ξ_n and η_n are mutually uncorrelated Gaussian white noises with variances σ_ξ^2 and σ_η^2, α and β are individual parameters, and c is a coupling coefficient. Here $T_{X\to Y}$ has been shown [22] to depend strongly on α and β, given c. To overcome this problem, a modified measure MIT has been suggested [22]: MIT at unit time lag in the direction $X \to Y$ is mutual information between $\mathbf{y}_n$ and $\mathbf{x}_{n-1}$, given $\mathbf{y}_{n-1}$ and $\mathbf{z}_{n-2}$. For the example (1), MIT is equal to mutual information between y_n and the "innovation" ξ_{n-1}, given y_{n-1}. It is an increasing function of $c\sigma_\xi/\sigma_\eta$, independent of α and β. Hence, MIT quantifies the coupling strength as a contribution of the "innovative part" of the coupling term cx_{n-1} relative to the noise term η_n. This quantification has been regarded natural and MIT has been used to compare coupling strengths across pairs of processes in the climate system [22,23].

Still, even MIT can hardly be considered a universally applicable and model-free coupling strength quantifier. First, the conditions for its independency of internal dynamics are quite restrictive [22]: the right-hand side of the discrete-time evolution equation for Y must consist of three additive terms as in Eq. (1), where the coupling term must be linear in $\mathbf{x}_{n-1}$ and the noise must be white. It means that, despite MIT is defined in a model-free manner, its interpretation is model based, since the systems under study are supposed to belong to the class (i.e., a preassumed model) specified by the above conditions. Moreover, MIT is related to the noise variances in the discrete-time representation of the dynamics, which may well depend on the corresponding (arbitrary) discrete time step.

Second, almost an opposite view of "natural coupling strength measure" is expressed in Ref. [45]. There, a set of two phase oscillators is considered with three terms in the right-hand side as in Eqs. (1). As a measure of coupling strength, the authors use the ratio of the coupling term to the individual dynamics term, not to the noise term. It is justified for self-oscillatory systems with weak noises, where the ratio of a coupling coefficient to a frequency mismatch determines stability of a synchronization regime. Concrete noise levels are irrelevant, so MIT is inappropriate to characterize such a coupling role.

Third, it remains unclear how to interpret concrete numerical values of MIT, e.g., given in nats [22]: Is a coupling characterized by a given MIT value strong and in what sense? The same difficulty of concrete interpretation relates to TE and many other TE-like measures. Indeed, in practice one often asks whether the coupling is important for certain properties of the dynamics to be observed, instead of any particular characterization of the strength of the coupling mechanism, which was one of the primary goals in developing MIT.

B. Spurious couplings and model-based interpretations

If scalar components x_n and y_n of higher-dimensional state vectors are observed, one defines an analog of the complete TE as $T_{X\to Y}^{\text{app}} = \langle D[\rho(y_n|\mathbf{x}_{n-1}^-,\mathbf{y}_{n-1}^-)||\rho(y_n|\mathbf{y}_{n-1}^-)]\rangle_{\mathbf{x}_{n-1}^-,\mathbf{y}_{n-1}^-}$, where $\mathbf{x}_{n-1}^- = (x_{n-1},x_{n-2},\dots)$ and $\mathbf{y}_{n-1}^- = (y_{n-1},y_{n-2},\dots)$. It is called TE [38,46] or apparent TE [42]. If the "full-history" time-delayed vectors $(\mathbf{x}_{n-1}^-,\mathbf{y}_{n-1}^-)$ determine the unobserved states $(\mathbf{x}_{n-1},\mathbf{y}_{n-1})$ well enough, then $T_{X\to Y}^{\text{app}}$ has a meaning similar to the complete TE and can serve as its approximation. The quantity $T_{X\to Y}^{\text{app}}$ is a concrete implementation of the general concept of Granger causality [25]. In practice, one often uses a simpler version based on comparing conditional means of the distributions $\rho(y_n|\mathbf{x}_{n-1}^-,\mathbf{y}_{n-1}^-)$ and $\rho(y_n|\mathbf{y}_{n-1}^-)$. This "causality in mean" [25,27] is quantified as $G^2_{X\to Y} = \frac{\mathrm{var}(y_n|\mathbf{y}^-_{n-1}) - \mathrm{var}(y_n|\mathbf{x}^-_{n-1},\mathbf{y}^-_{n-1})}{\mathrm{var}(y_n|\mathbf{y}^-_{n-1})}$, where var(·) stands for variance. These variances are mean-squared prediction errors of predictors with and without data from X, so $G^2_{X\to Y}$ is a prediction improvement. It was primarily called "strength of causality" [26]. For Gaussian processes, linear autoregressive predictive models are fitted to data in order to estimate $G^2_{X\to Y}$ [27].

It turns out that $G^2_{X\to Y}$ may strongly depend on the sampling interval. Thus, for simple linear stochastic oscillators with a unidirectional coupling $X \to Y$, one may observe nonzero opposite $G^2_{Y\to X}$ at reasonably large sampling intervals [35,36]. Under a naive interpretation, it would lead to a spuriously inferred influence $Y \to X$. Moreover, $G^2_{X\to Y}$ may be quite small as compared to the "spurious" $G^2_{Y\to X}$ [36]. There are exactly the same difficulties with $T^{\mathrm{app}}_{X\to Y}$ [38]. Thus, inferring the very existence of causal couplings from nonzero $G^2_{X\to Y}$ or $T^{\mathrm{app}}_{X\to Y}$ is valid only for certain classes of systems where the sparse sampling effects are negligibly small, i.e., it is again model based. To test for coupling bidirectionality [36], one also looks for a unidirectionally coupled *model* capable to reproduce relevant data properties.

To overcome such problems, the concept of "information flow" [41] is useful. It is formally the same as TE, but the conditional distributions correspond to the conditions imposed by interventions, not passively observed. Yet for its estimation one either needs to assume that complete state vectors are observed (which is again a model in a wider sense) and perform real interventions or to fit model equations to the data and perform "virtual interventions" in the model [3,42]. These examples additionally illustrate inevitably model-based interpretations even of such coupling characteristics, which are defined as apparently model-free.

C. Inquiries about longer-term effects of causal couplings

All of the above widely used approaches relate to one-step-ahead predictions (conditional distributions) and characterize short-term variations in the observed dynamics. However, longer-term effects may often be of greater interest [20]. In particular, having detected the coupling $X \to Y$, one often asks "To what extent is the observed variance of y determined by the influence from X?" [19] or "What is the contribution of X to the power spectrum of y at a particular frequency f?" In other words, one aims at assessing what would happen with the power spectrum of y at f (or variance of y) if an influence from the f component of x (or from x in total) decreased or vanished. There has been a good deal of research to find a decomposition of $G^2_{X\to Y}$ in the frequency domain [27,31,32]. The cornerstone work [31] has introduced a non-negative decomposition called "spectral Granger causality." Under the condition of mutually uncorrelated white noises in the autoregressive equations for x and y, this decomposition represents the ratio of the observed power spectrum of y to the power spectrum of y which would be observed if the noise in the autoregressive equation for x were zero [31,32]. One tries to interpret the spectral Granger causality in terms of "causal power contributions" [32], but it has only something in common with the idea of "contribution from X" and implies validity of an autoregressive model under different conditions.

As another example of longer-term characterization, the "long-term causality" [20] is based on a comparison of the behavior of an empirical model under various hypothetical conditions to assess whether changes in the behavior of a driving system induce any changes in the characteristic of interest (e.g., a linear trend) for the driven system.

To summarize the discussion of existing coupling characteristics, let us note two circumstances. On the one hand, even such common and advanced measures as TE and MIT are not sufficient to address a wide set of possible questions about the coupling role in the dynamics. On the other hand, the above multiple characteristics conceptually differ quite markedly and are either not easy to interpret (e.g., as spectral Granger causality) or too specific (e.g., as trend analysis) to be recommended as a consistent toolkit for causal coupling analysis in general. To cover and shape a broad field of causal coupling quantification, a general conceptual framework is introduced below, which considers diverse approaches from a single perspective, provides their numerical results with a definite unified interpretation, and can prompt a researcher the most appropriate tool for a problem at hand.

III. THE FRAMEWORK OF DYNAMICAL CAUSAL EFFECTS

The above state space process $\mathbf{z}(t)$ is a mathematical "embodiment" of the idea of causality for evolving systems [43]. It serves as a basis for the framework developed below. For a more detailed description, it is necessary to consider a parameter vector $\mathbf{a}$ (constant in time) along with the state $[\mathbf{x}(t),\mathbf{y}(t)]$. In physical models, $\mathbf{a}$ may include dissipation and coupling coefficients, and so on. Then, the probability densities for $\mathbf{x}(t)$ and $\mathbf{y}(t)$ at $t > 0$ read

$$\rho_t(\mathbf{x}|\mathbf{z}_0,\mathbf{a}) = L^X_t(\mathbf{z}_0,\mathbf{a}), \quad \rho_t(\mathbf{y}|\mathbf{z}_0,\mathbf{a}) = L^Y_t(\mathbf{z}_0,\mathbf{a}), \tag{2}$$

where the operators L^X_t and L^Y_t uniquely relate $\mathbf{z}_0$ to the respective distributions at $t > 0$, given $\mathbf{a}$. Let $\mathbf{a}$ consist of four components which characterize internal dynamics of the subsystems X ($\mathbf{a}_{xx}$) and Y ($\mathbf{a}_{yy}$) and the influences $Y \to X$ ($\mathbf{a}_{xy}$) and $X \to Y$ ($\mathbf{a}_{yx}$). Formally, it means that (i) X does not affect Y, i.e., $\partial L^Y_t(\mathbf{x}_0,\mathbf{y}_0,\mathbf{a})/\partial\mathbf{x}_0 \equiv \mathbf{0}$ for any $t > 0$, if and only if $\mathbf{a}_{yx} = \mathbf{0}$; (ii) Y does not affect X, i.e., $\partial L^X_t(\mathbf{x}_0,\mathbf{y}_0,\mathbf{a})/\partial\mathbf{y}_0 \equiv \mathbf{0}$ for any $t > 0$, if and only if $\mathbf{a}_{xy} = \mathbf{0}$; (iii) if the coupling parameter $\mathbf{a}_{yx} = \mathbf{0}$, then L^Y_t does not depend on the individual parameter $\mathbf{a}_{xx}$; and (iv) if $\mathbf{a}_{xy} = \mathbf{0}$, then L^X_t does not depend on $\mathbf{a}_{yy}$. A concrete example of such a formalism, widespread in physical theories and exploited throughout this work, is given by the Langevin-like stochastic differential equations

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{f}_x(\mathbf{x},\mathbf{y},\mathbf{a}_{xx},\mathbf{a}_{xy}) + \xi_x(t), \\ \dot{\mathbf{y}} &= \mathbf{f}_y(\mathbf{y},\mathbf{x},\mathbf{a}_{yy},\mathbf{a}_{yx}) + \xi_y(t), \end{aligned} \tag{3}$$

where $\mathbf{f}_x$ and $\mathbf{f}_y$ are drift terms, ξ_x and ξ_y are Gaussian white noises with covariances $\langle \xi_\mu(t_1)\xi_\nu(t_2)\rangle = \Gamma_{\mu\nu}\delta(t_1 - t_2)$, where μ and ν take the values "x" or "y," and δ is Dirac delta. The elements of Γ_{xx} and Γ_{yy} are included into the vectors $\mathbf{a}_{xx}$ and $\mathbf{a}_{yy}$, respectively. Here the operators L^X_t and L^Y_t are given by a solution to the Fokker-Planck equation.

All the results below relate only to the class of mathematical systems (2). However, the latter is a rather general description of two systems which evolve in time and can either be isolated from each other or interact, depending on their coupling parameters. Deterministic dynamical systems represent a specific subset of this class: system (3) with zero noise or system (2) with Dirac-delta conditional distributions.

In physics and engineering, in order to quantify a *causal effect* of a change in the value of a variable X (from x_1 to x_2) on a random variable Y, one uses a difference between the distributions of Y observed at $X = x_1$ and $X = x_2$ (e.g., Sec. 8.2.1 in Ref. [1]). Two circumstances are important: (i) the value of X is "set by intervention" rather than observed passively and (ii) all other relevant conditions are kept equal. Applying this concept to the evolving system (2), let us quantify the causal coupling $X \rightarrow Y$ by assessing how the dynamics of Y at $t > 0$ changes if "something in X or in coupling $X \rightarrow Y$" is varied at $t = 0$. Since there are two types of characterizing quantities (states and parameters), two kinds of "interventions" are possible. Let us call them (i) state-space intervention (SI) when the state $\mathbf{x}_0$ is changed to $\mathbf{x}_0^*$ and (ii) parametric intervention (PI) when the individual parameter $\mathbf{a}_{xx}$ is changed to $\mathbf{a}_{xx}^*$ or the coupling parameter $\mathbf{a}_{yx}$ is changed to $\mathbf{a}_{yx}^*$. "Other equal conditions" imply that one keeps unchanged an initial state $\mathbf{y}_0$ and either (i) a complete vector $\mathbf{a}$ or (ii) the state $\mathbf{x}_0$ and a part of $\mathbf{a}$ (excluding $\mathbf{a}_{xx}$ or $\mathbf{a}_{yx}$). In practice, such interventions can be called "virtual" [41] if they are performed only in a mathematical model of a real-world object. The term "intervention" is convenient but not compulsory: One just compares the difference of the behaviors of Y-phase orbits starting from the same $\mathbf{y}_0$ under the two different conditions. Let us define "a dynamical causal effect" of SI or PI via a difference between the two conditional distributions of Y. At finite $t > 0$, it can be called an "orbital effect" (OE), since it compares ensembles ("beams") of evolving phase orbits [Fig. 1(a)], or a "transient effect," since the beams are considered before they reach established (limit) distributions. Another possibility is to assess a difference between those limit distributions and, thereby, to define a stationary effect (SE). The two types of interventions and the two types of effects determine four families of coupling characteristics arranged in the classification scheme of Fig. 1(b).

The first novel point is that PIs are explicitly included into the framework and can be analyzed together with more traditional SI-based measures. The second one is the explicit distinction between SEs and short-term OEs, which is relevant since strong effects of one type may not imply strong effects of the other type, while either of them is often of interest.

A. State space interventions and orbital effects

The vector $\mathbf{a}$ is assumed unchanged throughout this subsection and omitted from all formulas for brevity. All coupling characteristics are denoted here by F with a subscript indicating the direction of influence and a superscript reflecting the type of distance used to compare the distributions $\rho_t(\mathbf{y}|\mathbf{x}_0,\mathbf{y}_0)$ and $\rho_t(\mathbf{y}|\mathbf{x}_0^*,\mathbf{y}_0)$. The superscript "KL" stands for the symmetrized KL distance $D_s(p,q) = [D(p||q) + D(q||p)]/2$. Let us define

$$F_{X\rightarrow Y}^{\mathrm{KL}}(t,\mathbf{y}_0,\mathbf{x}_0,\mathbf{x}_0^*) = \sqrt{D_s\left(\rho_t(\mathbf{y}|\mathbf{x}_0,\mathbf{y}_0),\rho_t(\mathbf{y}|\mathbf{x}_0^*,\mathbf{y}_0)\right)}. \quad (4)$$

This is a *local effect* resolved with respect to $\mathbf{y}_0$, $\mathbf{x}_0$, and $\mathbf{x}_0^*$. For nonlinear systems and small t, it may strongly depend on the

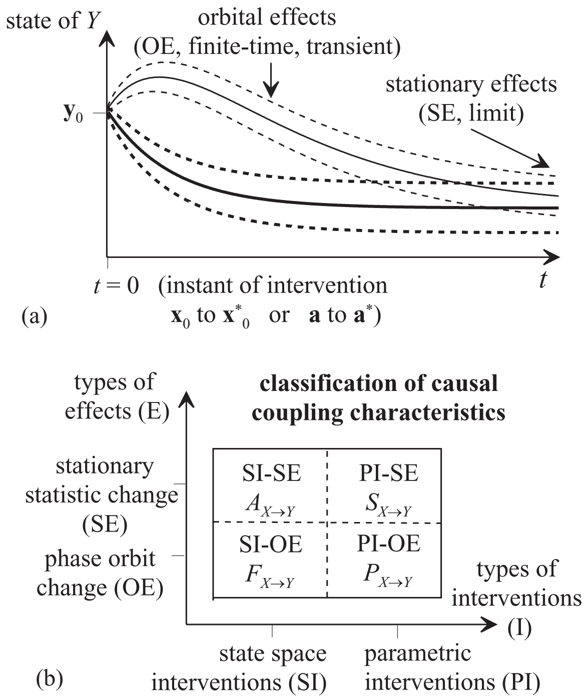
state of Y
orbital effects
(OE, finite-time, transient)
stationary effects
(SE, limit)
y0
t = 0 (instant of intervention
x0 to x*0 or a to a*)
t
(a)
types of effects (E)
classification of causal coupling characteristics
stationary statistic change (SE)
phase orbit change (OE)
SI-SE
A_{X→Y}
PI-SE
S_{X→Y}
SI-OE
F_{X→Y}
PI-OE
P_{X→Y}
types of interventions (I)
state space interventions (SI)
parametric interventions (PI)
(b)

FIG. 1. An illustration for causal coupling quantification: (a) two phase orbit beams (they may either coincide or differ at infinity) under two different conditions, where expectations are shown by solid lines and 95% intervals with dashed ones; (b) classification of coupling characteristics, where the notations are F to denote "effect," A to denote "attract," S to denote "stationary," and P to denote "parametric."

initial states. To have a more compact characterization, let us derive more global measures via averaging or maximization of the local effect.

Suppose that the system Z has a stationary distribution $\rho_{\mathrm{st}}(\mathbf{z})$. Then a phase orbit visits diverse Y states according to the stationary marginal distribution $\rho_{\mathrm{st}}(\mathbf{y})$. Returning to each $\mathbf{y}$ after a long time, the system faces different simultaneous X states according to the conditional distribution $\rho_{\mathrm{st}}(\mathbf{x}|\mathbf{y})$. One may say that in the course of time the system "experiences self-interventions" and "compares" evolutions from different initial X states, given a Y state. Let us draw $\mathbf{y}_0$ randomly from $\rho_{\mathrm{st}}(\mathbf{y})$, $\mathbf{x}_0$ from $\rho_{\mathrm{st}}(\mathbf{x}|\mathbf{y}_0)$, and $\mathbf{x}_0^*$ from $\rho_{\mathrm{st}}(\mathbf{x}|\mathbf{y}_0)$ independently of $\mathbf{x}_0$. Then averaging of (4) over $\mathbf{x}_0^*$ defines an orbital effect $F_{X\rightarrow Y}^{\mathrm{KL}}(t,\mathbf{y}_0,\mathbf{x}_0)$ at the reference point $\mathbf{x}_0,\mathbf{y}_0$. Subsequent averaging over $\mathbf{x}_0$ defines $F_{X\rightarrow Y}^{\mathrm{KL}}(t,\mathbf{y}_0)$ at $\mathbf{y}_0$. Finally, averaging over $\mathbf{y}_0$ yields one of the basic characteristics used below, which reads

$$F_{X\rightarrow Y}^{\mathrm{KL}}(t) = \sqrt{\left\langle\left(F_{X\rightarrow Y}^{\mathrm{KL}}(t,\mathbf{y}_0,\mathbf{x}_0,\mathbf{x}_0^*)\right)^2\right\rangle_{\mathbf{y}_0,\mathbf{x}_0,\mathbf{x}_0^*}}, \quad (5)$$

whose values at different t quantify spatially averaged shorter- or longer-term effects, i.e., manifestations of the coupling $X \rightarrow Y$ at different times. If it reaches a maximum at finite t, its temporal location shows how long it takes for a change in X to become most pronounced in the values of Y. For a compressed description of $F_{X\rightarrow Y}^{\mathrm{KL}}(t)$, let us use its maximum,

$$F_{X\rightarrow Y}^{\mathrm{KL}} = \sup_{t>0} F_{X\rightarrow Y}^{\mathrm{KL}}(t), \quad \tau_{X\rightarrow Y}^{F,\mathrm{KL}} = \arg\sup_{t>0} F_{X\rightarrow Y}^{\mathrm{KL}}(t). \quad (6)$$

For a more vivid interpretation of the KL distance and $F^{\mathrm{KL}}_{X\to Y}$ measure, consider Gaussian distributions $p(\mathbf{y})$ and $q(\mathbf{y})$ with expectations $\mathbf{m}_p$ and $\mathbf{m}_q$ and covariance matrices $\mathbf{C}_p$ and $\mathbf{C}_q$. Then, taking the integrals in the definition of $D_s(p||q)$, one finds

$$D_s(p,q) = D_{\mathrm{mean}}(p,q) + D_{\mathrm{var}}(p,q), \qquad (7)$$

where $D_{\mathrm{mean}}(p,q) = [(\mathbf{m}_p - \mathbf{m}_q)'(\mathbf{C}_p^{-1} + \mathbf{C}_q^{-1})(\mathbf{m}_p - \mathbf{m}_q)]/4$ (where all the vectors are columns and the prime denotes transposition) and $D_{\mathrm{var}}(p,q) = (\mathrm{tr}\{\mathbf{C}_p^{-1}\mathbf{C}_q\} + \mathrm{tr}\{\mathbf{C}_q^{-1}\mathbf{C}_p\} - 2d_Y)/4$ ($\mathrm{tr}\{\cdot\}$ stands for the trace of a matrix, d_Y is the dimension of $\mathbf{y}$). For $d_Y = 1$, one gets

$$D_s(p,q) = \frac{(m_p - m_q)^2(\sigma_p^{-2} + \sigma_q^{-2})}{4} + \frac{1}{4}\left(\frac{\sigma_p}{\sigma_q} - \frac{\sigma_q}{\sigma_p}\right)^2,$$

where σ_p^2 and σ_q^2 are variances. For $\sigma_p^2 = \sigma_q^2$, it further simplifies to $D_s(p,q) = (m_p - m_q)^2/(2\sigma_p^2)$. Thus, the first term in the right-hand side of Eq. (7) makes sense of a normalized squared difference of expectations and the second term means a normalized squared difference of variances. If one now defines

$$F^{\mathrm{KLmean}}_{X\to Y}(t,\mathbf{y}_0,\mathbf{x}_0,\mathbf{x}_0^*) = \sqrt{D_{\mathrm{mean}}(\rho_t(\mathbf{y}|\mathbf{y}_0,\mathbf{x}_0),\rho_t(\mathbf{y}|\mathbf{y}_0,\mathbf{x}_0^*))} \qquad (8)$$

and

$$F^{\mathrm{KLvar}}_{X\to Y}(t,\mathbf{y}_0,\mathbf{x}_0,\mathbf{x}_0^*) = \sqrt{D_{\mathrm{var}}(\rho_t(\mathbf{y}|\mathbf{y}_0,\mathbf{x}_0),\rho_t(\mathbf{y}|\mathbf{y}_0,\mathbf{x}_0^*))}, \qquad (9)$$

then it holds true $[F^{\mathrm{KL}}_{X\to Y}(t,\mathbf{y}_0,\mathbf{x}_0,\mathbf{x}_0^*)]^2 = [F^{\mathrm{KLmean}}_{X\to Y}(t,\mathbf{y}_0,\mathbf{x}_0,\mathbf{x}_0^*)]^2 + [F^{\mathrm{KLvar}}_{X\to Y}(t,\mathbf{y}_0,\mathbf{x}_0,\mathbf{x}_0^*)]^2$ for Gaussian conditional distributions. In a general case, one can use the two terms (8) and (9) as separate OE characteristics. The first term (8) measures how far a phase orbit beam shifts, if $\mathbf{x}_0$ is changed to $\mathbf{x}_0^*$. Roughly, its numerical value shows the ratio of the "beam shift" (difference of conditional expectations) to the "beam width" (standard deviation of the conditional distributions). The second term (9) measures the change in the diffusion rate of a beam. Both terms can be averaged similarly to (5) to define $F^{\mathrm{KLmean}}_{X\to Y}(t)$ and $F^{\mathrm{KLvar}}_{X\to Y}(t)$.

One can also use different normalizations of the expectation difference as compared to $F^{\mathrm{KLmean}}_{X\to Y}$ and different ways of defining a global effect. A useful combination appears to be provided by a fixed-size intervention $||\Delta\mathbf{x}_0|| = ||\mathbf{x}_0^* - \mathbf{x}_0|| = \mathrm{const}$ and the mean phase orbit shift $F^{\mathrm{mean}}_{X\to Y}(t,\mathbf{y}_0,\mathbf{x}_0,\mathbf{x}_0^*) = ||\mathbf{m}_Y(t,\mathbf{x}_0,\mathbf{y}_0) - \mathbf{m}_Y(t,\mathbf{x}_0^*,\mathbf{y}_0)||$, where $\mathbf{m}_Y(t,\mathbf{x}_0,\mathbf{y}_0) = \int \mathbf{y}\rho_t(\mathbf{y}|\mathbf{x}_0,\mathbf{y}_0)d\mathbf{y}$ and $||\cdot||$ denotes Euclidean distance. Dividing $F^{\mathrm{mean}}_{X\to Y}(t,\mathbf{y}_0,\mathbf{x}_0,\mathbf{x}_0+\Delta\mathbf{x}_0)$ by $||\Delta\mathbf{x}_0||$, one gets an OE "per unit SI." Imposing a unit $||\Delta\mathbf{x}_0||$ and maximizing $F^{\mathrm{mean}}_{X\to Y}(t,\mathbf{y}_0,\mathbf{x}_0,\mathbf{x}_0+\Delta\mathbf{x}_0)$ over all such $\Delta\mathbf{x}_0$, one finds the direction of a unit SI leading to the most pronounced OE. Let us denote it with an additional superscript "u" as $F^{u,\mathrm{mean}}_{X\to Y}(t,\mathbf{y}_0,\mathbf{x}_0) = \max_{||\Delta\mathbf{x}_0||=1} F^{\mathrm{mean}}_{X\to Y}(t,\mathbf{y}_0,\mathbf{x}_0,\mathbf{x}_0+\Delta\mathbf{x}_0)$, define $F^{u,\mathrm{mean}}_{X\to Y}(t)$ via averaging over $(\mathbf{x}_0,\mathbf{y}_0)$ with $\rho_{st}(\mathbf{z})$, and get $F^{u,\mathrm{mean}}_{X\to Y}$ and $\tau^{F,u,\mathrm{mean}}_{X\to Y}$ after maximization over t. Such a non-normalized measure is appropriate when a value of Y in physical units is important, e.g., a global surface temperature in degrees.

Many well-known characteristics, including TE and MIT, belong in essence to the SI-OE family and can be interpreted in terms of "intervention-effect," even though less directly than the above F measures as discussed in Appendix A.

B. Parametric interventions and orbital effects

In order to assess how the Y component of a phase beam changes in response to the PI "$\mathbf{a}$ is changed to $\mathbf{a}^*$," let us define the respective OE similarly to Eq. (4) as

$$P^{\mathrm{KL}}_{X\to Y}(t,\mathbf{z}_0,\mathbf{a},\mathbf{a}^*) = \sqrt{D_s[\rho_t(\mathbf{y}|\mathbf{z}_0,\mathbf{a}),\rho_t(\mathbf{y}|\mathbf{z}_0,\mathbf{a}^*)]}. \qquad (10)$$

This is a local effect which depends on the initial state $\mathbf{z}_0$. A global OE of the PI performed at a given value of $\mathbf{a}$ can be defined via averaging over $\mathbf{z}_0$ with the distribution $\rho^Z_{\mathrm{st}}(\mathbf{z}_0,\mathbf{a})$:

$$P^{\mathrm{KL}}_{X\to Y}(t,\mathbf{a},\mathbf{a}^*) = \sqrt{\left\langle\left[P^{\mathrm{KL}}_{X\to Y}(t,\mathbf{z}_0,\mathbf{a},\mathbf{a}^*)\right]^2\right\rangle_{\mathbf{z}_0}}. \qquad (11)$$

If the PI consists in changing the coupling parameter $\mathbf{a}_{yx}$ to $\mathbf{a}^*_{yx} = 0$, let us call it "coupling PI" (CPI). Then the quantity (10) assesses how different Y evolutions are from the same state for a given $\mathbf{a}_{yx}$ and zero coupling. Despite it being a very direct quantification of the coupling role, in practice other ideas are often used instead. Thus, for the system (3) with linear functions $\mathbf{f}_x$ and $\mathbf{f}_y$ one may consider a change in the individual X parameter Γ_{xx} to $\Gamma^*_{xx} = \mathbf{0}$ as assumed, in fact, in the definition of spectral Granger causality (Sec. II C). Let us call this intervention "noise level PI" (NPI). There is a similarity between the CPI and the NPI, e.g., their stationary effects coincide in case of the unidirectional coupling $X \to Y$ in the system (3) with linear functions $\mathbf{f}_x$ and $\mathbf{f}_y$. However, there are also significant differences between finite-time OEs of the CPI and NPI, especially for bidirectionally coupled systems, as shown below. The CPI measure is considered here as the basic one and denoted as just $P^{\mathrm{KL}}_{X\to Y}(t,\mathbf{a})$. The NPI is distinguished by a superscript as $P^{\mathrm{KL,noise}}_{X\to Y}(t,\mathbf{a})$.

One might also average the measure (11) over a range of $\mathbf{a}$ and $\mathbf{a}^*$ values. However, the simpler comparison of a given coupling to zero coupling seems to be sufficiently informative. For brevity, let us omit the $\mathbf{a}$ dependence and write just $P^{\mathrm{KL}}_{X\to Y}(t)$ and $P^{\mathrm{KL,noise}}_{X\to Y}(t)$ if it does not lead to confusion. The maximal OE is $P^{\mathrm{KL}}_{X\to Y} = \max_{t\geqslant 0} P^{\mathrm{KL}}_{X\to Y}(t)$ at $\tau^{P,\mathrm{KL}}_{X\to Y} = \arg\max_{t\geqslant 0} P^{\mathrm{KL}}_{X\to Y}(t)$. One can further define the CPI-OEs $P^{\mathrm{KLmean}}_{X\to Y}$ and $P^{\mathrm{KLvar}}_{X\to Y}$ by using the two-term representation (7) instead of the KL distance in (10), and the NPI-OEs $P^{\mathrm{KLmean,noise}}_{X\to Y}$ and $P^{\mathrm{KLvar,noise}}_{X\to Y}$ in the same way. Instead of averaging over $\mathbf{z}_0$ in (11), maximizing over $\mathbf{z}_0$ gives, e.g., a "unit initial state" PI-OE $P^{u,\mathrm{mean}}_{X\to Y}(t) = \max_{||\mathbf{z}_0||=1} ||\mathbf{m}_Y(t,\mathbf{z}_0,\mathbf{a}) - \mathbf{m}_Y(t,\mathbf{z}_0,\mathbf{a}^*)||$, which is an analog of $F^{u,\mathrm{mean}}_{X\to Y}(t)$ for linear systems as shown below.

PI-OE measures have in fact been used in studies of the global surface temperature evolution under different circumstances, e.g., simplified models are fitted to data in Refs. [20,47] and simulated under various CO_2 emission scenarios (PIs) over decades into the future to compare the model responses (OEs). However, PIs have not been systematically considered for coupling characterization, making the concept of virtual interventions incomplete.

C. Stationary effects

Finite-time OEs can be complemented with their limit (stationary) counterparts by taking $\lim_{t\to\infty}$ in the above formulas for local OEs. Thus, a stationary effect of a PI reads

$$S^{\mathrm{KL}}_{X\to Y}(\mathbf{z}_0,\mathbf{a},\mathbf{a}^*) = \lim_{t\to\infty} P^{\mathrm{KL}}_{X\to Y}(t,\mathbf{z}_0,\mathbf{a},\mathbf{a}^*). \tag{12}$$

If the system Z has a single stationary distribution for each of the two parameter values, then the quantity (12) does not depend on $\mathbf{z}_0$, so one may write just $S^{\mathrm{KL}}_{X\to Y}(\mathbf{a},\mathbf{a}^*)$ to quantify the difference between those distributions. It can be rewritten for Gaussian processes and the representation (7) as the sum of changes in expectation and covariance matrix of the stationary distribution of $\mathbf{y}$. In particular, the change in covariance matrix remains the only SE for zero mean Gaussian processes:

$$S^{\mathrm{KL}}_{X\to Y}(\mathbf{a},\mathbf{a}^*) = S^{\mathrm{KLvar}}_{X\to Y}(\mathbf{a},\mathbf{a}^*) \equiv \sqrt{D_{\mathrm{var}}\left(\rho^Y_{\mathrm{st}}(\mathbf{y}|\mathbf{a}),\rho^Y_{\mathrm{st}}(\mathbf{y}|\mathbf{a}^*)\right)}. \tag{13}$$

For a one-dimensional $\mathbf{y}$, $S^{\mathrm{KLvar}}_{X\to Y}(\mathbf{a},\mathbf{a}^*) = \frac{1}{2}\frac{\left|\sigma^2_y(\mathbf{a})-\sigma^2_y(\mathbf{a}^*)\right|}{\sigma_y(\mathbf{a})\sigma_y(\mathbf{a}^*)}$, where σ^2_y is the variance of y.

It is useful to consider a simpler normalization and a signed quantity to define a SE as $S^{\mathrm{var}}_{X\to Y}(\mathbf{a},\mathbf{a}^*) = \frac{\sigma^2_y(\mathbf{a})-\sigma^2_y(\mathbf{a}^*)}{\sigma^2_y(\mathbf{a}^*)}$. Under a CPI, $S^{\mathrm{var}}_{X\to Y}$ is positive (negative) if the variance of y at a given coupling $\mathbf{a}_{yx}$ is greater (less) than that in the uncoupled case, that is, $S^{\mathrm{var}}_{X\to Y}$ shows how strongly a given nonzero $\mathbf{a}_{yx}$ changes (either increases or decreases) the variance of Y as compared to its free dynamics at $\mathbf{a}^*_{yx} = \mathbf{0}$. Similarly, one can define $S^{\mathrm{KLvar,noise}}_{X\to Y}$ and $S^{\mathrm{var,noise}}_{X\to Y}$ to quantify SEs of NPI, i.e., to compare the variance of y at a given Γ_{xx} in Eq. (3) to that at $\Gamma^*_{xx} = \mathbf{0}$. Note that one often tries to assess "to what extent the observed variance of y is determined by the influence of X" [19], implying that the influence of X *increases* the variance of y as compared to the uncoupled dynamics. However, $S^{\mathrm{var}}_{X\to Y}$ may well be negative (Appendix D), which requires a reformulation of the very question. At that, the quantity $S^{\mathrm{KLvar,noise}}_{X\to Y}$ is positive under more general conditions but answers a different question regarding "to what extent the variance of y is determined by the noise source in the system X." The spectral Granger causality also compares the actual power spectrum of y to that for $\Gamma^*_{xx} = \mathbf{0}$ (Sec. II C) and is, therefore, similar to $S^{\mathrm{KLvar,noise}}_{X\to Y}$. Thus, one can interpret the former as a stationary effect of NPI, which assesses a difference between multidimensional probability distributions of the process $y(t)$. Such a clarification of meaning of any coupling measure in terms of "intervention-effect" is an advantage of the suggested perspective.

More generally, the PI-SE family relates to any analysis of changes in an established dynamical regime under variation of control parameters, including such basic concepts as dynamical regime charts and bifurcation or synchronization diagrams (e.g., Ref. [48]). The recent "chronotaxicity analysis" [8] also studies stationary characteristics (a fixed-point stability) under a specific PI (zeroing noise in the driven system). Thus, a bulk of theoretical characteristics well known in other contexts can be incorporated into the suggested framework.

Finally, stationary effects of SIs [$A_{X\to Y}$ in Fig. 1(b)] are defined as $A^{\mathrm{KL}}_{X\to Y}(\mathbf{y}_0,\mathbf{x}_0,\mathbf{x}^*_0) = \lim_{t\to\infty} F^{\mathrm{KL}}_{X\to Y}(t,\mathbf{y}_0,\mathbf{x}_0,\mathbf{x}^*_0)$ or $A^{u,\mathrm{mean}}_{X\to Y}(\mathbf{y}_0,\mathbf{x}_0,\mathbf{x}^*_0) = \lim_{t\to\infty} F^{u,\mathrm{mean}}_{X\to Y}(t,\mathbf{y}_0,\mathbf{x}_0,\mathbf{x}^*_0)$, similarly to the above considerations. If the system Z has a single stationary distribution (an ergodic invariant measure), these A quantities are zero. If several stationary regimes exist, the dependencies of A's on $\mathbf{y}_0,\mathbf{x}_0,\mathbf{x}^*_0$ show whether a change in $\mathbf{x}_0$ can "throw" Y into the "basin of attraction" of another regime (Appendix E). Thereby, the topics of basin boundaries [49] and stability to large perturbations [50] become related to the causal coupling quantification.

To summarize, diverse meaningful characteristics of causal couplings are introduced within the single framework and many others are shown to fit into it. Their numerical values acquire a definite interpretation as particular effects of certain interventions.

IV. NUMERICAL EXAMPLES

This section addresses the following concrete questions. Is it possible to replace diverse coupling characteristics with a single one? In other words, does there exist a coupling quantifier Q such that for any two pairs of systems (X_1,Y_1) and (X_2,Y_2) from class (2), the relationship $Q_{X_1\to Y_1} > Q_{X_2\to Y_2}$ implies that "the coupling $X_1 \to Y_1$ is stronger than that $X_2 \to Y_2$" according to any meaningful characteristic? If yes, Q can be used universally to compare coupling strengths across pairs of systems. If no, the second question arises: Are diverse characteristics from the four families closely linked at least within certain classes of systems? If yes, one can develop "relatively universal" quantifiers, applicable under the respective conditions. In order to answer these questions, the F, P, and S quantities are analyzed below for a benchmark class of stochastic systems:

$$\dot{x} = -\alpha_x x + k_{xy} y + \xi_x(t), \quad \dot{y} = -\alpha_y y + k_{yx} x + \xi_y(t), \tag{14}$$

where x and y are state variables, α_x and α_y determine characteristic (relaxation) times of the systems X and Y, k_{xy} and k_{yx} are coupling coefficients, and $[\xi_x(t),\xi_y(t)]$ is a bivariate Gaussian white noise with $\Gamma_{xy} = 0$. A unidirectional coupling $X \to Y$ ($k_{xy} = 0$) is considered in Secs. IV A, IV B, and IV C. For these systems, all the conditional distributions $\rho_t(\mathbf{z}|\mathbf{z}_0)$ are Gaussian, the dynamical causal effects are found exactly via solving ordinary differential equations for the first and second conditional moments (Appendix B), and the results appear sufficient to answer the above general questions (Sec. IV D). A higher-dimensional example is given in Appendix C, bidirectional coupling is discussed in Appendix D, and nonlinear Langevin-like equations are considered in Appendix E.

A. SI-OEs and PI-SEs: Irreducible to a single coupling quantifier

For the system (14) with $k_{xy} = 0$, one derives $F^{u,\mathrm{mean}}_{X\to Y}(t) = |k_{yx}|(e^{-\alpha_y t} - e^{-\alpha_x t})/(\alpha_x - \alpha_y)$ with a maximum time $\tau^{F,u,\mathrm{mean}}_{X\to Y} = [\ln(\alpha_x/\alpha_y)]/(\alpha_x - \alpha_y)$, see Figs. 2(b) and 2(d). If the two relaxation times differ strongly, one gets

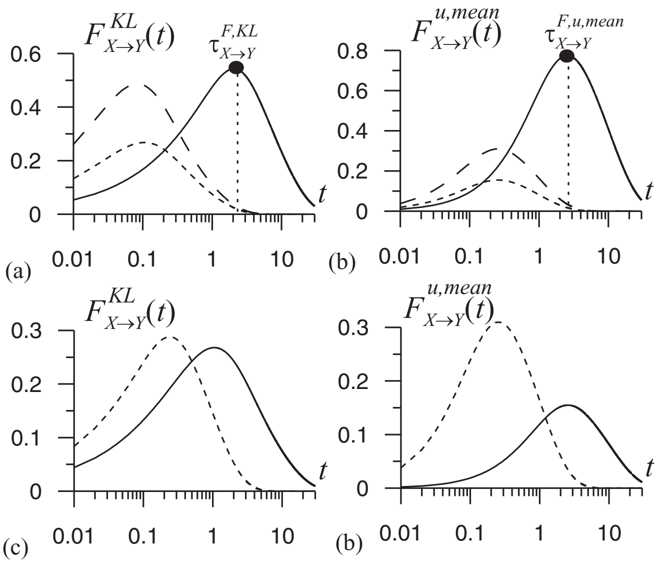

FIG. 2. Causal coupling characteristics for the system (14) with $k_{xy} = 0$ and the following other parameters: [(a) and (b)] $\alpha_y = 1$, $\Gamma_{yy} = 2$, with solid lines for $\alpha_x = 1$, $\Gamma_{xx} = 2$, $k_{yx} = 1$, short dashes for $\alpha_x = 10$, $\Gamma_{xx} = 20$, $k_{yx} = 2$, and long dashes for $\alpha_x = 10$, $\Gamma_{xx} = 20$, $k_{yx} = 4$; [(c) and (d)] $\alpha_x = 1$, $\Gamma_{xx} = 2$, with solid lines for $\alpha_y = 0.1$, $\Gamma_{yy} = 0.2$, $k_{yx} = 0.2$, and dashed lines for $\alpha_y = 10$, $\Gamma_{yy} = 20$, $k_{yx} = 4$.

$F^{u,\text{mean}}_{X\to Y} \approx |k_{yx}|/\max\{\alpha_x,\alpha_y\}$, while

$$\tau^{F,u,\text{mean}}_{X\to Y} \approx \frac{\ln(\max\{\alpha_x,\alpha_y\}/\min\{\alpha_x,\alpha_y\})}{\max\{\alpha_x,\alpha_y\}}$$

is closer to the relaxation time of the faster system. Since $F^{\text{KLvar}}_{X\to Y}(t) = 0$ for the stationary linear system, one gets $F^{\text{KL}}_{X\to Y}(t) = F^{\text{KLmean}}_{X\to Y}(t)$. The maximum time $\tau^{F,\text{KL}}_{X\to Y}$ is less than $\tau^{F,u,\text{mean}}_{X\to Y}$ [Figs. 2(a) and 2(b)], because $F^{\text{KL}}_{X\to Y}(t)$ is proportional to $F^{u,\text{mean}}_{X\to Y}(t)$ divided by the conditional variance which rises with t.

Figure 2 shows that the orbital effects depend on t, exhibiting clear maxima. Hence, if one computes the effects at a different time t (e.g., at a certain sampling interval Δt), the role of coupling may be evaluated rather differently. In particular, let us compare the results for the system (14) at two sets of parameter values. The *first case* is $\alpha_x = \alpha_y = 1$, $\Gamma_{xx} = \Gamma_{yy} = 2$, $k_{yx} = 1$ [solid lines in Figs. 2(a) and 2(b)]. The *second case* corresponds to a faster system X ($\alpha_x = 10$, $\Gamma_{xx} = 20$) and a greater coupling coefficient $k_{yx} = 2$ (short-dashed lines in Figs. 2(a) and 2(b)]. The stationary variance of the driving signal x is the same in both cases: $\sigma_x^2 = \Gamma_{xx}/(2\alpha_x) = 1$. Figures 2(a) and 2(b) show that $F^{\text{KL}}_{X\to Y}(t)$ and $F^{u,\text{mean}}_{X\to Y}(t)$ at $t \ll 0.1$ in the second case are twice as large as those in the first one. However, the OEs at $t \approx 1$ and the maximal OEs $F^{\text{KL}}_{X\to Y}$ and $F^{u,\text{mean}}_{X\to Y}$ are several times greater in the first case. Hence, the influence of the faster system X is less essential in a longer term despite a greater k_{yx} (at the same σ_x^2). This is even more evident in terms of SEs. Namely, one has $\sigma_{y,0}^2 = \Gamma_{yy}/(2\alpha_y) = 1$, where $\sigma_{y,0}^2$ denotes the variance of y in case of uncoupled X and Y. Then, the stationary variance of y reads

$$\sigma_y^2 = \sigma_{y,0}^2 + \frac{k_{yx}^2\Gamma_{xx}}{2\alpha_x\alpha_y(\alpha_x+\alpha_y)}.$$

Its relative increase as compared to $\sigma_{y,0}^2$ is $S^{\text{var}}_{X\to Y} = \frac{k_{yx}^2\Gamma_{xx}}{2\alpha_x\alpha_y(\alpha_x+\alpha_y)\sigma_{y,0}^2}$. Thus, one gets $S^{\text{var}}_{X\to Y} = 1/2$ in the first case and $4/11$ in the second one, i.e., the SE is considerably weaker in the second case, despite greater short-term OEs.

Larger maximal OEs $F^{\text{KL}}_{X\to Y}$ and $F^{u,\text{mean}}_{X\to Y}$ do not always imply a larger $S^{\text{var}}_{X\to Y}$ as well. Indeed, consider the *third case*: $k_{yx} = 4$ and all other parameters are the same as in the second case. Then $S^{\text{var}}_{X\to Y} = 16/11$ which is much greater than $1/2$ in the first case, while $F^{\text{KL}}_{X\to Y}$ and $F^{u,\text{mean}}_{X\to Y}$ [long-dashed lines, Figs. 2(a) and 2(b)] are still less than those in the first case.

A larger $F^{\text{KL}}_{X\to Y}$ at smaller $\tau^{F,\text{KL}}_{X\to Y}$ does not assure a larger $S^{\text{var}}_{X\to Y}$ as well. Indeed, consider the *fourth case* of $\alpha_x = 1, \Gamma_{xx} = 2$, $\alpha_y = 0.1, \Gamma_{yy} = 0.2$, $k_{yx} = 0.2$ [solid lines in Figs. 2(c) and 2(d)] and the *fifth case* of the same system with a faster Y ($\alpha_y = 10, \Gamma_{yy} = 20$) and a greater $k_{yx} = 4$ (dashed lines in Figs. 2(c) and 2(d)]. Then $F^{\text{KL}}_{X\to Y}$ and $F^{u,\text{mean}}_{X\to Y}$ are greater in the fifth case, but $S^{\text{var}}_{X\to Y} = 1.6/11$ in the fifth case is less than $4/11$ in the fourth one. Hence, the five cases are arranged in quite different orders according to different coupling characteristics.

Diverse possible orderings of the OEs are further illustrated by their dependencies on α for $\alpha_x = \alpha, \alpha_y = \alpha^{-2}, \Gamma_{xx} = 2\alpha_x$, $\Gamma_{yy} = 2\alpha_y$, $k_{yx} = 0.5$ [Fig. 3(a)] and all the same but fixed $\Gamma_{yy} = 2$ [Fig. 3(b)]. The quantities $F^{u,\text{mean}}_{X\to Y}$, $F^{\text{KL}}_{X\to Y}$, and $S^{\text{KL}}_{X\to Y}$ may all rise [Fig. 3(a), small α]. $F^{\text{KL}}_{X\to Y}$ and $S^{\text{KL}}_{X\to Y}$ may rise while $F^{u,\text{mean}}_{X\to Y}$ decreases [Fig. 3(a), large α]. $F^{u,\text{mean}}_{X\to Y}$ and $S^{\text{KL}}_{X\to Y}$ may rise while $F^{\text{KL}}_{X\to Y}$ decreases [Fig. 3(b), intermediate α]. Thus, a stronger effect of one type may be accompanied by weaker other effects. It means that these dynamical effects characterize different aspects of how the coupling $X \to Y$ manifests itself in the dynamics, evidencing irreducibility of diverse coupling characteristics to any single one.

B. SI-OEs and PI-SEs: Interpretations of numerical values

For the system (14) with $k_{xy} = 0$, the root-mean-squared value of the coupling term $k_{yx}\sigma_x$ can be called "driving amplitude," which is also a possible coupling characteristic. It can be shown to equal $\sqrt{\lim_{t\to 0}[F^{\text{mean}}_{X\to Y}(t)]^2/(2t)}$, where $F^{\text{mean}}_{X\to Y}(t)$ is the averaged $F^{\text{mean}}_{X\to Y}(t,\mathbf{y}_0,\mathbf{x}_0,\mathbf{x}_0^*)$, i.e., this driving amplitude also belongs to the SI-OE family. In practice, sensitivity of any coupling characteristic to changes in the driving amplitude is often an expected and desirable property. Let us check whether F's and S's rise with the driving amplitude or saturate.

Figures 3(e) and 3(f) show F's and S's versus k_{yx} and Figs. 3(i) and 3(j) show them versus $\sqrt{\Gamma_{xx}}$; recall that $\sqrt{\Gamma_{xx}/(2\alpha_x)} = \sigma_x$ for $k_{xy} = 0$. One can see that $F^{u,\text{mean}}_{X\to Y}$ is linear with respect to k_{yx} over the entire range but does not depend on Γ_{xx}, $\tau^{F,u,\text{mean}}_{X\to Y}$ being independent of both parameters. In contrast, $F^{\text{KL}}_{X\to Y}$ rises with $\sqrt{\Gamma_{xx}}$ [Fig. 3(i)]. Division by the conditional variance of y used in $F^{\text{KL}}_{X\to Y}$ (Sec. III A) is the reason for a slower increase of $F^{\text{KL}}_{X\to Y}$ at greater driving amplitudes [Figs. 3(e) and 3(i)], since the denominator rises with k_{yx} in Fig. 3(e) and with $\sqrt{\Gamma_{xx}}$ in

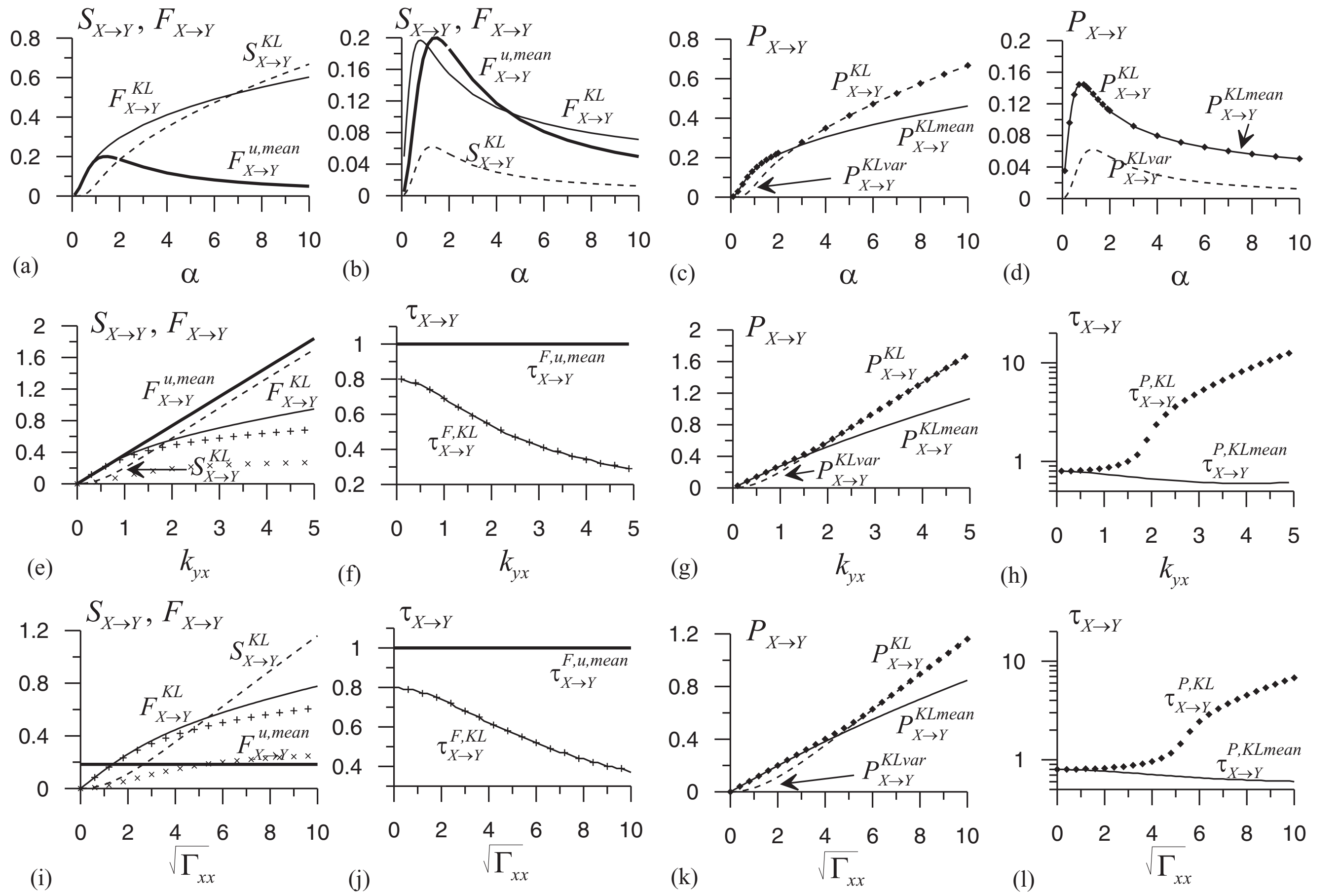

FIG. 3. Causal coupling characteristics for the system (14) with $k_{xy} = 0$: the first row for $k_{yx} = 0.5$, $\alpha_x = \alpha$, $\Gamma_{xx} = 2\alpha_x$, $\alpha_y = 1/\alpha^2$ and [(a) and (c)] $\Gamma_{yy} = 2\alpha_y$ or [(b) and (d)] $\Gamma_{yy} = 2$; the second row for $\alpha_x = \alpha_y = 1, \Gamma_{xx} = \Gamma_{yy} = 2$; the third row for $\alpha_x = \alpha_y = 1$, $\Gamma_{yy} = 2$, $k_{yx} = 0.5$. Thin solid lines show $F^{\text{KL}}_{X\to Y}$ and $P^{\text{KLmean}}_{X\to Y}$; dashed lines, $S^{\text{KL}}_{X\to Y}$ and $P^{\text{KLvar}}_{X\to Y}$; thick solid lines, $F^{u,\text{mean}}_{X\to Y}$; rhombs, $P^{\text{KL}}_{X\to Y}$. Pluses show the maximal $r_{X\to Y}$ and crosses the fixed-time $r_{X\to Y}(\Delta t = 1)$, see Appendix C.

Fig. 3(i). Therefore, $F^{\text{KL}}_{X\to Y}$ is less sensitive to changes in large driving amplitudes than in small ones. The maximum time $\tau^{F,\text{KL}}_{X\to Y}$ decreases with the driving amplitude [Figs. 3(f) and 3(j)] to provide the smaller denominator. The SE $S^{\text{KL}}_{X\to Y}$ is roughly linear in respect of the driving amplitude at its larger values, being weakly sensitive (quadratic) to small amplitudes [Fig. 3(e) and 3(i)]. Thus, the three quantities have again their own conditions for a higher sensitivity to the driving amplitude. None of them is always superior to the others.

To illustrate the last thesis with well-known approaches, consider the two above-mentioned (Sec. II A) coupling characteristics. First, MIT here equals $(1/2)k^2_{yx}(\Delta t)^2\Gamma_{xx}/\Gamma_{yy}$ for a sufficiently small sampling interval Δt. Thus, it depends on Δt, which can make comparison of the MIT values across different pairs of systems difficult to interpret in practice, where the ratios between sampling intervals and intrinsic time scales can be arbitrary. Second, the local effect of Eq. (12) in Ref. [45] represents the ratio of the coupling term to the internal dynamics term. In the integral form, it is equal to $k_{yx}\sigma_x/(\alpha_y\sigma_y)$ and does not depend on any sampling interval. However, consider an arbitrarily fast system X (almost infinite α_x) at a fixed driving amplitude ($\sigma_x = \text{const}$, no matter how large), which takes place for $\Gamma_{xx} \propto \alpha_x \to \infty$ (or just very large). Then one can easily show that all of the above SEs and finite-time OEs tend to zero; see, e.g., the formulas in the beginning of Sec. IV A. Hence, such coupling $X \to Y$ does not result in any visible effect in the dynamics of Y, so any intuitively clear characteristic should be negligibly small. However, the local effect considered exhibits a counterintuitive behavior. It does not tend to zero, but remains at a constant value, which can be kept arbitrarily large. Thus, despite both characteristics being considered meaningful and useful, they may be inappropriate in certain situations. It illustrates again that any single coupling quantifier cannot be universally applicable.

An important question is whether a concrete numerical value of a coupling characteristic, e.g., $F^{u,\text{mean}}_{X\to Y} \approx 0.8$ in Fig. 2(b), evidences a "strong" coupling. An obvious answer is that it depends on what we compare this value with. Still, a more informative reply can be achieved via a comparison of the observed value to all possible values of the characteristic within a certain class of systems. For example, let us consider how large $F^{u,\text{mean}}_{X\to Y}$ can be over all unidirectionally coupled pairs (14) with the same driven system Y ($\alpha_y = 1, \Gamma_{yy} = 2$), the same $k_{yx} = 1$, and the same $\sigma^2_x = \Gamma_{xx}/(2\alpha_x) = 1$, i.e., over all systems X with the same variance of x but different noise levels and relaxation times. The largest $F^{u,\text{mean}}_{X\to Y}$ is achieved at $\alpha_x \to 0$ (i.e., for the slowest X) and equals $|k_{yx}|/\alpha_y = 1$, while the smallest value is zero at $\alpha_x \to \infty$ (i.e., for the fastest X).

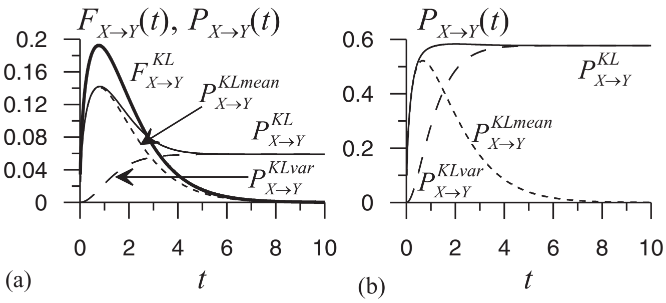

FIG. 4. Causal coupling characteristics for the system (14) with $k_{xy} = 0, \alpha_x = \alpha_y = 1, \Gamma_{xx} = \Gamma_{yy} = 2$ and (a) $k_{yx} = 0.5$ or (b) $k_{yx} = 3$. Thick solid line in panel (a) represents the coinciding plots for $F^{\mathrm{KL}}_{X\to Y}(t)$, $F^{u,\mathrm{mean}}_{X\to Y}(t)$, and $P^{u,\mathrm{mean}}_{X\to Y}(t)$; other plots are marked with the respective letters.

Thus, the value of $F^{u,\mathrm{mean}}_{X\to Y} \approx 0.8$ is closer to the maximal one, i.e., such coupling is strong enough *for the class of systems considered* with respect to the given characteristic. This is a constructive way to judge about the coupling "strength."

C. PI-OEs and a link between SI-OE and PI-SE families

Despite rather different meanings of various coupling characteristics, such as SI-OEs and PI-SEs, their plots demonstrate a similar monotone behavior in some domains, see, e.g., $\alpha > 6$ in Fig. 3(b) and $\alpha > 2.5$ in Fig. 3(e). It implies that quite different coupling quantifiers may be tightly interrelated under certain conditions. A deeper link among the SI-OE and PI-SE families through the PI-OE characteristics is revealed below.

Figure 4(a) presents F and P characteristics for the system (14) with $k_{xy} = 0$. The "unit initial state" OE of the CPI (i.e., of $k^*_{yx} = 0$) appears to equal $P^{u,\mathrm{mean}}_{X\to Y}(t) = F^{u,\mathrm{mean}}_{X\to Y}(t)$ as shown by the thick solid line. Indeed, the conditional expectation of $y(t)$ is the sum of two time-dependent functions with coefficients x_0 and y_0 [see Eq. (B2) in Appendix B], so zeroing k_{yx} at $x_0 = 1, y_0 = 0$ (corresponding to the maximal OE over all $||\mathbf{z}_0|| = 1$) produces the same mean phase orbit shift as that for the unit SI at fixed k_{yx}.

The quantity $P^{\mathrm{KL}}_{X\to Y}(t)$ [thin solid line in Fig. 4(a)] has contributions from the two components $P^{\mathrm{KLmean}}_{X\to Y}(t)$ and $P^{\mathrm{KLvar}}_{X\to Y}(t)$ [dashed lines in Fig. 4(a)]. $P^{\mathrm{KL}}_{X\to Y}(t)$ coincides with the first one at small t and with the second one at large t. In its turn, $P^{\mathrm{KLmean}}_{X\to Y}(t)$ behaves similarly to the SI-OE $F^{\mathrm{KL}}_{X\to Y}(t)$ [thick line in Fig. 4(a); numerical values of $P^{\mathrm{KLmean}}_{X\to Y}(t)$ and $F^{\mathrm{KL}}_{X\to Y}(t)$ differ just by the factor of $\sqrt{2}$ due to averaging over an initial state in P and over a difference of initial states in F], while $P^{\mathrm{KLvar}}_{X\to Y}(t)$ at large t approaches the PI-SE $S^{\mathrm{KL}}_{X\to Y}$. Thus, $P^{\mathrm{KL}}_{X\to Y}(t)$ traces a "transition" between the SI-OE and PI-SE families under an increase in t. Next, the maximal value $P^{\mathrm{KL}}_{X\to Y}$ is determined mainly by $P^{\mathrm{KLmean}}_{X\to Y}$ at small k_{yx} [solid line and short dashes in Fig. 4(a)] and approaches $P^{\mathrm{KLvar}}_{X\to Y} = S^{\mathrm{KL}}_{X\to Y}$ at greater k_{yx} [solid line and long dashes in Fig. 4(b)]. Thus, $P^{\mathrm{KL}}_{X\to Y}$ also links the SI-OE and PI-SE characteristics as its limiting cases. Therefore, $P^{\mathrm{KL}}_{X\to Y}$ combines the best sensitivities to the driving amplitude from both families and exhibits roughly linear sensitivity to k_{yx} over the whole range [Fig. 3(g), rhombs] coinciding with $P^{\mathrm{KLmean}}_{X\to Y}$ (solid line) at small k_{yx} and with $S^{\mathrm{KL}}_{X\to Y}$ (dashed line) at large k_{yx}. Similarly, $P^{\mathrm{KL}}_{X\to Y}$ combines the best sensitivities of $F^{\mathrm{KL}}_{X\to Y}$ and $S^{\mathrm{KL}}_{X\to Y}$ to changes in the noise level Γ_{xx} [Fig. 3(k)] and in α [Fig. 3(c)]. The maximum time $\tau^{P,\mathrm{KLmean}}_{X\to Y}$ [Figs. 3(h) and 3(l), solid lines) decreases with k_{yx} and Γ_{xx} similarly to $\tau^{F,\mathrm{KL}}_{X\to Y}$, but $\tau^{P,\mathrm{KL}}_{X\to Y}$ increases [Figs. 3(h) and 3(l), rhombs] since the contribution from $P^{\mathrm{KLvar}}_{X\to Y}$ exceeds that from $P^{\mathrm{KLmean}}_{X\to Y}$ at large k_{yx} and Γ_{xx}.

Note that the NPI ($\Gamma^*_{xx} = 0$) in any linear system gives $P^{\mathrm{KLmean,noise}}_{X\to Y}(t) = 0$, because it does not shift a phase orbit beam on average, but changes its diffusion rate [see Eqs. (B2) and (B3)]. Hence, $P^{\mathrm{KL,noise}}_{X\to Y}(t) = P^{\mathrm{KLvar,noise}}_{X\to Y}(t)$. The CPI-OE $P^{\mathrm{KL}}_{X\to Y}(t)$ includes a nonzero component $P^{\mathrm{KLmean}}_{X\to Y}(t)$ and therefore provides a richer coupling characterization than the NPI-OE $P^{\mathrm{KL,noise}}_{X\to Y}(t)$. Hence, the NPI-based measures, such as the widely used spectral Granger causality [31], may well be less informative than the CPI-based ones could be.

To summarize, this subsection has illustrated nontrivial links between apparently different coupling characteristics, which may be quite close to each other or differ strongly, depending on the properties of coupled systems under consideration.

D. Generality of the results

The set of F, P, and S characteristics presented for the class of systems (14) suffices to support and illustrate two rather general conclusions: irreducible diversity of coupling characteristics and nontrivial interrelations among them. Considering a larger set of coupling quantifiers or a broader class would only confirm and enrich these claims. However, even some concrete results should be the same for wider classes of systems as discussed below.

First, the "bell" shape of the temporal dependency $F^{u,\mathrm{mean}}_{X\to Y}(t)$ exhibiting an initial rise and a further decrease (Fig. 2) is common for all systems (2) with nonzero coupling $X \to Y$ and single stationary distribution, since $F^{u,\mathrm{mean}}_{X\to Y}(t)$ is then inevitably zero at $t = 0$ and at $t \to \infty$ being nonzero at least at some finite t. It is illustrated for higher-dimensional systems in Appendix C [Fig. 6(a)] and for bidirectional coupling in Appendix D (Fig. 7). Similarly, a bell-shaped dependence of the normalized quantity $F^{\mathrm{KL}}_{X\to Y}(t)$ holds for any Langevin-like system (3) with single stationary distribution, since the white noise provides normalization which gives vanishing $F^{\mathrm{KL}}_{X\to Y}(t)$ at $t \to 0$ as well.

Second, a tight mutual dependence between $P^{\mathrm{KL}}_{X\to Y}(t)$ and $F^{\mathrm{KL}}_{X\to Y}(t)$ at small t holds, at least, for any Langevin-like system (3) with additive coupling. As an illustration, consider one-dimensional y with $f_y(\mathbf{x},y) = h(y) + k_{yx}g(\mathbf{x})$ in Eq. (3). At small t and any k_{yx}, one has approximately Gaussian conditional distributions with $\mathrm{var}[y(t)|\mathbf{x}_0,y_0] \approx \Gamma_{yy}t$. Then one derives from (4) that $F^{\mathrm{KL}}_{X\to Y}(t) \approx |k_{yx}|\sqrt{\langle[g(\mathbf{x}_0) - g(\mathbf{x}^*_0)]^2\rangle_{\mathbf{x}_0,\mathbf{x}^*_0}/(2\Gamma_{yy}t)} = |k_{yx}|\sqrt{\mathrm{var}[g]/(\Gamma_{yy}t)}$. Similarly, one gets from (10): $P^{\mathrm{KL}}_{X\to Y}(t) \approx |k_{yx}|\sqrt{\langle g^2(\mathbf{x}_0)\rangle_{\mathbf{x}_0}/(2\Gamma_{yy}t)}$. If the coupling term has zero mean $\langle g(\mathbf{x}_0)\rangle_{\mathbf{x}_0} = 0$, then $F^{\mathrm{KL}}_{X\to Y}(t) \approx \sqrt{2}P^{\mathrm{KL}}_{X\to Y}(t)$. Everything is the same for higher-dimensional systems with additive coupling. For other classes, the correspondence between short-term $F(t)$ and $P(t)$ may be not so close but is still expected to hold approximately. Next, $P^{\mathrm{KL}}_{X\to Y}(t)$ tends to the PI-SE at $t \to \infty$ by definition. Thus, the PI-OE characteristic again links the SI-OE and PI-SE families for quite a broad class of systems.

Overall, the suggested framework applies to the basic class (2), which is rather general as noted in Sec. III. Therefore, it shapes the broad field of causal coupling quantification allowing to interpret various characteristics in a unified manner. In particular, the proposed set of time-resolved and stationary "dynamical causal effects" is shown to be a flexible toolkit, more informative than any single coupling quantifier.

V. DISCUSSION

To complete the consideration, let us discuss estimation issues and summarize both current gains from the suggested framework and its implications for a further research.

A. Estimation of dynamical causal effects

If a time series of complete state vectors $\mathbf{x}(t)$ and $\mathbf{y}(t)$ is recorded in an established regime at a given $\mathbf{a}$, the SI-OE measures such as $F^{\mathrm{KL}}_{X\to Y}$ can be estimated directly following their definition, where averaging over the stationary distribution is replaced by averaging over time. No mathematical model is needed, apart from the assumption that $\mathbf{x}(t)$ and $\mathbf{y}(t)$ specify a complete state of the combined system in the sense of Markov property.

Temporal averages may be insufficient to estimate other global characteristics. Thus, for the quantity $F^{u,\mathrm{mean}}_{X\to Y}$, some imposed values of $\mathbf{x}^*_0$ entering its definition (4) may be absent from observed data. An example is given by observations of a weakly perturbed synchronization regime if one is interested in SI-OEs for stronger perturbations. The first possibility is to record another time series under stronger interventions as it was done for coupling detection in Refs. [1,10,51]. The second possibility is to fit a mathematical model from a certain class to the data (e.g., with the aid of Markovian approximation [52], Bayesian inference [8,10,53], etc.) and estimate SI-OEs by using virtual SIs in the model. In such a case, the model equations are, in essence, extrapolated to unobserved domains of the state space. Validity of the extrapolation can be justified only by additional experiments or *a priori* theoretical knowledge. Still, it is worth noting that regimes close to synchrony represent a difficult case for causal coupling detection, see, e.g., Refs. [9,33]. The suggested framework may not help to detect couplings in such a case but may assist in avoiding misinterpretations of coupling estimation results. If incomplete states are observed, SI-OEs can be estimated again via model fitting and virtual SIs.

The SI-SE characteristics can be directly estimated from a time series, only if it contains perturbations throwing the state of the system between different basins of attraction. Otherwise, model fitting and extrapolation to the domains of unobserved attractors are necessary. The PI-OE and PI-SE characteristics cannot be directly estimated from a time series recorded at a single value of $\mathbf{a}$, since their definition compares the dynamics at two different parameter values. Hence, their estimation requires either real PIs (as implemented in Ref. [54] for coupling detection) or model fitting and virtual PIs. In the latter case, model equations are extrapolated to unobserved domains in the parameter space.

All such extrapolations of an empirical model can be a source of inaccuracies and errors. However, a model-based component is inevitably involved in interpretations of apparently model-free characteristics as well (Sec. II B). Explicit model fitting may be advantageous, since it clearly shows what can be potentially corrected in the analysis, if necessary.

B. Current gains and implications for further studies

The framework of dynamical causal effects integrates many previously known and newly introduced causal coupling characteristics. In particular, transfer entropy and other Granger causality measures, momentary information transfer, information flow, and phase dynamics modeling approaches [33] either belong directly to the SI-OE family or can be considered as approximations of its representatives (Appendix A). Spectral Granger causality [31], multiple regression-based methods [19], and various synchronization and bifurcation diagrams belong to (or may represent) the PI-SE family. The PI-OE family includes, in particular, the long-term causality approach [20] and various model-based studies such as assessments of climate sensitivity to emission scenarios [47]. The SI-SE family relates to the wide fields of basins of attractions [49] and stability to large perturbations [50]. The suggested perspective provides all these approaches with a unified interpretation in terms of "intervention-effect." Via an analysis of their distinctions and interrelations, it is revealed that no single quantity is universally applicable to characterize "coupling strength." In contrast, diversity of causal coupling characteristics is necessary to answer many practical questions about dynamical role of couplings. Thus, instead of coupling strength estimation, it is often more fruitful to study how the coupling works in the dynamics.

The suggested perspective also reveals that a completely model-free approach to causal coupling quantification universally applicable to compare "coupling strengths" across pairs of different systems is not feasible. First, the very detection of causal couplings as well as an interpretation of their quantitative measures imply certain limitations imposed on the underlying evolution laws, i.e., model assumptions (Sec. II). Second, in order to say whether a coupling is strong according to a certain measure, it is meaningful to compare an estimated value of that measure to its possible values over a set of situations, i.e., to perform an analysis within a model class (Sec. IV B). In general, any causal coupling inference seems to be inevitably model based. However, deeper relationships among various coupling characteristics for specific classes of systems may provide "relatively model-free" coupling quantifiers that are most informative and relevant under certain realistic conditions. The analysis of MIT in Ref. [22] may be considered a step in this direction.

It is worth noting that existence of the finite-time maximum of $F^{\mathrm{KL}}_{X\to Y}$ (5) at $\tau^{F,\mathrm{KL}}_{X\to Y}$ explains why one-step-ahead prediction-based techniques for coupling detection may be less sensitive to weak couplings than multi-step-ahead ones [14,24]. Indeed, a prediction time close to $\tau^{F,\mathrm{KL}}_{X\to Y}$ must provide maximal sensitivity due to maximal separation of the future conditional distributions. Such a temporal dependence of coupling quantifiers is often underestimated in time series analysis, while it is clearly interpretable from the suggested perspective.

More complicated relationships among various characteristics are expected for bidirectional couplings (Appendix D), where the suggested perspective promises to be especially useful for detailed coupling characterization and proper interpretation of the results.

Deterministic dynamical systems represent an important special case of the basic class (2): They exhibit Dirac-delta conditional distributions. As a result, the KL distance (4) becomes infinitely large so only the mean OEs take finite values and can be informative. However, if one considers initial states specified up to a certain finite error (coarse-grained state spaces), the conditional distributions become smeared and the above formalism applies. Still, deterministic systems may have their own peculiarities deserving a special study.

As for possible extensions of the class (2), which is the core of the suggested framework, the whole consideration readily generalizes to several coupled systems. Then the framework should be relevant to introduce clear distinctions between various possible interventions and effects, which are often lacking. Next, infinite-dimensional states can be considered in the same manner, as soon as norms and distributions in the infinite-dimensional state space are defined. The suggested framework provides a relevant perspective for such studies.

VI. CONCLUSIONS

This work suggests a general conceptual framework for quantification of causal couplings between evolving systems based on finite-dimensional state space representation. It unifies different approaches considering them from the "intervention-effect" perspective and arranging them into four families according to two types of virtual interventions (state space and parametric) and two types of dynamical causal effects (orbital and stationary). Multiple well-known measures (transfer entropy, momentary information transfer, spectral Granger causality, long-term causality, etc.) are shown to represent one of those families.

It is shown that a set of diverse characteristics is relevant to quantify different aspects of coupling manifestations in dynamics. As a representation, novel "intervention-dynamical effect" characteristics are consistently introduced within each family. Nontrivial interrelations among various characteristics and their important properties, such as nonmonotone temporal dependence of the orbital effects, are revealed. In general, different characteristics are not interchangeable and may not be unambiguously replaced with a single universal quantifier. Together, they reveal "how the coupling works in dynamics" rather than "how strong the coupling is," reformulating the very question about the coupling strength.

It is argued that any attempt to develop a model-free approach to causal coupling quantification is, in essence, oriented to a certain class of systems and inevitably involves model assumptions to interpret the results. To avoid misinterpretations originating from the lack of explicit formulation of such assumptions underlying an *apparently model-free* technique, a model-based approach seems to be relevant. Yet it is useful to search for "relatively universal" (applicable to a wider class) and still-informative characteristics [22].

Overall, the suggested framework shapes the broad field of causal coupling quantification and can guide a researcher through multiple ideas and techniques to choose an appropriate tool for a problem at hand, refusing the idea of a single, universal, and model-free "coupling strength" quantifier.

ACKNOWLEDGMENTS

The work is supported by the government of Russia (Agreement No. 14.Z50.31.0033 dated 03/24/2014), the Russian Foundation for Basic Research (Project No. 14-02-00492), and the president of Russia (Project No. NSh 1726.2014.2).

APPENDIX A: MEMBERS OF THE SI-OE FAMILY

This Appendix argues that several well-known measures, including TE-based ones, belong to the SI-OE family. First, from the definition of the complete TE (Sec. II A) and Eqs. (4) and (5), one can show that $F^{\mathrm{KL}}_{X\to Y}(t)$ over a sampling interval $t = \Delta t$ is equal to

$$F^{\mathrm{KL}}_{X\to Y}(\Delta t) = \sqrt{T_{X\to Y} + T^*_{X\to Y}}, \tag{A1}$$

where $T^*_{X\to Y} = \langle D(\rho(\mathbf{y}_n|\mathbf{y}_{n-1}))||\rho(\mathbf{y}_n|\mathbf{x}_{n-1},\mathbf{y}_{n-1})\rangle_{\mathbf{x}_{n-1},\mathbf{y}_{n-1}}$ is the averaged KL divergence between the same distributions as those in $T_{X\to Y}$ but taken in the reverse order. Hence, the TE can be derived as a component of $F^{\mathrm{KL}}_{X\to Y}(\Delta t)$, i.e., belongs to the SI-OE family.

If the relative difference between $\rho(\mathbf{x}_{n-1},\mathbf{y}_n|\mathbf{y}_{n-1})$ and $\rho(\mathbf{x}_{n-1}|\mathbf{y}_{n-1})\rho(\mathbf{y}_n|\mathbf{y}_{n-1})$ is uniformly small (much less than unity), one derives $T^*_{X\to Y} \approx T_{X\to Y}$ directly from the definitions and, thereby, the simple relation $F^{\mathrm{KL}}_{X\to Y}(\Delta t) \approx \sqrt{2T_{X\to Y}}$. However, $T^*_{X\to Y}$ gets large if there are domains in the space $(\mathbf{x}_{n-1},\mathbf{y}_n)$, where the joint probability density $\rho(\mathbf{x}_{n-1},\mathbf{y}_n|\mathbf{y}_{n-1})$ is close to zero while both individual densities $\rho(\mathbf{y}_n|\mathbf{y}_{n-1})$ and $\rho(\mathbf{x}_{n-1}|\mathbf{y}_{n-1})$ are not small. Hence, the value of $F^{\mathrm{KL}}_{X\to Y}(\Delta t)$ is more sensitive than $T_{X\to Y}$ to such couplings, which make some domains in $(\mathbf{x}_{n-1},\mathbf{y}_n)$ "depopulated" at a certain $\mathbf{y}_{n-1}$. In particular, an increase in coupling coefficient can make some domains less populated in such a way that $F^{\mathrm{KL}}_{X\to Y}(\Delta t)$ tends to infinity while $T_{X\to Y}$ saturates at a finite value or increases less rapidly (Appendix C). If the joint probability density is exactly zero in a certain domain, $F^{\mathrm{KL}}_{X\to Y}(\Delta t)$ becomes infinite. For such cases, one can use $F^{\mathrm{KLmean}}_{X\to Y}(\Delta t)$ (8) and $F^{\mathrm{KLvar}}_{X\to Y}(\Delta t)$ (9) instead of $F^{\mathrm{KL}}_{X\to Y}(\Delta t)$.

The apparent TE along with its modifications [13,24,30] as well as MIT, information flow [41], "local information flow" [42], and linear Granger causality measures [25] are justified as causal coupling characteristics, because they are versions or approximations of the complete TE. Hence, they relate to the SI-OE family as well. In particular, MIT corresponds to the distribution of $\mathbf{x}^*_0$ in Eq. (5) centered around $\mathbf{x}_0$ with variance determined by the noise ξ_x realization over an interval Δt. The phase dynamics modelling-based characteristics [33] are similar to the above approaches with a distinction that phases are used as the only state variables, which are sufficient to represent states of oscillatory systems under general conditions (see, e.g., Ref. [48]). Thus, one should attribute these measures to the SI-OE family.

Widely used coupling characteristics based on time-delay embedding and nearest-neighbor statistics [9] do not fit well into the "intervention-effect" framework. They use the fact that for a weak unidirectional coupling $X \to Y$ between low-dimensional deterministic systems [5], a time-delay-embedded vector from the driving system X is a unique function of a simultaneous time-delay-embedded vector from Y. Concrete measures of the $X \to Y$ influence quantify closeness of the X vectors corresponding to close Y vectors. Hence, their numerical values characterize quite a specific consequence of the causal coupling. They seem to be related to the SI-OE family, since only an interdependence between simultaneous pieces of the two time series is analyzed. However, such a relationship deserves further study.

APPENDIX B: COMPUTING DYNAMICAL CAUSAL EFFECTS

This Appendix gives concrete formulas to compute dynamical causal effects for a general linear stochastic system

$$\dot{\mathbf{z}} = \mathbf{A}\mathbf{z} + \xi(t), \tag{B1}$$

where ξ is Gaussian white noise with zero mean and covariance $\langle \xi(t_1)\xi'(t_2)\rangle = \Gamma\delta(t_1 - t_2)$, the prime denotes transposition. At $\mathbf{z}(0) = \mathbf{z}_0$, the conditional distribution $\rho_t(\mathbf{z}|\mathbf{z}_0)$ at any $t > 0$ is Gaussian with expectation $\mathbf{m}_{\mathbf{z}|\mathbf{z}_0}(t)$ and covariance matrix $\mathbf{C}_{\mathbf{zz}|\mathbf{z}_0}(t)$ given by

$$\dot{\mathbf{m}}_{\mathbf{z}|\mathbf{z}_0}(t) = \mathbf{A}\mathbf{m}_{\mathbf{z}|\mathbf{z}_0}(t) \tag{B2}$$

and

$$\dot{\mathbf{C}}_{\mathbf{zz}|\mathbf{z}_0}(t) = \mathbf{A}\mathbf{C}_{\mathbf{zz}|\mathbf{z}_0}(t) + \mathbf{C}_{\mathbf{zz}|\mathbf{z}_0}(t)\mathbf{A}' + \Gamma, \tag{B3}$$

where $\mathbf{m}_{\mathbf{z}|\mathbf{z}_0}(0) = \mathbf{z}_0$ and $\mathbf{C}_{\mathbf{zz}|\mathbf{z}_0}(0) = \mathbf{0}$. These are linear equations which can be solved via either algebraic tools or numerical integration. In particular, the conditional expectation is a linear function of the initial condition $\mathbf{m}_{\mathbf{z}|\mathbf{z}_0}(t) = \mathbf{B}(t)\mathbf{z}_0$, and the matrix $\mathbf{B}(t)$ can be found by integrating Eq. (B2) for several (d_Z) unit initial vectors. Separating two components $\mathbf{x}$ and $\mathbf{y}$ of the vector $\mathbf{z}$, one gets $\mathbf{B}(t)$ consisting of the blocks $\mathbf{B}_{\mathbf{xx}}(t)$, $\mathbf{B}_{\mathbf{xy}}(t)$, $\mathbf{B}_{\mathbf{yx}}(t)$, $\mathbf{B}_{\mathbf{yy}}(t)$, where $\mathbf{B}_{\mathbf{yx}}(t)$ determines a dependence of the expectation of $\mathbf{y}(t)$ on $\mathbf{x}_0$. $\mathbf{C}_{\mathbf{zz}|\mathbf{z}_0}(t)$ consists of the "on-diagonal" blocks $\mathbf{C}_{\mathbf{xx}|\mathbf{z}_0}(t)$ and $\mathbf{C}_{\mathbf{yy}|\mathbf{z}_0}(t)$ and "off-diagonal" $\mathbf{C}_{\mathbf{xy}|\mathbf{z}_0}(t)$ and $\mathbf{C}_{\mathbf{yx}|\mathbf{z}_0}(t)$, where $\mathbf{C}_{\mathbf{xx}|\mathbf{z}_0}(t)$ and $\mathbf{C}_{\mathbf{yy}|\mathbf{z}_0}(t)$ are covariance matrices for the distributions $\rho_t(\mathbf{x}|\mathbf{z}_0)$ and $\rho_t(\mathbf{y}|\mathbf{z}_0)$, respectively. The inverse of $\mathbf{C}_{\mathbf{yy}|\mathbf{z}_0}(t)$ is needed to calculate $F^{\mathrm{KL}}_{X\to Y}$ (4). For the stationary distribution, one derives $\mathbf{m}^{st}_{\mathbf{z}} = \mathbf{0}$ and covariance matrix $\mathbf{C}^{\mathrm{st}}_{\mathbf{zz}}$ satisfying $\mathbf{A}\mathbf{C}^{\mathrm{st}}_{\mathbf{zz}} + \mathbf{C}^{\mathrm{st}}_{\mathbf{zz}}\mathbf{A}' = -\Gamma$. The conditional stationary distribution $\rho_{\mathrm{st}}(\mathbf{x}|\mathbf{y})$ has covariance matrix $\mathbf{C}^{\mathrm{st}}_{\mathbf{xx}|\mathbf{y}} = (\mathbf{K}^{\mathrm{st}}_{\mathbf{xx}} - \mathbf{K}^{\mathrm{st}}_{\mathbf{xy}}\mathbf{C}^{\mathrm{st}}_{\mathbf{yy}}\mathbf{K}^{\mathrm{st}}_{\mathbf{yx}})^{-1}$, where $\mathbf{K}^{\mathrm{st}}_{\mathbf{zz}} = (\mathbf{C}^{\mathrm{st}}_{\mathbf{zz}})^{-1}$ and its blocks are denoted $\mathbf{K}^{\mathrm{st}}_{\mathbf{xx}}$, $\mathbf{K}^{\mathrm{st}}_{\mathbf{xy}}$, $\mathbf{K}^{\mathrm{st}}_{\mathbf{yx}}$, and $\mathbf{K}^{\mathrm{st}}_{\mathbf{yy}}$. Then one derives $[F^{\mathrm{KL}}_{X\to Y}(t,\mathbf{y}_0,\mathbf{x}_0,\mathbf{x}^*_0)]^2 = (\mathbf{x}_0 - \mathbf{x}^*_0)'\mathbf{B}'_{\mathbf{yx}}(t)\mathbf{C}^{-1}_{\mathbf{yy}|\mathbf{z}_0}(t)\mathbf{B}_{\mathbf{yx}}(t)(\mathbf{x}_0 - \mathbf{x}^*_0)/2$ and $[F^{\mathrm{KL}}_{X\to Y}(t)]^2 = \mathrm{tr}\{\mathbf{B}'_{\mathbf{yx}}(t)\mathbf{C}^{-1}_{\mathbf{yy}|\mathbf{z}_0}(t)\mathbf{B}_{\mathbf{yx}}(t)\mathbf{C}^{\mathrm{st}}_{\mathbf{xx}|\mathbf{y}}\}$, where $\mathrm{tr}\{\cdot\}$ denotes trace of a matrix.

Other measures are found similarly, e.g., $r^2_{X\to Y}(t)$ (Appendix C) is obtained from the formula for $[F^{\mathrm{KL}}_{X\to Y}(t)]^2$ via replacing $\mathbf{C}_{\mathbf{yy}|\mathbf{z}_0}(t)$ with $\mathbf{C}_{\mathbf{yy}|\mathbf{y}_0}(t) = \mathbf{C}_{\mathbf{yy}|\mathbf{z}_0}(t) + \mathbf{B}_{\mathbf{yx}}(t)\mathbf{C}^{\mathrm{st}}_{\mathbf{xx}|\mathbf{y}}\mathbf{B}'_{\mathbf{yx}}(t)$. The mean OE reads $[F^{\mathrm{mean}}_{X\to Y}(t,\mathbf{y}_0,\mathbf{x}_0,\mathbf{x}^*_0)]^2 = (\mathbf{x}_0 - \mathbf{x}^*_0)'\mathbf{B}'_{\mathbf{yx}|\mathbf{z}_0}(t)\mathbf{B}_{\mathbf{yx}|\mathbf{z}_0}(t)(\mathbf{x}_0 - \mathbf{x}^*_0)$. Then, $[F^{u,\mathrm{mean}}_{X\to Y}(t)]^2 = \max_{||\Delta\mathbf{x}_0||=1}\{(\mathbf{x}_0 - \mathbf{x}^*_0)'\mathbf{B}'_{\mathbf{yx}|\mathbf{z}_0}(t)\mathbf{B}_{\mathbf{yx}|\mathbf{z}_0}(t)(\mathbf{x}_0 - \mathbf{x}^*_0)\}$ can be found as the largest singular value of the matrix $\mathbf{B}'_{\mathbf{yx}|\mathbf{z}_0}(t)\mathbf{B}_{\mathbf{yx}|\mathbf{z}_0}(t)$. One gets $P(t,\mathbf{z}_0,\mathbf{a},\mathbf{a}^*)$ and other P quantities by solving Eqs. (B2) and (B3) at $\mathbf{a}$ and $\mathbf{a}^*$. S measures are obtained from the stationary distributions for the two parameter values.

APPENDIX C: GRANGER CAUSALITY AND ORBITAL EFFECTS

This Appendix compares Granger causality "in mean" [25] to F and P quantities. For one-dimensional Gaussian processes x and y (14), the complete TE reads $T_{X\to Y} = (1/2)\ln[1/(1 - r^2_{X\to Y}(\Delta t))]$, where $r^2_{X\to Y}(\Delta t) = \mathrm{cov}(x_n, y_{n+1}|y_n)/\sqrt{\mathrm{var}(x_n|y_n)\mathrm{var}(y_{n+1}|y_n)}$, $\mathrm{cov}(\cdot)$ is covariance. The quantity $r_{X\to Y}(\Delta t)$ is called "partial correlation" between x_n and y_{n+1}, being useful as a causality measure [23]. For the linear system, $r^2_{X\to Y}(\Delta t)$ equals $\mathrm{var}(y_{n+1}|x_n,y_n)/\mathrm{var}(y_{n+1}|y_n)$, i.e., a part of y_{n+1} variance explained by x_n, given y_n. In other words, $r^2_{X\to Y}(\Delta t)$ measures prediction improvement, which is achieved if x_n is taken into account in addition to y_n. A usual measure of Granger causality $G^2_{X\to Y}$ is very similar to $r^2_{X\to Y}(\Delta t)$, differing only by the full-history conditioning as discussed in Sec. II B.

One can derive that $r^2_{X\to Y}(\Delta t) = \frac{\langle(m_y(\Delta t,y_n,x_n) - m_y(\Delta t,y_n,x^*_n))^2\rangle_{x_n,x^*_n}}{2\mathrm{var}(y_{n+1}|y_n)}$ and, hence, it is a SI-OE measure differing from $[F^{\mathrm{KL}}_{X\to Y}(\Delta t)]^2 = \frac{\langle(m_y(\Delta t,y_n,x_n) - m_y(\Delta t,y_n,x^*_n))^2\rangle_{x_n,x^*_n}}{2\mathrm{var}(y_{n+1}|y_n,x_n)}$ only by the denominator. Denote a maximum of $r^2_{X\to Y}(t)$ over t as $r^2_{X\to Y}$. Then, for the linear system, the simple relation $[F^{\mathrm{KL}}_{X\to Y}]^2 = r^2_{X\to Y}/(1 - r^2_{X\to Y})$ holds true. However, $F^{\mathrm{KL}}_{X\to Y}$ is more sensitive to changes in large driving amplitudes. Indeed, $r_{X\to Y}$ for the system (14) with unidirectional coupling is shown by pluses in Figs. 3(e) and 3(i). It shows a stronger tendency to saturation with k_{yx} and Γ_{xx} than does $F^{\mathrm{KL}}_{X\to Y}$ (thin solid lines). The fixed-time $r_{X\to Y}(\Delta t = 1)$ saturates almost perfectly [Figs. 3(e) and 3(i), crosses] and is unable to distinguish among large coupling coefficients and among large noise levels in the driving system.

For a further comparison, Fig. 5(a) shows that under a decrease in the driven system noise level Γ_{yy}, $r_{X\to Y}$ saturates at unity (pluses), while $F^{\mathrm{KL}}_{X\to Y}$ (thin solid line) and $P^{\mathrm{KL}}_{X\to Y}$ (thick solid line) tend to infinity according to power laws $F^{\mathrm{KL}}_{X\to Y} \propto \Gamma^{-1/4}_{yy}$ and $P^{\mathrm{KL}}_{X\to Y} \propto \Gamma^{-1/2}_{yy}$. $P^{\mathrm{KLmean}}_{X\to Y} \propto \Gamma^{-1/2}_{yy}$ as well (dashed line) but the stationary effect $S^{\mathrm{KL}}_{X\to Y}$ appears stronger and gives the main contribution to $P^{\mathrm{KL}}_{X\to Y}$. Thus, $F^{\mathrm{KL}}_{X\to Y}$ and

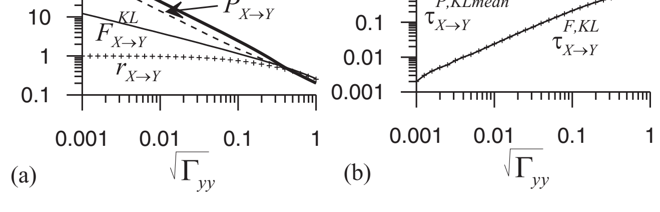

FIG. 5. Causal coupling characteristics for the system (14) with $k_{xy} = 0, \alpha_x = \alpha_y = 1, \Gamma_{xx} = 2, k_{yx} = 0.5$. Pluses show $r_{X\to Y}$ and its maximum times.

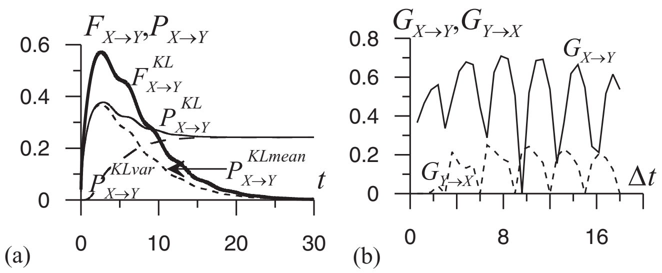

FIG. 6. Various coupling characteristics for the system (C1) with $\Gamma_{xx} = \Gamma_{yy} = 1$, $\omega_x^2 = 1.1$, $\omega_y^2 = 0.9$, and (a) $k_{yx} = 0.5$, $\alpha_x = 0.3$, $\alpha_y = 0.25$, or (b) $k_{yx} = 0.3$, $\alpha_x = \alpha_y = 0.05$. $G_{Y\to X}$ (dashed line) and $G_{X\to Y}$ (solid line) are computed with the exact method of Ref. [38].

especially PI-OEs are much more sensitive to changes in Γ_{yy} than is the partial correlation coefficient. The maximum time [Fig. 5(b)] decreases with decreasing noise for $F^{\mathrm{KL}}_{X\to Y}$ and for $r_{X\to Y}$ as $\tau^{F,\mathrm{KL}}_{X\to Y} \propto \Gamma_{yy}^{1/2}$. The maximum time saturates around the systems' relaxation time for $P^{\mathrm{KLmean}}_{X\to Y}$ and at a greater value for $P^{\mathrm{KL}}_{X\to Y}$. Note that for $\Gamma_{yy} \to 0$ one gets $r_{X\to Y} \to 1$ and $[F^{\mathrm{KL}}_{X\to Y}]^2 \approx 1/(1 - r^2_{X\to Y})$. It follows that $T_{X\to Y} \approx \ln F^{\mathrm{KL}}_{X\to Y}(\Delta t)$ in this case. The latter shows that $F^{\mathrm{KL}}_{X\to Y}$ is more sensitive than the TE to variations in small Γ_{yy}.

Another advantage of the model-based dynamical causal effects over the "model-free" $G_{X\to Y}$ and apparent TE concerns the very detection of causal couplings and avoiding false positives [36]. Consider stochastic linear oscillators

$$\begin{aligned} \dot{x}_1 &= x_2, \quad \dot{x}_2 = -\omega_x^2 x_1 - 2\alpha_x x_2 + \xi_x(t), \\ \dot{y}_1 &= y_2, \quad \dot{y}_2 = -\omega_y^2 y_1 - 2\alpha_y y_2 + k_{yx} x_1 + \xi_y(t), \end{aligned} \tag{C1}$$

where observables are $u = x_1$ and $v = y_1$ while state vectors are two dimensional, α_x and α_y are damping coefficients, oscillation frequencies are $\omega_x^2 = 1.1$ and $\omega_y^2 = 0.9$, noise intensities $\Gamma_{xx} = \Gamma_{yy} = 1$, and coupling is unidirectional $X \to Y$. Figure 6(a) shows that SI-OEs and PI-OEs in the "correct" direction $X \to Y$ for this higher-dimensional case exhibit a curve with a large-scale "bell shape" and superposed oscillations. All SI-OEs and PI-OEs in the opposite direction $Y \to X$ are zero as directly follows from their definition. In contrast, Fig. 6(b) shows that the one-step-ahead prediction improvement in the "spurious" direction $G_{Y\to X}$ (dashed line) can be positive and large depending on the sampling interval Δt, sometimes being comparable to the "true" $G_{X\to Y}$ (solid line) and even much greater, e.g., at $\Delta t = 9.6$.

APPENDIX D: BIDIRECTIONAL COUPLING

This Appendix shows that a nonzero coupling k_{xy} in the direction $Y \to X$ in Eq. (14) changes the causal effects $X \to Y$ in a complex way in comparison with the case of unidirectional coupling $X \to Y$. Thus, $F^{u,\mathrm{mean}}_{X\to Y}(t)$ at $k_{xy} = 0$ [Fig. 7(a), thick solid line] is less than that at $k_{xy} = 0.5$ (thin solid line) and greater than that at $k_{xy} = -1$ (dashed line). Indeed, the distribution of y at $t > 0$, given the initial state, is influenced by all intermediate values of x and y over the interval $(0,t)$, which depend on both coupling coefficients k_{xy} and k_{yx}.

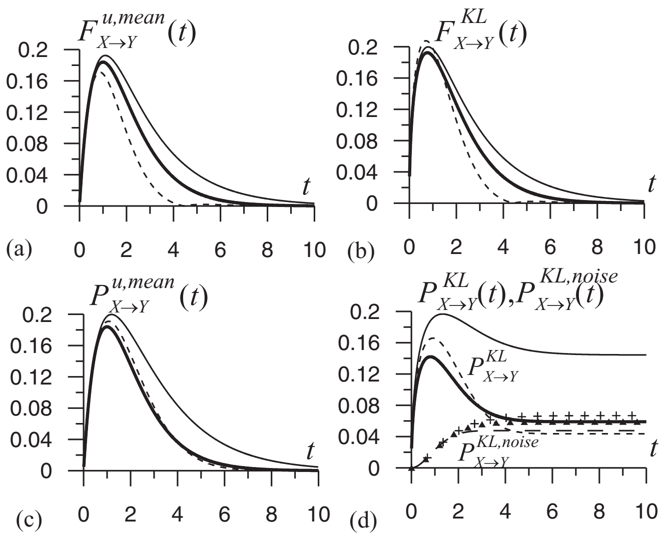

FIG. 7. Orbital effects for Eqs. (14) with $k_{yx} = 0.5$, $\alpha_x = \alpha_y = 1, \Gamma_{xx} = \Gamma_{yy} = 2$: [(a), (b), and (c)] thick solid lines for $k_{xy} = 0$, thin solid lines for $k_{xy} = 0.5$, dashed lines for $k_{xy} = -1$; (d) thick solid line [the coupling PI-OE $P^{\mathrm{KL}}_{X\to Y}(t)$] and triangles [the noise level PI-OE $P^{\mathrm{KL,noise}}_{X\to Y}(t)$] correspond to $k_{xy} = 0$, thin solid line and pluses to $k_{xy} = 0.5$, and short dashes and long dashes to $k_{xy} = -1$.

Dependencies of various dynamical causal effects $X \to Y$ on the opposite coupling coefficient k_{xy} may differ from each other. Thus, $F^{\mathrm{KL}}_{X\to Y}(t)$ [Fig. 7(b)] is maximal over the three cases at $k_{xy} = -1$, while $P^{u,\mathrm{mean}}_{X\to Y}$ is minimal at $k_{xy} = 0$ and maximal at $k_{xy} = 0.5$ [Fig. 7(c)]. Moreover, $P^{\mathrm{KL}}_{X\to Y}(t)$ and $P^{\mathrm{KL,noise}}_{X\to Y}(t)$ coincide at large t for a unidirectional coupling [Fig. 7(d), thick solid line and triangles] but differ from each other for a bidirectional coupling: Figure 7(d) shows that the CPI-OE $P^{\mathrm{KL}}_{X\to Y}(t)$ (thin solid line at large t) is greater than the NPI-OE $P^{\mathrm{KL,noise}}_{X\to Y}(t)$ (pluses) at $k_{xy} = 0.5$, while the situation is reversed at $k_{xy} = -1.0$ (short-dashed and long-dashed lines). Indeed, the CPI-SE $S^{\mathrm{var}}_{X\to Y}$ depends on the two coupling coefficients in a nonmonotone way since $\sigma_y^2 = \sigma_{y,0}^2 + \frac{\alpha_x k_{yx}(k_{yx}\sigma_{x,0}^2 + k_{xy}\sigma_{y,0}^2)}{(\alpha_x+\alpha_y)(\alpha_x\alpha_y - k_{xy}k_{yx})}$, where the denominator is always positive for stationary processes while the second term in the numerator may well be negative and give $\sigma_y^2 < \sigma_{y,0}^2$. The latter occurs in the last example [dashed lines in Fig. 7(d)], where $S^{\mathrm{var}}_{X\to Y}$ is negative and $S^{\mathrm{KL}}_{X\to Y}$ (a horizontal asymptote for short dashes) is less than $S^{\mathrm{KL,noise}}_{X\to Y}$ (an asymptote for long dashes).

Figure 8 shows F, P, and S characteristics versus k_{yx} at $k_{xy} = -1$. One can see that all F's and P's [Fig. 8(a)] reflect an increase in $|k_{yx}|$ reliably over that range. The behavior of S's is not as simple. The NPI-SE $S^{\mathrm{KL,noise}}_{X\to Y}$ is positive and exhibits a monotone increase with $|k_{yx}|$ [dashed line in Fig. 8(b)], being weakly sensitive (quadratic) to changes in small k_{yx}. The CPI-SE $S^{\mathrm{KL}}_{X\to Y}$ [thin solid line in Fig. 8(b)] is more sensitive to changes in small couplings and nonsymmetric in respect of the sign of k_{yx}. The signed quantity $S^{\mathrm{var}}_{X\to Y}$ may be even strongly negative [thick solid line in Fig. 8(b)]. Such a complex behavior is not a disadvantage of $S^{\mathrm{KL}}_{X\to Y}$ or $S^{\mathrm{var}}_{X\to Y}$, which reflects different aspects of how the coupling $X \to Y$ affects the dynamics of Y under other equal conditions including a nonzero k_{xy}.

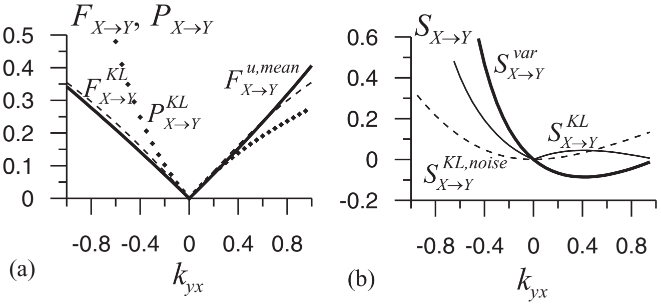

FIG. 8. Dynamical causal effects for the system (14) with $k_{xy} = -1$, $\alpha_x = \alpha_y = 1$, $\Gamma_{xx} = \Gamma_{yy} = 2$: (a) $F^{u,\text{mean}}_{X\to Y}$ (solid line), $F^{\text{KL}}_{X\to Y}$ (dashed line), $P^{\text{KL}}_{X\to Y}$ (rhombs); (b) $S^{\text{KL}}_{X\to Y}$ (thin solid line), $S^{\text{KL,noise}}_{X\to Y}$ (dashed line), $S^{\text{var}}_{X\to Y}$ (thick solid line).

APPENDIX E: SI-SE CHARACTERISTICS

All SI-SE measures A's (Sec. III C) are zero for stationary stochastic linear systems, since the latter possess a single stationary distribution. However, the situation differs for a nonlinear system with multiple stationary distributions. Consider a nonlinear example of a bistable damped oscillator Y driven by a linear relaxation system X:

$$\dot{x} = -\alpha_x x + \xi_x(t), \quad \dot{y} = \alpha_y(y - y^3) + k_{yx}x + \xi_y(t). \tag{E1}$$

Being isolated and noise-free, the system Y has two stable equilibria $y = \pm 1$ and an unstable equilibrium $y = 0$. From an initial state $y_0 < 0$, it evolves to the attractor $y = -1$. From $y_0 > 0$, it goes to $y = 1$. Let the noise ξ_y be sufficiently weak so there are two practically isolated stationary distributions around those two points. This is the case if the amplitude of fluctuations around the fixed point is small enough: $\sqrt{\Gamma_{yy}/\alpha_y} \ll 1$. Then the point $y = 0$ is an approximate boundary between the "basins of attraction" of those two distributions. Let the system X be much slower than Y, i.e., $\alpha_x \ll \alpha_y$. Then the state of X remains $x(t) \approx x_0$ for a relatively long time and enters the equation for y as a constant, which shifts the boundary. Hence, one easily derives that for $x_0 = 1$ and $\sqrt{\Gamma_{yy}/\alpha_y} \ll |k_{yx}|/\alpha_y \ll 1$, the system Y evolves to the negative attractor if $y_0 < -k_{yx}/\alpha_y$.

This consideration allows an easy calculation of the quantity $A^{u,\text{mean}}_{X\to Y}(y_0,x_0)$. If we take $x_0 = 0$ for brevity, then a unit SI means $x_0^* = \pm 1$. Since the expectation of y equals ± 1 for the two attracting stationary distributions, one gets $A^{u,\text{mean}}_{X\to Y}(y_0,0) \approx 2$, if $|y_0| < |k_{yx}|/\alpha_y$, and $A^{u,\text{mean}}_{X\to Y}(y_0,0) \approx 0$, otherwise. Thus, the quantity $A^{u,\text{mean}}_{X\to Y}(y_0,0)$ as a function of y_0 may serve as an indicator of the basin boundary. In essence, SI-SE characteristics assess whether a system X can throw the system Y from one basin of attraction to another one and how probable it is. They closely relate to the concept of stability to large perturbations, which is currently a subject of growing attention [50]. Such a link between the basins of attraction analysis [49,50] and causal coupling quantification enriches the latter concept.

[1] J. Pearl, *Causality: Models, Reasoning, and Inference* (Cambridge University Press, Cambridge, 2000).

[2] Y.-C. Hung and C.-K. Hu, Phys. Rev. Lett. **101**, 244102 (2008).

[3] M. Prokopenko, J. T. Lizier, and D. C. Price, Entropy **15**, 524 (2013); M. Prokopenko and J. T. Lizier, Sci. Rep. **4**, 5394 (2014).

[4] B. P. Bezruchko, V. I. Ponomarenko, A. S. Pikovsky, and M. G. Rosenblum, Chaos **13**, 179 (2003); I. T. Tokuda, S. Jain, I. Z. Kiss, and J. L. Hudson, Phys. Rev. Lett. **99**, 064101 (2007); B. Kralemann, L. Cimponeriu, M. Rosenblum, A. Pikovsky, and R. Mrowka, Phys. Rev. E **76**, 055201 (2007); **77**, 066205 (2008); B. P. Bezruchko and D. A. Smirnov, *Extracting Knowledge from Time Series: An Introduction to Nonlinear Empirical Modeling* (Springer-Verlag, Berlin, 2010).

[5] G. Sugihara, R. May, H. Ye *et al.*, Science **338**, 496 (2012).

[6] A. Porta, G. Baselli, F. Lombardi *et al.*, Biol. Cybern. **81**, 119 (1999); M. Palus and A. Stefanovska, Phys. Rev. E **67**, 055201(R) (2003); M. Palus and M. Vejmelka, *ibid.* **75**, 056211 (2007); A. Bahraminasab, F. Ghasemi, A. Stefanovska, P. V. E. McClintock, and H. Kantz, Phys. Rev. Lett. **100**, 084101 (2008).

[7] M. G. Rosenblum, L. Cimponeriu, A. Bezerianos, A. Patzak, and R. Mrowka, Phys. Rev. E **65**, 041909 (2002); A. S. Karavaev, M. D. Prokhorov, V. I. Ponomarenko *et al.*, Chaos **19**, 033112 (2009); B. Kralemann, M. Fruehwirth, A. Pikovsky *et al.*, Nat. Commun. **4**, 2418 (2013).

[8] Y. F. Suprunenko, P. T. Clemson, and A. Stefanovska, Phys. Rev. Lett. **111**, 024101 (2013); P. T. Clemson, Y. F. Suprunenko, T. Stankovski, and A. Stefanovska, Phys. Rev. E **89**, 032904 (2014).

[9] S. J. Schiff, P. So, T. Chang, R. E. Burke, and T. Sauer, Phys. Rev. E **54**, 6708 (1996); J. Arnhold, K. Lehnertz, P. Grassberger, and C. E. Elger, Physica D **134**, 419 (1999); R. Quian Quiroga, J. Arnhold, and P. Grassberger, Phys. Rev. E **61**, 5142 (2000); T. Kreuz, F. Mormann, R. G. Andrzejak *et al.*, Physica D **225**, 29 (2007); R. G. Andrzejak and T. Kreuz, Europhys. Lett. **96**, 50012 (2011).

[10] K. J. Friston, L. Harrison, and W. Penny, NeuroImage **19**, 1273 (2003); D. A. Pinotsis, R. J. Moran, and K. J. Friston, *ibid.* **59**, 1261 (2012).

[11] P. A. Tass, Biol. Cybern. **89**, 81 (2003); D. A. Smirnov, U. B. Barnikol, T. T. Barnikol *et al.*, Europhys. Lett. **83**, 20003 (2008); B. P. Bezruchko, V. I. Ponomarenko, M. D. Prokhorov *et al.*, Phys. Uspekhi **51**, 304 (2008); E. Sitnikova, T. Dikanev, D. Smirnov *et al.*, J. Neurosci. Methods **170**, 245 (2008).

[12] E. Pereda, R. Quian Quiroga, and J. Bhattacharya, Progr. Neurobiol. **77**, 1 (2005); B. Schelter, M. Winterhalder, M. Eichler *et al.*, J. Neurosci. Methods **152**, 210 (2006); J. Prusseit and K. Lehnertz, Phys. Rev. E **77**, 041914 (2008); G. Nolte, A. Ziehe, V. V. Nikulin, A. Schlogl, N. Kramer, T. Brismar, and K. R. Muller, Phys. Rev. Lett. **100**, 234101 (2008); A. N. Pavlov, A. E. Hramov, A. A. Koronovskii *et al.*, Phys. Uspekhi **55**, 845 (2012).

[13] M. Staniek and K. Lehnertz, Phys. Rev. Lett. **100**, 158101 (2008); S. Stramaglia, G. R. Wu, M. Pellicoro, and D. Marinazzo, Phys. Rev. E **86**, 066211 (2012).

[14] I. Vlachos and D. Kugiumtzis, Phys. Rev. E **82**, 016207 (2010); L. Faes, G. Nollo, and A. Porta, *ibid.* **83**, 051112 (2011); M. V. Sysoeva, E. Sitnikova, I. V. Sysoev *et al.*, J. Neurosci. Meth. **226**, 33 (2014).

[15] M. Wibral, B. Rahm, M. Rieder *et al.*, Prog. Biophys. Mol. Biol. **105**, 80 (2011); R. Vicente, M. Wibral, M. Lindner, and G. Pipa, J. Comp. Neurosci. **30**, 45 (2011).

[16] W. Wang, B. T. Anderson, R. K. Kaufmann, and R. B. Myneni, J. Climate **17**, 4752 (2004); T. J. Mosedale, D. B. Stephenson, M. Collins, and T. C. Mills, *ibid.* **19**, 1182 (2006); I. I. Mokhov and D. A. Smirnov, Geophys. Res. Lett. **33**, L03708 (2006); I. I. Mokhov, D. A. Smirnov, P. I. Nakonechny *et al.*, *ibid.* **38**, L00F04 (2011).

[17] U. Triacca, Theor. Appl. Climatol. **81**, 133 (2005); I. I. Mokhov and D. A. Smirnov, Izvestiya Atmos. Ocean. Phys. **44**, 263 (2008); Doklady Earth Sci. **427**, 798 (2009); E. Kodra, S. Chatterjee, and A. R. Ganguly, Theor. Appl. Climatol. **104**, 325 (2011); I. I. Mokhov, D. A. Smirnov, and A. A. Karpenko, Doklady Earth Sci. **443**, 381 (2012).

[18] P. F. Verdes, Phys. Rev. Lett. **99**, 048501 (2007); F. Donges, Y. Zou, N. Marwan, and J. Kurths, Europhys. Lett. **87**, 48007 (2009); M. van der Mheen, H. A. Dijkstra, A. Gozolchiani *et al.*, Geophys. Res. Lett. **40**, 2714 (2013); Y. Wang, A. Gozolchiani, Y. Ashkenazy, Y. Berezin, O. Guez, and S. Havlin, Phys. Rev. Lett. **111**, 138501 (2013).

[19] J. L. Lean and D. H. Rind, Geophys. Res. Lett. **35**, L18701 (2008); **36**, L15708 (2009).

[20] D. A. Smirnov and I. I. Mokhov, Phys. Rev. E **80**, 016208 (2009).

[21] J. Runge, J. Heitzig, V. Petoukhov, and J. Kurths, Phys. Rev. Lett. **108**, 258701 (2012).

[22] J. Runge, J. Heitzig, N. Marwan, and J. Kurths, Phys. Rev. E **86**, 061121 (2012).

[23] J. Runge, J. Kurths, and V. Petoukhov, J. Climate **27**, 720 (2014).

[24] M. Palus, Phys. Rev. Lett. **112**, 078702 (2014).

[25] C. W. J. Granger, J. Econ. Dynam. Control **2**, 329 (1980).

[26] C. W. J. Granger, Inform. Control **6**, 28 (1963).

[27] C. W. J. Granger, Econometrica **37**, 424 (1969).

[28] N. Ancona, D. Marinazzo, and S. Stramaglia, Phys. Rev. E **70**, 056221 (2004); D. Marinazzo, M. Pellicoro, and S. Stramaglia, Phys. Rev. Lett. **100**, 144103 (2008).

[29] T. Schreiber, Phys. Rev. Lett. **85**, 461 (2000); K. Hlavackova-Schindler, M. Palus, M. Vejmelka, and J. Bhattacharya, Phys. Rep. **441**, 1 (2007).

[30] A. B. Barrett, L. Barnett, and A. K. Seth, Phys. Rew E **81**, 041907 (2010); L. Barnett and T. Bossomaier, Phys. Rev. Lett. **109**, 138105 (2012); L. Barnett, J. T. Lizier, M. Harre, A. K. Seth, and T. Bossomaier, *ibid.* **111**, 177203 (2013).

[31] J. Geweke, J. Am. Stat. Assoc. **77**, 304 (1982).

[32] A. Brovelli, M. Ding, A. Ledberg *et al.*, Proc. Natl. Acad. Sci. USA **101**, 9849 (2004); M. Dhamala, G. Rangarajan, and M. Ding, Phys. Rev. Lett. **100**, 018701 (2008).

[33] M. G. Rosenblum and A. S. Pikovsky, Phys. Rev. E **64**, 045202(R) (2001); D. A. Smirnov and B. P. Bezruchko, *ibid.* **68**, 046209 (2003); J. Brea, D. F. Russell, and A. B. Neiman, Chaos **16**, 026111 (2006); D. A. Smirnov and B. P. Bezruchko, Phys. Rev. E **79**, 046204 (2009); B. Kralemann, M. Rosenblum, and A. Pikovsky, Chaos **21**, 025104 (2011).

[34] D. N. Reshef, Y. N. Reshef, H. K. Finucane, S. R. Grossman, G. McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher, and P. C. Sabeti, Science **334**, 1518 (2011).

[35] C. A. Sims, Econometrica **39**, 545 (1971).

[36] D. A. Smirnov and B. P. Bezruchko, Europhys. Lett. **100**, 10005 (2012).

[37] H. Nalatore, M. Ding, and G. Rangarajan, Phys. Rev. E **75**, 031123 (2007); D. W. Hahs and S. D. Pethel, Phys. Rev. Lett. **107**, 128701 (2011).

[38] D. A. Smirnov, Phys. Rev. E **87**, 042917 (2013).

[39] D. W. Hahs and S. D. Pethel, Entropy **15**, 767 (2013).

[40] *Climate Change 2007: The Physical Science Basis*, edited by S. Solomon *et al.* (Cambridge University Press, Cambridge, 2007).

[41] N. Ay and D. Polani, Adv. Complex Syst. **11**, 17 (2008).

[42] J. T. Lizier and M. Prokopenko, Eur. Phys. J. B **73**, 605 (2010).

[43] A. Kolmogoroff, Math. Ann. **104**, 415 (1931); I. I. Gihman and A. V. Skorohod, *The Theory of Stochastic Processes* (Springer, Berlin, 1975).

[44] T. Cover and J. Thomas, *Elements of Information Theory* (Wiley, NewYork, 1991); A. A. Borovkov, *Mathematical Statistics* (Gordon & Breach, New York, 1998).

[45] J. T. C. Schwabedal and A. Pikovsky, Phys. Rev. E **81**, 046218 (2010).

[46] L. Barnett, A. B. Barrett, and A. K. Seth, Phys. Rev. Lett. **103**, 238701 (2009).

[47] R. W. Bodman, P. J. Rayner, and D. J. Karoly, Nat. Clim. Change **3**, 725 (2013).

[48] A. S. Pikovsky, M. G. Rosenblum, and J. Kurths, *Synchronization: A Universal Concept in Nonlinear Sciences* (Cambridge University Press, Cambridge, 2001).

[49] H. E. Nusse and J. A. Yorke, Science **271**, 1376 (1996); T. M. Lenton, H. Held, E. Kriegler *et al.*, Proc. Natl. Acad. Sci. USA **105**, 1786 (2008).

[50] P. Menck, J. Heitzig, N. Marwan, and J. Kurths, Nat. Phys. **9**, 89 (2013); Y. Zou, T. Pereira, M. Small, Z. Liu, and J. Kurths, Phys. Rev. Lett. **112**, 114102 (2014); P. J. Menck, J. Heitzig, J. Kurths, and H. J. Schellnhuber, Nat. Commun. **5**, 3969 (2014).

[51] Z. Levnajic and A. Pikovsky, Phys. Rev. Lett. **107**, 034101 (2011).

[52] S. Kravtsov, D. Kondrashov, and M. Ghil, J. Climate **18**, 4404 (2005); D. Kondrashov, S. Kravtsov, and M. Ghil, *ibid.* **18**, 4425 (2005); A. J. Majda and J. Harlim, Nonlinearity **26**, 201 (2013); D. Kondrashov, M. D. Checkroun, and M. Ghil, Physica D (to be published).

[53] D. N. Mukhin, A. M. Feigin, E. M. Loskutov, and Ya. I. Molkov, Phys. Rev. E **73**, 036211 (2006); E. M. Loskutov, Ya. I. Molkov, D. N. Mukhin, and A. M. Feigin, *ibid.* **77**, 066214 (2008); Y. I. Molkov, E. M. Loskutov, D. N. Mukhin, and A. M. Feigin, *ibid.* **85**, 036216 (2012).

[54] M. Timme, Phys. Rev. Lett. **98**, 224101 (2007).