Information Flow Rate for Cross-Correlated Stochastic Processes

Dionissios T. Hristopulos, *Senior Member, IEEE*

***Abstract*—Causal inference seeks to identify cause-and-effect interactions in coupled systems. A recently proposed method by Liang detects causal relations by quantifying the direction and magnitude of information flow between time series. The theoretical formulation of information flow for stochastic dynamical systems provides a general expression and a data-driven statistic for the rate of entropy transfer between different system units. To advance understanding of information flow rate in terms of intuitive concepts and physically meaningful parameters, we investigate statistical properties of the data-driven information flow rate between coupled stochastic processes. We derive relations between the expectation of the information flow rate statistic and properties of the auto- and cross-correlation functions. Thus, we elucidate the dependence of the information flow rate on the analytical properties and characteristic times of the correlation functions. Our analysis provides insight into the influence of the sampling step, the strength of cross-correlations, and the temporal delay of correlations on information flow rate. We support the theoretical results with numerical simulations of correlated Gaussian processes.**

***Index Terms*—Information flow, causality analysis, Gaussian process, entropy, covariance kernel.**

I. Introduction

THERE is great interest in the inference of causal relations between variables based on time series data. Applications of causal inference extend across scientific disciplines such as neuroscience [1], [2], [3], bioinformatics [4], machine learning [5], [6], [7], climate and Earth system sciences [8], [9], [10]. Classical measures of statistical association, such as the linear Pearson and the rank order (Spearman) correlation coefficients fail to provide information about the directionality of interactions. Hence, they can not reveal cause-effect relations. In fact, it is well known that non-zero statistical correlations can be observed even in the absence of causal relations. The latter are characterized by two key properties: (i) temporal precedence, i.e., the cause precedes the effect, and (ii) physical influence, i.e., variations of the cause have an impact on the effect [11]. Causal analysis investigates methods that can determine cause-effect relationships [12], [13], [14], [15], [16], [17].

Manuscript received 13 June 2023; revised 9 November 2023; accepted 8 January 2024. Date of publication 25 January 2024; date of current version 7 February 2024. The publication of the article in OA mode was financially supported by HEAL-Link. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. George Atia.

The author is with the Department of Electrical and Computer Engineering, Technical University of Crete, 73100 Chania, Greece (e-mail: dchristopoulos@tuc.gr).

Digital Object Identifier 10.1109/TSP.2024.3358580

Various methods of causal inference are in use. They include *Wiener-Granger causality* (WGC) [18], [19], [20], [21] which is a standard tool for analyzing brain connectivity [22], [23], [24], [25], kernel WGC [25], [26] which generalizes WGC to nonlinear interactions, entropy-based methods [14], [27] such as transfer entropy (TE) and mutual information [28], [29], [30], [31], convergent cross mapping (CCM) which is based on the theory of dynamical systems [32], and PCMCI (Peter and Clark algorithm followed by momentary conditional independence test) which is based on Pearl's graphical model framework and is applicable to nonlinear time series [33], [34].

The *Liang information flow rate* (IFR), is a non-parametric causality measure between two potentially interacting time series [35], [36], [37], [38]. The IFR formulation is based on Shannon entropy and the theory of dynamical systems. It leads to general expressions for information flow between different variables which involve time-dependent expectations over joint probability density functions [35], [36], [37], [39]. The IFR has been formulated—at least for two dimensional systems—using both absolute and relative entropy [36], [37]. The relative entropy (Kullback–Leibler divergence) quantifies the amount of information added to a system with respect to the initial probability distribution. IFR was recently extended to describe information flow in quantum mechanical systems [40].

In order to derive an estimate of IFR based on observations, Liang used a linear stochastic dynamical system that satisfies the following first-order stochastic differential equation (also known as Langevin equation)

$$\mathrm{d}\mathbf{X}(t) = \mathbf{f}\,\mathrm{d}t + \mathbf{A}\,\mathbf{X}(t)\,\mathrm{d}t + \mathbf{B}\,\mathrm{d}\mathbf{W}(t)\,, \tag{1}$$

where $\mathrm{d}\mathbf{X}(t)$ is the differential of the two-dimensional stochastic process $\mathbf{X}(t)$, $\mathbf{f}$ is a 2×1 constant advection vector, $\mathbf{A}$ is a 2×2 matrix of (inverse) time constants, $\mathbf{B}$ is a 2×2 diffusion matrix, and $\mathrm{d}\mathbf{W}(t)$ is the differential of a 2×1 vector Wiener process (Brownian motion). Assuming that the observations represent a discretely sampled realization of $\mathbf{X}(t)$, he derived an estimate of IFR by maximizing the likelihood [37]. This data-driven IFR involves only the observed time series and their finite differences. Hence, in contrast with transfer entropy which requires estimating bivariate probability distributions, and WGC which requires estimating potentially large and computationally expensive autoregressive systems—the complexity scales as $\mathcal{O}(N^3D^3p^3)$, where N is the sample size, D is the number of components, and p is the autoregressive order [41]—IFR has lower computational complexity.

The data-driven IFR has been proven to recover causal relations in benchmark nonlinear systems [42]. It has also been used to infer brain connectivity from EEG recordings [43] and fMRI data [44], to determine causal relations between global temperature and CO2 concentration [45], to investigate interactions between climate modes (El Niño and the Indian Ocean Dipole) [37], to reconstruct sea surface height using satellite altimetry [46], and to discover causal relations between the stocks of companies traded in the stock market [38]. Nonetheless, there are still gaps in our understanding of the data-driven IFR.

Machine learning methods such as Gaussian process regression, employ non-parametric models that do not assume knowledge of dynamical equations. For applications in which directional connectivity matters, it is important to select models that reflect the causal relations of the system. This necessity further motivates studying connections between IFR and stochastic processes. In particular, the connection between different covariance kernel models and IFR properties has not been investigated. Knowledge of such relations can guide the selection of suitable cross-covariance kernels for data-driven models based on Gaussian processes. The dependence of IFR estimates and their uncertainty on statistical properties of the observed time series and different sampling steps are also important for practical applications. In addition, the dependence of IFR on covariance kernel parameters can provide intuitive understanding of physical factors that influence IFR.

We seek to understand how mathematical properties (of the covariance kernels) affect IFR. Kernel properties can be learned from time series data using statistical estimation methods and optimal model selection approaches [47]. Our analysis links the parameters and properties of the auto- and cross-covariance kernels with IFR. The results are based on ensemble expectations that provide accurate IFR estimates in the ergodic limit. We obtain an IFR expression in terms of spectral moments, and we investigate how the continuity and differentiability of the stochastic processes impact IFR. We also explore the dependence of IFR on the characteristic correlation times of the kernels and the sampling step which enables a deeper understanding of information flow rate measurements. Numerical simulations of cross-correlated Gaussian processes are also conducted to validate the theoretical results.

The manuscript is organized as follows: Section II presents notation and definitions. Section III focuses on permissibility conditions for the kernels of cross-correlated processes; these are necessary for constructing and simulating mathematically admissible models with specified properties. In Section IV we discuss Liang's data-driven IFR and its main mathematical properties. Section V derives explicit equations for the IFR of cross-correlated, second-order ergodic processes (which satisfy the conditions laid out in Section III). In Section VI we focus on mean-square differentiable stochastic processes: we obtain the continuous-sampling limit (which is applicable for very small sampling step) and leading-order corrections for finite time step. Section VI investigates IFR properties for mean-square-continuous (but non-differentiable) processes. These are suitable models for first-order stochastic systems driven by Gaussian white noise. Section VIII presents numerical simulations which validate the main results of the theoretical analysis in the preceding sections. Discussion related to IFR interpretation and limitations of the analysis is given in Section IX. Finally, Section X presents the main conclusions.

A. Summary of Main Results

The results obtained in this study for IFR between coupled (cross-correlated) stochastic processes are summarized below.

1) We define an expression for the equilibrium IFR which is valid under ergodic conditions (Theorem V.1). The equilibrium IFR allows connecting information flow with statistical properties of the processes involved that can be inferred from the data. We then calculate the equilibrium IFR in terms of spectral moments (Theorem V.2). The spectral formulation provides existence conditions for the equilibrium IFR that involve the spectral density tail.
2) For small sampling steps (compared to the characteristic correlation times), we obtain limit expressions of the equilibrium IFR (Theorems VI.1, VI.2, VII.1). These expressions depend on the regularity of the stochastic processes but are independent of the sampling step.
3) We establish that the equilibrium IFR vanishes for the popular class of separable cross-correlation models (Proposition V.1). Hence, such models are not suitable for causal analysis studies.
4) Our analysis proposes an interpretation of the IFR sign in the framework of stochastic processes. The interpretation is related to the regularity of the processes (Theorems VI.2, VII.1, VII.2).
5) For processes with time-delayed correlations, we show that the IFR is a non-monotonic function of the time delay over the characteristic correlation time: it peaks for a value of the ratio less than one and tends to zero as the ratio increases (Figs. 3 and 5).

II. Preliminaries

We use lowercase boldface symbols for vectors and uppercase boldfaced letters for matrices and vector stochastic processes (indexed by time). The transpose of matrix $\mathbf{A}$ is denoted by $\mathbf{A}^\top$, its inverse by $\mathbf{A}^{-1}$, and the determinant by $\det \mathbf{A}$. The sets of real and non-negative real numbers are denoted by $\mathbb{R}$ and $\mathbb{R}_{\geq 0}$ respectively. The set of natural numbers is denoted by $\mathbb{N}$ and that of all integers by $\mathbb{Z}$.

Definition 1 (Vector stochastic process): A vector stochastic process with $D \in \mathbb{N}$ components will be denoted by $\mathbf{X}(t) \triangleq [X_1(t), \ldots, X_D(t)]^\top$, where the symbol $\triangleq$ is used to define a mathematical entity. The stochastic process is defined on a probability space $(\Omega, \mathcal{F}, P)$ and indexed by the time $t \in \mathcal{T} \subset \mathbb{R}$, where $\mathcal{T}$ is an ordered set. The index $i = 1, \ldots, D$ in $X_i(t)$ selects a specific scalar stochastic process.

The *expectation* of $X_i(t)$ over the ensemble space Ω is denoted by $m_i(t) \triangleq \mathbb{E}[X_i(t)]$. The *fluctuation* is denoted by $X_i'(t) \triangleq X_i(t) - \mathbb{E}[X_i(t)]$. Specific vector states (realizations) are denoted by $\mathbf{x}(t)$ and their components by $x_i(t)$, for $i = 1, \ldots, D$.

Definition 2 (Auto- and cross-covariance functions): The covariance functions, $C_{i,j}(t,t+\tau) \triangleq \mathbb{C}\mathrm{ov}\,\{X_i(t), X_j(t+\tau)\}$, of two scalar processes $X_i(t)$ and $X_j(t)$ are defined by the expectation (for $i,j=1,\ldots,D$)

$$C_{i,j}(t,t+\tau) = \mathbb{E}[X_i'(t)\,X_j'(t+\tau)], \tag{2}$$

where τ is the temporal lag, for all $t \in T$. For $i=j$ the *auto-covariance* is obtained, while for $i \neq j$ (2) defines the *cross-covariance*. $\sigma_i^2 = C_{i,i}(0)$, also denoted as $\mathbb{V}\mathrm{ar}[X_i(t)]$, is the variance of the i-th component. The *auto-correlation functions (ACFs)* are then defined by $\rho_{i,i}(\tau) = C_{i,i}(\tau)/\sigma_i^2$ $(i=1,\ldots D)$, while the *cross-correlation functions (CCFs)* are given by $\rho_{i,j}(\tau) = C_{i,j}(\tau)/\sigma_i\sigma_j$, for $i \neq j = 1,\ldots,D$.

Definition 3 (Time series): A stochastic vector process sampled at a discrete set of times will be denoted by $\{\mathbf{X}_{t_n}\}_{n=1}^N$. Sampled values at t_n will be denoted by $\mathbf{x}_n$ or $x_{i;n}$, where $n \in \mathbb{N}$ is the time index. For a uniform time step δt, the sampling times are $t_n = n\delta t$. The set $\{x_{i,n}\}_{n=1}^N$, where $x_{i,n} \triangleq x_i(t_n)$ denotes a sample (time series) of the i-th stochastic process over $\{t_n\}_{n=1}^N$.

Definition 4 (Stationarity): A vector stochastic process $\mathbf{X}(t)$ is weakly or second-order stationary (henceforward, stationary) iff the following conditions hold: (i) $\mathbb{E}[\mathbf{X}(t)] = \mathbf{m} \in \mathbb{R}^D$ and (ii) $\mathbb{E}[X_i'(t-\tau)\,X_j'(t)] = C_{i,j}(\tau)$ for all $i,j=1,\ldots,D$. The minus sign preceding the time lag in $X_i(t-\tau)$ defines the CCF consistently with its use in the IFR formulation, see (6) below.

Definition 5 (Sample covariance functions): We use the "hat" symbol for sample-based estimates of statistical quantities. The sampling covariance function of the time series $X_i(t)$, $X_j(t)$ (where $i,j=1,\ldots,D$) is given by

$$\hat{C}_{i,j}(k\delta t) \triangleq \frac{1}{N-k}\sum_{n=k+1}^{N} x_{i,n-k}\,x_{j,n} - \overline{x_i}\,\overline{x_j}, \quad k \in \mathbb{Z}, \tag{3}$$

where the "overline" denotes the sample average, i.e., $\overline{x_1} = \frac{1}{N}\sum_{n=1}^N x_{1,n}$. Equation (3) gives the sample auto-covariance for $i=j$ and the cross-covariance for $i \neq j$. The *sampling correlation functions* are defined as follows

$$\hat{\rho}_{i,j}(k\delta t) \triangleq \frac{\hat{C}_{i,j}(k\delta t)}{\sqrt{\hat{C}_{i,i}(0)\,\hat{C}_{j,j}(0)}}. \tag{4}$$

Definition 6 (Nonnegative definiteness): A real, symmetric, $D\times D$ matrix $\mathbf{A}$ is nonnegative definite if for any real-valued vector $\mathbf{z} \in \mathbb{R}^D$ it holds that $\mathbf{z}^\top\mathbf{A}\mathbf{z} \geq 0$. A function $C(t,t') : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ is nonnegative definite iff $\sum_{i=1}^n\sum_{j=1}^n z_i C(t_i,t_j) z_j \geq 0$ for all $n \in \mathbb{N}$, all time vectors $(t_1,\ldots,t_N) \in \mathbb{R}^n$ and all vectors $(z_1,\ldots z_n) \in \mathbb{R}^n$.

Definition 7 (Fourier transforms): The *Fourier transform (FT)* of a function $C(\tau) : \mathbb{R} \to \mathbb{R}$ that is absolutely integrable over the interval $(-\infty,\infty)$ is given by

$$\widetilde{C}(\omega) = \int_{-\infty}^{\infty} \mathrm{d}\tau\, C(\tau)\, e^{-\imath\,\omega\tau},$$

where $\imath = \sqrt{-1}$ is the imaginary unit. In addition, if the function $C(\cdot)$ is of bounded variation in an interval which contains τ, i.e., if $C(\cdot)$ has at most a finite number of extrema and discontinuities within this interval, $C(\tau)$ is given by the inverse Fourier transform (IFT) [48]

$$C(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}\omega\, \widetilde{C}(\omega)\, e^{\imath\omega\tau}.$$

A continuous function $C(\tau)$ is a permissible covariance kernel for some stationary scalar stochastic process if and only if it is nonnegative definite. Bochner's theorem [49] provides easily testable permissibility conditions.

Theorem II.1 (Bochner's permissibility theorem): Let $C : \mathbb{R} \to \mathbb{R}$ be a continuous, absolutely integrable function. The function $C(\cdot)$ is non-negative definite if and only if its Fourier transform $\widetilde{C} \triangleq \mathrm{FT}[C]$ is nonnegative and integrable over $\mathbb{R}$.

The ergodic property allows replacing sample (temporal) with ensemble averages (i.e., expectations). It is useful in practical studies, because often only a single realization (sample) is available. We use ergodicity to investigate IFR. Slutsky's theorem provides necessary and sufficient conditions for second-order ergodic processes.

Theorem II.2 (Slutsky's theorem): A second-order stationary, *Gaussian* vector stochastic process $\mathbf{X}(t)$ is second-order ergodic if and only if $C_{i,j}(\tau) \to 0$ as $\tau \to \infty$ for all $i,j = 1,\ldots,D$ [50, pp. 526-533].

Remark 1: For non-Gaussian processes, second-order ergodicity requires conditions on higher-order moments than the covariance.

Notation (Information flow): We denote information flow from the *driver process* $X_1(t)$ to a *receiver process* $X_2(t)$ by means of $X_1 \to X_2$. $X_1 \to X_2$ implies that the equation which determines the dynamic evolution of X_2 depends on X_1. In the context of (1), such dependence is expressed by means of $A_{2,1} \neq 0$ and/or $B_{2,1} \neq 0$. Information flow $X_1 \to X_2$ does not imply information flow in the reverse direction $X_2 \to X_1$.

III. Permissibility of Covariance Kernels for Multivariate Processes

Bochner's theorem [49] applies to scalar, second-order stationary processes. For vector stochastic processes, permissibility requires the stricter conditions of Cramér's theorem [51]. This will be used below to construct valid separable and non-separable (time-delayed) covariance kernels for cross-correlated processes.

Theorem III.1 (Cramer's Theorem): The continuous matrix function $\mathbf{C} : \mathbb{R} \to \mathbb{R}^{D\times D}$ is a valid matrix covariance kernel for a continuous, stationary, stochastic vector process, if the following conditions hold for the matrix components $C_{i,j}(\tau)$, for $i,j \in \{1,\ldots,D\}$:

- (C1) The functions $C_{i,j} : \mathbb{R} \to \mathbb{R}$ are absolutely integrable (i.e., the FTs $\widetilde{C}_{i,j}$ exist) for all $i,j=1,\ldots,D$.
- (C2) The functions $C_{i,i} : \mathbb{R} \to \mathbb{R}$ satisfy Bochner's theorem for all $i=1,\ldots,D$.
- (C3) The *cross-spectral densities* have bounded variation, i.e., the integrals $\int_{\mathbb{R}} \mathrm{d}\omega \left|\widetilde{C}_{i,j}(\omega)\right|$ are finite for all $i \neq j = 1,\ldots,D$.
- (C4) The spectral density matrix $\widetilde{\mathbf{C}}(\omega)$, where $[\widetilde{\mathbf{C}}(\omega)]_{i,j} = \widetilde{C}_{i,j}(\omega)$, is nonnegative definite for all $\omega \in \mathbb{R}$.

The condition (C2) establishes that $\widetilde{C}_{i,i}(\omega)$ are permissible *auto-spectral densities* for scalar stochastic processes. Establishing the permissibility of $\widetilde{C}_{i,j}(\omega)$ for $i \neq j$ is not trivial due to condition (C4) which requires that all eigenvalues or all the principal minors of $\widetilde{\mathbf{C}}(\omega)$ are nonnegative for all $\omega \in \mathbb{R}$. To our knowledge, general methods for establishing the validity of (C4) are not available. The more restrictive concept of *diagonal dominance* is often used to derive sufficient conditions for matrix covariances [52].

To circumvent the permissibility problem, the so-called separable (or intrinsic) model [53], [54] is often used. It features a simple cross-correlation structure the permissibility of which is easily testable.

Definition 8 (Separable cross-correlation model): Let $\mathbf{c}$ be a $D \times D$ positive-definite matrix with entries in $\mathbb{R}$ and $\rho(\tau) : \mathbb{R} \to \mathbb{R}$ a non-negative definite function. Then, the matrix function $\mathbf{C}(\tau) = \mathbf{c}\,\rho(\tau) : \mathbb{R} \to \mathbb{R}^{D\times D}$ provides a permissible, separable (*intrinsic*) cross-correlation model [55].

While the separable model is demonstrably permissible, it is not very useful for studying information flow (see Proposition V.1 below.) Hence, we introduce a more flexible model which involves time-delayed cross correlations.

Lemma III.1 (Time-delayed cross-correlations): (i) Let the continuous functions $C_{i,i} : \mathbb{R} \to \mathbb{R}$, $i = 1, 2$, be non-negative definite. (ii) Let $C_0 : \mathbb{R} \to \mathbb{R}$ be a continuous, even function, i.e., $C_0(\tau) = C_0(-\tau)$, of bounded spectral variation (cf. Condition C3 in Theorem III.1) which has a global maximum at $\tau = 0$. (iii) Define $C_{1,2}(\tau) \triangleq C_0(\tau - \tau_*)$, and $C_{2,1}(\tau) \triangleq C_0(\tau + \tau_*)$, where $\tau_* > 0$ is the time delay. (iv) If the inequality $D(\omega) \triangleq \widetilde{C}_{1,1}(\omega)\widetilde{C}_{2,2}(\omega) - \widetilde{C}_0^2(\omega) > 0$ holds for all $\omega \in \mathbb{R}$, the matrix function $\mathbf{C}(\tau)$ with elements $C_{i,j}(\tau)$, $i, j \in \{1, 2\}$, is a valid matrix covariance function for the vector stochastic process $\mathbf{X}(t) = (X_1(t), X_2(t))^\top$ which comprises a leading (driver) series X_1 and a lagging (receiver) series X_2.

Proof: The conditions (C1)–(C3) of Cramer's theorem are satisfied by construction. The definitions for $C_{1,2}(\tau)$, and $C_{2,1}(\tau)$ imply that the symmetry $C_{1,2}(-\tau) = C_{2,1}(\tau)$ of the cross-covariance is satisfied since $C_0(-\tau - \tau_*) = C_0(\tau + \tau_*)$ due to the mirror symmetry of $C_0(\cdot)$. The absolute integrability of $C_0(\tau)$ ensures the existence of the Fourier transform $\widetilde{C}_0(\omega)$. Furthermore, based on the time shift property of the Fourier transform it holds that

$$\widetilde{C}_{1,2}(\omega) = e^{-\imath\omega\tau_*}\widetilde{C}_0(\omega), \tag{5a}$$

$$\widetilde{C}_{2,1}(\omega) = \widetilde{C}_{1,2}^{\dagger}(\omega) = e^{\imath\omega\tau_*}\widetilde{C}_0(\omega). \tag{5b}$$

Therefore, $|\widetilde{C}_{1,2}(\omega)| = |\widetilde{C}_{2,1}(\omega)| = |\widetilde{C}_0(\omega)|$ and thus Condition (C3) is satisfied. Finally, it holds that $\widetilde{C}_{1,1}(\omega) \geq 0$, $\widetilde{C}_{2,2}(\omega) \geq 0$, and $\widetilde{C}_{1,1}(\omega)\widetilde{C}_{2,2}(\omega) > \widetilde{C}_0^2(\omega)$ since $D(\omega) > 0$ for all $\omega \in \mathbb{R}$, according to (iv) above. In light of (5a), it also holds that $\widetilde{C}_0^2(\omega) = \widetilde{C}_{1,2}(\omega)\,\widetilde{C}_{2,1}(\omega)$. Thus condition (C4) is satisfied for all $\omega \in \mathbb{R}$. Furthermore, since $C_{1,2}(\tau) \triangleq \mathbb{E}[X_1(t - \tau)X_2(t)]$ attains its maximum value at $\tau = \tau_*$, the series X_1 leads and X_2 follows. □

IV. Data-Driven Information Flow Rate

Information flow has been rigorously defined by means of an *ab initio* approach [56]. The latter involves calculating expectations over joint probability density functions that evolve dynamically. However, in many cases (e.g., earth systems science, neuroscience, mathematical finance) the only information comes from available data because the underlying stochastic dynamical system is not known *a priori*. Liang developed a data-driven IFR estimate by maximizing the likelihood of a linear stochastic dynamical system with additive noise [37]. In the following, we investigate this IFR statistic for systems of bivariate stochastic processes. For systems involving $D > 2$ processes, one can consider all the pairwise combinations C_2^D.

For a pair of time series that represent realizations of two *stationary stochastic processes* $X_1(t)$ and $X_2(t)$, the Liang IFR from X_2 to X_1, denoted by $2 \to 1$, is given by [37]

$$\hat{T}_{2\to1}(\delta t) = \frac{\hat{r}}{1 - \hat{r}^2}\,\left[\,\hat{r}_{2,d1}(\delta t) - \hat{r}\,\hat{r}_{1,d1}(\delta t)\,\right], \tag{6a}$$

where $\hat{r} \triangleq \hat{r}_{1,2}$ is the linear (Pearson) correlation coefficient of $\{x_{1,n}\}_{n=1}^N$ and $\{x_{2,n}\}_{n=1}^N$. Note that $\hat{r} = \hat{\rho}_{1,2}(\tau = 0)$ where $\hat{\rho}_{1,2}(\tau)$ is the CCF defined in (4). The coefficients $\hat{r}_{2,d1}(\delta t)$ and $\hat{r}_{1,d1}(\delta t)$ represent correlations between the time series and their first-order finite differences, defined by means of

$$\hat{r}_{i,dj}(\delta t) \triangleq \frac{\hat{C}_{i,dj}(\delta t)}{\sqrt{\hat{C}_{i,i}(0)\,\hat{C}_{j,j}(0)}},\ i, j = 1, 2, \tag{6b}$$

where $\hat{C}_{i,dj}(\delta t) \triangleq \mathbb{E}[X_i'(t)\,\dot{X}_j'(t)]$ is the sample covariance of the time series $\{x_{i,n}\}$ and the *first-order difference* of $\{x_{j,n}\}$, defined as $\dot{x}_{j,n} \triangleq (x_{j,n+1} - x_{j,n})/\delta t$, for $n = 1, \ldots, N - 1$. Note that $\hat{r}_{i,dj}(\delta t)$ is not the standard correlation function between $X_i'(t)$ and $\dot{X}_j'(t)$. The correlation $\hat{r}_{i,dj}(\delta t)$ is equivalently expressed as

$$\hat{r}_{i,dj}(\delta t) = \frac{\hat{C}_{i,j}(\delta t) - \hat{C}_{i,j}(0)}{\delta t\,\sqrt{\hat{C}_{i,i}(0)\,\hat{C}_{j,j}(0)}} = \frac{\hat{\rho}_{i,j}(\delta t) - \hat{\rho}_{i,j}(0)}{\delta t}, \tag{6c}$$

where $\hat{C}_{i,j}(\delta t)$ and $\hat{\rho}_{i,j}(\delta t)$ are respectively the sampling covariance and correlation functions (auto- for $i = j$ and cross- for $i \neq j$) at lag δt defined by means of (3) and (4).

The general form of pair-wise IFR for a D-dimensional vector series $\mathbf{X}_t$ is given by (for $i, j = 1, \ldots, D$, $i \neq j$)

$$\hat{T}_{i\to j}(\delta t) = \frac{\hat{r}_{i,j}}{1 - \hat{r}_{i,j}^2}\,\left[\,\hat{r}_{i,dj}(\delta t) - \hat{r}_{i,j}\,\hat{r}_{j,dj}(\delta t)\,\right]. \tag{7}$$

Taking into account the cross-correlation expression (6c) and the identity $\hat{\rho}_{j,j}(0) = 1$, which is implied by (4), we obtain the following equation for the IFR (for $i, j = 1, \ldots, D$, $i \neq j$)

$$\hat{T}_{i\to j}(\delta t) = \frac{\hat{r}_{i,j}}{1 - \hat{r}_{i,j}^2}\,\frac{\hat{\rho}_{i,j}(\delta t) - \hat{r}_{i,j}\hat{\rho}_{j,j}(\delta t)}{\delta t}. \tag{8}$$

In (8), the $\hat{r}_{i,dj}$ terms are replaced by expressions that involve only the sampling correlations $\hat{\rho}_{i,j}(\cdot)$.

Remark 2 (General properties of $\hat{T}_{i\to j}$): The data-driven IFR satisfies the following general (independent of the covariance kernels) properties (for $i,j=1,\ldots,D,\ i\neq j$):

- For stationary processes, $\hat{T}_{i\to j}(\delta t)$ is independent of the time index t.
- Since $\hat{r}_{i,j}$ and $\hat{r}_{i,dj}(\delta t)$ are sample-based statistics, so are the $\hat{T}_{i\to j}(\delta t)$.
- The definition (7) implies that $\hat{T}_{i\to j}(\delta t)$ satisfies the following: (i) if $\hat{r}_{i,j}=0$ then $\hat{T}_{i\to j}=\hat{T}_{j\to i}=0$, and (ii) if $\hat{T}_{i\to j}\neq 0$, then $\hat{T}_{i\to j}(\delta t)\propto 1/\delta t$.
- The units of $\hat{T}_{i\to j}(\delta t)$ are natural units of information (nats) per unit time.
- $\hat{T}_{i\to j}=0$ does not imply that necessarily $\hat{T}_{j\to i}=0$.

V. Information Flow Rate for Second-Order Ergodic Processes

For stationary processes, the data-driven IFR (7) is a random variable that fluctuates between different system realizations. We investigate the properties of $\hat{T}_{i\to j}(\delta t)$ for cross-covariance-ergodic processes which satisfy Theorem II.2.

Practical use of Slutsky's theorem requires a sufficiently large sample to allow self-averaging. Hence, the length, $N\,\delta t$, of the observation window should be a large multiple of the longest correlation time among the functions $\{C_{i,j}\}_{i,j=1}^{d}$. This condition can be expressed as $N\delta t\gg\tau_d$, where

$$\tau_d=\max_{i,j=1,\ldots,d}\{\tau_{c;i,j}\}_{i,j=1}^{d},\quad \tau_{c;i,j}=\int_{-\infty}^{\infty}C_{i,j}(\tau)\,\mathrm{d}\tau\,,$$

is the largest of the ACF, $\tau_{c;i,i}$ and CCF, $\tau_{c;i,j}(i\neq j)$ correlation times (measured in units of δt). This condition implies measuring the time series length using an effective sample size (ESS) $N_{\text{eff}}<N$, which accounts for correlation effects [57], [58]. Different definitions of N_{eff} [59], [60], [61] agree on reducing N by a factor that reflects the correlation times. A typical estimate of the effective size is thus $N_{\text{eff}}=N\,\delta t/\tau_d$.

Theorem V.1 (Equilibrium IFR for ergodic processes): Let $\mathbf{X}(t)$ represent a second-order ergodic, D-vector, stochastic process with standard deviations σ_i, $i=1,\ldots,D$, auto- and cross-covariance functions $C_{i,j}(\tau)$, for $i,j=1,\ldots,D$, and respective correlation functions $\rho_{i,j}(\tau)=C_{i,j}(\tau)/\sigma_i\sigma_j$.

The *equilibrium IFR* between the process X_i and the process X_j, for $j\neq i$, is given by the $N\to\infty$ limit

$$T_{i\to j}(\delta t)\triangleq\lim_{N\to\infty}\hat{T}_{i\to j}(\delta t)=\frac{\rho_{i,j}(0)}{1-\rho_{i,j}^2(0)}\times\left[r_{i,dj}(\delta t)-\rho_{i,j}(0)\,r_{j,dj}(\delta t)\right],\tag{9}$$

where $r_{i,dj}(\delta t)$ is the slope of the correlation function $\rho_{i,j}(\cdot)$ near zero lag, expressed as

$$r_{i,dj}(\delta t)=\frac{1}{\delta t}\left[\,\rho_{i,j}(\delta t)-\rho_{i,j}(0)\,\right].\tag{10}$$

Proof: Second-order ergodicity implies that the sample mean as well as the auto-covariance and cross-covariance functions converge in the mean square sense to their ensemble counterparts as $N\to\infty$. If the ergodic conditions hold, the sample averages $\hat{r}_{i,j}$ in (7) can be replaced with their ensemble counterparts $\rho_{i,j}(0)$. For the $\hat{r}_{i,dj}(\delta t)$ terms, it holds that

$$\lim_{N\to\infty}\hat{r}_{i,dj}(\delta t)=\frac{1}{\sigma_i\,\sigma_j}\,\mathbb{E}\left[X_i'(t)\left(\frac{X_j'(t+\delta t)-X_j'(t)}{\delta t}\right)\right]=\frac{\left[\,C_{i,j}(\delta t)-C_{i,j}(0)\,\right]}{\delta t\,\sigma_i\sigma_j}=\frac{\rho_{i,j}(\delta t)-\rho_{i,j}(0)}{\delta t}.$$

The last step uses Definition 2 for the correlation functions. The above leads to (10) and concludes the proof for (9). □

Theorem V.1 allows replacing the data-driven IFR in the ergodic limit with $T_{i\to j}(\delta t)$. The latter involves only ensemble moments (correlation functions) and can thus be used for theoretical investigations of information flow between stochastic processes.

Corollary V.1 (Equivalent expression for equilibrium IFR): If the conditions and definitions of Theorem V.1 hold, the equilibrium IFR between two second-order ergodic processes $X_i(t)$ and $X_j(t)$ is given by

$$T_{i\to j}(\delta t)=\frac{\rho_{i,j}(0)}{1-\rho_{i,j}^2(0)}\,\frac{\rho_{i,j}(\delta t)-\rho_{i,j}(0)\rho_{j,j}(\delta t)}{\delta t}.\tag{11}$$

The above is expressed in terms of ACFs and CCFs of the two processes thus avoiding cross-correlations between the processes and their derivatives.

Proof: The equation (11) is obtained from (9) in view of the correlation slope (10) and the fact that $\rho_{j,j}(0)=1$. □

Next, we show that for processes described by a separable cross-correlation model the equilibrium IFR vanishes.

Proposition V.1 (IFR for separable cross-correlation models): Consider D cross-correlated processes which satisfy the covariance separability condition of Definition 8. The equilibrium IFR $T_{i\to j}(\delta t)$ vanishes between any two non-fully correlated processes, i.e., $T_{i\to j}(\delta t)=0$ for $i\neq j\in\{1,\ldots,D\}$ if $\rho_{i,j}(0)\neq\pm 1$.

Proof: Based on Definition 8, the correlation functions of separable models are given by

$$\rho_{i,j}(\tau)=\rho(\tau),\ \text{for } i=j,\tag{12a}$$

$$\rho_{i,j}(\tau)=\frac{c_{i,j}}{\sqrt{c_{i,i}c_{j,j}}}\,\rho(\tau),\ \text{for } i\neq j\,.\tag{12b}$$

The IFR satisfies $T_{i\to j}(\delta t)=A\,B(\delta t)/\delta t$, where $A\triangleq\rho_{i,j}(0)/(1-\rho_{i,j}^2(0))$ and $B(\delta t)\triangleq\rho_{i,j}(\delta t)-\rho_{i,j}(0)\rho_{j,j}(\delta t)$ according to Corollary V.1. If $\rho_{i,j}(0)\neq\pm 1$, then $A\in\mathbb{R}$. If we define $a_{i,j}\triangleq\frac{c_{i,j}}{\sqrt{c_{i,i}c_{j,j}}}$, it follows that $\rho_{i,j}(0)=a_{i,j}$, $\rho_{i,j}(\delta t)=a_{i,j}\rho(\delta t)$, and $\rho_{j,j}(\delta t)=\rho(\delta t)$, according to (12); from the above $B(\delta t)=0$ is obtained. □

Hence, there is no equilibrium information flow between stochastic processes with separable cross-correlation models. If the processes are fully correlated, i.e., $\rho_{i,j}(0)=\pm 1$, the IFR is not well-defined due to the zero term $1-\rho_{i,j}^2(0)$ in the denominator of (11). In this case, however, the two processes differ by a sign at most; therefore, they are essentially the same process.

Below we present a theorem which links the equilibrium IFR with the spectral moments of the CCFs. As a consequence, the

equilibrium IFR is shown to vanish for $\delta t \to 0$ if the CCF $\rho_{i,j}$ is an even function of τ.

Theorem V.2 (Spectral IFR expression for ergodic processes): Let $\{X_i(t)\}_{i=1}^D$ represent a set of stochastic processes with the properties specified in Theorem V.1. Assume that $\widetilde{\rho}_{i,j}(\omega)$ are the auto- and cross-spectral densities, defined as the FTs of the correlation functions $\rho_{i,j}(\tau)$. Furthermore, assume that the auto- $(i=j)$ and cross- $(i \neq j)$ spectral moments of order one, $\Lambda_{i,j}^{(1)}$, defined by means of the improper integrals

$$\Lambda_{i,j}^{(1)} \triangleq \frac{\imath}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}\omega\, \omega\, \widetilde{\rho}_{i,j}(\omega), \tag{13}$$

exist. Then, the equilibrium IFR is given by

$$T_{i\to j}(\delta t) = \frac{\rho_{i,j}(0)}{1-\rho_{i,j}^2(0)}\, \Lambda_{i,j}^{(1)} + \mathcal{O}(\delta t), \;\; \text{for } i \neq j\,. \tag{14}$$

Proof: Equation (14) follows from (9) by showing that $r_{i,dj}(\delta t) \propto \Lambda_{i,j}^{(1)} + \mathcal{O}(\delta t)$ for all i, j. We use (10) for $r_{i,dj}(\delta t)$, express $\rho_{i,j}(\cdot)$ in terms of its IFT, and replace $\exp(\imath\omega t)$ with its Taylor expansion around $\delta t = 0$, to obtain

$$\begin{aligned} r_{i,dj}(\delta t) &= \frac{1}{2\pi\delta t} \int_{-\infty}^{\infty} \widetilde{\rho}_{i,j}(\omega)\, \left(e^{\imath\omega\delta t} - 1\right) \mathrm{d}\omega \\ &= \frac{\imath}{2\pi} \int_{-\infty}^{\infty} \omega\, \widetilde{\rho}_{i,j}(\omega)\, \mathrm{d}\omega + \mathcal{O}(\delta t)\,. \end{aligned} \tag{15}$$

In general, $\widetilde{\rho}_{i,j}(\omega) = \widetilde{\rho}_{i,j}^{\mathrm{re}}(\omega) + \imath\, \widetilde{\rho}_{i,j}^{\mathrm{im}}(\omega)$. Since $\rho_{i,j}(\delta t)$ is a real-valued function, its Fourier transform respects $\widetilde{\rho}_{i,j}^{\mathrm{re}}(\omega) = \widetilde{\rho}_{i,j}^{\mathrm{re}}(-\omega)$ and $\widetilde{\rho}_{i,j}^{\mathrm{im}}(-\omega) = -\widetilde{\rho}_{i,j}^{\mathrm{im}}(\omega)$. The integral involving $\omega\widetilde{\rho}_{i,j}^{\mathrm{re}}(\omega)$ over the symmetric interval $(-\infty, \infty)$ vanishes because $\widetilde{\rho}_{i,j}^{\mathrm{re}}(\omega)$ is even, and thus the integrand $\omega\widetilde{\rho}_{i,j}^{\mathrm{re}}(\omega)$ is an odd function of ω. Hence, only the integral over the imaginary part survives in (15), leading to

$$r_{i,dj}(\delta t) = -\frac{1}{\pi} \int_0^{\infty} \mathrm{d}\omega\, \omega\, \widetilde{\rho}_{i,j}^{\mathrm{im}}(\omega) = \Lambda_{i,j}^{(1)}. \tag{16}$$

The calculation of $r_{j,dj}(\delta t)$ involves the density $\widetilde{\rho}_{j,j}^{\mathrm{im}}(\omega)$. The ACF $\rho_{j,j}(\tau)$ are real-valued, even functions. Hence, the respective FTs are also real-valued and even. Thus, $\widetilde{\rho}_{j,j}^{\mathrm{im}}(\omega) = 0$ for all $\omega \in \mathbb{R}$, and the diagonal spectral moments $\Lambda_{j,j}^{(1)}$ vanish. Therefore, only the $r_{i,dj}(\delta t)$ terms with $i \neq j$ give non-vanishing spectral integrals. Finally, the result (14) is obtained from (9) using the spectral integral (16). □

Remark 3 (Existence of spectral moments): The existence of $\Lambda_{i,j}^{(1)}$ requires that the spectral densities $\widetilde{\rho}_{i,j}(\omega)$ satisfy $\widetilde{\rho}_{i,j}^{\mathrm{im}}(\omega) \sim \omega^{-2-\epsilon}$, where $\epsilon > 0$ for $\omega \to \infty$. The Whittle-Matérn spectral density is $\widetilde{C}_0(\omega) \propto (1+\omega^2\lambda^2)^{-(\nu+1/2)}$, where $\nu > 0$ is the smoothness index and $\lambda > 0$ is a time constant [54], [62]. If $\nu = 1/2$ the spectral integral (16) has a logarithmic divergence and $\Lambda_{i,j}^{(1)}$ is not well-defined. However, for $\nu > 1/2$ the integral defining $\Lambda_{i,j}^{(1)}$ is finite.

Remark 4 (Vanishing of leading IFR term): As it follows from (14) (Theorem V.2), the leading IFR term vanishes if one the following conditions hold:

1) If $\rho_{i,j}(0) = 0$, i.e., if the processes are not cross-correlated at zero lag; then $T_{i\to j}(\delta t) = T_{j\to i}(\delta t) = 0$.
2) If $\widetilde{\rho}_{i,j}^{\mathrm{im}}(\omega) = 0$ for all $\omega > 0$, i.e., if the FT of the CCF is real-valued. This leads to $\Lambda_{i,j}^{(1)} = 0$ according to (16). A sufficient (but not necessary condition) is that $\rho_{i,j}(\tau)$ be an even function of τ; then $\widetilde{\rho}_{i,j}^{\mathrm{im}}(\omega) = 0$ for all $\omega \in \mathbb{R}$.

In general, $\widetilde{\rho}_{i,j}^{\mathrm{im}}(\omega) = 0$ for $\omega > 0$ does not imply $\widetilde{\rho}_{j,i}^{\mathrm{im}}(\omega) = 0$ *for* $\omega > 0$ since $\widetilde{\rho}_{j,i}^{\mathrm{im}}(\omega) = \widetilde{\rho}_{i,j}^{\mathrm{im}}(-\omega)$. So, if the leading-order term of $T_{i\to j}(\delta t)$ vanishes, this does not imply that the leading-order term of $T_{j\to i}(\delta t)$ also vanishes.

VI. IFR for Mean-Square Differentiable Processes

In the following we assume that the stochastic processes $\{X_i(t)\}_{i=1}^D$ are second-order ergodic, and mean-square differentiable. For example, the displacement of a damped, linear harmonic oscillator driven by white noise is a mean-square differentiable process [63]. We will need the following lemma [50], [64], [65].

Lemma VI.1 (Mean-square continuity and differentiability): A second-order stationary process $X_i(t)$ is continuous in the mean-square sense if its ACF $\rho_{i,i}(\tau)$ is continuous at $\tau = 0$. $X_i(t)$ is first-order differentiable (in the mean-square sense) if $\rho_{i,i}(\tau)$ admits a finite second-order derivative with respect to τ at zero lag, i.e., if $\rho_{i,i}^{(2)}(0)$ exists. This condition implies that the first derivative of $\rho_{i,i}(\tau)$ vanishes at $\tau = 0$ (extremum condition).

We will consider the continuous-sampling limit $\delta t \to 0$ where the sampling step is considerably smaller than the shortest correlation time and the delay τ_* (provided there is a finite delay τ_* between processes). At this limit the leading term of $T_{i\to j}(\delta t)$ is dominant; this term is independent of δt according to (14) in Theorem V.2. Hence, at this limit the IFR has a value which is independent of δt.

Theorem VI.1 (IFR continuous sampling limit): Let $\{X_i(t)\}_{i=1}^D$ be a set of second-order ergodic, and mean-square differentiable stochastic processes. In addition, assume that the CCFs $\rho_{i,j}(\tau) : \mathbb{R} \to \mathbb{R}$ are at least twice differentiable (for $i \neq j$). Let the *continuous-sampling limit of the IFR* be defined as follows:

$$\mathcal{T}_{i\to j} \triangleq \lim_{\delta t \to 0} T_{i\to j}(\delta t). \tag{17}$$

1) $\mathcal{T}_{i\to j}$ is a finite real number given by

$$\mathcal{T}_{i\to j} = \frac{\rho_{i,j}(0)\, \rho_{i,j}^{(1)}(0)}{1-\rho_{i,j}^2(0)}, \tag{18}$$

where $\rho_{i,j}^{(1)}(0)$ is the first-order derivative of the CCF evaluated at zero lag. Furthermore, the *sign* of $\mathcal{T}_{i\to j}$ is the sign of the first-order derivative of $\rho_{i,j}^2(\tau)$ *at* $\tau = 0$.
2) $\mathcal{T}_{i\to j}$ has odd symmetry with respect to interchange of the information flow indices, i.e., $\mathcal{T}_{i\to j} = -\mathcal{T}_{j\to i}$.

Proof: The proof is given in Appendix A. □

Remark 5 (Connection with spectral formulation): The continuous-sampling IFR (18) is equivalent to the leading term in the spectral expression (14) since $\Lambda_{i,j}^{(1)} = \rho_{i,j}^{(1)}(0)$ based on (14) and $\mathrm{FT}[\rho_{i,j}^{(1)}] = \imath\omega\, \mathrm{FT}[\rho_{i,j}]$.

Corollary VI.1 (Finite-time-step corrections): If δt is small but not negligible, the equilibrium IFR is given by $T_{i\to j}(\delta t) =$

$\mathcal{T}_{i\to j}+\delta\mathcal{T}_{i\to j}+\mathcal{O}(\delta t^2)$ for $i\neq j$. The first-order correction, $\delta\mathcal{T}_{i\to j}$, to the continuous sampling IFR $\mathcal{T}_{i\to j}$ is an $\mathcal{O}(\delta t)$ term given by

$$\delta\mathcal{T}_{i\to j}=\begin{cases}\mathcal{T}_{i\to j}\left[\dfrac{\rho^{(2)}_{i,j}(0)-\rho_{i,j}(0)\,\rho^{(2)}_{j,j}(0)}{2\rho^{(1)}_{i,j}(0)}\right]\delta t, & \text{if } \rho^{(1)}_{i,j}(0)\neq 0,\\ \dfrac{\rho_{i,j}(0)\,\left[\rho^{(2)}_{i,j}(0)-\rho_{i,j}(0)\rho^{(2)}_{j,j}(0)\right]\delta t}{2\left[1-\rho^2_{i,j}(0)\right]}, & \text{if } \rho^{(1)}_{i,j}(0)=0.\end{cases} \tag{19}$$

Proof: The proof is given in Appendix B. □

Above we provide general IFR expressions for mean-square differentiable processes. Next, we focus on models with time-delayed correlations.

A. Cross-Correlation Functions With Time Delay

Below we show that time-delayed cross-correlation models described in Lemma III.1 can sustain non-zero information flow $\mathcal{T}_{i\to j}$, in contrast with separable models (cf. Proposition V.1).

Theorem VI.2 (Continuous-sampling-limit IFR for time-delayed cross correlations): Let $\mathbf{C}(\tau):\mathbb{R}\to\mathbb{R}^2\times\mathbb{R}^2$ represent a matrix covariance function which satisfies the conditions of Lemma III.1. Assume that at least the second-order derivatives of $\rho_{i,i}(\tau)$, $i=1,2$, and $C_0(\tau)$ with respect to τ exist at $\tau=0$. Then, the following statements are true:

1) The continuous-sampling IFR is given by

$$\mathcal{T}_{i\to j}=\frac{C_0(-\epsilon_{i,j}\tau_*)\,C^{(1)}_0(u)\big|_{u=-\epsilon_{i,j}\tau_*}}{\sigma_i^2\sigma_j^2-C_0^2(-\epsilon_{i,j}\tau_*)}, \tag{20}$$

where $\sigma_i^2=C_{i,i}(0)>0$, $C^{(1)}_0(u)|_{u=-\epsilon_{i,j}\tau_*}$ is the first derivative of $C_0(u)$ at $u=-\epsilon_{i,j}\tau_*$ and $\epsilon_{i,j}$ is the *Levi-Civita symbol*: $\epsilon_{i,j}=1$ *if* (i,j) is an even and $\epsilon_{i,j}=-1$ *if* (i,j) is an odd permutation (e.g., $\epsilon_{1,2}=1;\epsilon_{2,1}=-1;\epsilon_{i,i}=0$).

2) $\mathcal{T}_{i\to j}$ is antisymmetric, i.e., $\mathcal{T}_{i\to j}=-\mathcal{T}_{j\to i}$.

Proof: The Levi-Civita symbol determines the sign of the time delay, i.e., $C_{i,j}(\tau)=C_0(\tau-\epsilon_{i,j}\tau_*)$. We use (18) from Theorem VI.1 for $\mathcal{T}_{i\to j}$. The value of $\rho_{i,j}(0)$ follows from $\rho_{i,j}(0)=C_0(-\epsilon_{i,j}\tau_*)/\sigma_i\sigma_j$. Using the change of variable $\tau\mapsto u\triangleq\tau-\epsilon_{i,j}\tau_*$, it holds that $\rho^{(1)}_{i,j}(0)=\alpha\frac{dC_0(\tau)}{d\tau}|_{\tau=0}=\alpha\frac{dC_0(u)}{du}|_{u=-\epsilon_{i,j}\tau_*}$, where $\alpha=1/(\sigma_i\sigma_j)$. Then, we obtain

$$\rho^{(1)}_{i,j}(0)=\frac{1}{\sigma_i\sigma_j}\left.\frac{dC_0(u)}{du}\right|_{u=-\epsilon_{i,j}\tau_*}, \tag{21}$$

which concludes the proof of (20). The antisymmetry of $\mathcal{T}_{i\to j}$ follows directly from (2) in Theorem VI.1. □

B. Square Exponential Covariance With Time Delays

We consider a bivariate stochastic process with square exponential auto-covariance functions $C_{i,i}(\tau)=\sigma_i^2\,\exp(-\tau^2/\tau_i^2)$ and time-delayed cross-covariances generated from the square exponential $C_0(\tau)=\sigma_0^2\,\exp(-\tau^2/\tau_0^2)$ as described in Lemma III.1. The respective spectral densities are given by $\widetilde{C}_i(\omega)=\sqrt{\pi}\,\tau_i\,\sigma_i^2\,\exp(-\omega^2\,\tau_i^2/4)$, for $i=0,1,2$. The permissibility condition (C4) of Theorem III.1 requires that $D(\omega)\geq 0$ for all $\omega\geq 0$, where

$$\begin{aligned}D(\omega)&\triangleq\widetilde{C}_{1,1}(\omega)\,\widetilde{C}_{2,2}(\omega)-\widetilde{C}_0^2(\omega)\\&=\pi\,\left[\sigma_1^2\,\sigma_2^2\tau_1\,\tau_2\,\mathrm{e}^{-\omega^2(\tau_1^2+\tau_2^2)/4}-\sigma_0^4\,\tau_0^2\,\mathrm{e}^{-\omega^2\tau_0^2/2}\right].\end{aligned} \tag{22}$$

The inequality $D(\omega)\geq 0$ is true for all $\omega\in\mathbb{R}$ provided that (i) $\sigma_1^2\sigma_2^2\tau_1\tau_2\geq\sigma_0^4\tau_0^2$ and (ii) $\tau_1^2+\tau_2^2\leqq 2\tau_0^2$. These conditions ensure that $D(\omega=0)>0$ and that $\widetilde{C}_{1,1}(\omega)\,\widetilde{C}_{2,2}(\omega)$ decays more slowly than $\widetilde{C}_0^2(\omega)$. The CCFs $(i\neq j)$ are expressed as $\rho_{i,j}(\tau)=C_0(\tau-\epsilon_{i,j}\tau_*)/\sigma_1\sigma_2$. Then, $\rho_{i,j}(0)$ and $\rho^{(1)}_{i,j}(0)$ are given according to (21) by

$$\rho_{1,2}(0)=\rho_{2,1}(0)=\frac{\sigma_0^2}{\sigma_1\sigma_2}\,\mathrm{e}^{-\tau_*^2/\tau_0^2}, \tag{23}$$

$$\rho^{(1)}_{i,j}(0)=\epsilon_{i,j}\frac{\sigma_0^2}{\sigma_1\sigma_2}\frac{2\tau_*}{\tau_0^2}\,\mathrm{e}^{-\tau_*^2/\tau_0^2},\ (i,j)=(1,2),(2,1). \tag{24}$$

Based on the above and (18) we obtain the continuous-sampling IFR limit for $(i,j)\in\{(1,2),(2,1)\}$:

$$\mathcal{T}_{i\to j}=\frac{\rho_{i,j}(0)\,\rho^{(1)}_{i,j}(0)}{1-\rho^2_{i,j}(0)}=\frac{2\tau_*\epsilon_{i,j}}{\tau_0^2}\frac{\sigma_0^4}{\sigma_i^2\sigma_j^2\,\mathrm{e}^{2\tau_*^2/\tau_0^2}-\sigma_0^4}. \tag{25}$$

Remark 6 (IFR properties of mean-square-differentiable processes): (1) The expression (25) for $\mathcal{T}_{i\to j}$ does not depend on the characteristic ACF times τ_1 and τ_2 because the continuous-time equilibrium IFR (18) does not include diagonal terms. (2) The expression (25) implies, in light of $\sigma_1\sigma_2>\sigma_0^2$ (i) positive information flow rate from the leading (driver) time series, X_1 to the lagging (receiver) time series, X_2, and (ii) an equal-magnitude but opposite sign IFR in the reverse direction $X_2\to X_1$. Furthermore, if we define the dimensionless variables $u\triangleq 2\tau_*^2/\tau_0^2$ and $w\triangleq\sigma_1^2\sigma_2^2/\sigma_0^4$, it follows that $\mathcal{T}_{i\to j}=\epsilon_{i,j}u/\tau_*\,(w\,e^u-1)$. Hence, $\mathcal{T}_{i\to j}$ (and $\mathcal{T}_{j\to i}$ respectively) depends just on three parameters, the dimensionless ratios u,w and the delay time τ_*.

C. Linear Regression Model With Delay

As a special case of the bivariate stochastic process in Section VI-B, we consider the regression model $X_2(t)=aX_1(t-\tau_*)+Y(t)$, where $0<a<1$ is a coupling factor, and τ_* is the time delay between the receiver process $X_2(t)$ and the driver process $X_1(t)$. The latter is assumed to be a zero-mean stationary process with square exponential covariance and characteristic time τ_1; $Y(t)$ is a zero-mean, stationary, mean-square differentiable process with variance $\sigma_y^2=b^2\sigma_1^2$, where $\mathbb{C}\mathrm{ov}\,\{X_1(t),Y(t')\}=0$ for all $t,t'\in\mathbb{R}$ and b is the relative (compared to X_1) amplitude of the uncorrelated component Y in the receiver. Based on the above definitions, the variance of X_2 is $\sigma_2^2=a^2\sigma_1^2+\sigma_y^2$, while $\sigma_0^2=a\sigma_1^2$ is the cross-covariance at zero lag. In addition, the CCF is given by

$$\rho_{1,2}(\tau)\triangleq\frac{\mathbb{E}[X_1(t-\tau)\,X_2(t)]}{\sigma_1\sigma_2}=\frac{a\,\rho_{1,1}(\tau-\tau_*)}{\sqrt{a^2+b^2}}. \tag{26}$$

The permissibility condition (C4) of Cramer's Theorem III.1, as specified in inequality (22), is satisfied because (i) $\sigma_1\sigma_2 = \sigma_1^2\sqrt{a^2+b^2} > \sigma_0^2 = a\sigma_1^2$ and (ii) $\tau_1^2 = \tau_2^2 = \tau_0^2$. Then, the expression (25) for the continuous-sampling, equilibrium IFR yields the following for $(i,j) \in \{(1,2),(2,1)\}$:

$$\mathcal{T}_{i\to j} = \frac{2\tau_*\epsilon_{i,j}}{\tau_0^2}\,\frac{1}{(1+b^2/a^2)\;\mathrm{e}^{2\tau_*^2/\tau_0^2}-1}. \tag{27}$$

According to (27), the magnitude of IFR is reduced as b/a increases. This reflects the decline of the cross-correlation between X_1 and X_2 with increasing b/a, as it follows from (26). The equilibrium IFR (27) is independent of the ACF of the "contaminating" signal $Y(t)$. The dependence of (27) is further studied in the numerical simulations of Section VIII-C.

VII. IFR for Mean-Square Continuous but Non-Differentiable Processes

First-order dynamical systems driven by Gaussian white noise generate stochastic processes which are mean-square continuous but not differentiable [66, p. 39]. If $X(t)$ is such a stochastic process, its ACF is continuous but non-differentiable at zero lag. A well-known example is the exponential (Ornstein-Uhlenbeck) model with $\rho(\tau) = \exp(-|\tau|/\tau_0)$; the lack of the first derivative at $\tau = 0$ is due to the change in the slope of $|\tau|$ from positive to negative. The ergodic expression (11) for the equilibrium IFR is still valid in this case; however, the Taylor expansions of $\rho_{i,j}(\delta t)$ (for $i = j$ and $i \neq j$) used in Theorem VI.1 are not. In this section we calculate IFR for correlation functions that are continuous but non-differentiable at the origin.

A. Cross-Correlation Model With Time Delay

We assume a time-delayed cross-correlation model as defined in Lemma III.1, with the additional constraint that the covariance functions are continuous but non-differentiable at the origin.

Theorem VII.1 (Mean-square continuous processes with time-delayed cross correlations): Let $\mathbf{C}(\tau) : \mathbb{R} \to \mathbb{R}^2 \times \mathbb{R}^2$ represent a matrix covariance function as defined in Lemma III.1. Assume that the ACFs $\rho_{i,i}(\tau)$, $i = 1, 2$, and the CCF generating function $C_0(\tau)$ are continuous and everywhere differentiable except at $\tau = 0$. In addition, assume that $C_0(\cdot)$ is an even function of bounded spectral variation as specified in condition (ii) of Lemma III.1. Then, the following statements are true for $i \neq j$:

1) The continuous-sampling ($\delta t \to 0$) IFR limit is given by
$$\mathcal{T}_{i\to j} = \frac{r}{1-r^2}\left[\rho_{i,j}^{(1)}(0_+) - r\,\rho_{j,j}^{(1)}(0_+)\right], \tag{28}$$
where $r \triangleq \rho_{i,j}(0)$, $\rho_{i,j}^{(1)}(0_+) \triangleq C_0^{(1)}(-\epsilon_{i,j}\tau_*)/\sigma_i\sigma_j$, and the Levi-Civita tensor $\epsilon_{i,j}$ is defined in Theorem VI.2.
2) In the continuous-sampling limit we obtain
$$\mathcal{T}_{i\to j} + \mathcal{T}_{j\to i} = \frac{-r^2\left[\rho_{i,i}^{(1)}(0_+) + \rho_{j,j}^{(1)}(0_+)\right]}{1-r^2} \geq 0\,. \tag{29}$$
Hence, the IFR antisymmetry $\mathcal{T}_{i\to j} = -\mathcal{T}_{j\to i}$ of mean-square differentiable processes (see Theorem VI.1) is broken.
3) For driver→receiver IFR ($T_{1\to 2}$), if $\delta t < \tau_*$ the expression (28) is the leading-order approximation in δt. For receiver → driver IFR ($T_{2\to 1}$), (28) is valid for small δt regardless of τ_*.
4) If $\delta t > \tau_*$ (slow sampling regime), the leading-order approximation of the driver → receiver IFR ($T_{1\to 2}$) is given by
$$T_{1\to 2}(\delta t) = \frac{r\left[\rho_{1,2}^{(1)}(0_+)\left(1-\frac{2\tau_*}{\delta t}\right) - r\,\rho_{2,2}^{(1)}(0_+)\right]}{1-r^2}. \tag{30}$$

Proof: The proof is given in Appendix C. □

Remark 7 (Zero IFR condition): It follows directly from (28) that $\mathcal{T}_{i\to j}$ vanishes (i) if $r = 0$, i.e., if $\rho_{i,j}(0_+) = 0$, or (ii) if $\rho_{i,j}^{(1)}(0_+) = r\,\rho_{j,j}^{(1)}(0_+)$. The slope $\rho_{j,j}^{(1)}(0_+)$ of the ACF at $\tau = 0_+$ is negative. Hence, if $r > 0$, condition (ii) is realized only if $\rho_{i,j}^{(1)}(0_+)$, or equivalently the first-order derivative of $C_0(\cdot)$, evaluated at $-\epsilon_{i,j}\tau_*$, is negative (provided that $\tau_* > 0$). Since $C_0(u)$ is a monotonically declining (increasing) function for $u > 0$ ($u < 0$), validity of condition (ii) requires $\epsilon_{i,j} < 0$. Therefore, IFR can vanish for $2 \to 1$ (lagging→leading) information flow but not in the opposite, $1 \to 2$, direction. The situation is reversed if $r < 0$; negative values for r can only be obtained if $C_0(\tau_*) < 0$.

A necessary (but not sufficient) condition for the separable correlation model is $\tau_* = 0$. The IFR $\mathcal{T}_{i\to j}$ in the limit $\tau_* \to 0$ is obtained from (28) for flow in both directions. Furthermore, in the separable case $\rho_{i,j}(\tau) = a_{i,j}\rho(\tau)$ (for $i \neq j$) while $\rho_{j,j}(\tau) = \rho(\tau)$ (cf. Proposition V.1). Hence, condition (ii) above is satisfied, and thus $\mathcal{T}_{i\to j} = \mathcal{T}_{j\to i} = 0$ follows from (28), in agreement with the more general result of Proposition V.1.

B. Ornstein-Uhlenbeck (O-U) Covariance Model

In this section we study a concrete example of a bivariate process with exponential auto-covariance and time-delayed exponential cross-covariance. The exponential covariance describes processes that obey the Ornstein-Uhlenbeck stochastic ordinary differential equation [67].

Corollary VII.1 (Permissibility of O-U cross-covariance model): Consider the exponential auto-covariance functions $C_i(\tau) = \sigma_i^2\,\exp(-|\tau/\tau_i|)$, where $\tau_i > 0$ for all $i \in \{1, 2\}$, and the time-delayed cross-covariances generated from $C_0(\tau) = \sigma_0^2\,\exp(-|\tau/\tau_0|)$ as specified in Lemma III.1. Then, sufficient permissibility conditions are as follows:

$$\text{(C1).}\quad \sigma_1\sigma_2\sqrt{\tau_1\tau_2} > \tau_0\sigma_0^2, \tag{31a}$$

$$\text{(C2).}\quad \frac{\tau_1^2+\tau_2^2}{2\tau_1\tau_2} \geq \frac{\sigma_1^2\sigma_2^2}{\sigma_0^4} \geq \frac{\tau_1\tau_2}{\tau_0^2}\,. \tag{31b}$$

Proof: The respective spectral densities for the O-U model are given by $\widetilde{C}_i(\omega) = 2\tau_i\sigma_i^2/(1+\omega^2\,\tau_i^2)$, for $i = 0, 1, 2$. The permissibility conditions (C4) of Theorem III.1 require that $\widetilde{C}_{1,1}(\omega) \geq 0$, $\widetilde{C}_{2,2}(\omega) \geq 0$, and $D(\omega) > 0$, for all $\omega \in \mathbb{R}$. The

first two inequalities are valid for $\sigma_i, \tau_i > 0$. The third inequality is expressed as

$$\frac{\tau_1\tau_2\sigma_1^2\sigma_2^2}{(1+\omega^2\,\tau_1^2)(1+\omega^2\,\tau_2^2)} > \frac{\tau_0^2\sigma_0^4}{(1+\omega^2\,\tau_0^2)^2} \equiv \frac{1}{D_1(\omega)} > \frac{1}{D_2(\omega)}\,,$$

for all $\omega \in \mathbb{R}$, where $D_i(\omega) = \alpha_i + \beta_i\omega^2 + \gamma_i\omega^4$ are quartic polynomials with coefficients

$$\alpha_1 = \frac{1}{\tau_1\tau_2\sigma_1^2\sigma_2^2},\ \beta_1 = \frac{\tau_1^2+\tau_2^2}{\tau_1\tau_2\sigma_1^2\sigma_2^2},\ \gamma_1 = \frac{\tau_1\tau_2}{\sigma_1^2\sigma_2^2}\,,$$
$$\alpha_2 = \frac{1}{\tau_0^2\sigma_0^4},\ \beta_2 = \frac{2}{\sigma_0^4},\ \gamma_2 = \frac{\tau_0^2}{\sigma_0^4}$$

Since $D_i(\omega) \geq 0$ for all $\omega \in \mathbb{R}$, the third inequality is equivalent to $D_2(\omega) > D_1(\omega)$ for all $\omega \in \mathbb{R}$. This condition is satisfied if $\alpha_2 > \alpha_1$, $\beta_2 \geq \beta_1$, and $\gamma_2 \geq \gamma_1$, which lead to the conditions (31a) by simple algebraic calculations. □

Theorem VII.2 (IFR for O-U covariance model with time delay): Assume that the O-U covariance model satisfies the permissibility criteria of Corollary VII.1. Then, the continuous-sampling IFR for the O-U model is given by

$$\mathcal{T}_{i\to j} = \frac{1}{\left(\frac{\sigma_1^2\sigma_2^2}{\sigma_0^4}\right)\,\mathrm{e}^{2\epsilon_{i,j}\tau_*/\tau_0}-1}\left(\frac{1}{\tau_j}+\frac{\epsilon_{i,j}}{\tau_0}\right). \tag{32}$$

Proof: According to (21), $\rho_{i,j}(0) = C_0(-\epsilon_{i,j}\tau_*)/\sigma_i\sigma_j$ and the derivatives $\rho_{i,j}^{(1)}(0)$, for $i,j=1,2$, are given by

$$\rho_{1,2}(0) = \frac{\sigma_0^2}{\sigma_1\sigma_2}\,\mathrm{e}^{-\tau_*/\tau_0},\ \rho_{2,1}(0) = \frac{\sigma_0^2}{\sigma_1\sigma_2}\,\mathrm{e}^{\tau_*/\tau_0} \tag{33a}$$
$$\rho_{i,j}^{(1)}(0_+) = \frac{-\epsilon_{i,j}}{\tau_0}\,\left(\frac{\sigma_0^2}{\sigma_1\sigma_2}\right)\mathrm{e}^{-\tau_*/\tau_0},\ i\neq j, \tag{33b}$$
$$\rho_{j,j}^{(1)}(0_+) = -\frac{1}{\tau_j}\,. \tag{33c}$$

The first derivative $\rho_{i,j}^{(1)}(0_+)$ (for $i \neq j$) is proportional to $\epsilon_{i,j}$. The Levi-Civita symbol controls the slope which is negative for $(i,j)=(1,2)$ and positive for $(i,j)=(2,1)$. If the permissibility criteria are satisfied, (32) is obtained by plugging the expressions (33) in the IFR (28), taking into account that $r = \rho_{i,j}(0)$. The expression (32) is tested with numerical simulations in Section VIII-D. □

Remark 8 (Asymmetric $\mathcal{T}_{i\to j}$ for O-U model): If $\tau_0 = \tau_1$, then $\mathcal{T}_{2\to1} = 0$ while $\mathcal{T}_{1\to2} \neq 0$. This asymmetric behavior of IFR is in contrast with the square exponential model in which both IFRs have the same magnitude and only differ in sign.

Finally, according to (30), for $\delta t > \tau_*$ (slow sampling) the leading-order with respect to δt leading→lagging IFR approximation is

$$T_{1\to2}(\delta t) = \frac{2\tau_*}{\tau_0\,\delta t}\,\frac{1}{\left(\frac{\sigma_1^2\sigma_2^2}{\sigma_0^4}\right)\mathrm{e}^{2\tau_*/\tau_0}-1}\,. \tag{34}$$

VIII. Numerical Experiments

In this section we test the theoretical analysis of the equilibrium IFR presented in Sections VI-VII with synthetic time series generated by simulating zero-mean, Gaussian stochastic processes with specified correlation properties.

A. Simulation Method

We use the multivariate normal (MVN) method to simulate bivariate time series [68, p. 50]. For time series comprising N sampling times $t_k \in [0,T]$ for $k=1,\ldots,N$, the temporal dependence is determined by the full covariance matrix $\mathbf{C}$:

$$\mathbf{C} = \begin{bmatrix}\mathbf{C}_{1,1} & \mathbf{C}_{1,2}\\ \mathbf{C}_{2,1} & \mathbf{C}_{2,2}\end{bmatrix}. \tag{35}$$

$\mathbf{C}$ is a $2N\times 2N$ symmetric block matrix. The block submatrices $\mathbf{C}_{i,j}$, $i,j=1,2$, include the $N\times N$ symmetric *auto-covariance matrices* $\mathbf{C}_{1,1}, \mathbf{C}_{2,2}$, and the $N\times N$ *cross-covariance matrices* $\mathbf{C}_{1,2}, \mathbf{C}_{2,1}$. The matrix elements $[\mathbf{C}_{i,j}]_{k,l} = \mathbf{K}_{i,j}(t_k - t_l)$ for $i,j=1,2$ and $k,l=1,\ldots,N$ are obtained from a 2×2 matrix function $\mathbf{K}(\tau)$ [e.g., see (36) below]. If $\mathbf{C} = \mathbf{A}\mathbf{A}^T$ is a factorization of $\mathbf{C}$, and $\mathbf{z}$ is a $2N\times1$ vector of independent random numbers drawn from the standard normal distribution $\mathcal{N}(0;1)$, then the $2N\times1$ vector $\mathbf{x} = \mathbf{A}\,\mathbf{z}$ is a realization of the bivariate time series with said covariance structure. We use the principal square root factorization of $\mathbf{C}$ for numerical stability [69].

B. Square Exponential Covariance With Time Delay

We simulate two Gaussian processes governed by the delayed square exponential covariance model (see Section VI-B). We consider a sampling window of length $T=10$ and step $\delta t = 0.002$, leading to $N = 5\times10^3$ sampling points. The covariance matrix function $\mathbf{K}$ is given by

$$\mathbf{K}(\tau) = \begin{bmatrix}\sigma_1^2\,\mathrm{e}^{-\tau^2/\tau_1^2} & \sigma_0^2\,\mathrm{e}^{-(\tau-\tau_*)^2/\tau_0^2}\\ \sigma_0^2\,\mathrm{e}^{-(\tau+\tau_*)^2/\tau_0^2} & \sigma_2^2\,\mathrm{e}^{-\tau^2/\tau_2^2}\end{bmatrix}, \tag{36}$$

where $\tau_0 = T/200$, $\tau_1=\tau_2=\tau_0$, $\tau_* = 0.008$, $\sigma_1^2=\sigma_2^2=1.1$ and $\sigma_0^2=1$. For these parameters, the CCF at zero lag is $\rho_{1,2}(0)\approx0.89$ according to (23). The small ratio τ_0/T is selected in order to ensure that $N_{\mathrm{eff}} \triangleq N\,\delta t/\tau_0 = T/\tau_0 \gg 1$ in compliance with ergodic conditions (see Remark 1).

An ensemble of $N_{\mathrm{sim}} = 100$ realizations is generated using MVN simulation (outlined in Section VIII-A). The ACFs and CCFs along with a typical realization are shown in Fig. 1. The equilibrium IFR is obtained from (25). The data-driven IFR is calculated from each realization using (7). The histograms of $\mathcal{T}_{1\to2}$ and $\mathcal{T}_{2\to1}$ calculated from the ensemble are shown in Fig. 2. The two histograms are nearly identical mirror images as expected from Theorem VI.2. The theoretical values (marked by continuous vertical lines in Fig. 2) lie in the middle of the respective histograms. According to (25), $\mathcal{T}_{1\to2}\approx23.4$ whereas the simulation-average yields $\overline{\mathcal{T}}_{1\to2}\approx23.75$. In addition, the sample estimates of $\mathcal{T}_{1\to2}$ ($\mathcal{T}_{2\to1}$) obtained from the simulated states are consistently positive (negative).

1) **IFR Dispersion:** The IFR values exhibit broad distributions with a coefficient of variation ≈0.15 (in absolute value). This behavior is due to between-samples fluctuations of $\hat{r}$, and the $1-\hat{r}^2$ dependence of the IFR [cf. (7)]. The IFR variance thus tends to increase as $\hat{r}\to\pm1$. In spite of the fluctuations, the IFR probability distributions preserve the signs of $T_{1\to2}$ and $T_{2\to1}$ (see Fig. 2),

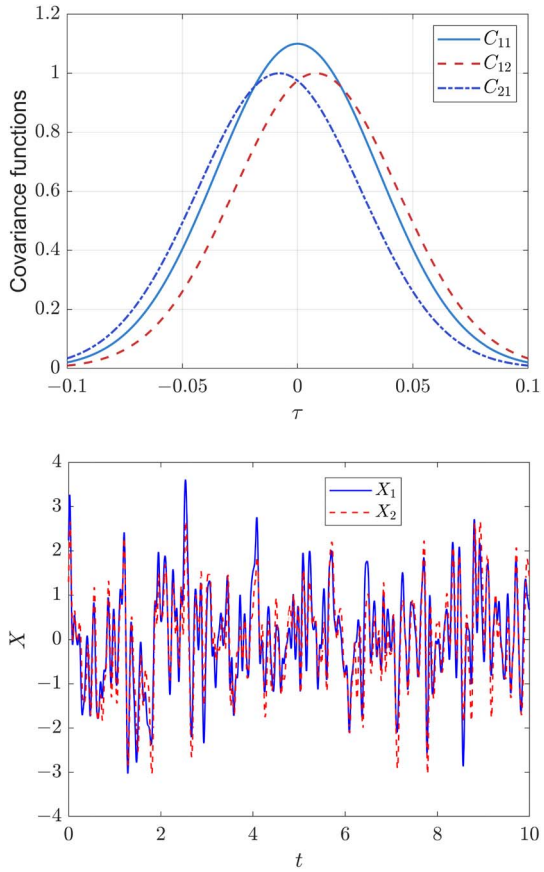

Fig. 1. Auto- and cross-covariance functions (top) and simulated samples (bottom) of two Gaussian processes representing a driver time series $X_1(t)$ (continuous line, blue online) and a receiver time series $X_2(t)$ (broken line, red online). The plots are based on the bivariate Gaussian process with the delayed square exponential covariance model (36). In the top frame, the auto-covariance is marked by the continuous line (cyan online), the cross-covariance $C_{1,2}(\tau)$ by the broken line (red online), and the cross-covariance $C_{2,1}(\tau)$ by the dash-dot line (blue online).

implying that the data-driven IFR correctly captures the direction of information flow from the leading to the lagging series.

2) **Finite-Sampling-Step Corrections:** To obtain accurate IFR estimates based on the continuous sampling expression (25), it is necessary that the $\mathcal{O}(\delta t)$ correction, $\delta\mathcal{T}_{i\to j}$, be negligible. Based on (19) the magnitude of the relative correction is

$$\left|\frac{\delta\mathcal{T}_{i\to j}}{\mathcal{T}_{i\to j}}\right| = \left|\frac{\rho^{(2)}_{i,j}(0) - \rho_{i,j}(0)\,\rho^{(2)}_{j,j}(0)}{2\rho^{(1)}_{i,j}(0)}\right| \delta t\,.$$

Since $\rho_{j,j}(\tau) = \exp(-\tau^2/\tau_j^2)$ and for $i \neq j$ it holds that $\rho_{i,j}(\tau) = (\sigma_0^2/\sigma_1\sigma_2)\,\exp\left[-(\tau - \epsilon_{i,j}\tau_*)^2\right]$, using straightforward algebraic operations, it follows that

$$\left|\frac{\delta\mathcal{T}_{i\to j}}{\mathcal{T}_{i\to j}}\right| = \frac{\delta t}{2\tau_*}\left(\frac{\tau_0^2}{\tau_j^2} - 1\right) + \frac{\delta t\,\tau_*}{\tau_0^2}\,. \tag{37}$$

The first term on the right-hand side of (37) vanishes since $\tau_0 = \tau_1 = \tau_2$. Using the specified values of τ_0, τ_* and δt, we obtain $|\delta\mathcal{T}_{i\to j}/\mathcal{T}_{i\to j}| \approx 6.4 \times 10^{-3}$ which is indeed negligible.

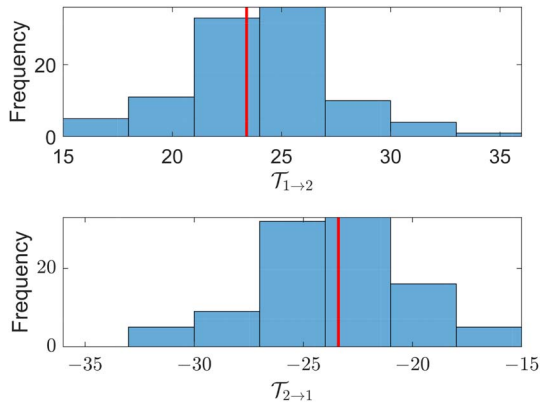

Fig. 2. Histograms of continuous sampling IFR $\mathcal{T}_{1\to 2}$ (top) and $\mathcal{T}_{2\to 1}$ (bottom) generated from an ensemble of 100 realizations of a bivariate cross-correlated Gaussian stochastic process with square exponential covariance model (36). The continuous line in the middle (red online) of both plots marks the theoretical estimate of the equilibrium IFR (25). The model parameters for the stochastic processes (cf. Section VIII-B), are: $T = 10$ (length of sampling window) $\delta t = 0.002$ (sampling step), $N = 5 \times 10^3$ (number of sampling points), $\tau_1 = \tau_2 = \tau_0 = 0.05$ (correlation time), $\tau_* = 0.008$ (time delay), $\sigma_1^2 = \sigma_2^2 = 1.1$ and $\sigma_0^2 = 1$.

C. Linear Regression Model With Time Delay

We test the validity of the theoretical IFR (27) for the linear regression model of mean-square differentiable processes presented in Section VI-C. The study design is described below.

1) The processes $X_1(t)$ and $X_2(t)$ are coupled by means of the delayed square exponential covariance (36). The model parameters are $\tau_0 = T/200$, $\tau_1 = \tau_2 = \tau_0 = 1$, $\sigma_1^2 = 1.1$ and $\sigma_0^2 = 1$. Note that $a = \sigma_0^2/\sigma_1^2 = (1.1)^{-1}$, while $\sigma_2^2 = a^2\sigma_1^2\left(1 + b^2/a^2\right)$ varies with b/a.
2) The equilibrium IFR is calculated for 100 values of the ratio $\tau_*/\tau_0 \in [0, 1]$ and for $b/a \in \{0.1, 0.2, 0.4, 0.6, 0.8\}$.
3) An ensemble of $N_{\text{sim}} = 100$ simulations is generated by means of the MVN method for each combination $(\tau_*/\tau_0, b/a)$ (see Section VIII-A).
4) Each pair of time series in the ensemble comprises $N = 1000$ time instants $t \in [0, 100]$.
5) For each realization, we calculate $\mathcal{T}_{1\to 2}$ and $\mathcal{T}_{2\to 1}$ based on the data-driven IFR (7).
6) We generate parametric plots of the equilibrium IFRs versus τ_*/τ_0 for the different b/a values.

Fig. 3 compares the ensemble averages of the IFR $\mathcal{T}_{1\to 2}$ with the equilibrium IFR (27). The error bars are based on $2\sigma_{\text{IFR}}$, where σ_{IFR} is the IFR standard deviation estimated from the simulation ensemble. Near-perfect agreement is observed between the theoretical and the simulation-based estimates. The IFR peaks at a τ_*/τ_0 value which depends on b/a. The IFR (both theoretical and ensemble estimates) tend to zero as $\tau_*/\tau_0 \to 0$ in agreement with Proposition V.1. The dispersion of simulation-based IFR values declines with increasing τ_*/τ_0 (for fixed b/a) and with increasing b/a (for fixed τ_*/τ_0). This behavior agrees with (26) which shows that $\rho_{1,2}(0)$ declines with increasing b/a (under fixed τ_*) and with increasing τ_* (under fixed b/a). Values of $\rho_{1,2}(0)$ (and thus of $\hat{r}$) that approach ± 1 inflate the $\mathcal{T}_{1\to 2}$ variance. Finally, according to (37) the

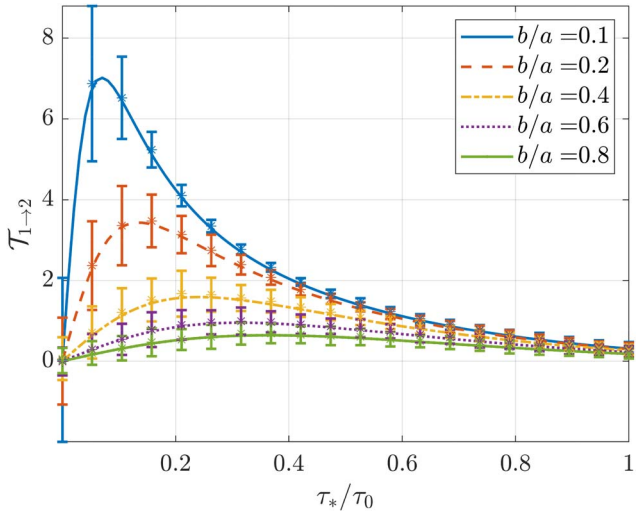

Fig. 3. Parametric plots of $\mathcal{T}_{1\to2}$ for the regression model $X_2(t) = aX_1(t-\tau_*) + Y(t)$ (see Section VI-C) versus τ_*/τ_0 for different b/a ratios. $X_1(t), X_2(t)$ and $Y(t)$ are Gaussian processes with square exponential auto- and cross-covariances. It is assumed that $\sigma_0^2 = 1$, $\sigma_1^2 = 1.1$, $a = \sigma_0^2/\sigma_1^2$ and $\sigma_2^2 = a^2\,\sigma_1^2(1 + b^2/a^2)$. The time constants are $T = 100$, $\tau_1 = \tau_2 = \tau_0 = 1$, while the time delay τ_* is determined from the ratio τ_*/τ_0. The continuous curves are based on the theoretical equilibrium IFR expression (27). For every realization (pair of length $N = 1000$ time series) from an ensemble of $N_{\rm sim} = 100$ simulations, the IFR is estimated based on the data-driven estimator (7). Lower values of b/a imply higher cross correlation between X_1 and X_2. The star markers denote ensemble averages of IFR estimates, while the associated error bars have a width of two standard deviations (as estimated from the ensemble).

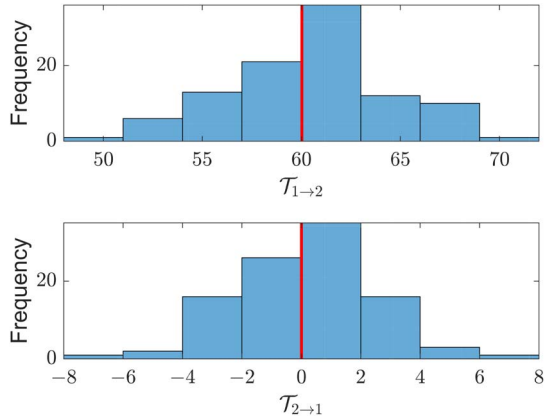

Fig. 4. Histograms of continuous sampling IFR $\mathcal{T}_{1\to2}$ (top) and $\mathcal{T}_{2\to1}$ (bottom) generated from an ensemble of 100 realizations comprising two Gaussian stochastic processes governed by the time-delayed, exponential (Ornstein-Uhlenbeck) model which is defined in Section VII-B. The vertical lines in the middle of the histograms (red online) near 60 and 0 respectively, mark the theoretical, equilibrium IFR estimate (32). The model parameters for the stochastic processes are: $T = 10$ (length of sampling window) $\delta t = 0.002$ (sampling step), $N = 5 \times 10^3$ (number of sampling points), $\tau_1 = \tau_2 = \tau_0 = 0.05$ (correlation time), $\tau_* = 0.008$ (time delay), $\sigma_1^2 = \sigma_2^2 = 1.1$ and $\sigma_0^2 = 1$.

relative magnitude of second-order corrections to the theoretical estimate is $\sim \delta t\,\tau_*/\tau_0^2$. Given the values of $\delta t, \tau_0$ and the range of τ_*/τ_0, such corrections are negligible. The results for $\mathcal{T}_{2\to1}$ are not shown since they are practically mirror images (with reversed sign) of those for $\mathcal{T}_{1\to2}$.

D. Ornstein-Uhlenbeck Covariance With Time Delay

We study two coupled, mean-square continuous Gaussian processes governed by an exponential (O-U) covariance model with time delay. The process parameters N, τ_l and σ_l^2, where $l = 0, 1, 2$, the time step δt, the observation window T and the number $N_{\rm sim}$ of simulated states take the values used in Section VII-B. The auto-covariance functions are given by $C_{i,i}(\tau) = \sigma_i^2\,\exp(-|\tau|/\tau_i)$, for $i = 1, 2$, and the cross-covariance functions are given by the exponentials $C_{i,j}(\tau) = \sigma_0^2\exp(-|\tau - \epsilon_{i,j}\tau_*|/\tau_0)$, for $(i, j) = (1, 2), (2, 1)$.

The IFR histograms obtained from 100 realizations are shown in Fig. 4. Both plots agree with the theoretical result (32). In addition, both $\mathcal{T}_{1\to2}$ and $\mathcal{T}_{2\to1}$ exhibit significant dispersion (due to the fluctuations of $1/(1-\hat{r}^2)$ as discussed in Section VIII-B). However, $\mathcal{T}_{1\to2}$ is persistently non-negative, demonstrating information flow from the leading to the lagging time series, while $\mathcal{T}_{2\to1}$ fluctuates around zero as expected based on (32) and the fact that $\tau_1 = \tau_2 = \tau_0$.

Next, we use the study design of Section VIII-C to calculate the IFR for different combinations of τ_*/τ_0 and b/a in the context of the linear regression model. The only difference is that herein the exponential model is used instead of the square exponential covariance in Section VIII-C. The results are shown in Fig. 5. The top panel shows $T_{1\to2}$ and the bottom panel shows $T_{2\to1}$. The $T_{1\to2}$ curves are calculated from (32) for $\delta t < \tau_*$ (i.e., for $\tau_*/\tau_0 > 0.1$) and from (34) for $\delta t \geq \tau_*$ ($\tau_*/\tau_0 \leq 0.1$). The $T_{2\to1}$ curves are calculated from (32) for all τ_*. The $T_{1\to2}$ curves show the same tendencies, albeit different shapes, with respect to τ_*/τ_0 and b/a as the respective curves in Fig. 3. The ensemble-based estimates are marked by stars with $\pm 2\sigma_{\rm IFR}$ error bars. A small, yet systematic difference is observed between the theoretical and ensemble-based IFR values in the $\delta t \geq \tau_*$ (slow sampling) regime. This is caused by $\mathcal{O}\left((\delta t - 2\tau_*)^2\right)$ terms in the expansion (44) for $\rho_{1,2}(\tau)$, which are not included in the leading-order theoretical estimate. The difference is more pronounced for smaller b/a due to the amplification caused by the $(1-r^2)^{-1}$ factor for $r \approx \pm 1$. The $T_{2\to1}$ are consistently close to the theoretical estimate (i.e., zero), while the dispersion is reduced for higher τ_*/τ_0 and b/a.

IX. Discussion

A. IFR Interpretation

In real-world problems, the absolute magnitude of the data-driven IFR (or its normalized counterpart [38]) are used to determine causal relations, e.g. [37], [43], [44], [45]. Values of $|\hat{T}_{i\to j}| \neq 0$ indicate information flow in the direction $X_i \to X_j$. The statistical significance of non-zero $\hat{T}_{i\to j}$ estimates should be tested. A parametric test based on confidence intervals [37] is derived from estimates of the coefficients of system (1). A non-parametric test of $\hat{T}_{i\to j}$ uses random permutations $X_i^{(p)}$ of the series X_i (i.e., the signal being investigated as potential driver) to obtain a baseline for non-significant deviations of $\hat{T}_{i\to j}$ from zero [43]. The sign of $T_{i\to j}$ however, does not have a clear meaning. Our analysis provides a possible physical interpretation of the IFR sign.

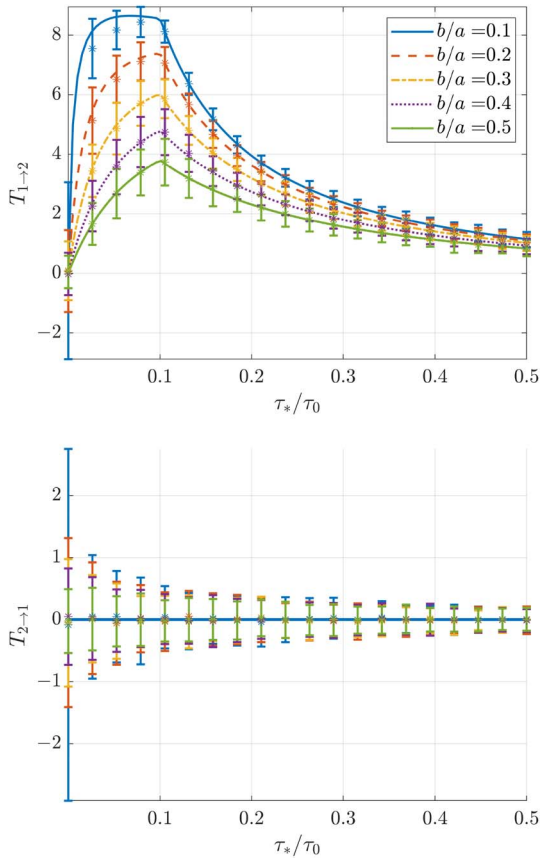

Fig. 5. Parametric plots (continuous curves) of $T_{1\to2}$ (top) and $T_{2\to1}$ (bottom) versus τ_*/τ_0 for different b/a ratios. Two Gaussian processes, X_1 and X_2, with time-delayed, Ornstein-Uhlenbeck model—defined in (33), are simulated. It is assumed that $\sigma_0^2 = 1$, $\sigma_1^2 = 1.1$, $a = \sigma_0^2/\sigma_1^2$ and $\sigma_2^2 = a^2\,\sigma_1^2(1 + b^2/a^2)$. Lower b/a values imply higher cross correlation between X_1 and X_2. The time constants are $T = 100$, $\tau_1 = \tau_2 = \tau_0 = 1$, while τ_* is determined from the ratio τ_*/τ_0. The continuous curves are based on the theoretical equilibrium estimates, i.e., (32) for $\delta t < \tau_*$ (top and bottom panels), and (34) for $T_{1\to2}$ if $\delta t \geq \tau_*$ (top). For every realization (pair of length $N = 1000$ time series) from an ensemble of $N_{\text{sim}} = 100$ simulations, the IFR is obtained using the data-driven estimator (7). Ensemble-based averages are marked by stars, while the associated error bars have a width of two standard deviations (based on the ensemble estimates).

More generally, we have shown that the interpretation of $\hat{T}_{i\to j}$ for coupled stochastic processes depends on the smoothness (regularity) of the processes. For mean-square-differentiable processes with lead-lag correlations, the IFR is positive for leading→lagging information flow and negative in the opposite direction. If the processes are mean-square continuous but non-differentiable, the information flow does not share the same magnitude in both directions: the leading→lagging IFR is positive, while the lagging→leading IFR can take zero or negative values. For linear dynamical systems that satisfy (1) with diagonal diffusion matrix $\mathbf{B}$, the $T_{i\to j}$ is exactly zero for any pair of indices (i, j) such that $A_{j,i} = 0$ [37]. Inspection of (1) shows that the differential of the stochastic Wiener process forces discontinuous first-order derivatives for $\mathbf{X}(t)$. Such systems should be modeled with mean-square continuous (non-differentiable) stochastic processes to avoid erroneous information flow estimates. For mean-square differentiable processes (such as the displacement from equilibrium of coupled, linear, stochastic harmonic oscillators), the information flow is non-zero in both directions due to the odd symmetry of the IFR (see Theorem VI.1). In these cases, analysis of the derivatives (i.e., displacement rates), which represent mean-square continuous but non-differentiable processes, may be more revealing regarding the direction of information flow.

To apply the results of our analysis to real-world data, one needs information about the process regularity. According to Lemma VI.1, mean-square continuity and differentiability of a stochastic process is determined by the regularity of its ACF at the origin. An optimal correlation model can be selected from the data using likelihood-based [70] or Bayesian model selection, or the cross validation approach [71].

Our analysis also suggests that in order to accurately identify the direction of information flow, the sampling step should be smaller than the temporal delay and the latter be a fraction (optimally ≈ 0.1) of the correlation time. In addition, if the time delay τ_* of interaction between processes is much smaller or much larger than the correlation time τ_0, the observed IFR tends to zero. Nonetheless, the IFR is able to detect information flow over a wide range of τ_*/τ_0 (cf. Figs. 3 and 5). In contrast, WGC analysis requires specifying an autoregressive order which should be equal or higher than the interaction time lag between the processes in order to capture causal dependence [24], [72]. Estimates of τ_0 and τ_* can be obtained by means of model fitting and model selection approaches mentioned in the preceding paragraph.

B. Limitations of the Analysis

Herein we focus on second-order ergodic (therefore, also stationary) processes. In practice, conditions of ergodicity require that time series be sufficiently long compared to the correlation times of the processes involved. The regularity conditions (Lemma VI.1) do not require distributional assumptions for the stochastic processes. However, for the numerical simulations we used Gaussian processes because (1) ergodicity conditions can be established by means of Slutsky's theorem based exclusively on second-order moments and (2) the simulation of Gaussian time series is straightforward.

In many problems of interest (e.g., EEG recordings of brain activity at rest or during task execution), the observed time series are non-stationary. Since the calculation of data-driven IFR estimates is straightforward, in non-stationary systems one can estimate time-dependent IFR measures within contiguous time windows that are locally stationary. In such cases, one can derive the information flow using non-overlapping or overlapping time windows, e.g., as in bootstrap subsampling [73] and windowed Fourier transforms [74]. One of the goals of neuroscience is to identify effective brain connectivity within and across different spectral activity bands [27]. IFR has so far been used to investigate brain connectivity without spectral segmentation [43], [44]. This question could be pursued by extending the results of Theorem V.2, using suitably modified IFR spectral expressions that involve band-limited spectral moments over different frequency bands of interest.

Liang has recently generalized the expression for the bivariate, data-driven IFR to multivariate systems containing $N \geq 2$ units (time series) [75]. For $N = 2$, this formulation yields the same IFR between a pair of time series $x_i(t)$ and $x_j(t)$ with $i \neq j$ as the bivariate expression (7). The multivariate expression also quantifies the self-interaction of each time series which is missing in (7). Extension of the present analysis to the multivariate formulation is not straightforward, because the latter involves the cofactors of $N \times N$ covariance matrices. In addition, verifying the admissibility of an N-variate cross-covariance model (where $N \gg 2$) requires controlling a large set of parameters.

Based on (7), IFR vanishes if $\hat{r}_{i,j} = 0$, i.e., if there are no cross correlations. This makes sense, because in the linear limit causation implies correlation (but not vice versa); hence, if two processes are causally linked, their cross-correlation is expected to be non-zero. A pathological case involves two deterministic cosine processes, one of which leads the other by $\pi/2$. Note that causality in this system is not detectable by the WGC method which is applicable to stochastic systems. In spite of the clear causal link, $\hat{r}_{i,j}$ vanishes if it is estimated by integrating over a multiple of the signals' period; therefore, IFR also vanishes. However, it was recently shown that if the cosine functions are modeled as a stochastic bi-harmonic system with additive noise, the normalized IFR tends to unity as the noise variance tends to zero, thus recovering the causal relation [76].

One may ask if IFR is related to other data-driven methods of causal inference such as WGC and transfer entropy [28]. The latter two were shown to be entirely equivalent under the Gaussian assumption [77]. To our knowledge, there are no systematic comparisons of IFR and WGC in the literature; a recent study focuses on differences in brain connectivity derived from the analysis of fMRI data [44]. The two methods have quite different origins. WGC is based on a statistical test of the null hypothesis that the "driver" does not impact the "response" in a statistically significant manner. WGC assumes that the system under study is described by means of a linear, stationary, vector autoregressive model (although there are extensions that relax these assumptions [24], [25]).

Liang's IFR is obtained by calculating the flow of information (expressed in terms of entropy) between subspaces of a stochastic, nonlinear dynamical system. This leads to IFR equations that involve multidimensional integrals of joint probability density functions which are not in general amenable to explicit solution. The data-driven IFR (7), studied herein, was derived by applying maximum likelihood estimation to time series data, assuming a linear stochastic system that satisfies (1) [37]. Nonetheless, the data-driven IFR has been successfully applied to benchmark dynamical systems such as the Baker, Hénon and Kaplan–Yorke maps, and Rössler oscillators [42], [56], [75]. A deeper physical understanding of IFR's magnitude and sign is needed for nonlinear systems. In the linear case, it remains to be investigated if there is a deeper connection between IFR and WGC. IFR has a practical advantage over WGC, because IFR's computational complexity is controlled by the calculation of correlations for two time lags (zero and δt)—therefore it scales linearly with the size of the time series—in contrast with the cubic scaling of WGC (cf. Section I).

X. Conclusion

We investigated the data-driven Liang information flow rate between coupled, second-order ergodic stochastic processes. We defined the equilibrium IFR (Theorem V.1), and we developed a spectral formulation for the equilibrium IFR in terms of spectral moments of the coupled processes (Theorem V.2). We showed that the continuous sampling limit ($\delta t \to 0$) of the equilibrium IFR can be defined for both mean-square differentiable (Theorem VI.1) and mean-square continuous (Theorem VII.1) processes. We also derived leading-order finite-step corrections in the case of mean-square differentiable processes (Corollary VI.1). Furthermore, we established that the equilibrium IFR vanishes for separable cross-correlation models $\mathbf{C}(\tau) = \mathbf{c}\rho(\tau)$, where $\mathbf{c}$ is a $D \times D$ covariance matrix and $\rho(\tau)$ an admissible scalar correlation function. This result holds for second-order ergodic processes independently of regularity properties (Proposition V.1).

For mean-square differentiable processes with non-separable covariance kernels, we found that the IFR has the same magnitude in both directions but is positive in one direction and negative in the other (Theorem VI.1). We also investigated the equilibrium IFR for cross-correlation models featuring time delays. In the case of differentiable covariance kernels, the IFR exhibits an odd symmetry: it is positive in the direction from the leading to the lagging series and negative in the opposite direction (Theorem VI.2). These general results were explored by means of a simple regression model (Section VI-C). The IFR antisymmetry is broken for mean-square continuous (but non-differentiable) processes (Theorem VII.1). For the Ornstein-Uhlenbeck correlation model, we have shown that the IFR from the leading to the lagging series is positive, while the IFR from lagging to the leading series vanishes provided that the characteristic time constants of the ACFs and CCFs match (Remark 8). Finally, we showed that for mean-square differentiable processes, as well as mean-square continuous processes with sampling step smaller than the time delay ($\delta t < \tau_*$), the leading-order equilibrium IFR is independent of δt. In the case of mean-square continuous processes with $\tau_* \leq \delta t$ (slow sampling regime), the leading-order term is $\mathcal{O}(1/\delta t)$.

Appendix A
Proof of Theorem VI.1: IFR Continuous Sampling Limit

Proof: (1) According to (9), $T_{i\to j}(\delta t)$ involves the product of $\rho_{i,j}(0)/[1-\rho_{i,j}^2(0)]$, which is independent of δt, with $r_{i,dj}(\delta t) - \rho_{i,j}(0)\, r_{j,dj}(\delta t)$. Hence, to determine $\lim_{\delta t\to 0} T_{i\to j}(\delta t)$, the limit $r_{i,dj}(\delta t)$ as $\delta t \to 0$ needs to be evaluated, where $r_{i,dj}(\delta t)$ is defined in (10). Assuming mean-square differentiability, the limit $\delta t \to 0$ becomes

$$\lim_{\delta t\to 0} r_{i,dj}(\delta t) = \lim_{\delta t\to 0} \frac{1}{\delta t}\left[\rho_{i,j}(\delta t) - \rho_{i,j}(0)\right]. \tag{38}$$

We consider the terms $r_{i,dj}$ $(i \neq j)$ and $r_{j,dj}$ separately.

Term 1 ($r_{i,dj}$ for $i \neq j$): The CCF $\rho_{i,j}(\tau)$ does not necessarily peak at $\tau = 0$; in fact, the peak appears at $\tau \neq 0$ if the influence of $X_i(t)$ on $X_j(t)$ occurs after a finite time delay. Then, the first-order derivative of $\rho_{i,j}(\tau)$ at zero does not vanish. The Taylor expansion of $\rho_{i,j}(\delta t)$ around $\delta t = 0$ leads to

$$\rho_{i,j}(\delta t) = \rho_{i,j}(0) + \rho_{i,j}^{(1)}(0)\,\delta t + \frac{\rho_{i,j}^{(2)}(0)}{2}\,\delta t^2 + \mathcal{O}(\delta t^3). \quad (39)$$

It follows from (38) and (39) that $\lim_{\delta t\to 0} r_{i,dj}(\delta t) = \rho_{i,j}^{(1)}(0)$ for $i \neq j$.

Term 2 ($r_{i,dj}$ for $i = j$): According to Lemma VI.1 it holds that $\rho_{j,j}^{(1)}(0) = 0$. Since $\rho_{j,j}(0) = 1$, $\rho_{j,j}(\delta t)$ admits the following Taylor expansion around $\delta t = 0$:

$$\rho_{j,j}(\delta t) = 1 + \frac{1}{2}\rho_{j,j}^{(2)}(0)\,\delta t^2 + \mathcal{O}(\delta t^4). \quad (40)$$

Based on the above expansion, the limit (38) is given by

$$\lim_{\delta t\to 0} r_{j,dj}(\delta t) = \lim_{\delta t\to 0}\left[\frac{\delta t}{2}\,\rho_{j,j}^{(2)}(0) + \mathcal{O}(\delta t^3)\right] = 0.$$

Therefore, the term $\propto r_{j,dj}(\delta t)$ in (9) vanishes at the limit $\delta t \to 0$. This result signifies that differentiable processes $X(t)$ are uncorrelated with $\dot{X}(t)$. Hence, only $\lim_{\delta t\to 0} r_{i,dj}(\delta t) = \rho_{i,j}^{(1)}(0)$ for $i \neq j$ enters in $\mathcal{T}_{i\to j}$ leading to (18).

According to (18), the sign of IFR is determined by the sign of the product $\rho_{i,j}(0)\rho_{i,j}^{(1)}(0)$. The latter is the first-order derivative of $\rho_{i,j}^2(\tau)$ with respect to τ evaluated at $\tau = 0$.

(2) The antisymmetry of $\mathcal{T}_{i\to j}$ follows from (18). Since $\rho_{i,j}(0) = \rho_{j,i}(0)$, the sign of $\mathcal{T}_{j\to i}$ is determined from $\rho_{j,i}^{(1)}(0)$. CCFs respect the *reflection symmetry* $\rho_{i,j}(\delta t) = \rho_{j,i}(-\delta t)$ [50]. Using the Taylor series expansion of both terms around $\delta t = 0$ and equating terms of the same order in δt, it follows that the n-order derivative satisfies $\rho_{j,i}^{(n)}(0) = (-1)^n\,\rho_{i,j}^{(n)}(0)$ and thus $\rho_{j,i}^{(1)}(0) = -\,\rho_{i,j}^{(1)}(0)$. Hence, $\mathcal{T}_{j\to i} = -\mathcal{T}_{i\to j}$. □

Appendix B
Proof of Corollary VI.1: Finite-Time-Step Corrections

Proof: We use the equilibrium IFR (11) and define $g_{i,j}(\delta t) \triangleq \frac{1}{\delta t}\,[\rho_{i,j}(\delta t) - \rho_{i,j}(0)\rho_{j,j}(\delta t)]$. Then (11) is expressed as

$$T_{i\to j}(\delta t) = \frac{\rho_{i,j}(0)}{1-\rho_{i,j}^2(0)}\,g_{i,j}(\delta t)\,. \quad (41)$$

Using (39)-(40) and the above definition of $g_{i,j}(\delta t)$, the following Taylor series expansion is obtained for $g_{i,j}(\delta t)$

$$g_{i,j}(\delta t) = \rho_{i,j}^{(1)}(0) + \frac{\delta t}{2}\left[\rho_{i,j}^{(2)}(0) - \rho_{i,j}(0)\rho_{j,j}^{(2)}(0)\right] + \mathcal{O}(\delta t^2)\,. \quad (42)$$

Case 1: $\rho_{i,j}^{(1)}(0) = 0$. If this condition holds, according to (18) the continuous sampling IFR vanishes, i.e. $\mathcal{T}_{i\to j} = 0$. By setting $\rho_{i,j}^{(1)}(0) = 0$ in (42) we obtain from (41) the second branch of (19).

Case 2: $\rho_{i,j}^{(1)}(0) \neq 0$. In this case (42) leads to $g_{i,j}(\delta t) = \rho_{i,j}^{(1)}(0) + \mathcal{O}(\delta t)/2$. In addition, $\mathcal{T}_{i\to j} \neq 0$ according to (18). The leading correction is thus given by $\delta\mathcal{T}_{i\to j} = T_{i\to j}(\delta t) - \mathcal{T}_{i\to j} - \mathcal{O}(\delta t^2)$. Based on (41) and (42) we obtain

$$\delta\mathcal{T}_{i\to j} = \frac{\rho_{i,j}(0)}{1-\rho_{i,j}^2(0)}\,\frac{\delta t\left[\rho_{i,j}^{(2)}(0) - \rho_{i,j}(0)\rho_{j,j}^{(2)}(0)\right]}{2}\,.$$

Finally, if we multiply and divide the right-hand side of the above with $\rho_{i,j}^{(1)}(0) \neq 0$ and recall (18) for $\mathcal{T}_{i\to j}$, the first branch of (19) is recovered. □

Appendix C
Proof of Theorem VII.1: Mean-Square Continuous Processes With Time-Delayed Correlations

Proof: For $i \neq j$, $\rho_{i,j}(0) = C_0(-\epsilon_{i,j}\tau_*)/\sigma_i\sigma_j$; given the symmetry of $C_0(\cdot)$ it holds that $C_0(-\epsilon_{i,j}\tau_*) = C_0(\tau_*)$ and thus $\rho_{i,j}(0) = r$. We use a unified notation for $i = j$ and $i \neq j$, by introducing functions $\phi_{i,j}(u) : \mathbb{R}_{\geq 0} \to \mathbb{R}$ such that $\rho_{i,j}(\tau) = \phi_{i,j}(|\tau - \tilde{\tau}|)$; for $i = j$, $\phi_{i,i}(u) = \rho_{i,i}(\tau)$, where $u = |\tau - \tilde{\tau}|$, while $\phi_{i,j}(u) = C_0(\tau)/\sigma_i\sigma_j$ for $i \neq j$. The temporal offset $\tilde{\tau} \in \mathbb{R}$ is given by $\tilde{\tau} = \epsilon_{i,j}\tau_*$ so that it produces the correct sign for the leading/lagging series and vanishes for $i = j$.

Based on the conditions specified above for $C_0(\cdot)$, the functions $\phi_{i,j}(u)$ are differentiable for all $u \in \mathbb{R}$ and have a global maximum at $u = 0$. For the Ornstein-Uhlenbeck model all of these functions are of the form $\varphi(u) = \exp(-u/\tau_0)$; a Taylor expansion of $\varphi(\cdot)$ around $\tau = 0_+$ (thus $u = |\tilde{\tau}|$) yields

$$\varphi(|\tau - \tilde{\tau}|) = \varphi(|\tilde{\tau}|) + \varphi^{(1)}(|\tilde{\tau}|)\,(\,|\,\tau - \tilde{\tau}| - |\tilde{\tau}|\,) + \mathcal{O}\,(\,|\,\tau - \tilde{\tau}| - |\tilde{\tau}|\,)^2, \quad (43)$$

where $\varphi^{(1)}(|\tilde{\tau}|) \triangleq \varphi^{(1)}(u)\big|_{u=|\tilde{\tau}|}$; thus, if $\tilde{\tau} \to 0$ the derivative $\varphi^{(1)}(0_+)$ is evaluated. For example, the following first-order Taylor approximation is easily confirmed for $|\tau - \tilde{\tau}| \ll \tau_0$:

$$\mathrm{e}^{-|\tau-\tilde{\tau}|/\tau_0} \approx \mathrm{e}^{-\tilde{\tau}/\tau_0}\left(1 - \frac{|\tau-\tilde{\tau}|}{\tau_0} - \frac{|\tilde{\tau}|}{\tau_0}\right).$$

In light of (43) and recalling that $\tilde{\tau} = \epsilon_{i,j}\tau_*$, the Taylor expansion for the *forward finite difference* $\rho_{i,j}(\delta t)$ around $\delta t = 0$ is as follows:

$$\rho_{i,j}(\delta t) = \rho_{i,j}(0) + \rho_{i,j}^{(1)}(0_+)\,(|\delta t - \epsilon_{i,j}\tau_*| - |\epsilon_{i,j}\tau_*|) + \frac{1}{2}\rho_{i,j}^{(2)}(0_+)\,(|\delta t - \epsilon_{i,j}\tau_*| - |\epsilon_{i,j}\tau_*|)^2 + \mathcal{O}(\delta t^3), \quad (44)$$

where $\rho_{i,j}^{(n)}(0_+) = \phi_{i,j}^{(n)}(|\tilde{\tau}|)$ is the limit of the nth-order derivative of $\phi_{i,j}(u)$ with respect to u as $\delta t \to 0_+$.

Based on the equilibrium IFR expression (11) and the Taylor expansion (44), and taking into account that $|\epsilon_{i,j}\tau_*| = \tau_*$, the leading-order in δt IFR approximation for $i \neq j$ is given by

$$T_{i\to j}(\delta t) = \frac{r}{1-r^2}\,\frac{1}{\delta t}\left[\rho_{i,j}^{(1)}(0_+)\,(\,|\delta t - \epsilon_{i,j}\tau_*| - \tau_*) - \rho_{i,j}(0)\,\rho_{j,j}^{(1)}(0_+)\delta t\right]. \quad (45)$$

Based on (45) and $\epsilon_{2,1} = -1$, the leading-order $T_{2\to 1}(\delta t)$ approximation for information flow in the receiver→driver ($2 \to 1$) direction becomes

$$T_{2\to 1}(\delta t) = \frac{r}{1-r^2}\left[\rho_{2,1}^{(1)}(0_+) - r\,\rho_{1,1}^{(1)}(0_+)\right].$$

Hence, $T_{2\to1}(\delta t)$ is independent of δt and the continuous-sampling limit $\mathcal{T}_{2\to1}$ is well defined. This proves the expression (28) for receiver→driver IFR as well as statement (3).

The expression (45) for $T_{1\to2}(\delta t)$ involves $\epsilon_{1,2}=1$. For information flow in the driver→receiver $(1\to2)$ direction, we need to consider two cases: (1) for $\delta t>\tau_*$ we obtain

$$T_{1\to2}(\delta t)=\frac{r}{1-r^2}\left[\rho^{(1)}_{1,2}(0_+)\left(1-\frac{2\tau_*}{\delta t}\right)-r\,\rho^{(1)}_{2,2}(0_+)\right], \tag{46a}$$

and (2) for $\delta t\le\tau_*$

$$T_{1\to2}(\delta t)=\frac{r}{1-r^2}\left[-\rho^{(1)}_{1,2}(0_+)-r\,\rho^{(1)}_{2,2}(0_+)\right]. \tag{46b}$$

The second branch of (46) proves the continuous sampling limit $\mathcal{T}_{1\to2}$ (28) for IFR in the driver→receiver direction (statement 1). It also proves statement (4) in Theorem VII.1 which applies if the sampling step exceeds the delay.

To prove (29), i.e., statement (2) in Theorem VII.1, we evaluate the sum $\mathcal{T}_{i\to j}+\mathcal{T}_{j\to i}$ using (45). We employ the symmetry $C_0(\tau_*)=C_0(-\tau_*)$, and the fact that $C^{(1)}(u)$ is an odd function of u, i.e., $C^{(1)}(u)=-C^{(1)}(-u)$. The latter implies that $\rho^{(1)}_{i,j}(0_+)+\rho^{(1)}_{j,i}(0_+)$ vanishes in $\mathcal{T}_{i\to j}+\mathcal{T}_{j\to i}$ (due to the odd symmetry of the first derivative of the CCF, cf. Appendix A). Thus, the expression (29) follows by adding the second terms on the right-hand side of (45). Finally, the non-negativity of $\mathcal{T}_{i\to j}+\mathcal{T}_{j\to i}$ is due to the fact that $\rho^{(1)}_{j,j}(0_+)<0$ and $0\le r^2\le1$. □

ACKNOWLEDGMENT

We acknowledge constructive email communications with X. San Liang (Fudan University and Southern Marine Science and Engineering Guangdong Laboratory, China) and Arif Babul (University of Victoria, Canada).

REFERENCES

[1] K. Friston, "Causal modelling and brain connectivity in functional magnetic resonance imaging," *PLoS Biol.*, vol. 7, no. 2, 2009, Art. no. e1000033.
[2] S. Hu, G. Dai, G. A. Worrell, Q. Dai, and H. Liang, "Causality analysis of neural connectivity: Critical examination of existing methods and advances of new methods," *IEEE Trans. Neural Netw.*, vol. 22, no. 6, pp. 829–844, Jun. 2011.
[3] L. Ning and Y. Rathi, "A dynamic regression approach for frequency-domain partial coherence and causality analysis of functional brain networks," *IEEE Trans. Med. Imag.*, vol. 37, no. 9, pp. 1957–1969, Sep. 2018.
[4] S. S. Ahmed, S. Roy, and J. Kalita, "Assessing the effectiveness of causality inference methods for gene regulatory networks," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 17, no. 1, pp. 56–70, Jan./Feb. 2020.
[5] Y. Huang, Z. Fu, and C. L. E. Franzke, "Detecting causality from time series in a machine learning framework," *Chaos Interdisciplinary J. Nonlinear Sci.*, vol. 30, no. 6, 2020, Art. no. 063116.
[6] B. Schölkopf et al., "Toward causal representation learning," *Proc. IEEE*, vol. 109, no. 5, pp. 612–634, May 2021.
[7] A. R. Nogueira, A. Pugnana, S. Ruggieri, D. Pedreschi, and J. Gama, "Methods and tools for causal discovery and causal inference," *Wiley Interdiscip. Rev. Data Mining Knowl. Discovery*, vol. 12, no. 2, 2022, Art. no. e1449.
[8] J. Runge et al., "Inferring causation from time series in Earth system sciences," *Nature Commun.*, vol. 10, no. 1, pp. 1–13, 2019.
[9] M. Kretschmer et al., "Quantifying causal pathways of teleconnections," *Bull. Amer. Meteorol. Soc.*, vol. 102, no. 12, pp. E2247–E2263, 2021.
[10] E. Díaz, J. E. Adsuara, Á. M. Martínez, M. Piles, and G. Camps-Valls, "Inferring causal relations from observational long-term carbon and water fluxes records," *Sci. Rep.*, vol. 12, no. 1, pp. 1–12, 2022.
[11] M. Eichler, "Causal inference with multiple time series: Principles and problems," *Philos. Trans. Roy. Soc. A Math. Physical Eng. Sci.*, vol. 371, no. 1997, 2013, Art. no. 20110613.
[12] L. Sommerlade, M. Eichler, M. Jachan, K. Henschel, J. Timmer, and B. Schelter, "Estimating causal dependencies in networks of nonlinear stochastic dynamical systems," *Phys. Rev. E*, vol. 80, no. 5, 2009, Art. no. 051128.
[13] B. He et al., "Electrophysiological brain connectivity: Theory and implementation," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 7, pp. 2115–2137, Jul. 2019.
[14] K. Hlaváčková-Schindler, M. Palus, M. Vejmelka, and J. Bhattacharya, "Causality detection based on information-theoretic approaches in time series analysis," *Phys. Rep.*, vol. 441, no. 1, pp. 1–46, 2007.
[15] J. Pearl, *Causality: Models, Reasoning and Inference*, 2nd ed. New York, NY, USA: Cambridge Univ. Press, 2009.
[16] J. Peters, D. Janzing, and B. Schölkopf, *Elements of Causal Inference: Foundations and Learning Algorithms*. Cambridge, MA, USA: MIT Press, 2017.
[17] D. A. Smirnov, "Transient and equilibrium causal effects in coupled oscillators," *Chaos Interdisciplinary J. Nonlinear Sci.*, vol. 28, no. 7, 2018, Art. no. 075303.
[18] N. Wiener, "The theory of prediction," in *Modern Mathematics for Engineers*, vol. 1, E. Beckenbach, Ed. New York, NY, USA: McGraw-Hill, 1956.
[19] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica: J. Econometric Soc.*, vol. 37, no. 3, pp. 424–438, Jul. 1969.
[20] E. Siggiridou and D. Kugiumtzis, "Granger causality in multivariate time series using a time-ordered restricted vector autoregressive model," *IEEE Trans. Signal Process.*, vol. 64, no. 7, pp. 1759–1773, Apr. 2016.
[21] E. Siggiridou and D. Kugiumtzis, "Dimension reduction of polynomial regression models for the estimation of Granger causality in high-dimensional time series," *IEEE Trans. Signal Process.*, vol. 69, pp. 5638–5650, 2021.
[22] M. Kamiński, M. Ding, W. A. Truccolo, and S. L. Bressler, "Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance," *Biol. Cybern.*, vol. 85, no. 2, pp. 145–157, 2001.
[23] W. Hesse, E. Möller, M. Arnold, and B. Schack, "The use of time-variant EEG Granger causality for inspecting directed interdependencies of neural assemblies," *J. Neurosci. Methods*, vol. 124, no. 1, pp. 27–44, 2003.
[24] S. L. Bressler and A. K. Seth, "Wiener–Granger causality: A well established methodology," *NeuroImage*, vol. 58, no. 2, pp. 323–329, 2011.
[25] D. Marinazzo, W. Liao, H. Chen, and S. Stramaglia, "Nonlinear connectivity by Granger causality," *NeuroImage*, vol. 58, no. 2, pp. 330–338, 2011.
[26] D. Marinazzo, M. Pellicoro, and S. Stramaglia, "Kernel method for nonlinear Granger causality," *Phys. Rev. Lett.*, vol. 100, no. 14, 2008, Art. no. 144103.
[27] M. X. Cohen, *Analyzing Neural Time Series Data: Theory and Practice*. Cambridge, MA, USA: MIT Press, 2014.
[28] T. Schreiber, "Measuring information transfer," *Phys. Rev. Lett.*, vol. 85, no. 2, pp. 461–464, 2000.
[29] R. Salvador, M. Anguera, J. Gomar, E. Bullmore, and E. Pomarol-Clotet, "Conditional mutual information maps as descriptors of net connectivity levels in the brain," *Front. Neuroinf.*, vol. 4, pp. 115–123, Nov. 2010.
[30] R. Vicente, M. Wibral, M. Lindner, and G. Pipa, "Transfer entropy—A model-free measure of effective connectivity for the neurosciences," *J. Comput. Neurosci.*, vol. 30, no. 1, pp. 45–67, 2011.
[31] M. H. I. Shovon, D. N. Nandagopal, R. Vijayalakshmi, J. T. Du, and B. Cocks, "Transfer entropy and information flow patterns in functional brain networks during cognitive activity," in *Proc. Int. Conf. Neural Inf. Process.*, New York, NY, USA: Springer-Verlag, 2014, pp. 1–10.
[32] G. Sugihara et al., "Detecting causality in complex ecosystems," *Science*, vol. 338, no. 6106, pp. 496–500, 2012.
[33] J. Runge, "Causal network reconstruction from time series: From theoretical assumptions to practical estimation," *Chaos Interdisciplinary J. Nonlinear Sci.*, vol. 28, no. 7, 2018, Art. no. 075310.

[34] J. Runge, P. Nowack, M. Kretschmer, S. Flaxman, and D. Sejdinovic, "Detecting and quantifying causal associations in large nonlinear time series datasets," *Sci. Adv.*, vol. 5, no. 11, 2019, Art. no. eaau4996.
[35] X. San Liang, "Information flow within stochastic dynamical systems," *Phys. Rev. E*, vol. 78, no. 3, 2008, Art. no. 031113.
[36] X. San Liang, "Local predictability and information flow in complex dynamical systems," *Physica D*, vol. 248, pp. 1–15, Apr. 2013.
[37] X. San Liang, "Unraveling the cause-effect relation between time series," *Phys. Rev. E*, vol. 90, no. 5, 2014, Art. no. 052150.
[38] X. San Liang, "Normalizing the causality between time series," *Phys. Rev. E*, vol. 92, no. 2, 2015, Art. no. 022126.
[39] X. San Liang, "The Liang-Kleeman information flow: Theory and applications," *Entropy*, vol. 15, no. 1, pp. 327–360, 2013.
[40] B. Yi and S. Bose, "Quantum Liang information flow as causation quantifier," *Phys. Rev. Lett.*, vol. 129, no. 2, 2022, Art. no. 020501.
[41] B. Lindner, L. Auret, M. Bauer, and J. W. D. Groenewald, "Comparative analysis of Granger causality and transfer entropy to present a decision flow for the application of oscillation diagnosis," *J. Process Control*, vol. 79, pp. 72–84, Jul. 2019.
[42] X. San Liang, "Causation and information flow with respect to relative entropy," *Chaos Interdisciplinary J. Nonlinear Sci.*, vol. 28, no. 7, 07 2018, Art. no. 075311.
[43] D. T. Hristopulos, A. Babul, S. Babul, L. R. Brucar, and N. Virji-Babul, "Disrupted information flow in resting-state in adolescents with sports related concussion," *Front. Hum. Neurosci.*, vol. 13, 2019, Art. no. 00419.
[44] J. Cong et al., "Altered default mode network causal connectivity patterns in autism spectrum disorder revealed by Liang information flow analysis," *Human Brain Mapping*, vol. 44, no. 6, pp. 2279–2293, 2023.
[45] A. Stips, D. Macias, C. Coughlan, E. García-Górriz, and X. S. Liang, "On the causal structure between CO_2 and global temperature," *Sci. Rep.*, vol. 6, 2016, Art. no. 21691.
[46] Y. Rong and X. S. Liang, "An information flow-based sea surface height reconstruction through machine learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–9, 2022.
[47] Z. Chen, B. Wang, and A. N. Gorban, "Multivariate Gaussian and Student-t process regression for multi-output prediction," *Neural Comput. Appl.*, vol. 32, pp. 3005–3028, Apr. 2020.
[48] M. B. Priestley, *Spectral Analysis and Time Series: Probability and Mathematical Statistics*, vol. 1 (Probability and Mathematical Statistics). London, U.K.: Academic, 1981.
[49] S. Bochner, *Lectures on Fourier Integrals*. Princeton, NJ, USA: Princeton Univ. Press, 1959.
[50] A. Papoulis and S. U. Pillai, *Probability Random Variables and Stochastic Processes*, 4th ed. Boston, MA, USA: McGraw-Hill, 2002.
[51] H. Cramér, "On the theory of stationary random processes," *Ann. Math.*, vol. 41, no. 1, pp. 215–230, 1940.
[52] D. T. Hristopulos and E. Porcu, "Multivariate Spartan spatial random field models," *Probab. Eng. Mech.*, vol. 37, pp. 84–92, Jul. 2014.
[53] K. V. Mardia and C. R. Goodall, "Spatial-temporal analysis of multivariate environmental monitoring data," *Multivariate Environ. Statist.*, vol. 6, no. 76, pp. 347–385, 1993.
[54] M. G. Genton and W. Kleiber, "Cross-covariance functions for multivariate geostatistics," *Statist. Sci.*, vol. 30, no. 2, pp. 147–163, 2015.
[55] S. Banerjee and A. E. Gelfand, "On smoothness properties of spatial processes," *J. Multivariate Anal.*, vol. 84, no. 1, pp. 85–100, 2003.
[56] X. San Liang, "Information flow and causality as rigorous notions ab initio," *Phys. Rev. E*, vol. 94, no. 5, 2016, Art. no. 052201.
[57] H. J. Thiébaux and F. W. Zwiers, "The interpretation and estimation of effective sample size," *J. Appl. Meteorol. Climatol.*, vol. 23, no. 5, pp. 800–811, 1984.
[58] R. E. Kass, B. P. Carlin, A. Gelman, and R. M. Neal, "Markov chain Monte Carlo in practice: A roundtable discussion," *Amer. Statistician*, vol. 52, no. 2, pp. 93–100, 1998.
[59] M. S. Bartlett, "On the theoretical specification and sampling properties of autocorrelated time-series," *Suppl. J. Roy. Statist. Soc.*, vol. 8, no. 1, pp. 27–41, 1946.
[60] M. H. Quenouille, "Notes on the calculation of autocorrelations of linear autoregressive schemes," *Biometrika*, vol. 34, no. 3/4, pp. 365–367, 1947.
[61] S. Afyouni, S. M. Smith, and T. E. Nichols, "Effective degrees of freedom of the Pearson's correlation coefficient under autocorrelation," *NeuroImage*, vol. 199, pp. 609–625, Oct. 2019.
[62] D. T. Hristopulos, *Random Fields for Spatial Data Modeling: A Primer for Scientists and Engineers*. Dordrecht, The Netherlands: Springer, 2020.
[63] D. T. Hristopulos, "Non-separable covariance kernels for spatiotemporal Gaussian processes based on a hybrid spectral method and the harmonic oscillator," *IEEE Trans. Inf. Theory*, vol. 70, no. 2, pp. 1268–1283, Feb. 2024.
[64] R. J. Adler, *The Geometry of Random Fields*. New York, NY, USA: Wiley, 1981.
[65] H. Cramér and M. R. Leadbetter, *Stationary and Related Stochastic Processes: Sample Function Properties and Their Applications*. Mineola, NY, USA: Dover, 2004.
[66] S. Särkkä and A. Solin, *Applied Stochastic Differential Equations*, vol. 10. Cambridge, U.K.: Cambridge University Press, 2019.
[67] G. E. Uhlenbeck and L. S. Ornstein, "On the theory of the Brownian motion," *Phys. Rev.*, vol. 36, no. 5, pp. 823–841, Sep. 1930.
[68] M. E. Johnson, *Multivariate Statistical Simulation*. New York, NY, USA: Wiley, 1987.
[69] N. J. Higham, "Computing real square roots of a real matrix," *Linear Algebra Appl.*, vols. 88–89, pp. 405–430, Apr. 1987.
[70] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods* (Statistics), 2nd ed. New York, NY, USA: Springer-Verlag, 2009.
[71] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006. Accessed: Feb. 1, 2024. [Online]. Available: https://direct.mit.edu/books/book/2320/Gaussian-Processes-for-Machine-Learning
[72] P. A. Stokes and P. L. Purdon, "A study of problems encountered in Granger causality analysis from a neuroscience perspective," *Proc. Nat. Acad. Sci.*, vol. 114, no. 34, pp. E7063–E7072, 2017.
[73] D. N. Politis and J. P. Romano, "Large sample confidence regions based on subsamples under minimal assumptions," *Ann. Statist.*, vol. 89, no. 428, pp. 2031–2050, 1994.
[74] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. Cambridge, U.K.: Cambridge Univ. Press, 2007.
[75] X. San Liang, "Normalized multivariate time series causality analysis and causal graph reconstruction," *Entropy*, vol. 23, no. 6, 2021, Art. no. 679.
[76] X. S. Liang and X.-Q. Yang, "A note on causation versus correlation in an extreme situation," *Entropy*, vol. 23, no. 3, 2021, Art. no. 316.
[77] L. Barnett, A. B. Barrett, and A. K. Seth, "Granger causality and transfer entropy are equivalent for Gaussian variables," *Phys. Rev. Lett.*, vol. 103, no. 23, Dec. 2009, Art. no. 238701.

Dionissios T. Hristopulos (Senior Member, IEEE) received the Diploma degree in electrical engineering from the National Technical University of Athens and the Ph.D. degree in physics from Princeton University. He is a Professor with the School of Electrical and Computer Engineering, Technical University of Crete (TUC), Greece and the Director of the master's program in machine learning and data science since 2022. He has held appointments in the Department of Environmental Sciences and Engineering, University of North Carolina at Chapel Hill, and the Pulp and Paper Research Institute of Canada (currently FPInnovations). He returned to Greece in 2022 as an Associate Professor in geostatistics with the School of Mineral Resources Engineering, TUC. He is a member of the American Physical Society and life member of the International Association of Mathematical Geosciences (IAMG). He serves on the editorial boards of Computers & Geosciences and Stochastic Environmental Research and Risk Assessment. He is the author of *Random Fields for Spatial Data Modeling: A Primer for Scientists and Engineers* (Springer, 2020). His current research interests include spatiotemporal statistics, statistical physics, Gaussian processes, and applications in problems related to neuroscience, the environment, energy, and natural resources. In 2003, he received (jointly with Tetsu Uesaka) the Johannes A. Van den Akker International Prize for Advances in Paper Physics, and he was awarded the 2024 Georges Matheron lectureship by the IAMG.