PHYSICAL REVIEW E **96**, 042150 (2017)

Multiscale Granger causality

Luca Faes and Giandomenico Nollo
Bruno Kessler Foundation, Trento, Italy
and BIOtech, Department of Industrial Engineering, University of Trento, Italy

Sebastiano Stramaglia
Dipartimento di Fisica, Universitá degli Studi Aldo Moro, Bari, Italy
and Istituto Nazionale di Fisica Nucleare, Sezione di Bari, Italy

Daniele Marinazzo
Data Analysis Department, Ghent University, Ghent, Belgium

(Received 27 March 2017; revised manuscript received 18 August 2017; published 25 October 2017)

In the study of complex physical and biological systems represented by multivariate stochastic processes, an issue of great relevance is the description of the system dynamics spanning multiple temporal scales. While methods to assess the dynamic complexity of individual processes at different time scales are well established, multiscale analysis of directed interactions has never been formalized theoretically, and empirical evaluations are complicated by practical issues such as filtering and downsampling. Here we extend the very popular measure of Granger causality (GC), a prominent tool for assessing directed lagged interactions between joint processes, to quantify information transfer across multiple time scales. We show that the multiscale processing of a vector autoregressive (AR) process introduces a moving average (MA) component, and describe how to represent the resulting ARMA process using state space (SS) models and to combine the SS model parameters for computing exact GC values at arbitrarily large time scales. We exploit the theoretical formulation to identify peculiar features of multiscale GC in basic AR processes, and demonstrate with numerical simulations the much larger estimation accuracy of the SS approach compared to pure AR modeling of filtered and downsampled data. The improved computational reliability is exploited to disclose meaningful multiscale patterns of information transfer between global temperature and carbon dioxide concentration time series, both in paleoclimate and in recent years.

DOI: 10.1103/PhysRevE.96.042150

I. INTRODUCTION

Granger causality (GC) is a powerful tool for assessing directional interactions from time-series data according to the notion of time-lagged influence first proposed by Wiener [1] and then formalized by Granger [2] and Geweke [3,4] in the framework of vector autoregressive (AR) modeling of stochastic processes. Since its formulation, GC has gained increasing popularity and is nowadays ubiquitously employed in several scientific fields ranging from econometrics to social and climate sciences, neuroscience and physiology [5–8]. The great success of this measure comes from its conceptual simplicity, data-driven nature, and relative ease of implementation. An additional appealing property of GC is the principled interpretation of its generalized probabilistic formulation, which is closely related to the the information-theoretic concept of transfer entropy [9].

Many processes in physics, biology, and other fields exhibit dynamics spanning multiple temporal scales [10–14]. The multiscale properties of an observed stochastic process can be explored first by resampling at different temporal scales the originally measured realization of the process, and then assessing the dynamical complexity of the rescaled series [15]. This approach has been followed with great success to quantify the multiscale behavior of the individual dynamics of scalar processes [11–14]. However, its extension to the multiscale computation of the information transfer between processes, though attempted in empirical studies [16,17], is far less straightforward. In fact, the multiscale evaluation of lagged influence measures such as the GC is severely complicated by theoretical and practical issues [18,19]. These issues arise from the rescaling procedure, which essentially consists of a filtering step eliminating the fast temporal scales (classically performed by averaging [15]) followed by a downsampling step that coarse-grains the time series around the selected scale. The filtering step leaves theoretically unchanged the GC values, but degrades severely their estimation affecting reliability, stability, and data demand [20]. The downsampling step is even more problematic, as it alters GC values in a way that was unknown until very recently [18,19] and impacts consistently detectability and accuracy of GC estimates. For these reasons, research on the data-driven inference of the multiscale structure of coupled processes, though holding a great potential, is still largely undeveloped.

To provide a formal extension of GC to multiscale analysis, here we introduce an analytical frame for its computation on linear multivariate stochastic processes subjected to averaging and downsampling. We exploit the theory of state space (SS) models [18,19,21] to yield exact GC values for for coupled processes observed at different time scales. The high computational reliability of the associated multiscale GC estimator is demonstrated in simulated AR processes and then used to explore, spanning a wide range of time scales, the patterns of information transfer between anthropogenic emissions and global temperatures.

II. GRANGER CAUSALITY

A. Definition

To lay the groundwork for multiscale GC computation and introduce notations, we start resuming the calculation of GC for general multivariate processes [2,4]. Let us consider a discrete-time, stationary vector stochastic process composed of M real-valued zero-mean scalar processes, $Y_n = [y_{1,n} \cdots y_{M,n}]^T$, $-\infty < n < \infty$, and assume the process y_j as the *target* and the process y_i as the *driver* (the remaining $M-2$ processes form the vector Y_k, where $k = \{1,\ldots,M\}\backslash\{i,j\}$). Then, denoting the present and the past of vector and scalar variables respectively as Y_n, y_n, and $Y_n^- = [Y_{n-1}^T Y_{n-2}^T \cdots]$, $y_n^- = [y_{n-1} y_{n-2} \cdots]$, GC from y_i to y_j (conditional on Y_k) quantifies the extent to which $y_{i,n}^-$ improves the prediction of $y_{j,n}$ above and beyond the extent to which $y_{j,n}$ is predicted by $y_{j,n}^-$ and $Y_{k,n}^-$. This definition is assessed in the time domain performing a regression of the present of the target on the past of all processes, yielding the prediction error $e_{j|ijk,n} = y_{j,n} - \mathbb{E}[y_{j,n}|Y_n^-]$, and on the past of all processes except the driver, yielding the prediction error $e_{j|jk,n} = y_{j,n} - \mathbb{E}[y_{j,n}|y_{j,n}^-, Y_{k,n}^-]$ ($\mathbb{E}$ is the expectation operator). The prediction error variances resulting from these "full" and "restricted" regressions, $\lambda_{j|ijk} = \mathbb{E}[e_{j|ijk,n}^2]$ and $\lambda_{j|jk} = \mathbb{E}[e_{j|jk,n}^2]$ are then combined to yield GC from y_i to y_j as [4]

$$F_{i\to j} = \ln \frac{\lambda_{j|jk}}{\lambda_{j|ijk}}. \tag{1}$$

The measure (1) is the log-likelihood ratio for the two linear regressions associated with the projections $\mathbb{E}[y_{j,n}|Y_n^-]$ and $\mathbb{E}[y_{j,n}|y_{j,n}^-, Y_{k,n}^-]$ [18]. According to the axiomatic definition of transfer entropy and its equivalence (up to a factor of 2) to Eq. (1) for Gaussian variables [9], the GC can be interpreted as the rate of "information transfer" from driver to target.

B. State-space formulation

Following the derivations of a recent work by Barnett *et al.* [18], now we move to describe the computation of GC for state space (SS) processes. The very well-known linear SS representation of an observed multivariate process Y is given by [22]

$$X_{n+1} = \mathbf{A}X_n + W_n, \tag{2a}$$

$$Y_n = \mathbf{C}X_n + V_n, \tag{2b}$$

where X is the state (unobserved) process, and W and V are zero-mean white-noise processes with covariances $\mathbf{\Xi} \equiv \mathbb{E}[W_n W_n^T]$ and $\mathbf{\Psi} \equiv \mathbb{E}[V_n V_n^T]$, and cross-covariance $\mathbf{\Theta} \equiv \mathbb{E}[W_n V_n^T]$.

The SS process has an equivalent representation, referred to as "innovations form" SS (ISS), evidencing the *innovations* $E_n = Y_n - \mathbb{E}[Y_n|Y_n^-]$, i.e., the residuals of the linear regression of Y_n on its infinite past Y_n^-, whose covariance matrix is $\mathbf{\Phi} \equiv \mathbb{E}[E_n E_n^T]$. The ISS representation, which is typically associated with Kalman filtering, is characterized by the state process $Z_n = \mathbb{E}[X_n|Y_n^-]$ and by the Kalman Gain matrix $\mathbf{K}$:

$$Z_{n+1} = \mathbf{A}Z_n + \mathbf{K}E_n, \tag{3a}$$

$$Y_n = \mathbf{C}Z_n + E_n. \tag{3b}$$

The SS and ISS representations share the state and observation matrices $\mathbf{A}$ and $\mathbf{C}$, and differ in the noise matrices $(\mathbf{\Xi},\mathbf{\Psi},\mathbf{\Theta})$ and $(\mathbf{K},\mathbf{\Phi})$. To find the ISS parameters $(\mathbf{A},\mathbf{C},\mathbf{K},\mathbf{\Phi})$ from the SS parameters $(\mathbf{A},\mathbf{C},\mathbf{\Xi},\mathbf{\Psi},\mathbf{\Theta})$ it is necessary to solve a so-called discrete algebraic Ricatti equation (DARE), formulated in terms of the state error variance matrix $\mathbf{P}$:

$$\mathbf{P} = \mathbf{APA}^T + \mathbf{\Xi} - (\mathbf{APC}^T + \mathbf{\Theta})(\mathbf{CPC}^T + \mathbf{\Psi})^{-1}(\mathbf{CPA}^T + \mathbf{\Theta}^T), \tag{4}$$

from which $\mathbf{K}$ and $\mathbf{\Phi}$ are obtained as

$$\mathbf{\Phi} = \mathbf{CPC}^T + \mathbf{\Psi},$$

$$\mathbf{K} = (\mathbf{APC}^T + \mathbf{\Theta})\mathbf{\Phi}^{-1}. \tag{5}$$

Then, GC can be computed from the ISS parameters as follows [18]. The error variance of the full regression is simply the jth diagonal element of the innovation covariance, $\lambda_{j|ijk} = \mathbf{\Phi}(j,j)$. The error of the restricted regression is obtained by forming a *submodel* that excludes the driver process, i.e., considering a state space model with state equation (3a) and observation equation

$$Y_n^{(jk)} = \mathbf{C}^{(jk)}Z_n + E_n^{(jk)}, \tag{6}$$

where the superscript $^{(a)}$ denotes selection of the rows with indices a of a matrix. Of note, this technique represents one of the key elements in the derivation of a rigorous formalism for defining the information flow between the states of both discrete stochastic mappings and continuous time stochastic systems [23]. Here, we have that the submodel (3a) and (6) is an SS model with parameters $[\mathbf{A},\mathbf{C}^{(jk)},\mathbf{K\Phi K}^T,\mathbf{\Phi}(jk,jk),\mathbf{K\Phi}(:,jk)]$, which can be converted to an ISS model with innovation covariance $\mathbf{\Phi}^R$ solving the DARE (4) and (5), so that the restricted error variance becomes $\lambda_{j|jk} = \mathbf{\Phi}^R(j,j)$. This shows that GC can be computed numerically from the ISS parameters $(\mathbf{A},\mathbf{C},\mathbf{K},\mathbf{\Phi})$ of an observed process Y.

III. MULTISCALE GRANGER CAUSALITY

In this section we develop our framework for the multiscale computation of GC for linear multivariate processes. Here we consider the most common operalization of GC, i.e., that grounded on the AR representation of multivariate processes [2,7–9]

$$Y_n = \sum_{k=1}^{p} \mathbf{A}_k Y_{n-k} + U_n, \tag{7}$$

where p is the model order, $\mathbf{A}_k$ are $M \times M$ matrices of coefficients, and $U_n = [u_{1,n} \cdots u_{M,n}]^T$ is a vector of M zero-mean Gaussian innovation processes with covariance matrix $\Sigma \equiv \mathbb{E}[U_n U_n^T]$. To study the observed process Y at the temporal scale identified by the scale factor τ, we apply to each constituent process $y_m, m = 1,\ldots,M$, the following transformation which performs a weighted average of q consecutive samples of the process:

$$\bar{y}_{m,n} = \sum_{l=0}^{q} b_l y_{m,n\tau-l}. \tag{8}$$

This rescaling operation corresponds to transform the original process Y through a two-step procedure that consists of

the following *filtering* and *downsampling* steps, performed, respectively, with a filter of order q and a rate of downsampling equal to τ. The filtering and downsampling steps yield, respectively, the processes $\tilde{Y}$ and $\bar{Y}$ defined as

$$\tilde{Y}_n = \sum_{l=0}^{q} b_l Y_{n-l}, \tag{9a}$$

$$\bar{Y}_n = \tilde{Y}_{n\tau}. \tag{9b}$$

The change of scale in Eq. (8) generalizes the averaging procedure originally proposed in Ref. [15], which sets $q = \tau - 1$ and $b_l = 1/\tau$. In this study we identify the b_l as the coefficients of a linear lowpass filter with cutoff frequency set at $f_\tau = 1/2\tau$ to avoid aliasing in the subsequent downsampling step. Here, we develop a FIR filter of order q using the window method and implementing a Hamming window [24]. The design of a causal filter which performs one-side filtering was chosen on purpose to avoid that past and future samples mix up in the filtering process with potentially harmful consequences on the evaluation of causality. The use of one-side filtering is in agreement with the adoption of a Euler forward scheme, rather than a central differencing scheme, in the definition of causal measures of information flow (see, e.g., [25,26]). Moreover, the use of a FIR filter achieves better elimination of the fast temporal scales with respect to the averaging procedure commonly adopted in multiscale complexity analysis [15]; in our preliminary work [27] we showed indeed that the use of simple averaging may induce the detection of spurious causal influences over uncoupled directions.

Substituting Eq. (7) in Eq. (9a), the filtering step leads to the process representation

$$\tilde{Y}_n = \sum_{k=1}^{p} \mathbf{A}_k \tilde{Y}_{n-k} + \sum_{l=0}^{q} \mathbf{B}_l U_{n-l}, \tag{10}$$

where $\mathbf{B}_l = b_l \mathbf{I}_M$ ($\mathbf{I}_M$ is the $M \times M$ identity matrix). Hence, the change of scale introduces a moving average (MA) component of order q in the original AR(p) process, transforming it into an ARMA(p,q) process. Then, exploiting the close relation between ARMA and SS models [21], the process (10) is turned into an ISS model by defining the state process $\tilde{Z}_n = [Y_{n-1}^T \cdots Y_{n-p}^T U_{n-1}^T \cdots U_{n-q}^T]^T$ that, together with $\tilde{Y}_n$, obeys the state equations (3) with parameters $(\tilde{\mathbf{A}}, \tilde{\mathbf{C}}, \tilde{\mathbf{K}}, \tilde{\mathbf{\Phi}})$, where

$$\tilde{\mathbf{C}} = [\mathbf{A}_1 \cdots \mathbf{A}_p \quad \mathbf{B}_1 \cdots \mathbf{B}_q],$$

$$\tilde{\mathbf{A}} = \begin{bmatrix} \tilde{\mathbf{C}} & & \\ \mathbf{I}_{M(p-1)} & \mathbf{0}_{M(p-1)\times M(q+1)} & \\ \mathbf{0}_{M\times M(p+q)} & & \\ \mathbf{0}_{M(q-1)\times Mp} & \mathbf{I}_{M(q-1)} & \mathbf{0}_{M(q-1)\times M} \end{bmatrix},$$

$$\tilde{\mathbf{K}} = [\mathbf{I}_M \quad \mathbf{0}_{M\times M(p-1)} \quad \mathbf{B}_0^{-T} \quad \mathbf{0}_{M\times M(q-1)}]^T,$$

and where $\tilde{\mathbf{\Phi}} = \mathbf{B}_0 \Sigma \mathbf{B}_0^T$ is the covariance of the innovations $\tilde{E}_n = \mathbf{B}_0 U_n$. Moreover, the downsampled process $\bar{Y}_n$ can be put in ISS form directly from the ISS formulation of the filtered process $\tilde{Y}_n$: exploiting a recent result (theorem III in Ref. [19]), we find that $\bar{Y}_n = \tilde{Y}_{n\tau}$ has an ISS representation with state process $\bar{Z}_n = \tilde{Z}_{n\tau}$, innovation process $\bar{E}_n = \tilde{E}_{n\tau}$, and parameters $(\bar{\mathbf{A}}, \bar{\mathbf{C}}, \bar{\mathbf{K}}, \bar{\mathbf{\Phi}})$, where $\bar{\mathbf{A}} = \tilde{\mathbf{A}}^\tau$, $\bar{\mathbf{C}} = \tilde{\mathbf{C}}$, and where $\bar{\mathbf{K}}$ and $\bar{\mathbf{\Phi}}$ are obtained solving the DARE (4,5) for the SS model $(\bar{\mathbf{A}}, \bar{\mathbf{C}}, \mathbf{\Xi}_\tau, \tilde{\mathbf{\Phi}}, \mathbf{\Theta}_\tau)$ with

$$\begin{aligned} \mathbf{\Theta}_\tau &= \tilde{\mathbf{A}}^{\tau-1} \tilde{\mathbf{K}} \tilde{\mathbf{\Phi}}, \\ \mathbf{\Xi}_\tau &= \tilde{\mathbf{A}} \mathbf{\Xi}_{\tau-1} \tilde{\mathbf{A}}^T + \tilde{\mathbf{K}} \tilde{\mathbf{\Phi}} \tilde{\mathbf{K}}^T, \quad \tau \geqslant 2, \\ \mathbf{\Xi}_1 &= \tilde{\mathbf{K}} \tilde{\mathbf{\Phi}} \tilde{\mathbf{K}}^T, \quad \tau = 1. \end{aligned} \tag{11}$$

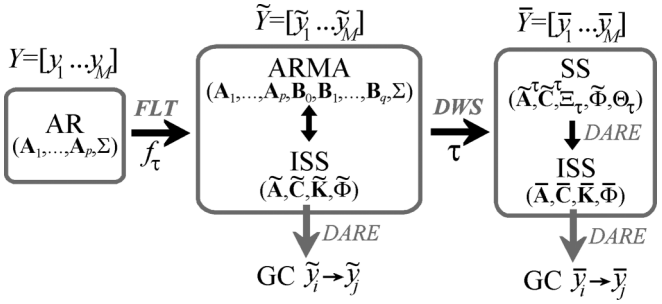

FIG. 1. Schematic representation of a linear multivariate AR process (left) and of its multiscale representation obtained through filtering (FLT) and downsampling (DWS) steps. At each step, an innovation state space (ISS) model can be defined which describes the multivariate process; then ISS submodels can be formed and, after solving a discrete algebraic Ricatti equation (DARE), GC can be computed for any scale factor $\tau \geqslant 1$ ($\tau = 1$ yields GC for the nonrescaled processes).

The overall procedure for multiscale analysis is depicted in Fig. 1: filtering with cutoff f_τ the AR(p) process Y yields an ARMA(p,q) process, which is equivalent to an ISS process; the subsequent downsampling yields a different SS process, which, in turn, can be converted to the ISS form solving the DARE. Thus, both filtered and downsampled processes are described by ISS models, whose parameters can be used to compute GC by forming a submodel in which the target is observed without considering the driver process [Eq. (6)] and solving the DARE for this submodel. This procedure allows analytical computation of GC measures for multiscale (filtered and downsampled) processes, which is illustrated in the following for simulated and real time series.

IV. SIMULATION STUDY

A. VAR process with time-delayed causal interactions

Theoretical analysis and simulations are first performed for the bivariate AR process with equations

$$y_{1,n} = c_{11} y_{1,n-d_{11}} + c_{12} y_{2,n-d_{12}} + u_{1,n}, \tag{12a}$$

$$y_{2,n} = c_{22} y_{2,n-d_{22}} + c_{21} y_{1,n-d_{21}} + u_{2,n}, \tag{12b}$$

with iid noise processes $u_{1,n}, u_{2,n} \sim \mathcal{N}(0,1)$. The parameters in Eq. (12) are set to generate autonomous dynamics with strength c_{ii} and lag d_{ii} for each scalar process y_i, and causal interactions with strength c_{ij} and lag d_{ij} from y_j to y_i $(i,j = 1,2)$. We consider two parameter configurations: unidirectional interaction at lag 2 from y_1 to y_2, obtained setting $c_{12} = 0$ and $c_{21} = 0.5, d_{21} = 2$, where also autonomous dynamics are generated for y_1 but not for y_2 ($c_{11} = 0.5, d_{11} = 1, c_{22} = 0$); bidirectional interactions with different lags and strengths ($c_{12} = 0.75, d_{12} = 2; c_{21} = 0.5, d_{21} = 7$) in the

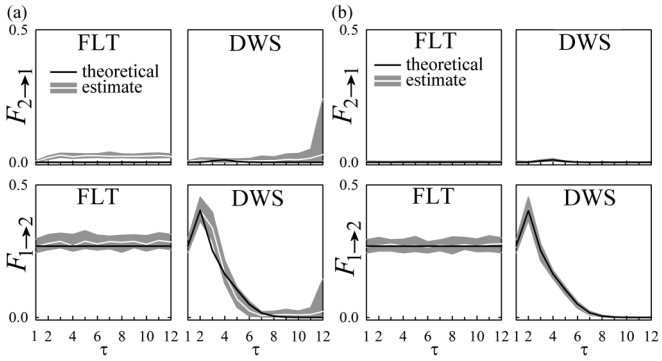

FIG. 2. Multiscale GC analysis of the AR process (12) configured to unidirectional coupling from y_1 to y_2. Plots depict the theoretical values (black lines) and distribution of estimates (median: white lines; interquartile range: gray areas) of GC computed as a function of the time scale after filtering (FLT) and downsampling (DWS). Estimates are obtained using (a) the naïve AR approach and (b) the proposed framework implemented using a lowpass FIR filter of order $q = 6$.

presence of autonomous dynamics for both processes ($c_{11} = c_{22} = 0.5, d_{11} = d_{22} = 1$). First, we study the exact values of multiscale GC obtained from the true AR parameters. The theoretical trends depicted in Figs. 2 and 3 (black solid lines) document from the perspective of SS modeling the invariance of GC under filtering, already proven in Ref. [20]. The behavior of the information transfer across multiple temporal scales is thus shaped by the downsampling step, revealing the tendency of GC to peak at scales corresponding with the lag of the imposed causal interactions: maximal information transfer is found at $\tau = 2$ for $F_{1\to2}$ in the unidirectional scheme [Figs. 3(a) and 3(b)], and at $\tau = 7$ for $F_{1\to2}$ and $\tau = 2$ for $F_{2\to1}$ in the bidirectional scheme [Fig. 3(c)]. The behavior is general, in the sense that it was observed also for different parameter configurations.

Next, we test the reliability of multiscale GC estimates obtained from finite-length realizations of Eq. (12) (a) following a standard approach whereby GC is computed performing full and restricted regressions on the filtered and downsampled time series, and (b) performing AR identification on the original time series and then applying the alternate proposed framework to the estimated AR parameters [28]. The application of the two approaches to 100 process realizations of 500 points is depicted in Figs. 2(a), and 3(a) and in Figs. 2(b), and 3(b), respectively, and indicates the need of state-space analysis: while the computation of GC after filtering and downsampling returns strongly biased and highly variable estimates, the new framework yields accurate detection of GC across multiple scales.

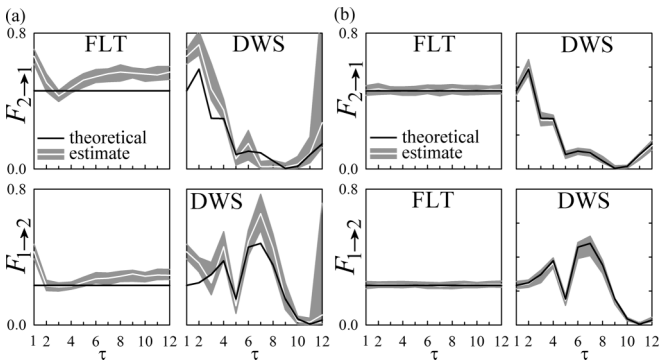

FIG. 3. Multiscale GC analysis of the AR process (12) configured to bidirectional coupling between y_1 and y_2. Plots and symbols are as in Fig. 2.

B. VAR process with multiscale causal interactions

As a second simulation example, we consider the case of coupled stochastic processes displaying multiscale structure and scale-dependent causal interactions. Specifically, we consider the bivariate process $Y_n = [y_{1,n} y_{2,n}]^T$ obtained as the instantaneous mixing of pairs of scalar processes taken from the two bivariate AR processes $X_n = [x_{1,n} x_{2,n}]^T$ and $Z_n = [z_{1,n} z_{2,n}]^T$:

$$x_{1,n} = 1.9x_{1,n-1} - 0.9025x_{1,n-2} + u_{1,n}, \tag{13a}$$

$$x_{2,n} = 0.5x_{1,n-1} + u_{2,n}, \tag{13b}$$

$$z_{1,n} = 1.6929z_{1,n-1} - 0.9025z_{1,n-2} + w_{1,n}, \tag{13c}$$

$$z_{2,n} = z_{1,n-1} + w_{2,n}, \tag{13d}$$

$$y_{1,n} = x_{1,n} + z_{2,n}, \tag{13e}$$

$$y_{2,n} = x_{2,n} + z_{1,n}. \tag{13f}$$

In the bivariate processes X and Z, autonomous dynamics are set for the subprocesses x_1 and z_1 according to Eqs. (13a) and (13c); the autonomous rhythms are obtained by placing a pole with modulus $\rho_{x_1} = \rho_{z_1} = 0.95$ in the complex plane representation of each individual subprocess, and the phase of the poles is varied to obtain slow oscillations for x_1 ($\phi_{x_1} = 0$) and faster oscillations for z_1 ($\phi_{z_1} = 0.47$). Moreover, causal interactions are imposed from x_1 to x_2 and from z_1 to z_2 according to Eqs. (13b) and (13d); in this study, the variances of the uncorrelated innovations are set to $\lambda_{u_1} = 0.25, \lambda_{u_2} = 0.5$, for the process U, and to $\lambda_{w_1} = 1, \lambda_{w_2} = 0.5$, for the process W. Then, the mixing obtained with Eqs. (13e) and (13f) is such that the bivariate process Y exhibits causal interactions from y_2 to y_1 visible at small time scales for the faster oscillations, as well as causal interactions from y_1 to y_2 visible at larger time scales for the slower oscillations.

To test the ability of our framework to detect these multiscale behaviors, we performed VAR identification on realizations of 1000 data points of the observed process Y and then computed the GC for temporal scales ranging from 1 to 15; the model order was optimized using the BIC criterion, and the order of the lowpass FIR filter was set to $q = 6$. The results of the analysis are reported in Fig. 4. The exemplary realizations shown in Fig. 4(a) display multiscale patterns characterized by a slow rhythm (cycle of ~ 70 points) superimposed to faster oscillations (cycle of ~ 13 points); the multiscale GC analysis reveals that the direction of interaction is from y_2 to y_1 for the fast oscillations ($F_{y_2\to y_1}$ is maximum at low time scales), and from y_1 to y_2 for the slower rhythm ($F_{y_1\to y_2}$ emerges at higher time scales when fast oscillations are filtered out). These results are confirmed by the analysis extended to several process realizations reported in Fig. 4(b), which indicates that GC peaks at scale $\tau = 2$ along the direction $y_2 \to y_1$, and

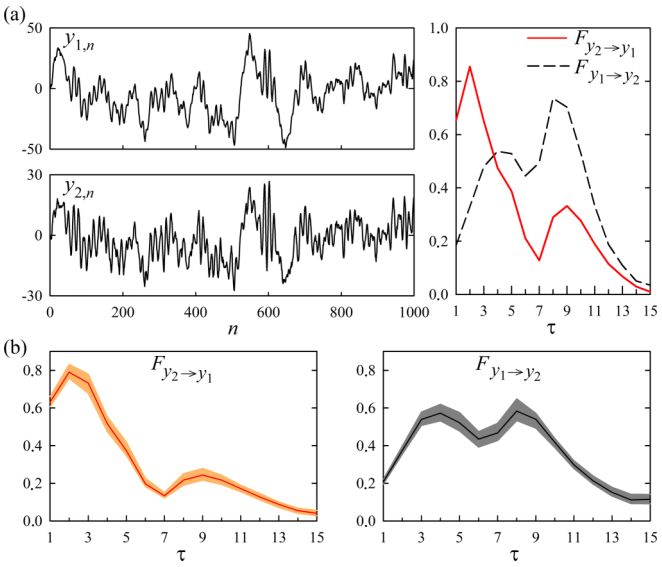

FIG. 4. Multiscale GC analysis of the bivariate process Y defined as in Eq. (13). Plots depict (a) two exemplary realizations of the observed process $Y_n = [y_{1,n}\, y_{2,n}]$ together with the GC computed as a function of the time scale τ along the two directions of interaction and (b) the distribution of estimates (median: solid lines; interquartile range: shaded areas) obtained over 100 realizations of the process.

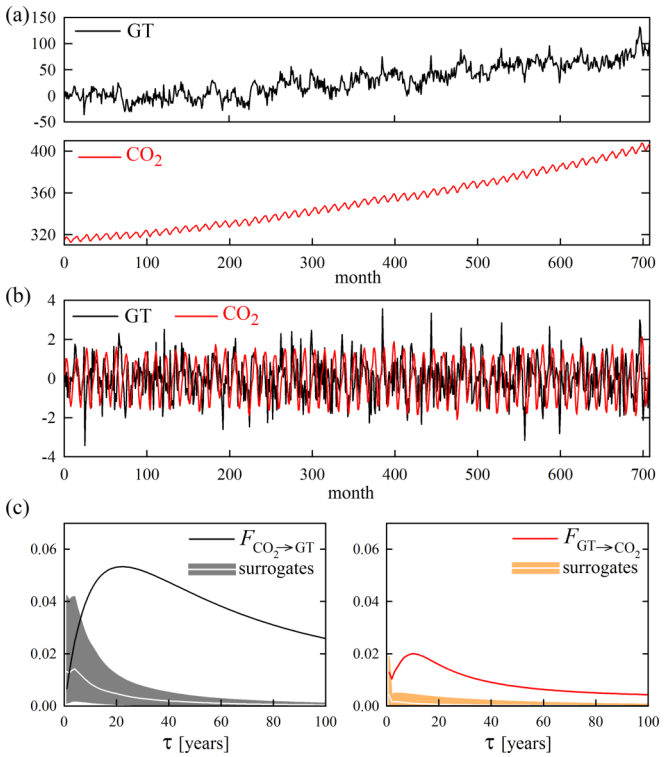

FIG. 5. Multiscale GC analysis of global temperature (GT) and CO_2 concentration for modern climate data. Plots depict (a) the original GT and CO_2 time series; (b) the time series superimposed after detrending and normalization; and (c) the multiscale GC computed on the normalized time series (solid lines) and over 100 IAAFT surrogates (median: white lines; 5th to 95th percentiles: shaded areas). Computations are performed using the proposed framework implemented with an order-6 FIR lowpass filter. The AR model order, set by the BIC criterion, is $p = 14$.

at scales $\tau = 4$ and $\tau = 8$ along the direction $y_1 \to y_2$, thus detecting the multiscale patterns of bidirectional interaction imposed in the simulation.

V. PRACTICAL APPLICATION

As a practical application, we consider the multiscale analysis of GC between carbon dioxide concentration (CO_2) and global temperature (GT). It is widely considered that the raise of CO_2 is a main cause of global warming [29], although the validity of such causal relation is still under debate. The problem of understanding the causes of climate change is usually tackled by numerical experiments using Global Climate Models [30] which aim at catching the complexity of climate dynamics. However, data-driven approaches, such as GC, are also fruitful in assessing cause-effect relationships between temperature and external forcings. In Ref. [31] it has been shown that CO_2 Granger-causes temperature, based on data from 1860 to 2008, partly from ice cores, and analyzing second differences of both CO_2 and GT. Similar conclusions were found in Ref. [32], using GC, in Ref. [33] by estimating the time rate of information flowing from one time series to the other, and in Ref. [6] using a physical approach.

Here, we first analyze the global land-ocean temperature index [34] and CO_2 concentration [35] measured at monthly resolution from March 1958 to February 2017. The measured time series, lasting 708 data points, are shown in Fig. 5(a). To fulfill stationarity criteria, we detrended the two series applying an L1 norm filter. The analyzed time series, normalized to zero mean and unit variance, are shown in Fig. 5(b). We applied the proposed framework to compute GC along the two directions of interaction for time scales ranging from 1 to 100 years. To test the statistical significance of the estimated multiscale patterns of causality, the analysis was performed both for the original time series and for a set of 100 pairs of uncoupled time series sharing the autocorrelation and amplitude distribution of the original series; these surrogate series are generated using the iterative amplitude-adjusted Fourier transform (IAAFT) algorithm [36].

The results depicted in Fig. 5(c) show that the GC along both directions is not distinguishable from the corresponding surrogate counterparts at τ equal one year, i.e., when standard GC analysis not encompassing multiple time scales is performed. On the other hand, the multiscale approach reveals, at longer time scales (> 10 years), that GC is significantly higher than the surrogate threshold along both directions of interaction, thus showing the need of a multiscale approach to put in evidence this mutual interdependency between CO_2 and GT. Moreover we remark that the GC along the direction $CO_2 \to$ GT is characterized by a higher value of the statistics computed on the original time series, but also by higher values for the surrogate time series, compared with the direction GT $\to CO_2$.

Next, we dramatically change the time scales and turn to consider paleoclimatological data, so as to analyze the GC between GT and CO_2 concentration on the Vostok Ice Core data from 400 000 to 6000 years ago, extended by

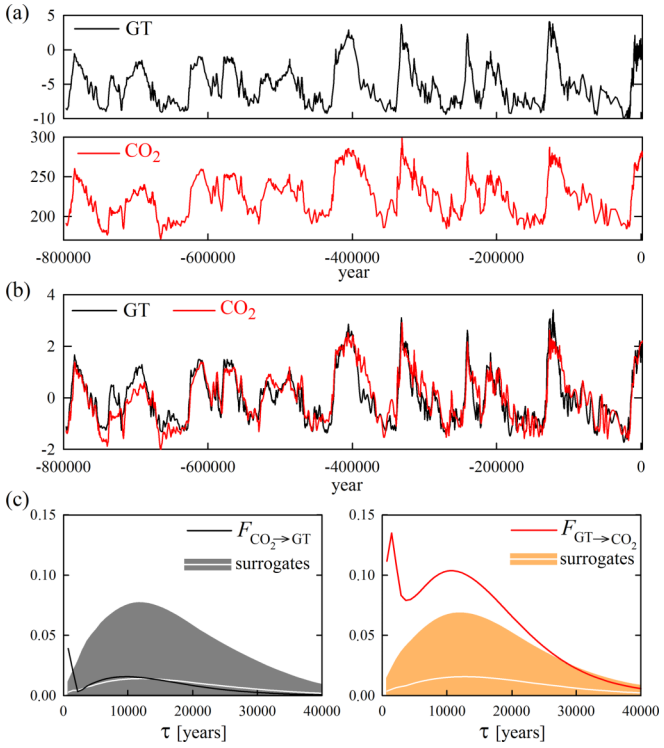

FIG. 6. Multiscale GC analysis of global temperature (GT) and CO_2 concentration for paleoclimate data. Plots depict (a) the original paleoclimatological time series; (b) the time series superimposed after uniform resampling of the time axis and normalization; and (c) the multiscale GC computed on the normalized time series (solid lines) and over 100 IAAFT surrogates (median: white lines; 5th to 95th percentiles: shaded areas). Computations are performed using the proposed framework implemented with an order-6 FIR lowpass filter. The AR model order, set by the BIC criterion, is $p = 3$.

the EPICA Dome C data which go back to 800 000 years ago [37]. The two time series, which are sampled with nonuniform time spacing, are shown in Fig. 6(a). Here, we studied the data resampled to an uniform time spacing of 729.77 years, corresponding to a time series length of 1095 points, and after normalization to zero mean and unit variance [Fig. 6(b)]; the results of multiscale GC analysis did not change substantially if the original nonuniformly sampled time series were considered, or applying slightly different uniform resampling. In Ref. [38], empirical evidences for the existence of Granger causal influences along both directions $CO_2 \rightarrow$ GT and GT $\rightarrow CO_2$ have been found after correcting for deterministic trends on the same data. Here, applying our framework for multiscale causality analysis we obtained the GC curves reported in Fig. 6(c). We find that, at paleolithic time scales, the GC GT $\rightarrow CO_2$ is highly significant and peaks around 1000 and 10 000 years. This result may be related to the lags between Antarctic deglacial warming and CO_2 increase reported in Ref. [39], and also confirms the good evidences reported on the fact that higher global temperatures do promote a rise of greenhouse gas levels [40]. The opposite causal influence from CO_2 to GT is much less pronounced and exceeds the IAAFT threshold for statistical significance only at very small time scales.

Summarizing, our results show that carbon dioxide and temperature changes are interdependent at multiple time scales, with a predominance of GT $\rightarrow CO_2$ effects at paleolithic scales, and the presence of bidirectional causal interactions between GT and CO_2 at the time scales of modern climate. These results support the expectations that changing temperatures could be held responsible for changes in greenhouse gas concentrations on paleolithic time scales, while during the last 60 years the effect of human activities becomes evident as anthropogenic radiactive forcings are seemingly driving the global temperature changes. These causal relationships between CO_2 and global warming have been recently demonstrated in Ref. [33]: in that work, the use of the rigorous formulation of information flow provided by [25] led to evidence a clear unidirectional nature for the causal relation GT $\rightarrow CO_2$ in paleoclimatological data, and for the causal relation $CO_2 \rightarrow$ GT in modern climate data; in the same work, the application of the standard GC index to modern climate data suggested the presence of bidirectional effects $CO_2 \rightarrow$ GT and GT $\rightarrow CO_2$, thus pointing to some ambiguity in the assessment of a predominant direction of interaction using GC. Our results agree with this interpretation, as we do not find a prevalent causal direction using the classical GC index computed at the smallest time scale, and the use of surrogate data indicates the lack of statistical significance (Fig. 5). Nevertheless, the analysis performed at higher time scales reveals the existence of significant GC $CO_2 \rightarrow$ GT and a nontrivial GC GT $\rightarrow CO_2$ also in the modern climate. Although this result needs to be confirmed by the implementation of more robust measures of information flow, it may be of great relevance for climate studies as it is indicative of a positive feedback which will increase the effect of anthropogenic emissions on global temperatures.

VI. CONCLUSION

The present study makes a step toward the theoretical understanding of multiscale causal relations between coupled stochastic processes, and opens the way to the reliable estimation of these relations starting from simple AR identification. This will likely boost new impetus for research in the area of data-driven causality analysis, both in physics and in a wide variety of applicative fields. The proposed framework is flexible enough to encompass more general model representations that may unveil important multiscale features of coupled processes. For instance, integrating the standard AR representation with the fractional integrated (FI) innovation modeling [41] would be straightforward as ARFI models have an SS representation, and would easily lead to assess multiscale GC in the presence of long-range correlations.

The proposed setting provides also the basis to expand the applicability of multiscale GC to nonstationary and nonlinear SS processes [42], and to formalize the exact computation of cross-scale information transfer within and between multivariate processes [17], thus opening new avenues of research in the evaluation of causal interactions among coupled processes. Of particular interest in this context is the recent formalization of the notion of information flow based on first principles, rather than axiomatic postulates or empirical proposals, implemented in Ref. [23]. The later work completes the

rigorous formalism introduced in Refs. [43,44] and provides a well-principled alternative to the operational implementation of GC, and of transfer entropy intended as its nonparametric generalization, which are known to be complicated in many practical settings to an extent that spurious causalities may be revealed (e.g., in the presence of unobserved variables, measurement noise, or inappropriate time resolution) [45–48]. Hence, the availability of a rigorous derivation of the information flowing among the components of discrete time stochastic mapping, provided in Ref. [23] and extended therein to continuous time stochastic mappings and to deterministic systems, certainly constitutes a firm basis for the design of a more faithful analysis of causality between dynamical system components operated across multiple temporal scales.

[1] N. Wiener, Modern Mathematics for Engineers **1**, 125 (1956).
[2] C. W. Granger, Econometrica **37**, 424 (1969).
[3] J. Geweke, J. Am. Stat. Assoc. **77**, 304 (1982).
[4] J. F. Geweke, J. Am. Stat. Ass. **79**, 907 (1984).
[5] J. R. Freeman, American Political Sc. **27**, 327 (1983).
[6] D. A. Smirnov and I. I. Mokhov, Phys. Rev. E **80**, 016208 (2009).
[7] S. L. Bressler and A. K. Seth, Neuroimage **58**, 323 (2011).
[8] A. Porta and L. Faes, Proc. IEEE **104**, 282 (2016).
[9] L. Barnett, A. B. Barrett, and A. K. Seth, Phys. Rev. Lett. **103**, 238701 (2009).
[10] P. Ivanov, L. Nunes Amaral, A. Goldberger, S. Havlin, M. Rosenblum, Z. Struzik, and H. Stanley, Nature **399**, 461 (1999).
[11] X. Kang, X. Jia, R. Geocadin, and N. V. Thakor, IEEE Trans. Biomed. Eng. **56**, 1023 (2009).
[12] J. Valencia, A. Porta, M. Vallverdú, F. Clariá, R. Baranowski, E. Orłowska-Baranowska, and P. Caminal, IEEE Trans. Biomed. Eng. **56**, 2202 (2009).
[13] C.-M. Chou, Entropy **13**, 241 (2011).
[14] J. Wang, P. Shang, X. Zhao, and J. Xia, Int. J. Mod. Phys. C **24**, 1350006 (2013).
[15] M. Costa, A. L. Goldberger, and C.-K. Peng, Phys. Rev. Lett. **89**, 068102 (2002).
[16] M. Lungarella, A. Pitti, and Y. Kuniyoshi, Phys. Rev. E **76**, 056117 (2007).
[17] M. Paluš, Entropy **16**, 5263 (2014).
[18] L. Barnett and A. K. Seth, Phys. Rev. E **91**, 040101 (2015).
[19] V. Solo, Neural Comput. **28**, 914 (2016).
[20] L. Barnett and A. Seth, J. Neurosci. Methods **201**, 404 (2011).
[21] M. Aoki and A. Havenner, Econ. Rev. **10**, 1 (1991).
[22] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Information and System Sciences Series, edited by Thomas Kailath (Prentice-Hall, Englewood Cliffs, NJ, 1979), Vol. 21.
[23] X. S. Liang, Phys. Rev. E **94**, 052201 (2016).
[24] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1975).
[25] X. S. Liang, Phys. Rev. E **90**, 052150 (2014).
[26] X. S. Liang, Phys. Rev. E **92**, 022126 (2015).
[27] L. Faes, A. Montalto, S. Stramaglia, G. Nollo, and D. Marinazzo, arXiv:1602.06155.
[28] S. L. Marple, *Digital Spectral Analysis: With Applications*, Vol. 5 (Prentice-Hall, Englewood Cliffs, NJ, 1987).
[29] B. B. B. Booth, C. D. Jones, M. Collins, I. J. Totterdell, P. M. Cox, S. Sitch, C. Huntingford, R. A. Betts, G. R. Harris, and J. Lloyd, Environ. Res. Lett. **7**, 024002 (2012).
[30] T. L. Delworth *et al.*, J. Clim. **19**, 643 (2006).
[31] E. Kodra, S. Chatterjee, and A. R. Ganguly, Theor. Appl. Climatol. **104**, 325 (2011).
[32] A. Attanasio, Theor. Appl. Climatol. **110**, 281 (2012).
[33] A. Stips, D. Macias, C. Coughlan, E. Garcia-Gorriz, and X. S. Liang, Sci. Rep. **6**, 21691 (2016).
[34] Nasa, Goddard Institute for Space Studies, https://data.giss.nasa.gov/, accessed: 2017-03-23.
[35] Earth System Research Laboratory, https://www.esrl.noaa.gov/, accessed: 2017-03-23.
[36] T. Schreiber and A. Schmitz, Phys. Rev. Lett. **77**, 635 (1996).
[37] National Oceanic and Atmospheric Administration, https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets/ice-core, accessed: 2017-03-23.
[38] J. Kang and R. Larsson, Theor. Appl. Climatol. **116**, 537 (2014).
[39] N. Caillon, Science **299**, 1728 (2003).
[40] M. Scheffer, V. Brovkin, and P. M. Cox, Geophys. Res. Lett. **33**, L10702 (2006).
[41] R. J. Sela and C. M. Hurvich, J. Time Ser. Anal. **30**, 631 (2009).
[42] G. Kitagawa, J. Am. Stat. Assoc. **82**, 1032 (1987).
[43] X. S. Liang and R. Kleeman, Physica D: Nonlinear Phenomena **231**, 1 (2007).
[44] X. S. Liang and R. Kleeman, Physica D: Nonlinear Phenomena **227**, 173 (2007).
[45] D. A. Smirnov, Phys. Rev. E **87**, 042917 (2013).
[46] D. W. Hahs and S. D. Pethel, Phys. Rev. Lett. **107**, 128701 (2011).
[47] H. Nalatore, M. Ding, and G. Rangarajan, Phys. Rev. E **75**, 031123 (2007).
[48] D. Chicharro and S. Panzeri, Frontiers Neuroinformatics **8**, 64 (2014).