Recurrence measure of conditional dependence and applications

Antônio M. T. Ramos,[1,2,*] Alejandro Builes-Jaramillo,[3,4] Germán Poveda,[3] Bedartha Goswami,[2,5] Elbert E. N. Macau,[1] Jürgen Kurths,[2,6] and Norbert Marwan[2]

[1]*National Institute for Space Research - INPE, 12227-010 São José dos Campos, São Paulo, Brazil*

[2]*Potsdam Institute for Climate Impact Research, Potsdam 14473, Germany*

[3]*Universidad Nacional de Colombia, Sede Medellín, Department of Geosciences and Environment, Facultad de Minas, Carrera 80 No 65-223, Bloque M2. Medellín, Colombia*

[4]*Facultad de Arquitectura e Ingeniería, Institución Universitaria Colegio Mayor de Antioquia, Carrera 78 65 - 46, Edificio patrimonial. Medellín, Colombia*

[5]*Institute of Earth and Environmental Science, University of Potsdam, Karl-Liebknecht Str. 2425, Potsdam 14476, Germany*

[6]*Department of Physics, Humboldt University Berlin, Berlin, Germany*

(Received 6 March 2017; published 11 May 2017)

Identifying causal relations from observational data sets has posed great challenges in data-driven causality inference studies. One of the successful approaches to detect direct coupling in the information theory framework is transfer entropy. However, the core of entropy-based tools lies on the probability estimation of the underlying variables. Here we propose a data-driven approach for causality inference that incorporates recurrence plot features into the framework of information theory. We define it as the recurrence measure of conditional dependence (RMCD), and we present some applications. The RMCD quantifies the causal dependence between two processes based on joint recurrence patterns between the past of the possible driver and present of the potentially driven, excepting the contribution of the contemporaneous past of the driven variable. Finally, it can unveil the time scale of the influence of the sea-surface temperature of the Pacific Ocean on the precipitation in the Amazonia during recent major droughts.

DOI: 10.1103/PhysRevE.95.052206

I. INTRODUCTION

Recurrence-based methods have been proposed to detect coupling between two processes [1–3]. Another modern approach to infer linkages among variables is applying an information theory functional in time series analysis [4]. However, in most information theoretic approaches, the required estimation of the probability distribution underlying the variables can be challenging [5]. In a recent study Goswami *et al.* [6] defined what has been called the recurrence measure of dependence (RMD), evaluating the joint recurrences between two time series by introducing a relevant lag in the possible driver. Although RMD successfully measures the lagged coupling between variables, it cannot distinguish if this relation is due to autocorrelations of the potentially driven variable. In other words, the information could have been shared in the common past of both processes, which can lead to inaccurate conclusions about information transfer.

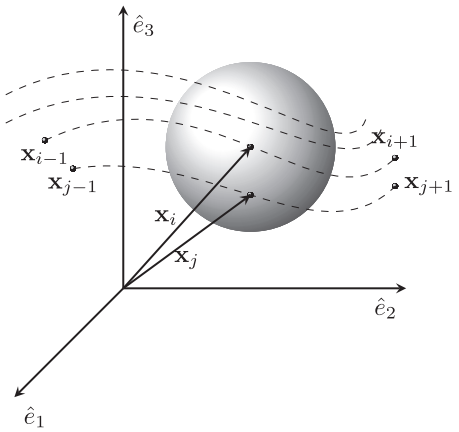
To overcome this drawback, we extend the concept of recurrence-based connectivity measurement and propose a functional called the recurrence measure of conditional dependence (RMCD), which enables us to detect lagged coupling between variables. It incorporates recurrence quantification analysis [7,8] into the concept of transfer entropy (TE) [9–12]. RMCD unveils causality accessing the recurrence dependence between two variables. It quantifies the recurrence information shared by the past of the potential driver and the present of the potentially driven variable, while conditioning to the past of the driven variable. RMCD helps to rule out any self-influence of the driven variable ensuring nonsymmetrical property that reveals directional coupling. This definition is much in line with the well-known Granger causality [13,14], however with a recurrence-based interpretation.

To demonstrate the efficacy of RMCD, we first apply it to three classes of paradigmatic model systems where, in each example, we consider different lagged coupling schemes. After that, we implement it to a real-world scenario to detect the possibly lagged coupling between the sea-surface temperature (SST) in the Pacific Ocean region Niño3.0 and precipitation in Amazonia. According to the current state of knowledge, the oceans are assumed to force the precipitation over the southwestern Amazon river basin during selected anomalous drought periods in the Amazon [15–17].

The characterization of intrinsic probabilities using fixed binning can lead to some drawbacks in the case of variables with nonstationary properties or with nonstandard distribution. However, the recurrence approach provides a dynamic binning that can cope with such variables. Analogously as the permutation entropy has contributed the ordinal pattern procedure to the information theory [18–21], the RMCD introduces the concept of recurrences into the transfer entropy definition.

This article is organized in the following way: Sec. II briefly summarizes the theoretical framework of recurrence plot. Section II B defines the recurrence measure of conditional dependence. Section II C presents a method to evaluate the coupling via a statistical significance test. Section III presents the applications of the RMCD to some paradigmatic coupled models and to a climatological study case. Finally, in Sec. IV we draw some concluding remarks about the main findings.

*antonio.ramos@inpe.br



$$\mathbf{R}^X = \begin{array}{c} \\ i-1 \\ i \\ i+1 \\ j-1 \\ j \\ j+1 \end{array} \begin{array}{c} \begin{array}{ccccccc} i-1 & i & i+1 & & j-1 & j & j+1 \end{array} \\ \begin{pmatrix} 1 & \vdots & & & \vdots & 0 & \vdots \\ \vdots & 1 & \vdots & \cdots & 0 & 1 & 0 \\ & \vdots & 1 & & \vdots & 0 & \vdots \\ \vdots & 0 & \vdots & & 1 & \vdots & \\ 0 & 1 & 0 & \cdots & \vdots & 1 & \vdots \\ \vdots & 0 & \vdots & & & \vdots & 1 \end{pmatrix} \end{array}$$

FIG. 1. Recurrence matrix $\mathbf{R}^X$ definition. Consider the system at instant i, when the trajectory at instant j recurs to the vicinity of $\mathbf{x}_i$ we write 1 in the element i,j of $\mathbf{R}^X$. In contradistinction, at instant $j+1$ the trajectory is outside the vicinity of $\mathbf{x}_i$ therefore we set 0 in $i,j+1$ of $\mathbf{R}^X$. Notice that the recurrence matrix is symmetrical.

II. THEORETICAL BACKGROUND

A. Recurrence and joint recurrence plots

Consider an observable X from which we obtain measured values represented as the scalar time-discrete series $\{x_1,x_2,\ldots,x_n\}$. Using Takens' theorem [22], we can reconstruct the phase space of its underlying dynamics. The reconstruction follows the time delay method, which states that a vector $\mathbf{x}_i$ belonging to the phase-space trajectory is defined by

$$\mathbf{x}_i = \sum_{j=1}^{m} x_{i+(j-1)d}\hat{e}_j, \tag{1}$$

where m is the embedding dimension, d is the embedding delay, and the vectors $\hat{e}_j$ are unit vectors that span an orthogonal coordinate system $\mathbb{R}^m$, such that $\mathbf{x}_i \in \mathbb{R}^m$ for discrete time points $i = 1,\ldots,N$. The analysis of time series via phase-space reconstruction requires a suitable choice for the embedding dimension m and the time delay d parameters [8]. Once the embedding parameters have been set appropriately, one can investigate the recurrence behavior of the system.

The recurrence plot (RP) [8,23] is a powerful tool to visualize and quantify the recurrences that take place in dynamical systems. The recurrences in the trajectory of system X are represented by the $N \times N$ recurrence matrix $\mathbf{R}$,

$$R_{ij}^X = \Theta(\varepsilon - \|\mathbf{x}_i - \mathbf{x}_j\|), \quad i,j = 1,\ldots,N, \tag{2}$$

where $\Theta(\cdot)$ is the Heaviside function, $\|\cdot\|$ is a suitably chosen norm, N is the number of sample points of the embedding trajectory, and ε is a threshold distance of the given norm. The issue of setting ε is discussed in detail in Ref. [8]. Figure 1 illustrates the idea behind the definition (2); a nonzero element i,j in the recurrence matrix represents a vector in the phase space $\mathbf{x}_j$ that falls into the neighborhood of vector $\mathbf{x}_i$ (sphere in the Euclidean norm).

To investigate the dynamical relation between X and another variable, we use a multivariate approach called joint recurrence matrix [24]. This matrix is the entrywise product of two or more recurrence matrices whose joint feature unveils the instances at which a recurrence in X occurs simultaneously to a recurrence in a second dynamical system. For instance, suppose another variable Y such that a given proxy $\{y_1,y_2,\ldots,y_n\}$ generates, via embedding procedure, a set of vector $\mathbf{y}_i$. Then the joint recurrence matrix between X and Y is defined as

$$R_{ij}^{X,Y} = R_{ij}^X R_{ij}^Y = \Theta(\varepsilon_X - \|\mathbf{x}_i - \mathbf{x}_j\|)\Theta(\varepsilon_Y - \|\mathbf{y}_i - \mathbf{y}_j\|). \tag{3}$$

Joint recurrence analysis can be expanded to a more general definition for n systems through the multivariate joint recurrence plot,

$$R_{ij}^{X^1,\ldots,X^n} = \prod_{k=1}^{n} \Theta\left(\varepsilon_{X^k} - \left\|\mathbf{x}_i^k - \mathbf{x}_j^k\right\|\right). \tag{4}$$

One can also use the recurrence plot approach to estimate the probability of a system to recur [8], more precisely, the probability that a particular state recurs to its ε neighbourhood in phase space. From this standpoint, RP works as an intermediary step towards the estimation of the probability distribution function associated with the recurrence behavior of the system. In Sec. II B, this probabilistic interpretation of recurrence plots allied with the framework of information theory is used to define causal measures.

B. Recurrence measure of conditional dependence

Based on the recurrence matrix $\mathbf{R}^X$, the estimation of the probability for the system X to recur to the vicinity of a particular state x_i at time i is defined as

$$p_i^X = \frac{1}{N}\sum_{j=1}^{N} R_{ij}^X \tag{5}$$

and naturally the probability estimation of a variable X to recur in the whole time interval is defined as

$$p^X = \frac{1}{N}\sum_{i=1}^{N} p_i^X, \tag{6}$$

which is analogous to the recurrence rate definition [8]. Similarly, the joint probability that system X recurs to value

x_i and system Y recurs to y_i at time i is

$$p_i^{X,Y} = \frac{1}{N}\sum_{j=1}^{N} R_{ij}^{X,Y}. \tag{7}$$

For the sake of simplicity, we omit the summation interval from now on. Following this definition, the probability dependence of X to recur conditionally to Y in a given time i is $p_i^{X|Y}$ such that

$$p_i^{X|Y} = \frac{p_i^{X,Y}}{p_i^{Y}} = \frac{\sum_j R_{ij}^{X,Y}}{\sum_j R_{ij}^{Y}}, \tag{8}$$

accordingly, the probability of a variable X to recur conditionally to Y during the whole time interval is defined as,

$$p^{X|Y} = \frac{1}{N}\sum_i \frac{\sum_j R_{ij}^{X,Y}}{\sum_j R_{ij}^{Y}}. \tag{9}$$

Inspired by the conditional mutual information [25], we propose the recurrence measure of conditional dependence (RMCD). It can quantify the recurrence relation between X and Y given Z, such that,

$$\mathcal{I}_{\mathrm{RMCD}}(X,Y|Z) = \frac{1}{N}\sum_i p_i^{X,Y,Z}\log\left(\frac{p_i^{X,Y|Z}}{p_i^{X|Z}\,p_i^{Y|Z}}\right), \tag{10}$$

the probabilities above are defined by

$$p_i^{X,Y,Z} = \frac{1}{N}\sum_j R_{ij}^{X,Y,Z}, \tag{11}$$

$$p_i^{X,Y|Z} = \frac{\sum_j R_{ij}^{X,Y,Z}}{\sum_j R_{ij}^{Z}}, \tag{12}$$

$$p_i^{X|Z} = \frac{\sum_j R_{ij}^{X,Z}}{\sum_j R_{ij}^{Z}}, \tag{13}$$

$$p_i^{Y|Z} = \frac{\sum_j R_{ij}^{Y,Z}}{\sum_j R_{ij}^{Z}}. \tag{14}$$

Substitute the above equations in (10) and we end up with the following expression

$$\mathcal{I}_{\mathrm{RMCD}}(X,Y|Z) = \frac{1}{N}\sum_i\left[\frac{1}{N}\sum_j R_{ij}^{X,Y,Z} \times \log\left(\frac{\sum_j R_{ij}^{X,Y,Z}\sum_j R_{ij}^{Z}}{\sum_j R_{ij}^{X,Z}\sum_j R_{ij}^{Y,Z}}\right)\right]. \tag{15}$$

Analogously to conditional mutual information, $\mathcal{I}_{\mathrm{RMCD}}(X,Y|Z)$ is non-negative, in particular $\mathcal{I}_{\mathrm{RMCD}}(X,Y|Z) = 0$ when $Z = X$ or $Z = Y$; another possibility is if X, Y, and Z are mutually independent.

Like transfer entropy [9], we can quantify the causal dependence of X on Y based on joint recurrence patterns between the past of the potential driver $X^\tau = \{x_{i-\tau}\}$ and the present of the potential driven Y, excluding any contribution that has already been in the contemporaneous past of the driven variable $Y^{\tau'}$, such that,

$$\mathcal{I}_{\mathrm{RMCD}}(X^\tau,Y|Y^{\tau'}) = \frac{1}{N}\sum_i\left[\frac{1}{N}\sum_j R_{ij}^{X^\tau,Y,Y^{\tau'}} \times \log\left(\frac{\sum_j R_{ij}^{X^\tau,Y,Y^{\tau'}}\sum_j R_{ij}^{Y^{\tau'}}}{\sum_j R_{ij}^{X^\tau,Y^{\tau'}}\sum_j R_{ij}^{Y,Y^{\tau'}}}\right)\right]. \tag{16}$$

Notice that the lags τ and τ' represent how many times the sample is shifted back into the variable's past. Whenever $\tau = \tau'$, RMCD incorporates the recurrence behavior into the transfer entropy definition, i.e., it captures the influence of the past of X^τ into the present of Y by excluding any contemporaneous self-influence of Y^τ.

C. Statistical test

Due to the finiteness of the sampled data, we resort on a null hypothesis test to determine the significance of RMCD in signaling a possible causal relation between the X and Y variables. The null hypothesis assumes the trajectories in the embedding space are independent realizations of the system, corresponding to a different initial condition. The statistical significance is tested using a twin surrogate hypothesis test [26,27]. The procedure finds points in the trajectory that are neighbors and also share the same neighborhood in phase space. These points are called twins, and although the points are indistinguishable regarding their recurrent neighbors, their past and future are different. Surrogates of the original trajectories are produced by replacing the future of the current point of the trajectory by either its original subsequent points or by the subsequent points of its twin (if it has one). The N_{surr} surrogates have the same dynamical invariants as the original sample, but with different recurrence structure. The 99.9% percentile of the distribution of the surrogates RMCD values is defined as the confidence interval.

The RMCD value of the original time series is compared with its respective confidence interval at different lags τ. If the RMCD is higher than the confidence interval, we reject the null hypothesis, i.e., the variables are not independent with respect to the surrogate test with significance 0.001. Otherwise, the hypothesis is accepted, meaning the variables are independent in the recurrence sense. Therefore, the rejection of the null hypothesis for a particular τ indicates a causal interaction with the time scale τ.

Finally, we perform a multiple comparison analysis between all lags investigated using the Dunn-Sidák test [28]. The significance per comparison $\alpha = 0.001$ yields a familywise error rate around 0.03.

III. RESULTS

As a proof of concept, we apply RMCD to some paradigmatic models to validate the causal inference. The models consider variables coupled with a given time lag. The recurrence calculations were made using the CRP TOOLBOX [8,29,30].

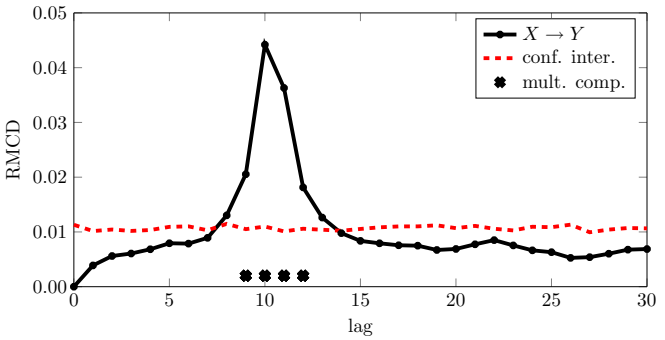


FIG. 2. RMCD value (solid black) and the respective confidence interval (dashed red) versus the lag for the ARMA[1,1] model, Eq. (17). Cross marks pinpoint the lags that pass the multiple comparison analysis.

A. Coupled ARMA model

Consider two coupled autoregressive-moving-average models ARMA[1,1] X and Y such that X influence Y after $\tau = 10$ time steps,

$$x_i = \tfrac{2}{5}x_{i-1} + \tfrac{1}{2}\xi^x_{i-1},$$
$$y_i = \tfrac{2}{5}y_{i-1} + \tfrac{1}{2}\xi^y_{i-1} + \eta x_{i-10}, \qquad (17)$$

where $\eta = 0.5$ is the coupling parameter and ξ is an uniformly distributed random number in the interval $(0,1)$. In this example, we use the following embedding parameters: $m = 2$, $d = 27$, and ε defined as 20% the size of the phase space. The embedding parameters are chosen to maximize the RMCD between the variables.

Figure 2 reveals that RMCD can pinpoint the lag of the coupling. It shows the RMCD values (solid black) and the respective confidence interval (dashed red).

The confidence interval is calculated from the 99.9th percentile of 2000 twin surrogates. The cross marks represent lags that pass the multiple comparison analysis with a familywise error rate of 0.03. With a sharp maximum at lag $\tau = 10$, the method accurately pinpoints the interaction time scale. However, the null hypothesis procedure presents some uncertainty. In this particular case, the multiple comparison analysis highlights the lag interval from 9–12 as significant outcome, having $\tau = 10$ the maximum of the flow of information between X and Y.

RMCD also plays an important role in detecting causalities in multivariate time series. For instance, consider three coupled ARMA[1,1] X, Y, and Z such that X influences Y after $\tau = 10$ time steps, analogous to Eq. (17); and Y influences Z after $\tau = 5$ time steps,

$$z_i = \tfrac{2}{5}z_{i-1} + \tfrac{1}{2}\xi^z_{i-1} + \eta y_{i-5}. \qquad (18)$$

Figure 3(a) shows the RMD unveiling the influence between X and Z, which in fact is indirect. To distinguish direct from indirect influence, one should resort to RMCD. Figure 3(b) shows the $\mathcal{I}_{\mathrm{RMCD}}(X^\tau, Z|Y^5)$ rules out this indirect influence of X over Z by conditioning the lagged Y.

B. Coupled logistic maps

Analogously to the previous section, we construct two coupled logistic maps inspired by the study of Gyllenber [31] and Hastings [32] to describe migration dynamics within two populations with a normalized carrying capacity.

In this model, the X and Y population have two governing terms. The first is the local logistic growing rate, and the second one represents the migration of individuals from one population to the other. We modify the latter by adding a lagged migration influx so that one variable influences the other after ten time steps. Both populations have the same growth rate r, and the two governing terms are balanced by the migration parameter η, i.e.,

$$x_i = (1-\eta)r x_{i-1}(1-x_{i-1}) + \eta y_{i-10},$$
$$y_i = (1-\eta)r y_{i-1}(1-y_{i-1}) + \eta x_{i-10}, \qquad (19)$$

where the coupling or migration parameter is $\eta = 0.5$ and the growth rate is $r = 3.9$. For the recurrence analysis, we employ the following embedding parameters: $m = 2$, $d = 27$, and ε is 20% the size of the phase space. The RMCD and its respective confidence interval is evaluated from lag $\tau = 0$ up to $\tau = 30$. The embedding parameters are chosen to maximize the RMCD between the variables.

Figure 4 shows the result of the RMCD of the original series X and Y (solid black) and the confidence interval (dashed red). The solid line is the RMCD calculated from a sample of the time series of the coupled logistic map model. The cross marks at the bottom represent the lags that pass the multiple comparison analysis with a familywise error rate of 0.03. Once more, the null hypothesis is rejected in a range of lags, and the multiple comparison analysis points out the ones between 9 and 11 as significant. Nevertheless, the resulting interval correctly includes the real time scale of coupling, and the

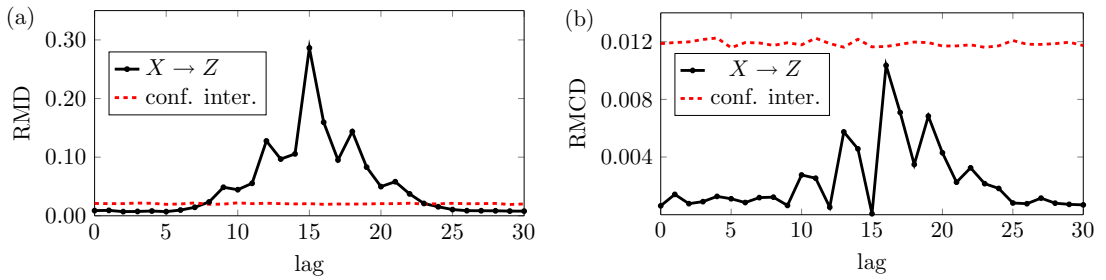


FIG. 3. Analysis of indirect influence between X and Z, Eqs. (17) and (18), using: (a) RMD and (b) RMCD, versus the lag. The solid-black line is the RMD or RMCD value and the dashed red line is the respective confidence interval.

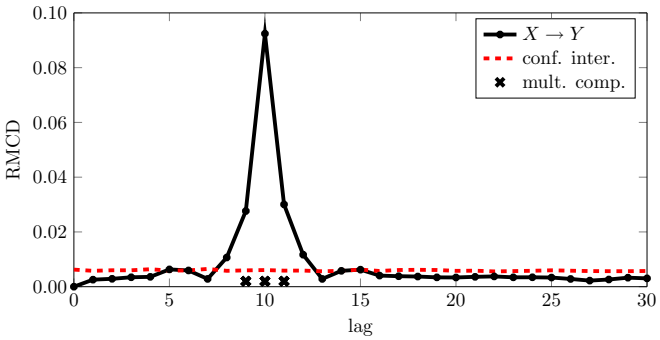


FIG. 4. RMCD value (solid black) and the respective confidence interval (dashed red) versus the lag for two coupled logistic maps whose coupling is lagged by $\tau = 10$ from one another, Eq. (19). Cross marks pinpoint the lags that pass the multiple comparison analysis.

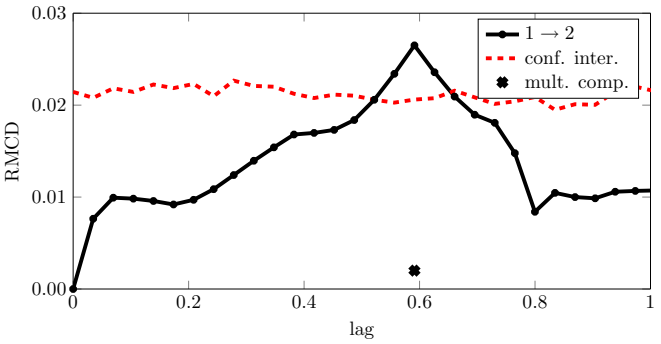


FIG. 5. RMCD value (solid black) and the respective confidence interval (dashed red) versus the lag for two coupled Lorenz system, Eq. (20). Cross mark pinpoints the lags that pass the multiple comparison analysis.

maximum value of the RMCD pinpoints the true coupling at $\tau = 10$.

One possible way to avoid such false positives is excluding not only contemporaneous past X^τ and Y^τ, but different lags that may produce indirect pathways of interactions, i.e., $\mathcal{I}_{\mathrm{RMCD}}(X^\tau; Y|Y^\tau, Y^{\tau-1}, Y^{\tau+1}, \ldots)$. But this approach leads us to the so-called curse of dimensionality, which makes the computation highly expensive. To avoid this problem, one can apply the preselected causal algorithm suggested by Runge *et al.* [33]. Despite this technical improvement, it is not the purpose of this article to dwell upon complementary algorithms that enhance the method outcome. So we focus on the recurrence properties and thus explore only past contemporaneous influence.

C. Coupled Lorenz system

Inspired by the study of Frenzel and Pompe [34], we now consider two coupled Lorenz systems such that system 1 influences system 2 with a time lag $\tau_{21} = 0.6$ unit of time; more precisely,

$$\begin{aligned}
\dot{x_i}(t) &= \sigma[y_i(t) - x_i(t)], \\
\dot{y_i}(t) &= r\, x_i(t) - y_i(t) - x_i(t) z_i(t) + \sum_{j \neq i} K_{ij} y_j^2(t - \tau_{ij}), \\
\dot{z_i}(t) &= x_i(t) y_i(t) - b\, z_i(t)
\end{aligned} \tag{20}$$

with $i, j = 1,2$ and the Lorenz systems parameters are $\sigma = 10$, $r = 28$, $b = 8/3$; the coupling parameters are $k_{21} = 1.5$ and $k_{12} = 0$, this way, system 1 influences 2 but not the other way around. We integrate the equations (20) using the adaptative Bogacki-Shampine method [35] and interpolate the final solution to get a convenient uniform integrated sample. After the transient dynamics has been discarded, we record every fifth point leading to the time step $\Delta t \approx 0.0348$. For the recurrence analysis, we use the following embedding parameters: $m = 3$, $d = 23$, and ε of 20% the size of the phase space. The embedding parameters are chosen to maximize the RMCD between the variables.

Figure 5 shows the result of the RMCD applied to time series of the variables y_2 and y_1 (solid black) and the 99.9th percentile confidence interval (dashed red) generated by the twin surrogates method. The cross mark at the bottom represents the lags that pass the multiple comparison analysis with a familywise error rate of about 0.03. We find that the null hypothesis is rejected by three lags around $\tau \approx 0.5912$ time units. However, according to the multiple comparison analysis, only the lag $\tau \approx 0.5912$ is significant, close to the lag $\tau = 0.6$.

D. Climate data set

According to historical data, the Amazon River basin has experienced several droughts in the past hundred years [36]. The droughts during 2005 and 2010 have been acknowledged as the most intense and harmful ones for the communities that depend on the river. Extreme events in the Amazon River have distinctive characteristics, and not all can be attributed or related to the same climatic forcing. The reason is that interannual precipitation variability in the Amazon basin is mainly associated with the sea-surface temperature (SST) in the Pacific and Atlantic oceans [37–41].

Some droughts have been related to the tropical Pacific Ocean macroclimatic system called El Niño Southern Oscillation (ENSO), such as the ones of 1926, 1983, 1998, and 2010 [36]. Others, such as the droughts of 1964, 1980, and 2005 have shown to be unrelated to this macroclimatic system [42]. In particular, the severe drought of 2005 is suggested to be explained by warmer SST anomalies in the Tropical North Atlantic, while the 2010 drought is related not only with anomalous SST in the Tropical North Atlantic but also with the ENSO event [16,43–45].

We apply the RMCD to investigate the possible influence of the Pacific into the South Amazon region during the years 2005 and 2010. Therefore, we use the anomalous temperature in the Pacific as the driver variable and the precipitation anomalies in South Amazon region as the driven variable. This investigation considers lags τ up to 240 days (eight months). We create an ensemble of 500 surrogates for each lag and calculate the 95th percentile.

In this study, we use daily precipitation from the Tropical Rainfall Measuring Mission (TRMM) whose product 3B42 provides satellite-measured precipitation corrected with rain gauge information [46]. Precipitation data is obtained from

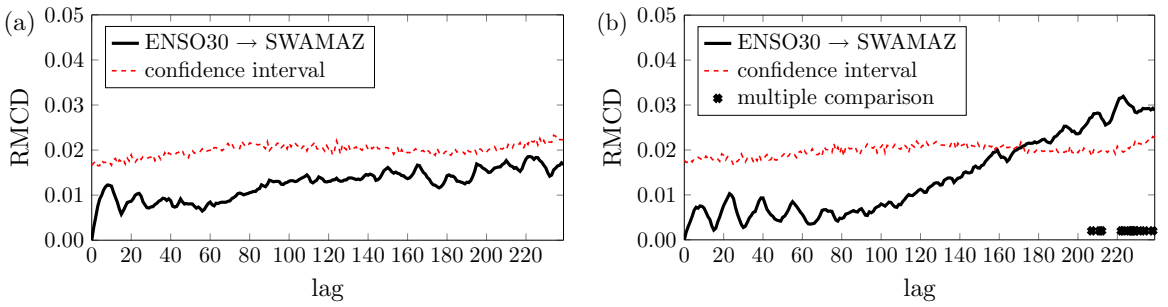


FIG. 6. Time lag influence of temperature anomaly of the Pacific Ocean (ENSO30) over the precipitation anomaly in the Southwest Amazon during (a) 2005, and (b) 2010, respectively.

the Southwest Amazon (75°W - 50°W, 15°S - 4°S) region, as defined by Zou *et al.* [15]. Precipitation records are averaged over the region, and precipitation anomalies are then estimated based on a climatology computed for the period 1999–2014. The daily SST is obtained from the NOAA OI SST High Resolution Dataset with spatial resolution of 0.25° × 0.25° and spanning from 1985–2014 [47]. The SST is averaged over a region in the Pacific Ocean called Niño3.0, which is bounded by 90°W - 150°W and 5°S - 5°N. The seasonality is computed from the climatology throughout 1982–2014. Finally, the SST anomalies are obtained after removing seasonal effects.

Figures 6(a) and 6(b) show the result of the RMCD of the original data ocean's temperature and Amazon's precipitation (solid black) and the confidence interval (dashed red) during 2005 and 2010 respectively. Figure 6(a) exhibits no significant refusal of the null hypothesis in 2005 [48,49]. So, accordingly to the RMCD method, in the year 2005, the anomalous SST in the Pacific region Niño3.0 does not play a role in the Southwest Amazon precipitation anomaly (up to a lag of eight months). This result agrees with the literature conjectures [17]. On the other hand, Fig. 6(b) exhibits an indication that the anomalous Pacific's SST influences the anomalous precipitation in the Southwest Amazon from 170 days on—around five months. The multiple comparison analysis pinpoint the lags between seven and eight months as significant outcomes. So, the anomaly in the SST responsible for causing the anomaly in the precipitation of the Amazonia dates back May and June 2009, which are months preceding the moderate Niño of 2009–2010. This result supports the premise that the Pacific has significantly influenced Amazon's drought in 2010 [36,44]. RMCD analysis agrees with a recent study about the difference between the oceanic influence in the years 2005 and 2010 [50]. However, the time scale of the influence in 2010 found here differs from that analysis of about two months. Further studies are necessary to understand such difference that may be related to the intrinsic nature of the probability estimation, for instance, the embedding parameter definition prior to the estimation.

IV. CONCLUSIONS

In this paper, we present a tool that employs recurrence properties within the transfer entropy framework. This approach can reveal the coupling time scale of several paradigmatic models. RMCD is able to identify the causal relation between variables in the discrete linear coupled ARMA model, the discrete nonlinear coupled logistic maps, and the continuous nonlinear Lorenz coupled model. Additionally, the RMCD results agree with the literature about the role of the Pacific's macroclimatic system in the Southwest Amazonia region. RMCD is indeed able to distinguish the absence (presence) of a Niño3.0's anomalous temperature influence on the anomalous precipitation in Southwest Amazon during 2005 (2010), and, in the case of 2010, it also reveals the interaction time scale. The RMCD functional reveal the causal interaction based on the recurrent behavior of the underlying dynamic system. The probability estimation procedure based on recurrence plot analysis offers a dynamic binning that can outperform other methods in a nonstationary scenario.

ACKNOWLEDGMENTS

This research was supported by the São Paulo Research Foundation FAPESP (Grants No. 2015/50122-0, No. 2014/14229-2, and No. 2015/07373-2); The Deutsche Forschungsgemeinschaft in the context of the project DFG-IRTG 1740/2, and the project DFG RTG 2043/1 "Natural hazards and risks in a changing world." The authors thank the NASA/Goddard Space Flight Center's Mesoscale Atmospheric Processes Laboratory and PPS, which provided TMPA data, and develop and compute the TMPA as a contribution to TRMM project. A.B.J. was partially supported by the program Research Grants - Short-Term Grants, 2015 (57130097) of the Deutscher Akademischer Austauschdienst (DAAD) and by the Humboldt University of Berlin. The work of G.P. was supported by Universidad Nacional de Colombia at Medellín, as a contribution to the AMAZALERT research programme, funded by the European Commission. N.M. Cross Recurrence Plot Toolbox for MATLAB®, Ver. 5.21 (R31.2) [51].

[1] M. C. Romano, M. Thiel, J. Kurths, and C. Grebogi, Estimation of the direction of the coupling by conditional probabilities of recurrence, Phys. Rev. E **76**, 036211 (2007).

[2] M. Thiel N. Marwan Y. Zou, M. C. Romano, and J. Kurths, Inferring indirect coupling by means of recurrences, Int. J. Bifurcation Chaos **21**, 1099 (2011).

[3] Y. Zou, M. C. Romano, M. Thiel, N. Marwan, and J. Kurths, Inferring indirect coupling by means of recurrences, Int. J. Bifurcation Chaos **21**, 1099 (2011).

[4] G. Balasis, R. V. Donner, S. M. Potirakis, J. Runge, C. Papadimitriou, I. A. Daglis, K. Eftaxias, and J. Kurths, Statistical mechanics and information-theoretic perspectives on complexity in the earth system, Entropy **15**, 4844 (2013).

[5] A. Kraskov, H. Stögbauer, and P. Grassberger, Estimating mutual information, Phys. Rev. E **69**, 066138 (2004).

[6] B. Goswami, N. Marwan, G. Feulner, and J. Kurths, How do global temperature drivers influence each other? Eur. Phys. J. Special Topics **222**, 861 (2013).

[7] J. P. Zbilut and C. L. Webber, Embeddings and delays as derived from quantification of recurrence plots, Phys. Lett. A **171**, 199 (1992).

[8] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, Recurrence plots for the analysis of complex systems, Phys. Rep. **438**, 237 (2007).

[9] T. Schreiber, Measuring Information Transfer, Phys. Rev. Lett. **85**, 461 (2000).

[10] J. G. Orlandi, O. Stetter, J. Soriano, T. Geisel, and D. Battaglia, Transfer entropy reconstruction and labeling of neuronal connections from simulated calcium imaging, PLoS ONE **9**, e98842 (2014).

[11] R. E. Spinney, M. Prokopenko, and J. T. Lizier, Transfer entropy in continuous time, with applications to jump and neural spiking processes, Phys. Rev. E **95**, 032319 (2017).

[12] J. Sun, C. Cafaro, and E. M. Bollt, Identifying the coupling structure in complex systems through the optimal causation entropy principle, Entropy **16**, 3416 (2014).

[13] C. W. J. Granger, Investigating causal relations by econometric models and cross-spectral methods, Econometrica **37**, 424 (1969).

[14] L. Barnett, A. B. Barrett, and A. K. Seth, Granger Causality and Transfer Entropy are Equivalent for Gaussian Variables, Phys. Rev. Lett. **103**, 238701 (2009).

[15] Y. Zou, E. E. N. Macau, G. Sampaio, A. M. T. Ramos, and J. Kurths, Do the recent severe droughts in the Amazonia have the same period of length? Clim. Dyn. **46**, 3279 (2016).

[16] J.-H. Yoon and N. Zeng, An atlantic influence on amazon rainfall, Clim. Dyn. **34**, 249 (2009).

[17] J. A. Marengo and J. C. Espinoza, Extreme seasonal droughts and floods in Amazonia: causes, trends and impacts, Int. J. Climatol. **36**, 1033 (2015).

[18] M. Matilla-García, M. Ruiz Marín, and M. I. Dore, A permutation entropy based test for causality: The volume-stock price relation, Physica A **398**, 280 (2014).

[19] H. Dickten and K. Lehnertz, Identifying delayed directional couplings with symbolic transfer entropy, Phys. Rev. E **90**, 062706 (2014).

[20] R. Monetti, W. Bunk, T. Aschenbrenner, S. Springer, and J. M. Amigó, Information directionality in coupled time series using transcripts, Phys. Rev. E **88**, 022911 (2013).

[21] K. Keller, T. Mangold, I. Stolz, and J. Werner, Permutation entropy: New ideas and challenges, Entropy **19**, 134 (2017).

[22] F. Takens, *Detecting Strange Attractors in Turbulence* (Springer, Berlin, 1981), pp. 366–381.

[23] J.-P. Eckmann, S. O. Kamphorst, and D. Ruelle, Recurrence plots of dynamical systems, Europhys. Lett. **4**, 973 (1987).

[24] M. C. Romano, M. Thiel, J. Kurths, and W. von Bloh, Multivariate recurrence plots, Phys. Lett. A **330**, 214 (2004).

[25] A. D. Wyner, A definition of conditional mutual information for arbitrary ensembles, Inf. Control **38**, 51 (1978).

[26] M. Thiel, M. C. Romano, J. Kurths, M. Rolfs, and R. Kliegl, Twin surrogates to test for complex synchronisation, Europhys. Lett. **75**, 535 (2006).

[27] M. Thiel, M. C. Romano, J. Kurths, M. Rolfs, and R. Kliegl, Generating surrogates from recurrences, Philos. Trans. R. Soc. London A **366**, 545 (2008).

[28] Z. Šidák, Rectangular confidence regions for the means of multivariate normal distributions, J. Am. Stat. Assoc. **62**, 626 (1967).

[29] N. Marwan, N. Wessel, U. Meyerfeldt, A. Schirdewan, and J. Kurths, Recurrence-plot-based measures of complexity and their application to heart-rate-variability data, Phys. Rev. E **66**, 026702 (2002).

[30] N. Marwan and J. Kurths, Nonlinear analysis of bivariate data with cross recurrence plots, Phys. Lett. A **302**, 299 (2002).

[31] M. Gyllenberg, G. Söderbacka, and S. Ericsson, Does migration stabilize local population dynamics? analysis of a discrete metapopulation model, Math. Biosci. **118**, 25 (1993).

[32] A. Hastings, Complex interactions between dispersal and dynamics: Lessons from coupled logistic equations, Ecology **74**, 1362 (1993).

[33] J. Runge, J. Heitzig, V. Petoukhov, and J. Kurths, Escaping the Curse of Dimensionality in Estimating Multivariate Transfer Entropy, Phys. Rev. Lett. **108**, 258701 (2012).

[34] S. Frenzel and B. Pompe, Partial Mutual Information for Coupling Analysis of Multivariate Time Series, Phys. Rev. Lett. **99**, 204101 (2007).

[35] P. Bogacki and L. F. Shampine, A 3(2) pair of Runge-Kutta formulas, Appl. Math. Lett. **2**, 321 (1989).

[36] J. A. Marengo, J. Tomasella, L. M. Alves, W. R. Soares, and D. A. Rodriguez, The drought of 2010 in the context of historical droughts in the amazon region, Geophys. Res. Lett. **38**, L12703 (2011).

[37] P. Aceituno, On the functioning of the southern oscillation in the south american sector. part i: Surface climate, Monthly Weather Rev. **116**, 505 (1988).

[38] J. A. Marengo, Interannual variability of surface climate in the amazon basin, Int. J. Clim. **12**, 853 (1992).

[39] G. Poveda and O. J. Mesa, Feedbacks between hydrological processes in tropical south america and large-scale ocean-atmospheric phenomena, J. Clim. **10**, 2690 (1997).

[40] L. Gimeno, A. Stohl, R. M. Trigo, F. Dominguez, K. Yoshimura, L. Yu, A. Drumond, A. M. Durán-Quesada, and R. Nieto, Oceanic and terrestrial sources of continental precipitation, Rev. Geophys. **50** (2012).

[41] R. J. van der Ent and H. H. G. Savenije, Oceanic sources of continental precipitation and the correlation with sea surface temperature, Water Resour. Res. **49**, 3993 (2013).

[42] J. A. Marengo, J. Tomasella, W. R. Soares, L. M. Alves, and C. A. Nobre, Extreme climatic events in the amazon basin, Theor. Appl. Climatol. **107**, 73 (2011).

[43] N. Zeng, J.-H. Yoon, J. A. Marengo, A. Subramaniam, C. A. Nobre, A. Mariotti, and J. D. Neelin, Causes and impacts of the 2005 Amazon drought, Environment. Res. Lett. **3**, 014002 (2008).

[44] S. L. Lewis, P. M. Brando, O. L. Phillips, G. M. F. van der Heijden, and D. Nepstad, The 2010 amazon drought, Science **331**, 554 (2011).

[45] C. A. S. Coelho, I. A. F. Cavalcanti, S. M. S. Costa, S. R. Freitas, E. R. Ito, G. Luz, A. F. Santos, C. A. Nobre, J. A. Marengo, and A. B. Pezza, Climate diagnostics of three major drought events in the amazon and illustrations of their seasonal precipitation predictions, Meteorol. Appl. **19**, 237 (2012).

[46] G. J. Huffman, D. T. Bolvin, E. J. Nelkin, D. B. Wolff, R. F. Adler, G. Gu, Y. Hong, K. P. Bowman, and E. F. Stocker, The trmm multisatellite precipitation analysis (tmpa): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, J. Hydrometeorol. **8**, 38 (2007).

[47] R. W. Reynolds, T. M. Smith, L. Chunying, D. B. Chelton, K. S. Casey, and M. G. Schlax, Daily high-resolution-blended analyses for sea surface temperature, J. Clim. **20**, 5473 (2007).

[48] J. A. Marengo, C. A. Nobre, J. Tomasella, M. D. Oyama, G. S. de Oliveira, R. de Oliveira, H. Camargo, L. M. Alves, and F. I. Brown, The drought of Amazonia in 2005, J. Clim. **21**, 495 (2008).

[49] J. A Marengo, C. A. Nobre, J. Tomasella, M. F. Cardoso, and M. D. Oyama, Hydro-climatic and ecological behavior of the drought of amazonia in 2005, Philos. Trans. R. Soc. London B **363**, 1773 (2008).

[50] A. M. de T. Ramos, Y. Zou, G. S. de Oliveira, J. Kurths, and E. E. N. Macau, Unveiling non-stationary coupling between Amazon and ocean during recent extreme events, Clim. Dyn., 1 (2017).

[51] Available at available at http://tocsy.pik-potsdam.de/CRP toolbox.