Universal window size-dependent transition of correlations in complex systems

Tao Wu, Feng An, Xiangyun Gao, Siyao Liu, Xiaotian Sun, Zhigang Wang, Zhen Su, Jürgen Kurths

Originally published as
Wu, T., An, F., Gao, X., Liu, S., Sun, X., Wang, Z., Su, Z., Kurths, J. (2023): Universal window size-dependent transition of correlations in complex systems. - Chaos, 33, 2, 023111.
https://doi.org/10.1063/5.0134944

Chaos
An Interdisciplinary Journal of Nonlinear Science

RESEARCH ARTICLE | FEBRUARY 13 2023

Universal window size-dependent transition of correlations in complex systems

Tao Wu ; Feng An ; Xiangyun Gao ; Siyao Liu; Xiaotian Sun ; Zhigang Wang ; Zhen Su ; Jürgen Kurths

Chaos 33, 023111 (2023)
https://doi.org/10.1063/5.0134944

Universal window size-dependent transition of correlations in complex systems

Cite as: Chaos **33**, 023111 (2023); doi: 10.1063/5.0134944
Submitted: 15 November 2022 · Accepted: 23 January 2023 ·
Published Online: 13 February 2023

Tao Wu,[1,2] Feng An,[3,a)] Xiangyun Gao,[1,4,a)] Siyao Liu,[5] Xiaotian Sun,[1] Zhigang Wang,[6,7] Zhen Su,[2,8] and Jürgen Kurths[2,9,a)]

AFFILIATIONS

[1]School of Economics and Management, China University of Geosciences, Beijing 100083, China
[2]Potsdam Institute for Climate Impact Research (PIK)—Member of the Leibniz Association, Potsdam 14473, Germany
[3]School of Economics and Management, Beijing University of Chemical Technology, Beijing 100029, China
[4]Key Laboratory of Carrying Capacity Assessment for Resource and Environment, Ministry of Land and Resources, Beijing 100083, China
[5]Institutes of Science and Development, Chinese Academy of Sciences, 100190 Beijing, China
[6]International Academic Center of Complex Systems, Beijing Normal University, Zhuhai 519087, China
[7]School of Systems Science, Beijing Normal University, Beijing 100875, China
[8]Department of Computer Science, Humboldt University at Berlin, Berlin 12489, Germany
[9]Department of Physics, Humboldt University at Berlin, Berlin 12489, Germany

[a)]**Authors to whom correspondence should be addressed:** af15910602135@126.com; gxy5669777@126.com; and juergen.kurths@pik-potsdam.de

ABSTRACT

Correlation analysis serves as an easy-to-implement estimation approach for the quantification of the interaction or connectivity between different units. Often, pairwise correlations estimated by sliding windows are time-varying (on different window segments) and window size-dependent (on different window sizes). Still, how to choose an appropriate window size remains unclear. This paper offers a framework for studying this fundamental question by observing a critical transition from a chaotic-like state to a nonchaotic state. Specifically, given two time series and a fixed window size, we create a correlation-based series based on nonlinear correlation measurement and sliding windows as an approximation of the time-varying correlations between the original time series. We find that the varying correlations yield a state transition from a chaotic-like state to a nonchaotic state with increasing window size. This window size-dependent transition is analyzed as a universal phenomenon in both model and real-world systems (e.g., climate, financial, and neural systems). More importantly, the transition point provides a quantitative rule for the selection of window sizes. That is, the nonchaotic correlation better allows for many regression-based predictions.

Complex connections between different units can be simply approximated by correlation analysis between corresponding time series. When the complete information (the entire time series) is considered for analysis, dynamic connections are aggregated into a single value, reflecting the overall macro linkage. When segmented information (a sliced time series) is combined with sliding windows, the underlying dynamic connections can be approximated by time-varying correlations. Intuitively, the longer the segments are, the more likely to capture cyclic behavior. A typical example is that in climate science, large-scale climate phenomena, such as seasonal changes induced by the annual cycle of solar radiation, are not observable on the timescale of diurnal cycles. Similarly, for correlation analysis, choosing a suitable window scale to capture the necessary patterns hidden in the time series is fundamental; yet, how to do so is unclear. We intend to address this issue in our work.

I. INTRODUCTION

Correlation analysis based on time series is a fundamental approach for studying the underlying mechanisms in various real-world systems, including life science, ecological, climate, and financial systems.[1–5] For example, asset correlation is often selected

as a basic indicator in financial markets where positively correlated assets have a poor ability to hedge risk.[6] Furthermore, correlation-based networks have been widely used in many fields.[7–9] However, we often find unequal or even opposite correlation relationships (e.g., Pearson coefficient values) of the same variables using data from different periods. Moreover, some inconsistent findings are identified on different scales (sampling rates, e.g., daily, weekly, and monthly observations), even using the same variables from the same periods. Generally, in a nonlinear system, variable correlations are often not constant but time-varying. Scholars usually use the sliding window framework to capture the varying correlations between time series, in which pairwise correlations are calculated on continuous window segments.[10,11] Often, the window size is chosen empirically or simply assigned a larger value to ensure that the underlying patterns can be captured.[12] Instead, this work intends to provide a quantitative selection rule for it. To this end, an intuitive approach is to analyze how different window sizes affect the dynamics of the approximated varying correlations.

Specifically, given two time series and a fixed window size, we create a correlation-based series based on pairwise nonlinear correlations (e.g., mutual information) from continuous window segments. This correlation-based series approximately embeds the dynamics of correlations between the original two time series under a specific window size, which enables the indirect detection of the dynamics of variable correlations via the correlation-based series. Moreover, similar to generic time series, correlation-based series can be investigated using many traditional time series analysis tools. This enables us to study the window size-dependent dynamics of variable correlations through their correlation-based series. Correlation-based series with different window sizes will show quite different dynamics. Taking a two-species ecological system as an example[13] [see Eq. (1) and Fig. 1(a)], we estimate its correlation-based series with different window sizes; see Figs. 1(b) and 1(c). Notably, the correlation-based series with a window size of 100 [Fig. 1(b)] exhibits lower complexity and is more regular than that with a window size of 10 [Fig. 1(c)],

$$\begin{aligned} X(t+1) &= X(t)[3.8 - 3.8X(t) - 0.02Y(t)], \\ Y(t+1) &= Y(t)[3.5 - 3.5Y(t) - 0.1X(t)]. \end{aligned} \tag{1}$$

The correlation-based series from nonlinear systems shows varying magnitudes of complexity on different window sizes. According to information theory, short series contain less information and show higher uncertainty.[14] Thus, short window segments contain less information, leading to large variations in the measured correlations from continuous segments. This may lead to correlation-based series with shorter windows having higher complexity than those with longer windows. One can speculate that (1) the decrease in complexity from a high to a low level with increasing window size may be a universal phenomenon in various real-world systems and (2) this transition in complexity may be useful for quantitatively choosing the window size.

Thus, we can detect the influence of window size on variable correlations via the complexity of the correlation-based series. The complexity can be measured by the largest Lyapunov exponent, which identifies the chaotic behaviors by quantifying the exponential divergence of initially close state-space trajectories.[15] Chaotic behavior suggests that small differences in initial conditions can yield widely diverging outcomes for a dynamical system, rendering long-term prediction of its behavior impossible in general.[15] In this work, our objective is to determine whether the correlation-based series shows chaotic behaviors under short windows and whether there exist transitions from chaotic states to nonchaotic states with increasing window size. The state transitions of the correlation-based series may facilitate the selection of a suitable window size for capturing the varying correlations in a complex system.

In this work, we propose a data-driven framework to resolve to the above issues. Our framework has roots in measuring the complexity of varying correlations concerning the window size. Specifically, we aim to find a window size-dependent state transition of varying correlations from the time series. Several benchmark models and real-world systems are employed in this work. The results show a universal phenomenon in which the varying correlations experience a critical transition from a chaotic-like state to a nonchaotic state under a specific window size. Hopefully, the transition points can be considered appropriate values in practical scenarios for sliding windows.

The remainder of this paper is organized as follows. Section II describes the methods. In Sec. III, we detect the window size-dependent transitions for both model and real-world systems. Section IV discusses the results and their implications. Section V summarizes our main findings.

II. METHODS

To uncover the window size-dependent state transitions of correlations between time series, our framework has three steps:

A. Step 1: Embedding the pairwise correlations into a correlation-based series using sliding windows

Given two time series $X(t)$ and $Y(t)$ ($t = 1, \ldots, L$, where L is the length of the series), we use sliding windows to obtain k segments by sliding 1 step (sliding other numbers of steps is discussed in Sec. III). The total number of segments k is obtained as

$$k = L - m + 1, \quad 0 < m \leq L, \tag{2}$$

where m is the window size. Since a window size that is too short is not suitable for calculating the window-wise correlation, we start with $m = 10$.

For each window segment, we measure the nonlinear correlations between two variables, i.e., mutual information. The mutual information is expressed as[2]

$$R(X(t), Y(t+d)) = \sum_{uv} p(u,v) \log \frac{p(u,v)}{p(u)p(v)}, \quad d \text{ is a nonnegative integer}, \tag{3}$$

where $p(u)$ is the probability density function of the time series $X(t)$ and $p(u,v)$ is the joint probability density function of $X(t)$ and $Y(t+d)$. We use a simple histogram tool with equally sized bins to estimate the probability densities. The mutual information $R(X(t), Y(t+d))$ can be interpreted as the excess amount of information generated by falsely assuming the two time series $X(t)$ and $Y(t+d)$ to be independent. By

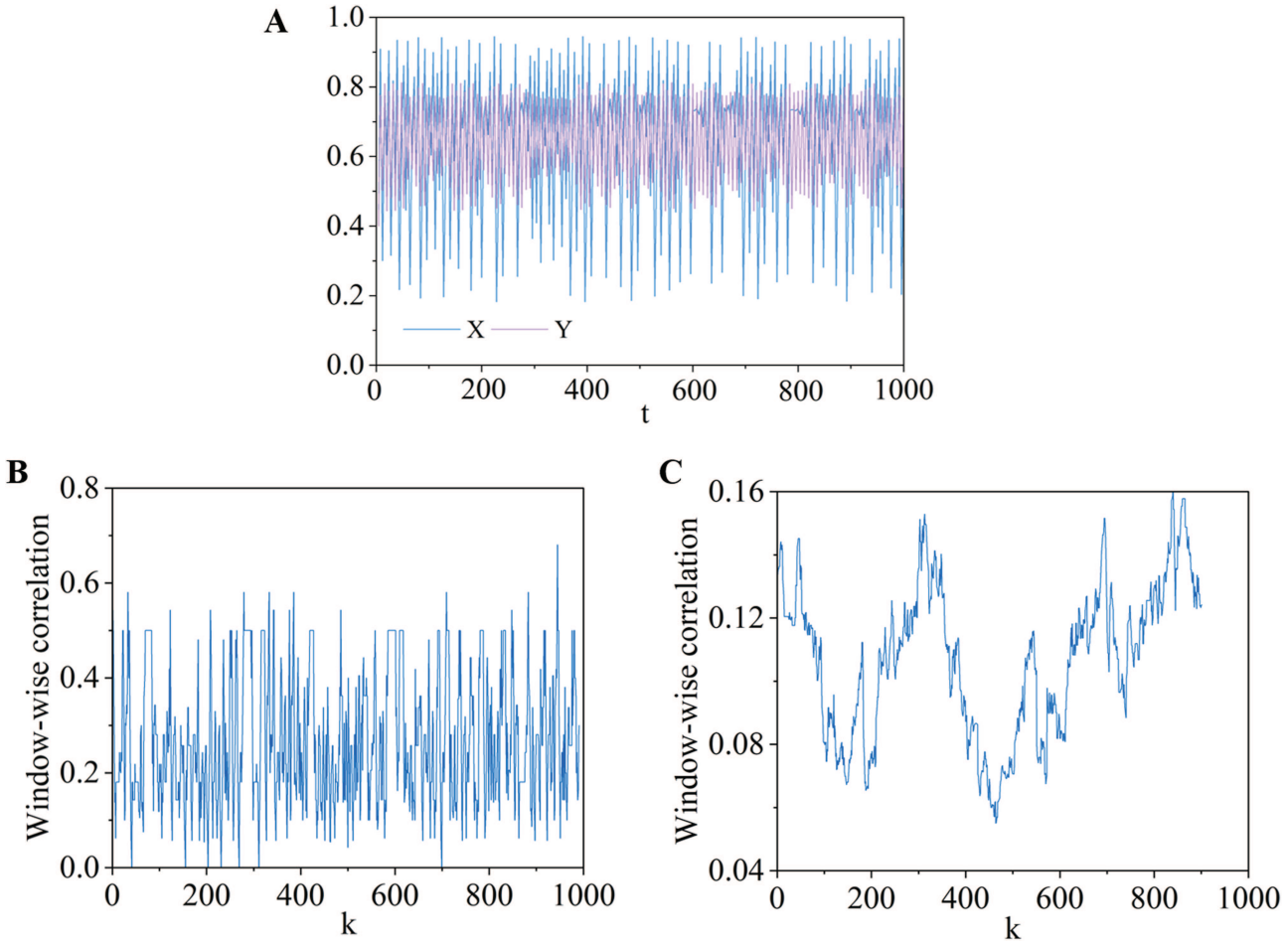

FIG. 1. Correlation-based series from the ecological system [Eq. (1)]. The original discrete series (populations of species X and Y), where the initial values are $X(0) = Y(0) = 0.4$ and the output series has a length of 1000. (b) and (c) The correlation-based series between variables X and Y with window sizes of 10 and 100, where the estimated largest Lyapunov exponents are 7.5 and −0.44, respectively. k represents the kth window, and the sliding step size is 1. The correlations are measured by mutual information (see Sec. II).

definition, $R(X(t), Y(t+d))$ is nonnegative and symmetric such that $R(X(t), Y(t+d)) = R(Y(t+d), X(t))$. If the two time series are independent, i.e., $p(u, v) = p(u)p(v)$, then $R(X(t), Y(t+d)) = 0$, and if the two time series are highly nonlinearly correlated, then $R(X(t), Y(t+d))$ will be larger.

All the estimated correlations on continuous windows form a correlation-based series $\tilde{X}(k)$ (with length $L - m + 1$), which approximates the dynamical evolutions of the correlations between variables $X(t)$ and $Y(t)$,

$$\tilde{X}(k) = \widehat{\rho}_k(X(t), Y(t+d)), \quad k = 1, \ldots, L - m + 1, \tag{4}$$

where $\widehat{\rho}_k$ represents the nonlinear correlation between the two variables at the kth window.

B. Step 2: Measuring the dynamics of the correlation-based series by the largest Lyapunov exponent

The correlation-based series $\tilde{X}(k)$ approximates the dynamics of the variable correlations; therefore, we can investigate the dynamics of the variable correlations between the original variables $X(t)$ and $Y(t)$ through $\tilde{X}(k)$. Specifically, we use the traditional Lyapunov exponent to measure the complexity of the correlation-based series in relation to the window size. In this work, the largest Lyapunov exponent is based on Rosenstein *et al.*[15] and is calculated using the Predictive Maintenance Toolbox of MATLAB R2018a. The embedding dimension E and delay τ are two critical parameters for the estimation of the largest Lyapunov exponent, and they are identified using the false nearest neighbor (FNN) method[16] and mutual information function,[17] respectively. It is well accepted that a positive largest Lyapunov exponent value indicates chaos (higher

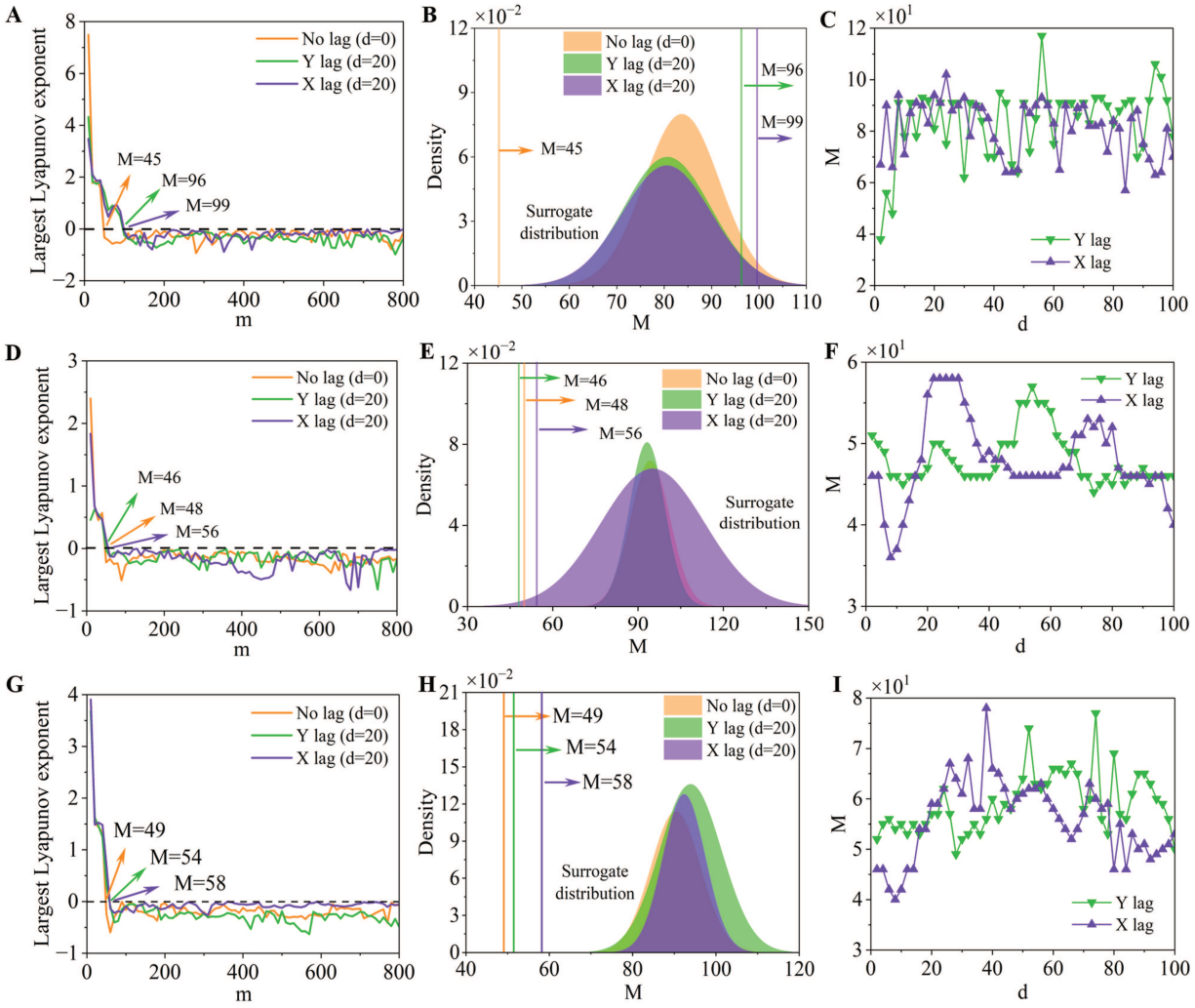

FIG. 2. Window size-dependent transitions of pairwise correlations from three model systems, namely, an ecological system (a)–(c), the ordinary Lorenz system (with a constant parameter) (d)–(f), and the time-varying Lorenz system (with a varying parameter) (g)–(i). (a) The estimated largest Lyapunov exponents under different window sizes (m) for nonlagged correlations ($d = 0$) and lagged correlations ($d = 20$), where $E = 2, \tau = 1$. (b) Comparisons of the transitions (M) from the original series (vertical lines) and surrogate series (Gaussian-fitted distributions), where the original transition points (M), as in (a), are all smaller than the fifth percentile of the surrogate distributions. (c) The transitions of lagged correlations ($d = 2, 4, \ldots, 100$). (d)–(f) The window size-dependent transitions of the correlations in the ordinary Lorenz system, where $E = 3, \tau = 1$, and the experiments are the same as those in (a)–(c). (g)–(i) The window size-dependent transitions of the correlations in the time-varying Lorenz system, where $E = 3, \tau = 1$, and the experiments are the same as those in (a)–(c). Note: We repeat the experiments 100 times in the surrogate tests, and the output 100 transitions (M) are approximately fitted by a Gaussian distribution. The transition scale (M) corresponds to the first negative largest Lyapunov exponent since our experiments show nonzero largest Lyapunov exponents. For the Lorenz system, variables X and Y are used as our sample.

complexity) in a system and a negative value indicates convergence (low complexity). If the measured largest Lyapunov exponent is equal to 0, it indicates a "state transition" between chaos and nonchaos. Thus, we estimate the largest Lyapunov exponents of correlation-based series using different window sizes.

C. Step 3: Significance tests on surrogate data

To guarantee the significance of our findings, we conduct tests on surrogate data, which are generated by randomly shuffling the sequences of the original time series.[18] By repeating Steps 1 and 2 for the surrogate data 100 times, a null-model distribution of the transition points is obtained. If we can obtain transition information from the surrogate data, our findings are independent of the order of time series, which in turn verifies the universality of the phenomenon. If the transition sizes from the original data are positioned at the tail part of the null-model distribution, we can determine how significant the model and real-world data are in generating this phenomenon. We present an example of the surrogate test in Subsection 1 of the Appendix (Fig. 6).

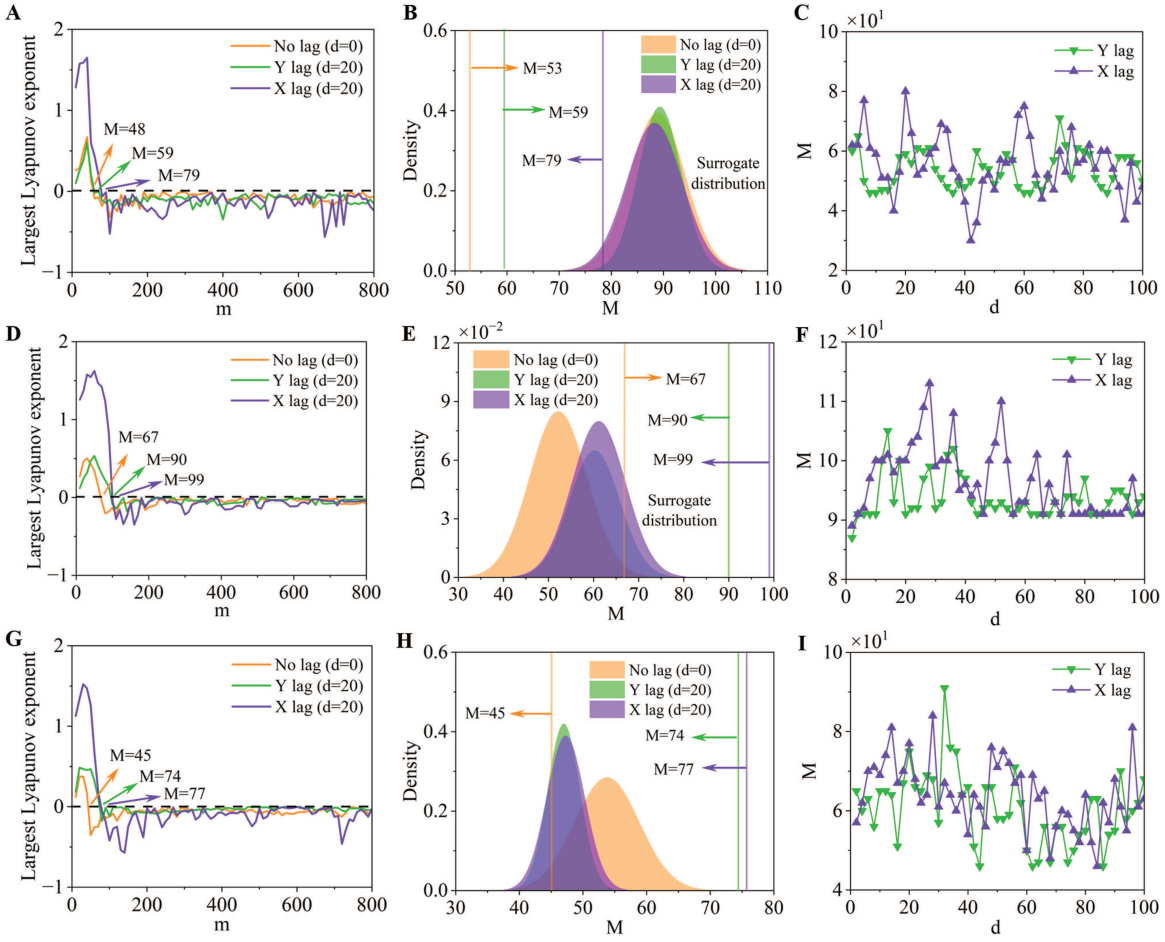

FIG. 3. Window size-dependent transitions of correlations in three real-world systems, namely, a climate system (X and Y represent the weekly temperatures in the Niño 1 + 2 and Niño 3.4 regions) (a)–(c), a stock market system (X and Y represent the daily prices of NASDAQ and US30) (d)–(f), and a neural system (X and Y represent the EEG signals from two regions in the brain) (g)–(i). (a) The estimated largest Lyapunov exponents under different window sizes (m) for nonlagged correlations ($d = 0$) and lagged correlations ($d = 20$). (b) Comparisons of the transitions (M) from the original series (vertical lines) and the surrogate series (Gaussian-fitted distributions), where the original transition points (M), as in (a), are all smaller than the fifth percentile of the surrogate distributions. (c) The transitions of lagged correlations ($d = 2, 4, \ldots, 100$). (d)–(f) The window size-dependent transitions of correlations in the climate system, where the experiments are the same as those in (a)–(c). (g)–(i) The window size-dependent transitions of correlations in the nervous system, where the experiments are the same as those in (a)–(c). Note: We repeat the experiments 100 times on the surrogate data, and the output 100 transitions (M) are approximately fitted by a Gaussian distribution. The embedding dimension and delay are $E = 3, \tau = 1$.

III. RESULTS

A. Model data

To study the dynamics of varying correlations in complex systems, we start with three benchmark model systems, i.e., a two-species ecological system [the details are provided in Eq. (1) and Fig. 1], an ordinary Lorenz system (chaotic system) [Eq. (5a)],[19] and a time-varying Lorenz system (the structure changes temporally).[20] For the ordinary Lorenz system, we consider the classic chaotic behaviors by setting the parameters as follows: $a_1 = 10$, $a_2 = 28$, and $a_3 = -8/3$. Then, we obtain 1081 discrete data points for each variable, where the initial state is $(X(0), Y(0), Z(0)) = (1, 2, 3)$ and the integration interval is $[10, 20]$. For the time-varying Lorenz system, the parameter a_1 changes from $a_1 = 10$ to $a_1 = 10.5$ over time [Eq. (5b)], and the other parameters are fixed to their traditional values,

$$\text{(a)} \begin{cases} \dot{X} = a_1(Y - X), \\ \dot{Y} = a_2 X - Y - XZ, \\ \dot{Z} = a_3 Z + XY, \end{cases} \qquad \text{(b)} \begin{cases} a_1 = 10, t \in [10, 15), \\ a_1 = 10.5, t \in [15, 20]. \end{cases} \tag{5}$$

The simulated ecological system is a typical nonlinear system. From its output time series, we study the dynamics of varying correlations under different window sizes. The results show that the estimated largest Lyapunov exponents from the correlation-based series experience a shift from positive values to negative values under

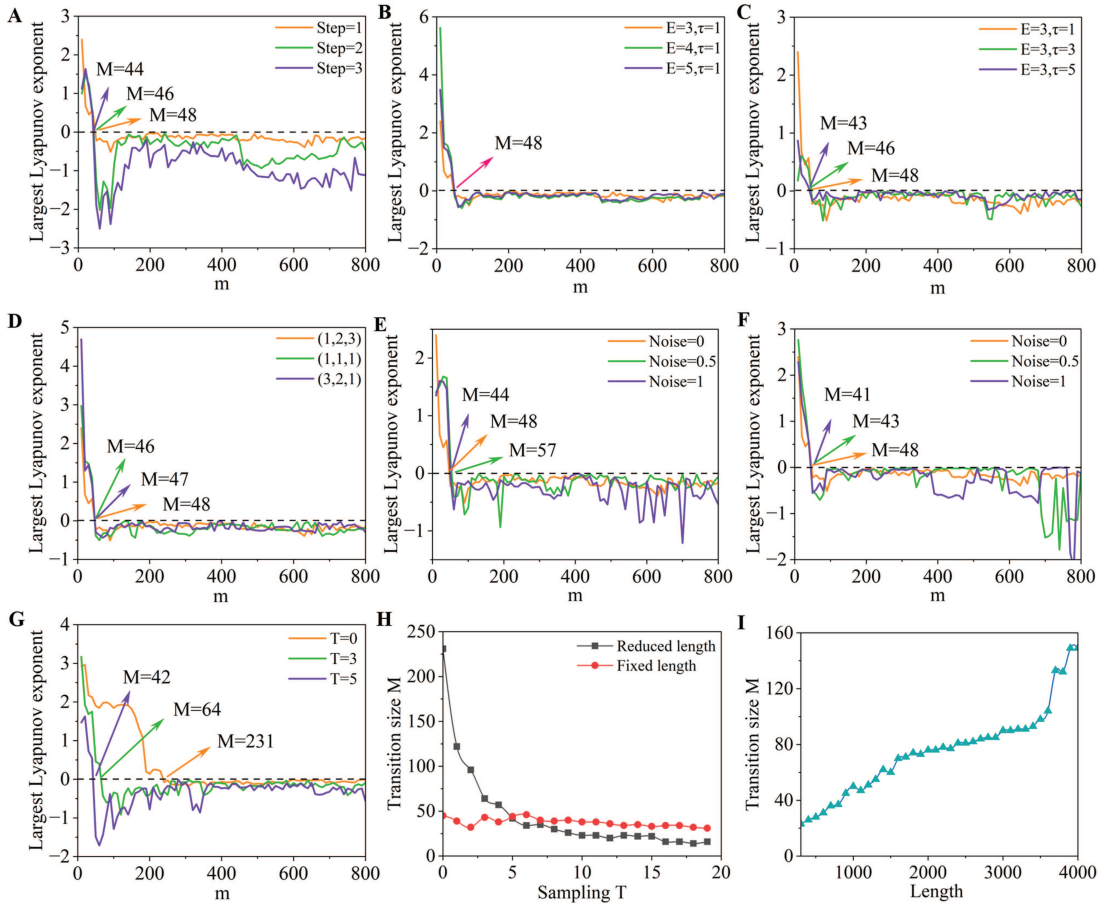

FIG. 4. Robustness tests of window size-dependent transitions with respect to the sliding step (a), embedding dimension (b), time delay (c), initial state (d), additive noise (e), multiplicative noise (f), sampling interval [(g) and (h)], and series length (i). Note: The variables X and Y from the ordinary Lorenz system are used here. T represents the sampling interval; thus, $1/T$ is the sampling frequency.

the window size $M = 60$ and that they remain negative as the window size is increased [see the yellow line in Fig. 2(a)]. This indicates a window size-dependent transition of correlations from a chaotic-like state to a nonchaotic state. Additionally, we also find window size-dependent transitions of correlation-based series from lagged correlations; i.e., the Y-lagged (e.g., 20 steps) correlation-based series has a transition at $M = 71$ [see the green line in Fig. 2(a)], and the X-lagged (e.g., 20 steps) correlation-based series has a transition at $M = 89$ [see the purple line in Fig. 2(a)]. Moreover, these transitions are also uncovered from various surrogate data (universality), and the transition sizes identified from the original data are significantly different from those identified from the surrogate data (significance) [Fig. 2(b)]. Even with different time lags (d), we also find window size-dependent transitions of correlation-based series [see Fig. 2(c)].

Interestingly, the window size-dependent transitions of correlation-based series are examined from both the ordinary and the time-varying Lorenz systems; see Figs. 2(d)–2(i) and Fig. 7 in Subsection 2 of the Appendix.

B. Real-world data

To examine the window size-dependent transitions of variable correlations in real-world systems, we select data from climate, financial, and neural systems (the time series data are depicted in Fig. 8 of Subsection 3 in the Appendix).

(a) For the climate system, we consider the El Niño–Southern Oscillation (ENSO) cycle,[21] which originates in the tropical Pacific through interactions between the ocean and the atmosphere. Many scholars have reported that ENSO-generated tropical Pacific sea surface temperature (SST) anomalies have significant impacts on climate conditions worldwide.[22] SST is widely used to monitor and forecast this abnormal phenomenon.[23] We select the weekly SST values from two regions, i.e., Niño 1 + 2 (90 °W–80 °W, 0–10 °S) and Niño 3.4 (170 °W–120 °W, 5 °N–5 °S). Each observation has a length of 1633, where the interval is from 04/01/1990 to 14/04/2021. These data are freely available from https://www.cpc.ncep.noaa.gov/data/indices/. To study the dynamics of varying correlations in relation to window size, we start

with nonlagged correlations. The estimated largest Lyapunov exponents from the correlation-based series experience a shift from positive values to negative values at the window size $M = 48$, and they remain negative with increasing window size [Fig. 3(a)]. This suggests a state transition of the correlation-based series from a chaotic-like state to a nonchaotic state. This window size-dependent transition is also found in their lagged (e.g., 20 steps) correlations [Fig. 3(a)]. Moreover, this phenomenon is determined to be universal and significant through surrogate tests [Fig. 3(b)]. Such window size-dependent transitions are also uncovered from the correlation-based series with various time lags [see Fig. 3(c)].

(b) For the financial system, we select the prices of NASDAQ and US30 (stock indices), which are representative indicators in the U.S. stock markets. The daily closing prices from 04/01/2016 to 17/09/2021 (with a length of 1439 for each observation) are selected as our sample. All the data are downloaded from the Choice database. We also find state transitions of correlation-based series under specific window sizes, i.e., the nonlagged, Y-lagged (20 steps) and X-lagged (20 steps) correlation-based series experience transitions when $M = 67$, 90, and 99, respectively; see Figs. 3(d)–3(f).

(c) In neuroscience, electroencephalographic (EEG) signals are widely used to uncover the complex mechanism between neurons and diagnosis related diseases.[24–26] We use two segments of electroencephalographic (EEG) signals (per second) collected from the brain (electrode 1 and electrode 2) of a healthy man (for more details, see Ref. 24), and each observation has a length of 1000. These data are freely available on http://archive.ics.uci.edu. From the observed EEG signal series, we uncover window size-dependent transitions of correlation-based series under specific window sizes [see Figs. 3(g)–3(i)].

Thus, it is a universal phenomenon that the varying correlations experience critical transitions from chaotic-like states to nonchaotic states under specific window sizes. Specifically, correlation-based series show chaotic-like dynamics under small window sizes and transition to nonchaotic states with increasing window size. Chaotic series are sensitive to initial perturbations, showing higher complexity and unpredictability. Thus, a chaotic correlation-based series has higher complexity, which makes it more difficult to uncover the system's mechanisms via the correlations among components. Therefore, a window size smaller than the transition size is not optimal. On the other hand, a correlation-based series with a long window shows little fluctuation, providing less useful information for studying the temporal patterns between different units. Consequently, the transition size can be practically selected as an appropriate window size for detecting the varying correlations between time series.

C. Robustness tests

In this section, we test the robustness of our findings concerning several issues: (a) the sliding step, (b) the embedding dimension E and delay τ, (c) the initial state, (d) additive and multiplicative noise, (e) the sampling rate, and (f) the length of the input series.

(a) The sliding step is associated with the number of separated windows and determines the length of the correlation-based series.[25] Naturally, we can capture the most information by setting the step size to 1 (with a fixed window size) since the output correlation-based series is the longest. However, the length of the correlation-based series will decrease under a longer sliding step (longer than 1), leading to some information being lost. Nevertheless, we also find transitions of correlation-based series under several sliding step sizes, e.g., at $M = 46$ with a sliding step size of 2 and at $M = 44$ with a sliding step size of 3 [see Fig. 4(a)].

(b) The embedding dimension E and delay τ are two critical parameters for the reconstruction of the dynamic attractor in the state space,[26] which will influence the estimation of the largest Lyapunov exponent. Although several criteria have been developed for identifying the embedding dimension E and delay τ,[28,29] they often yield different values for these two parameters. Generally, an optimal embedding dimension E is required to satisfy $E > 2D$ (D is the box dimension of the original attractor). Some scholars suggest that low dimensions are always preferable in practical applications.[30] We conduct tests with embedding dimensions $E = 4$ and $E = 5$, where equal transitions ($M = 48$) are identified [Fig. 4(b)]. However, the correlation-based series shows different transition sizes with different delays (e.g., $\tau = 3, 5$) [Fig. 4(c)].

(c) Chaotic systems are sensitive to initial perturbations, and they may output quite different time series with different initial states. We conduct tests with several initial states, e.g., $(X(0), Y(0), Z(0)) = (1, 1, 1)$ and $(X(0), Y(0), Z(0)) = (3, 2, 1)$. The results also show window size-dependent transitions of correlation-based series from different initial states [see Fig. 4(d)].

(d) Noise is inevitable in real-world systems, and we test the window size-dependent transitions of varying correlations from the series with additive and multiplicative white noise (e.g., with noise strengths of 0.5 and 1, respectively). The results also show state transitions of varying correlations from the noisy data [see Fig. 4(e)].

(e) The sampling frequency $1/T$ (T is the sampling interval) represents the scale of the observed time series, such as daily, weekly, or monthly. The lower the sampling frequency from a fixed period is, the shorter the observed series will be. Thus, we first study the series sampling from a fixed period (e.g., with an integration interval of $[10, 60]$ and an output series length of 5885). As shown in Fig. 4(g), we identify window size-dependent transitions with several sampling intervals. Moreover, the transition sizes decrease with increasing sampling interval [see the black line in Fig. 4(h)]. To reduce the influence of the sample length, we consider a series with a fixed length after sampling (e.g., 1081). The transition sizes are maintained at approximately 48 with various sampling frequencies [see the red line in Fig. 4(h)]. It seems that the length of the series, rather than the sampling rate, has a typical impact on the identified transition sizes.

(f) The length of the input series will influence the estimation of the largest Lyapunov exponent from the reconstructed attractor since an attractor reconstructed from a short series appears sparse.[30] As shown in Fig. 4(i), the transition size increases sharply with increasing length.

Consequently, our framework is reliable for finding the universal window size-dependent transitions of correlations in complex

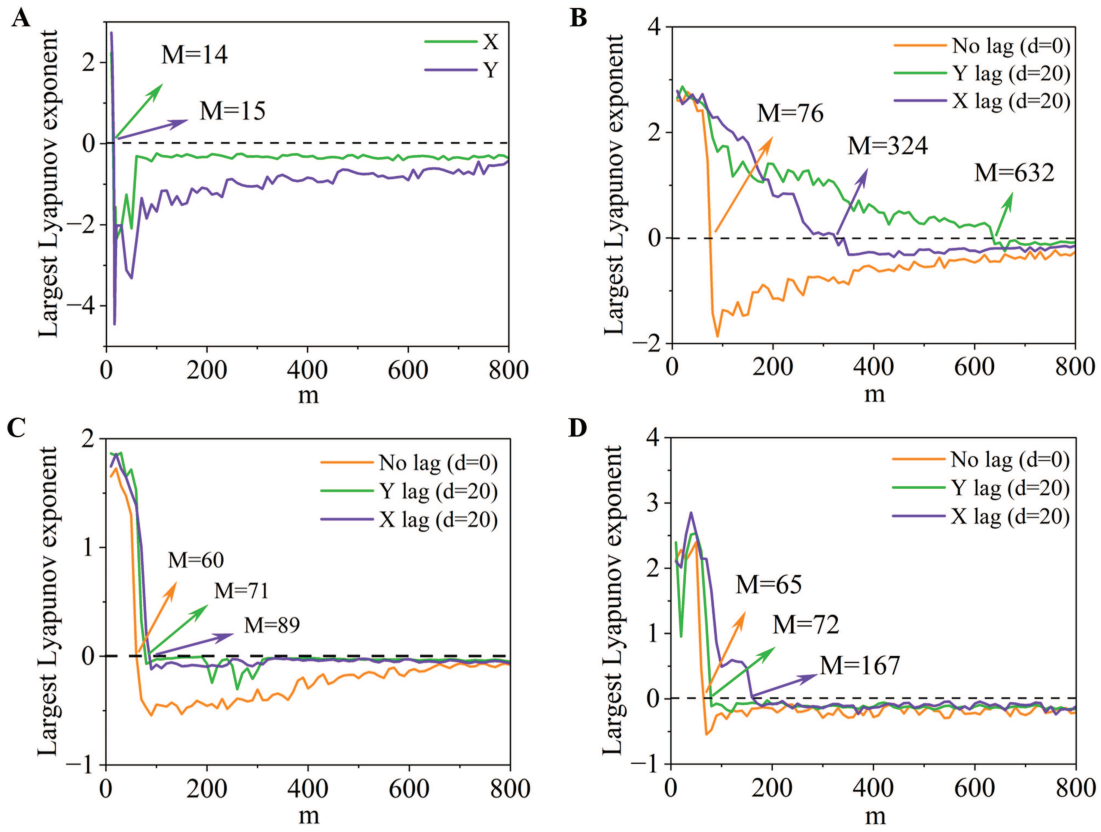

FIG. 5. Window size-dependent transitions of correlations estimated from different correlation measurements, including autoregression [AR(1): $X(t) = aX(t-1) + \varepsilon$] (a), bivariate linear regression (b), Pearson correlation (c), and Spearman correlation (d). Note: For linear regression, we test the X lag by establishing a model $Y(t) = aX(t-d) + b + \varepsilon$, and the correlation-based series is approximated by the fitted parameter a on consecutive windows. The ordinary Lorenz system is used as an example.

systems, which can be practically applied to select an appropriate window size for correlation analysis.

IV. DISCUSSION

To improve the universality of our findings, we study the state transitions of correlations through several classic correlation measurements, including autoregression [AR(1)], bivariate linear regression, Pearson correlation, and Spearman correlation.[31] We also find window size-dependent state transitions of correlations under specific window sizes. However, the transition sizes obtained through different correlation measurements are often different (Fig. 5). Thus, the window size-dependent transition is shown to be a universal phenomenon from both linear and nonlinear correlations. However, it is still unclear how the transition size changes with different measurements. In total, the transition sizes identified from mutual information are often less than 150 in various systems (Figs. 2 and 3), whereas the transition sizes identified with Pearson correlation may vary greatly, see Fig. 9 in Subsection 4 of the Appendix.

It is a universal phenomenon that the estimated correlations from nonlinear time series will shift from chaotic-like states to nonchaotic states under a specific window size. Generally, chaotic dynamics show higher uncertainty and complexity due to the sensitivity to the initial conditions. The window size-dependent transition of correlations from nonlinear time series can be interpreted based on changes in the uncertainty or complexity. This can be explained via information theory.[32] That is, short time series often contain less information and have higher uncertainty. Thus, the estimated correlations from short window segments may exhibit higher uncertainties, leading to higher uncertainty of the correlation-based series. With increasing window size, the uncertainty of a correlation-based series may decrease since more information is used from each window segment. However, this is not absolute since the length of a correlation-based series will decrease with increasing window size. This will influence the uncertainty of the correlation-based series, as demonstrated in both model and real-world systems, where numerous fluctuations of the estimated largest Lyapunov exponents are identified under larger window sizes (e.g., larger than the transition size).

The universal window size-dependent transitions of correlations may have several underlying implications: (a) Traditional regression-based prediction and classification are devoted to finding historical correlations among different components. Nonconstant

correlations between time series may drive the fitted parameters to be unstable. Moreover, the universal chaotic-like behaviors of variable correlations aggravate this issue. Rather than short input series, longer series may be preferable for training a regression model. Hopefully, it will be useful if we can achieve advance predictions of the varying correlations since correlation-based series can be studied as generic time series. (b) It is widely reported that long-term prediction is challenging for chaotic systems, and the chaotic behaviors of correlations between chaotic series provide additional evidence for this. (c) Numerous works have studied the complexity of time series based on the time series themselves (e.g., nonlinearity and chaos),[27] and our findings extend this issue by uncovering the complexity of the connectivity between time series.

V. CONCLUSIONS

This work proposed a quantitative framework for studying the window size-dependent transitions of correlations in various model and real-world systems. The estimated correlations show chaotic-like behaviors under small window sizes and transition to nonchaotic behaviors under large window sizes. This phenomenon is demonstrated to be universal and significant in various real-world systems. The identified transition size can be practically used as an appropriate window size to study the temporal correlation patterns behind the time series.

However, our work focuses on the correlation between time series, which fails to consider their causal interactions. Many works have reported time-varying causality in financial and economic systems, and it is unknown whether such time-varying causality also shows window size-dependent transitions.

ACKNOWLEDGMENTS

This research was supported by the National Natural Science Foundation of China (Grant Nos. 71991481, 71991485, 71991480, 42001242, and 41871202), the Beijing Natural Science Foundation (Grant No. 9202013), and the Fundamental Research Funds for the Central Universities (Grant No. 265208247).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Tao Wu: Conceptualization (equal); Methodology (equal); Software (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Feng An:** Conceptualization (equal); Investigation (equal); Supervision (equal). **Xiangyun Gao:** Conceptualization (equal); Formal analysis (equal); Funding acquisition (equal); Supervision (equal); Writing – review & editing (equal). **Siyao Liu:** Data curation (equal); Formal analysis (equal); Validation (equal). **Xiaotian Sun:** Data curation (equal); Formal analysis (equal); Investigation (equal); Visualization (equal). **Zhigang Wang:** Data curation (equal); Formal analysis (equal); Resources (equal). **Zhen Su:** Data curation (equal); Formal analysis (equal); Resources (equal); Writing – review & editing (equal). **Jürgen Kurths:** Conceptualization (equal); Methodology (equal); Supervision (equal); Validation (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

APPENDIX: SUPPLEMENTARY INFORMATION

1. Example of the surrogate test

To demonstrate the simulation process in detail on surrogate data, we consider a two-species ecological system as an example. The surrogate data are generated by randomly shuffling the entire

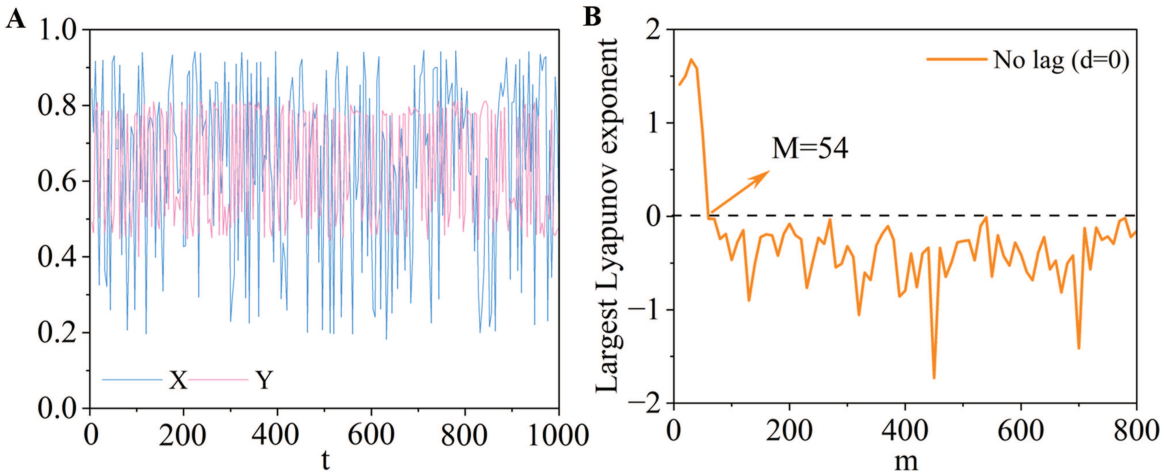

FIG. 6. Surrogate test. (a) The surrogate series from the two-species ecological model. (b) The measured largest Lyapunov exponents from the correlation-based series under different window sizes.

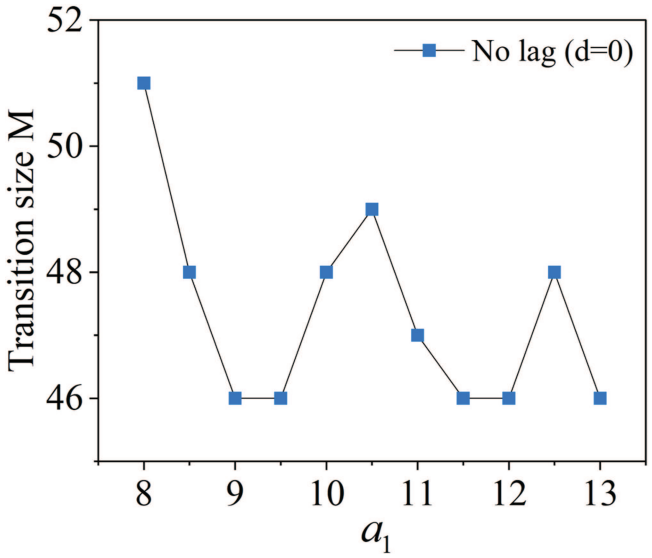

FIG. 7. Transition size for the time-varying Lorenz system, where $a_1 = 8, 8.5, \ldots, 13$ when $t \in [15, 20]$ and $a_1 = 10$ for $t \in [10, 15]$.

series of variables X and Y separately [Fig. 6(a)]. Then, we embed their window-wise correlations (e.g., mutual information) into a correlation-based series. Finally, we estimate the largest Lyapunov exponents of the correlation-based series under different window sizes. The results show a transition at the window size $M = 54$, where the estimated largest Lyapunov exponents shift from positive to negative values; see Fig. 6(b).

2. Window size-dependent transition for a time-varying Lorenz system

Transition size for the time-varying Lorenz system.

3. Real-world data

Real-world data.

4. Window size-dependent transition of Pearson correlations in complex systems

Window size-dependent transitions of pairwise Pearson correlations in the ordinary Lorenz system (a), the two-species ecological system (b), the climate system (c), the financial system (d), and the neural system (e).

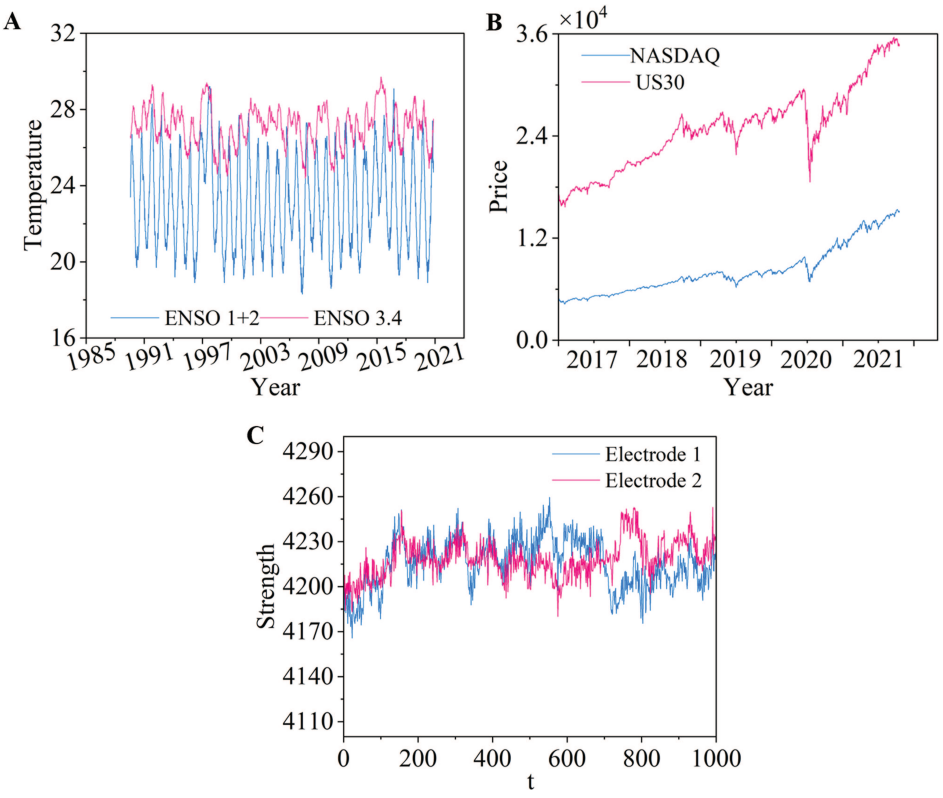

FIG. 8. Real-world data. (a) The weekly sea surface temperatures from regions ENSO 1 + 2 and ENSO 3.4. (b) The daily closing prices of NASDAQ and US30. (c) EEG signals from two regions of the brain.

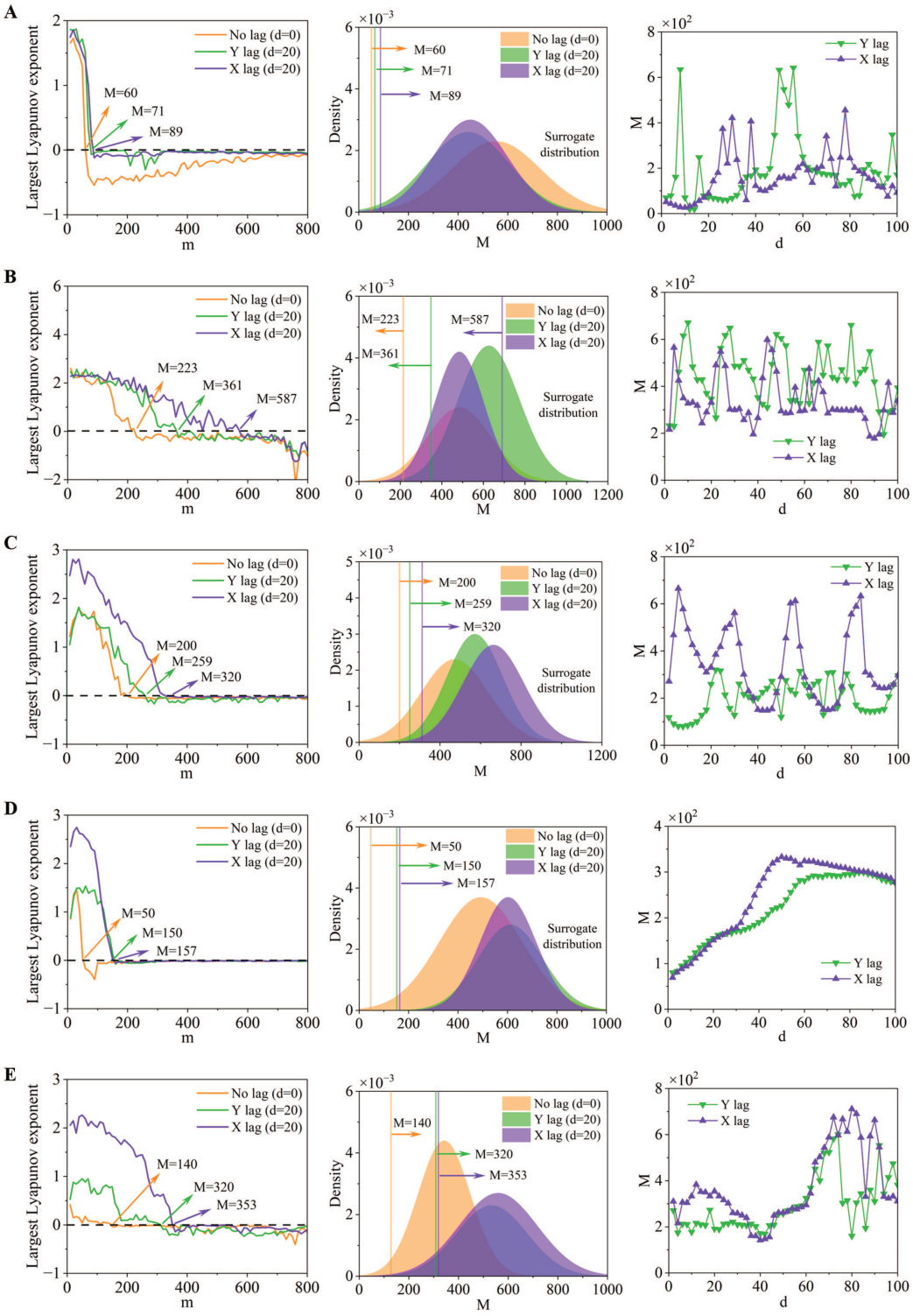

FIG. 9. Window size-dependent transitions of pairwise Pearson correlations in the ordinary Lorenz system (a), the two-species ecological system (b), the climate system (c), the financial system (d), and the neural system (e). All the experiments are analogous to those in Figs. 2 and 3.

REFERENCES

[1] J. Schulte, F. Policielli, and B. Zaitchik, "A skewed perspective of the Indian rainfall-El Nino-Southern Oscillation (ENSO) relationship," Hydrol. Syst. Sci. **24**, 5473–5489 (2020).
[2] J. F. Donges, Y. Zou, N. Marwan, and J. Kurths, "Complex networks in climate dynamics," Eur. Phys. J. **174**, 157–179 (2009).
[3] E. L. Sander, J. T. Wootton, and S. Allesina, "Ecological network inference from long-term presence-absence data," Sci. Rep. **7**, 7154 (2017).
[4] P. Schober, C. Boer, and L. A. Schwarte, "Correlation coefficients: Appropriate use and interpretation," Anesth. Analg. **126**, 1763–1768 (2018).
[5] A. Haluszczynski, I. Laut, H. Modest, and C. Rath, "Linear and nonlinear market correlations: Characterizing financial crises and portfolio optimization," Phys. Rev. E **96**, 062315 (2017).
[6] M. Bardoscia *et al.*, "The physics of financial networks," Nat. Rev. Phys. **3**, 490–507 (2021).
[7] J. Fan *et al.*, "Statistical physics approaches to the complex earth system," Phys. Rep. **896**, 1–84 (2021).
[8] X. Gao, H. An, and W. Fang, "Transmission of linear regression patterns between time series: From relationship in time series to complex networks," Phys. Rev. E **90**, 012818 (2014).
[9] M. Laib, L. Telesca, and M. Kanevski, "Long-range fluctuations and multifractality in connectivity density time series of a wind speed monitoring network," Chaos **28**, 033108 (2018).
[10] T. Wu, X. Gao, S. An, and S. Liu, "Time-varying pattern causality inference in global stock markets," Int. Rev. Financial Anal. **77**, 101806 (2021).
[11] R. Carmona-Cabezas *et al.*, "A sliding window-based algorithm for faster transformation of time series into complex networks," Chaos **29**, 103121 (2019).
[12] F. Kwasniok, "Detecting, anticipating, and predicting critical transitions in spatially extended systems," Chaos **28**, 033614 (2018).
[13] G. Sugihara *et al.*, "Detecting causality in complex ecosystems," Science **338**, 496–500 (2012).
[14] K. Hlavackova-Schindler, M. Palus, M. Vejmelka, and J. Bhattacharya, "Causality detection based on information-theoretic approaches in time series analysis," Phys. Rep. **441**, 1–46 (2007).
[15] M. T. Rosenstein, J. J. Collins, and C. J. Deluca, "A practical method for calculating largest Lyapunov exponents from small data sets," Physica D **65**, 117–134 (1993).
[16] A. Krakovska, K. Mezeiova, and H. Budacova, "Use of false nearest neighbours for selecting variables and embedding parameters for state space reconstruction," J. Complex Syst. **2015**, 932750.
[17] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractor from mutual information," Phys. Rev. A **33**, 1134–1140 (1986).
[18] T. Nakamura and M. Small, "Small-shuffle surrogate data: Testing for dynamics in fluctuating data with trends," Phys. Rev. E **72**, 056216 (2005).
[19] E. N. Lorenz, "Deterministic nonperiodic flow," J. Atmos. Sci. **20**, 130–141 (1963).
[20] J.-W. Hou *et al.*, "Harvesting random embedding for high-frequency change-point detection in temporal complex systems," Natl. Sci. Rev. **9**, nwab228 (2022).
[21] M. J. McPhaden *et al.*, "ENSO as an integrating concept in earth science," Science **314**, 1740–1745 (2006).
[22] D. Chen, N. Smith, and W. Kessler, "The evolving ENSO observing system," Natl. Sci. Rev. **5**, 805–807 (2018).
[23] Y. Tang *et al.*, "Progress in ENSO prediction and predictability study," Natl. Sci. Rev. **5**, 826–839 (2018).
[24] S. M. Fernandez-Fraga *et al.*, "Feature extraction of EEG signal upon BCI systems based on steady-state visual evoked potentials using the ant colony optimization algorithm," Discrete Dyn. Nat. Soc. **2018**, 2143873.
[25] J. D. Hamilton, *Time Series Analysis* (Princeton University Press, 1994), Vol. 2.
[26] F. Takens, "Detecting strange attractors in turbulence," Mathematics **898**, 366–381 (1981).
[27] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, 2004), Vol. 7.
[28] K. H. Kraemer *et al.*, "A unified and automated approach to attractor reconstruction," New J. Phys. **23**, 033017 (2021).
[29] J. Stark *et al.*, "Delay embeddings for forced systems: II. Stochastic forcing," J. Nonlinear Sci. **13**, 519–577 (2003).
[30] H. F. Ma *et al.*, "Randomly distributed embedding making short-term high-dimensional data predictable," Proc. Natl. Acad. Sci. U.S.A. **43**, E9994–E10002 (2018).
[31] T. C. Mills, *Time Series Techniques for Economists* (Cambridge University Press, 1990).
[32] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. (Wiley, 2005).