Constructing networks from a dynamical system perspective for multivariate nonlinear time series

Tomomichi Nakamura,[1,*] Toshihiro Tanizawa,[2,†] and Michael Small[3,4,‡]

[1]*Graduate School of Simulation Studies, University of Hyogo, 7-1-28 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan*
[2]*Kochi National College of Technology, Monobe-Otsu 200-1, Nankoku, Kochi 783-8508, Japan*
[3]*School of Mathematics and Statistics, The University of Western Australia, 35 Stirling Hwy., Crawley, WA 6009, Australia*
[4]*Mineral Resources, CSIRO, Kensington, WA 6151, Australia*

We describe a method for constructing networks for multivariate nonlinear time series. We approach the interaction between the various scalar time series from a deterministic dynamical system perspective and provide a generic and algorithmic test for whether the interaction between two measured time series is statistically significant. The method can be applied even when the data exhibit no obvious qualitative similarity: a situation in which the naive method utilizing the cross correlation function directly cannot correctly identify connectivity. To establish the connectivity between nodes we apply the previously proposed small-shuffle surrogate (SSS) method, which can investigate whether there are correlation structures in short-term variabilities (irregular fluctuations) between two data sets from the viewpoint of deterministic dynamical systems. The procedure to construct networks based on this idea is composed of three steps: (i) each time series is considered as a basic node of a network, (ii) the SSS method is applied to verify the connectivity between each pair of time series taken from the whole multivariate time series, and (iii) the pair of nodes is connected with an undirected edge when the null hypothesis cannot be rejected. The network constructed by the proposed method indicates the intrinsic (essential) connectivity of the elements included in the system or the underlying (assumed) system. The method is demonstrated for numerical data sets generated by known systems and applied to several experimental time series.

I. INTRODUCTION

To understand the nature of ongoing interaction in real-world complex systems it is first necessary to deduce the interconnection between the components of the system under study [1]. Hence, the relationship between the elements (the intrinsic connectivity) needs to be determined before we can understand the system—usually this is done by either mapping physical (for example, anatomical) connections onto a network or, as we will consider here, by inferring functional connectivity from multivariate time series data sets. Once the connectivity has been determined the effect of that connectivity is frequently studied using the concept of complex networks [2–7]. There have been several recent works for constructing networks from scalar time series such as recurrence networks [8], cycle networks for pseudoperiodic time series [9,10], and the horizontal visibility algorithm [11]. A distinctive local feature in the scalar time series is treated as a node of a network in these approaches. There are also approaches for constructing networks for multivariate time series [12–18]. Especially the work by Walker *et al.* is noteworthy because they applied a surrogate method to construct refined networks from multivariate time series [18]. We focus our attention on constructing networks from multivariate time series in this paper. In these approaches each time series is considered as a basic node of a network. Nodes are connected if the dynamics of the corresponding scalar time series are

sufficiently *similar*. The naive (and usual) way to measure "similarity" is with a thresholded linear correlation. While this naive approach is expeditious, it is also flawed when one is looking at nonlinear (possibly temporally delayed) interaction in complex systems. We describe the naive approach in detail in Sec. II.

In particular, as experimental time series will typically show some irregular fluctuations, it is difficult to determine the precise relationship among the various variables and simple linear correlation (with a constant threshold) will often be insufficient. Nonetheless, for constructing a network from multivariate time series the cross correlation function remains the most commonly used approach [12–17]. When the cross correlation exhibits strong peaks or has large absolute values between −1 and +1 at some fixed time lags the result is argued to be a good indication that the data have "similarities"—and hence, that the corresponding nodes should be connected. Hence, we suppose that there are correlation structures between the two signals (or that similar factors may influence both systems). On the other hand, when the cross correlation does not have strong peaks one usually concludes that there is no similarity. In this case, we conclude that there is no correlation structure and that the dynamics of the measured components are unconnected [19].

However, the cross correlation function is only a useful measure of *linear* similarity [19]. That is to say, even when two signals are not similar, there are still possibilities that these two data sets have some kind of correlation structures (that is, the two data sets are interconnected or interrelated). Here, when we use the term "correlation structures" we mean any functional connectivity, irrespective of whether the structures are linear or nonlinear. This means that it is not enough to use

*tomo@sim.u-hyogo.ac.jp
†tanizawa@ee.kochi-ct.ac.jp
‡michael.small@uwa.edu.au

only linear statistics such as the cross correlation function. To investigate correlation structures more completely, we adopt the viewpoint of deterministic dynamical systems and nonlinear time series analysis [20]. The purpose of our method is to construct networks for multivariate time series, even if there are nonlinearities in these time series.

We now consider the meaning of correlation structures and the meaning of the network properties. We consider that if an element y is included in a function of x (in other words, y is one of the variables in a function of x) and the influence of y on x cannot be ignored, there is some kind of correlation structures between x and y, irrespective of whether the correlation is linear or nonlinear. Hence, we infer that connecting elements with correlation structures indicates the intrinsic connectivity of the elements included in the system or the underlying (assumed) system.

In this paper, we introduce a method to construct such a network for multivariate nonlinear time series. To verify the intrinsic (essential) connection between two data sets from the viewpoint of a deterministic dynamical system, we apply the previously proposed small-shuffle surrogate method [19]. The method we propose does not estimate the system itself (equations or models). The method estimates the intrinsic connectivity (interconnection or interrelation) of the elements among which the influences are considerable.

We first describe the current technique briefly and then we describe our approach. After describing these techniques, we will apply this algorithm to several cases using simulated multivariate time series of three systems, linear and nonlinear systems, and the coupled Lorenz systems [21]. Based on the numerical experiments, we apply our method to real-world data. We select two specific systems of particular interest to us: hourly meteorological time series and daily exchange rate data. Although meteorological physical quantities under controlled situations in laboratories are well described by the equation of state of gases in thermodynamics, actual situations in weather are fairly complicated and far from controlled. Similarly, although exchange rate markets are built by humans, there are many market participants and the deals are intricate. Hence, we consider that the relationship among meteorological data and that among exchange rate data are not trivial.

Alternative general nonlinear similarity measures have been proposed in the literature—the most common approach is cross mutual information [22]. However, unlike mutual information based methods (which compute probabilistic independence from statistical estimates of histograms), the approach we propose is grounded in unraveling the deterministic interaction among the components—rather than looking for statistical or probabilistic dependence, we are seeking to describe nonlinear determinism between the components. We also note that there are many alternative methods to which we could compare our approach. However, such an exhaustive comparison would be beyond this work and the purpose of this work. We have chosen to focus on mutual information as it is a particularly well understood, and technically well motivated approach. Moreover, mutual information is widely used and has been seen to perform extremely well in many situations. We use mutual information as a widely accepted benchmark, and we anticipate that the reader can then infer the relative performance of their own favorite measure.

II. THE NAIVE APPROACH TO NETWORK CONSTRUCTION

The most extensively used method to construct networks from multivariate time series using the cross correlation can be reduced to the following three steps [12–17].

(1) Each time series is considered as a basic node of a network.

(2) To investigate the relationship among multivariate time series, the cross correlation between each pair of time series (i.e., two time series) taken from the whole multivariate time series is estimated.

(3) The pair of nodes corresponding to the chosen two time series is connected with an undirected edge when the value of the cross correlation is larger than an appropriately chosen threshold.

We refer to this method as "the naive method." The basic idea behind the naive method is as follows. When signals are similar, we expect that there may be some sort of relationship between the corresponding nodes, and hence the pair is considered to be connected with an undirected link. On the other hand, there are cases where time series are not similar enough. In this case, as we may have the impression that these are independent or have no relationship, we do not connect them. This approach relies on one selecting an appropriate threshold, and this is usually done in an *ad hoc* manner (often arguments are made that a suitable value of threshold can be selected such that one ensures graph connectivity, or some minimum number of edges).

Although the naive method has been proved to be effective in various cases [12–17], the range of applicability might be restrictive because "no similarity" is not equivalent to "no correlation" [19]. In the next section, we describe an approach to reduce this problem.

III. A DIFFERENT APPROACH TO CONSTRUCT NETWORKS

The approach of the proposed method is basically the same as the naive method described in Sec. II. The difference is the way of verifying the connection between two data sets. As mentioned above, only the cross correlation function is used in the naive method. To determine whether two nodes should be connected statistically and to make the result rigorous, we apply the concept of the surrogate data method [23], and in particular, we apply the recently proposed small-shuffle surrogate (SSS) method [19]. The approach employed to construct a network for multivariate time series and the method to generate surrogate data are not novel, but the combination of network construction with the small-shuffle surrogate method improves the accuracy of the network connectivity in the system or the underlying (assumed) system.

To investigate correlation structures, the random-shuffle surrogate (RSS) method has been proposed in [23]. The RSS method can test for whether data can be fully described by independent and identically distributed random variables.

When a system is composed of many variables we anticipate that each variable is influenced by some of these variables. Hence, we consider that the relationship (correlation structures) among the variables might be very complicated.

The cross correlation function is intrinsically a statistic to investigate linear similarity between two data channels. To investigate whether two data channels selected from many variables are somehow connected, we construct artificial time series, destroying the correlation structure between the two data while preserving all independent features of each independent data set. That is, we generate an ensemble of artificial realizations that explicitly *do not* have a correlation between them (either linear or nonlinear), but otherwise appear like the original data.

The RSS method destroys any correlation structure in data. That is, not only the relationship between two data sets but also all relationships among any other variables that may be present is also destroyed. Hence, by examining the cross correlation function of the (decorrelated) RSS data, and comparing this to the original, we can test whether the original variables exhibit significant dependencies.

However, the RSS method assumes global stationarity and performs a pairwise linear decoupling between channels. In many typical examples the individual channels are also influenced by local drift and other nonstationary variation. To account for these effects, we propose that the small-shuffled method is preferable to the RSS method.

A. The small-shuffle surrogate method

To investigate whether temporal correlations in time series data are absent or if the data are independently distributed random variables, the SSS method is often used [19]. The SSS method destroys local structures or correlations in irregular fluctuations (short-term variabilities) and preserves the global behaviors by shuffling the data index on a small (local) scale.

SSS data are generated as follows. Let the original data be $x(t)$, let $i(t)$ be the index of $x(t)$ [that is, $i(t) = t$, and so $x[i(t)] = x(t)$], let $g(t)$ be Gaussian random numbers, and $s(t)$ will be the surrogate data.

(i) Obtain $i'(t) = i(t) + Ag(t)$, where A is an amplitude.

(ii) Sort $i'(t)$ by the rank order and let the index of $i'(t)$ be $\hat{i}(t)$.

(iii) Obtain the surrogate data $s(t) = x[\hat{i}(t)]$.

It has been found that choosing $A = 1.0$ is adequate for nearly all purposes [19]—although this parameter choice remains heuristic. Further details of the method and the mechanism are provided in [19]. When we apply the SSS method to multivariate data, the null hypothesis (NH) is that there is no short-term correlation structure between the data or that the irregular fluctuations are independent [19].

When we use multivariate data exhibiting irregular fluctuations (short-term variabilities) the behaviors are varied. This means that we need to treat data with different trends in some cases. When we want to know some kind of relationship between two data sets, the correlation coefficient is a simple approach because this statistic is effective to investigate similarities of long-term trends between two data sets. However, to know the detailed relationship, we need to investigate the local relationship (short-term dynamics) among data. A requisite null hypothesis to accomplish this investigation in these situations is the null hypothesis described above.

B. When to reject a null hypothesis

Discriminating statistics are necessary for surrogate data hypothesis testing. The SSS method changes the flow of information in the data. It is preferable to use discriminating statistics which can accurately reflect features of the surrogate method. Hence, we choose to use the cross correlation (CC) function and the average mutual information (AMI) as discriminating statistics. These statistics can determine, on average, how much one learns about one signal by observing the other [24].

After the calculation of these statistics, we need to determine whether a NH should be rejected. We employ Monte Carlo hypothesis testing and determine whether the estimated statistics of the original data fall within or outside the statistical distribution of the surrogate data [25]. When the statistics fall within the distributions of the surrogate data, we conclude that the hypothesis may not be rejected. In this paper, we generate 99 SSS data and hence the nonparametric significance level is between 0.01 and 0.02 for a one-sided test with two nonindependent statistics [26].

It should be noted that although the multiple comparison problem is common in surrogate data applications—it is exacerbated in the multivariate network setting. We use the CC and the AMI as complementary statistics. This is because some of the test systems are robust to one or the other of our two primary statistical tests (the CC and the AMI) [19]. We show plots of both the CC and the AMI as a function of time lag. In fact, we expect that it is usually a meaningful test statistic for relatively small lags (for example, between −10 and 10), because the CC and the AMI of the original and surrogate data will coincide for large lag. For more details see [19].

We also note that we apply a surrogate data based hypothesis test as a part of our program. This is a Monte Carlo resampling scheme to test against a specific NH (in this case encapsulated in the small-shuffled surrogate scheme). A necessary component of this hypothesis testing is a test statistic with which we compare the observed data and the surrogate generated null distribution. We choose the CC and the AMI as they are robust, independent of the underlying surrogate generation scheme and also pivotal (in the sense of yielding the same distribution of values for a wide range of seed realization consistent with the null). Nonetheless, this choice is not critical and one is free to utilize other suitable statistics at this point in our procedure.

C. Statistical rigor of the proposed method

In the proposed method the SSS method is used and the SSS method is a statistical hypothesis test. Applying a statistical hypothesis test can result in two outcomes: either the NH is rejected, or it is not. In the former case there is a possibility that the NH is rejected even though it is true, which is called type I error; in the latter case there is a possibility that we will fail to reject the NH when it is in fact false, which is called type II error [27]. As the proposed method relies on a statistical hypothesis test, we cannot avoid this problem.

Also, when the influence of the observational noise level is more significant than that of data from the systems, the SSS method cannot properly investigate the features of the data [28]. Hence, when the observational noise is large
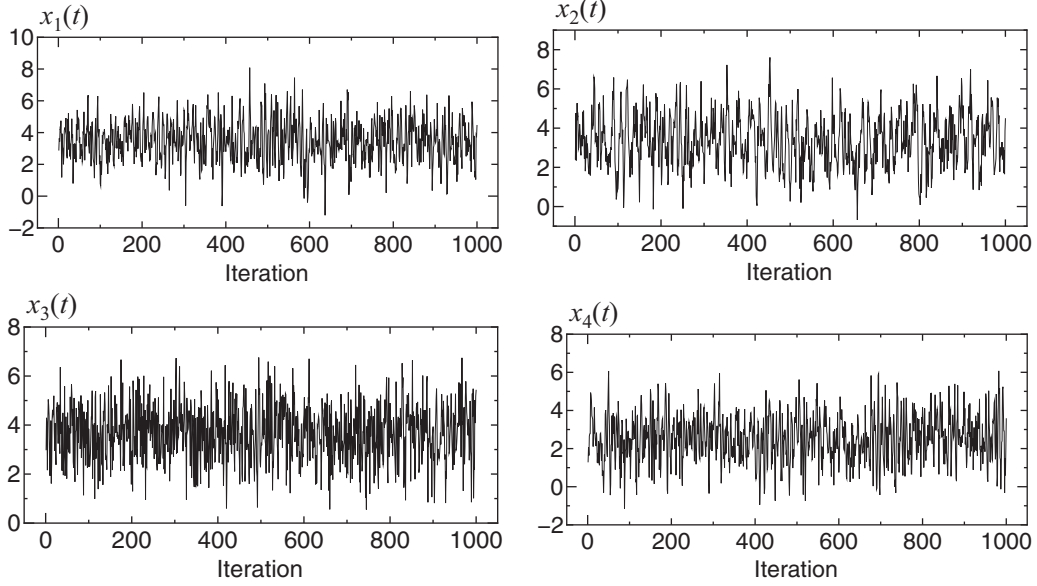
FIG. 1. Time series data generated by the linear system, Eqs. (1)–(4). We use the data to construct the network.

the network constructed by the proposed method might not indicate the connectivity in the system or the underlying (assumed) system.

IV. NUMERICAL EXAMPLES

We now demonstrate the application of our algorithm to various simulated multivariate time series data sets, and confirm our theoretical arguments with several examples. For comparison we also apply the naive method to the data sets. Broadly speaking, we consider two situations, in which the precise relationship among the variables is either explicit or not. For the former case, we use two systems: a linear system and a nonlinear system [29]. As the systems are in actual time, the relationship among the variables is clear. For the latter case, we use the coupled Lorenz systems presented in the form of a differential equation [21].

In all cases, we use $A = 1.0$ for generating SSS data, generate 99 SSS data, and the data is 1000 points with Gaussian
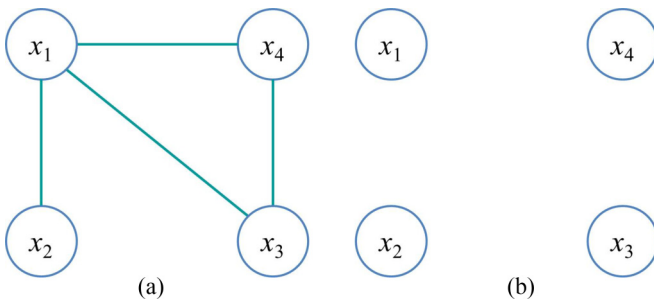

FIG. 2. The linkage of network: (a) the connectivity of Eqs. (1)–(4) and Eqs. (9)–(12). The same network is obtained when the proposed method is applied to the data shown in Figs. 1 and 4. We show the results later in Secs. IV A and IV B. (b) The network when we apply the naive method to the data. As shown in this figure, there is no link among the nodes.

observational noise with the mean zero and the standard deviation 0.01.

A. Data from a linear system

The first application is to time series from a linear system. To verify the correlation structures between the elements, we use the system which consists of four dynamical variables, $x_1(t)$, $x_2(t)$, $x_3(t)$, and $x_4(t)$, and the models are described by the following expressions:

$$x_1(t) = 1.3 + 0.4\, x_1(t-1) - 0.2\, x_1(t-3)$$
$$+ 0.3\, x_2(t-4) + 0.2\, x_4(t-7) + \varepsilon_1(t), \quad (1)$$
$$x_2(t) = 2.0 + 0.6\, x_2(t-1) - 0.2\, x_2(t-6) + \varepsilon_2(t), \quad (2)$$
$$x_3(t) = 2.2 + 0.2\, x_1(t-2) + 0.3\, x_4(t-9) + \varepsilon_3(t), \quad (3)$$
$$x_4(t) = 1.3 + 0.2\, x_1(t-2) + 0.5\, x_4(t-1)$$
$$- 0.3\, x_4(t-3) + \varepsilon_4(t), \quad (4)$$

where $\varepsilon_i(t)$ ($i = 1,2,3,4$) are dynamic noise, independent and identically distributed Gaussian random variables with mean zero and standard deviation 1.0. The behaviors of the four time series generated by these models are shown in Fig. 1. The

TABLE I. The largest absolute values of the cross correlation function of all possible pairs between the time lag -10 and 10, where the number in parentheses is the time lag when the cross correlation function has the largest absolute value. The data are generated by the linear system, Eqs. (1)–(4), and the values are estimated using 1000 data points.

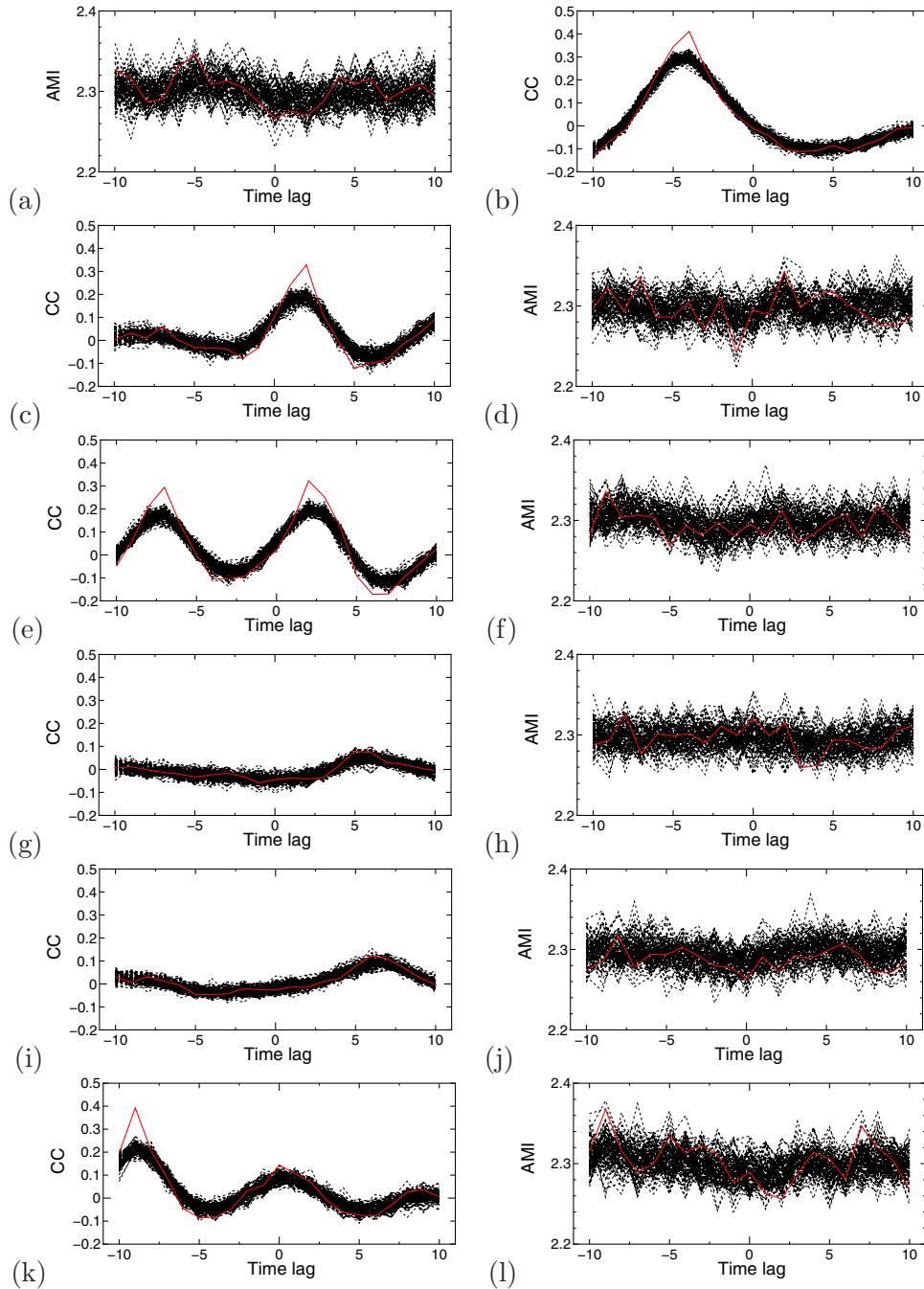|	x_1	x_2	x_3	x_4
x_1	1.0000	–	–	–
x_2	0.4111 (−4)	1.0000	–	–
x_3	0.3285 (2)	0.0788 (5)	1.0000	–
x_4	0.3225 (2)	0.1230 (6)	0.3942 (−9)	1.0000

FIG. 3. The result of the linear system, Eqs. (1)–(4). A plot of (a), (c), (e), (g), (i) and (k) the cross correlation function (CC), and (b), (d), (f), (h), (j) and (l) the average mutual information (AMI), (a) and (b) are result of x_1 and x_2, (c) and (d) are result of x_1 and x_3, (e) and (f) are result of x_1 and x_4, (g) and (h) are result of x_2 and x_3, (i) and (j) are result of x_2 and x_4, and (k) and (l) are result of x_3 and x_4. The solid line is the original data and the dotted lines are the SSS data.

behaviors show irregular fluctuations and it is difficult to know the relationship among the data by visual inspection.

In this paper, we distinguish between "component" and "variable" as different technical terms. We use the term "component" to represent x_i, and the term "variable" when it takes a particular value $x_i(t - l)$. We treat the components as the nodes of the network. That is, Eq. (1) has three components (x_1, x_2, and x_4) and four variables, $x_1(t - 1)$, $x_1(t - 3)$, $x_2(t - 4)$, and $x_4(t - 7)$. As shown in Eqs. (1)–(4),

each dynamical variable at time t is determined by various other dynamical variables. We consider the connectivity of the linear system, Eqs. (1)–(4). Equation (1) shows that the component x_1 is influenced by three components, x_1, x_2, and x_4. That is, other components which connect the component x_1 are x_2 and x_4. Similarly, as Eq. (2) shows that x_2 is driven by only x_2, there is no connection with x_2. As Eq. (3) shows that x_3 is driven by x_1 and x_4, x_1 and x_4 connect x_3. As Eq. (4) shows that x_4 is driven by x_1 and x_4, x_1 connects x_4. Based on
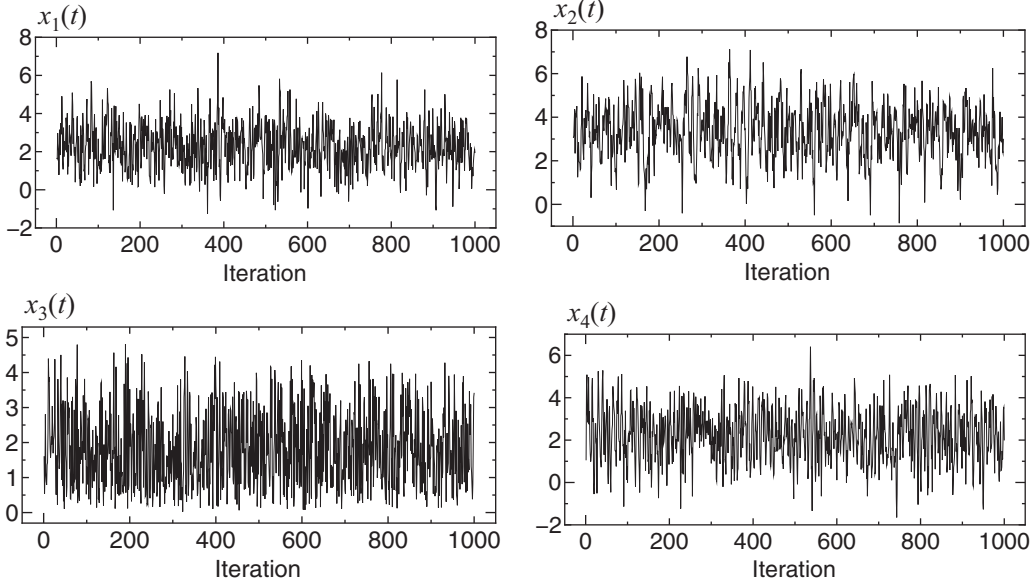
FIG. 4. Time series data generated by the nonlinear system, Eqs. (9)–(12). We use the data to construct the network.

this, the connectivity expressions of the linear system become

$$x_1 = C_1(x_2, x_4), \quad (5)$$

$$x_2 = 0, \quad (6)$$

$$x_3 = C_3(x_1, x_4), \quad (7)$$

$$x_4 = C_4(x_1), \quad (8)$$

where C_i stands for the function representing connectivity of the ith component, x_i, and zero means that there is no connection. The network structure constructed based on this idea is shown in Fig. 2(a).

We first estimate the cross correlation function to apply to the naive method. All the values are shown in Table I. We need to determine the threshold value to decide whether a link is present between two components. If we set the value 0.5, as shown in Table I, we cannot connect any link between nodes [30]. The network structure constructed by the naive method is shown in Fig. 2(b) and Fig. 2(b) shows that there is no link among any node on this network. However, we note that as Eqs. (1)–(4) show, there are correlation structures among the components. This result clearly indicates that only the application of the cross correlation function is not effective. We consider that the network we really want and need is that shown in Fig. 2(a), because the network reflects the connectivity of components in the system of Eqs. (1)–(4).

We apply the SSS method to the data of all possible pairs to verify the connection between two data sets. Figure 3 shows the result. Figures 3(a) and 3(b) show that the CC of the original data falls outside the distribution of the SSS data, although the AMI of the original data falls inside the distribution of the SSS data. As a result, we can consider that there is some sort of correlation structure between x_1 and x_2, and this diagnosis is correct. On the other hand, Figs. 3(g)–3(j) show that both of the CC and AMI of the original data falls inside the distribution of the SSS data. Hence, we do not connect the link between x_2 and x_3 and between x_2 and x_4. In this way when we examine the connectivity between the components, we can obtain the

same result as Eqs. (5)–(8). This result indicates that we can discriminate correctly whether there are correlation structures between two signals. Based on this we can construct the same network as shown in Fig. 2(a) [31].

B. Data from a nonlinear system

To investigate whether the proposed method works even if there is nonlinearity, we use the following models instead of Eqs. (1) and (3) where Eqs. (10) and (12) are the same as Eqs. (2) and (4).

$$x_1(t) = 1.3 + 0.2\, x_1(t-1) - 0.1\, x_1(t-3)$$
$$+ 0.1\, x_2(t-4)x_4(t-7) + \varepsilon_1(t), \quad (9)$$

$$x_2(t) = 2.0 + 0.6\, x_2(t-1) - 0.2\, x_2(t-6) + \varepsilon_2(t), \quad (10)$$

$$x_3(t) = h[2.2 + 0.2\, x_1(t-2)$$
$$+ 0.3\, x_4(t-9) + \varepsilon_3(t)], \quad (11)$$

$$x_4(t) = 1.3 + 0.2\, x_1(t-2) + 0.5\, x_4(t-1)$$
$$- 0.3\, x_4(t-3) + \varepsilon_4(t). \quad (12)$$

TABLE II. The largest absolute values of the cross correlation function of all possible pairs between the time lag -10 and 10, where the number in parentheses is the time lag when the cross correlation function has the largest absolute value. The data are generated by the nonlinear system, Eqs. (9)–(12), and the values are estimated using 1000 data points.

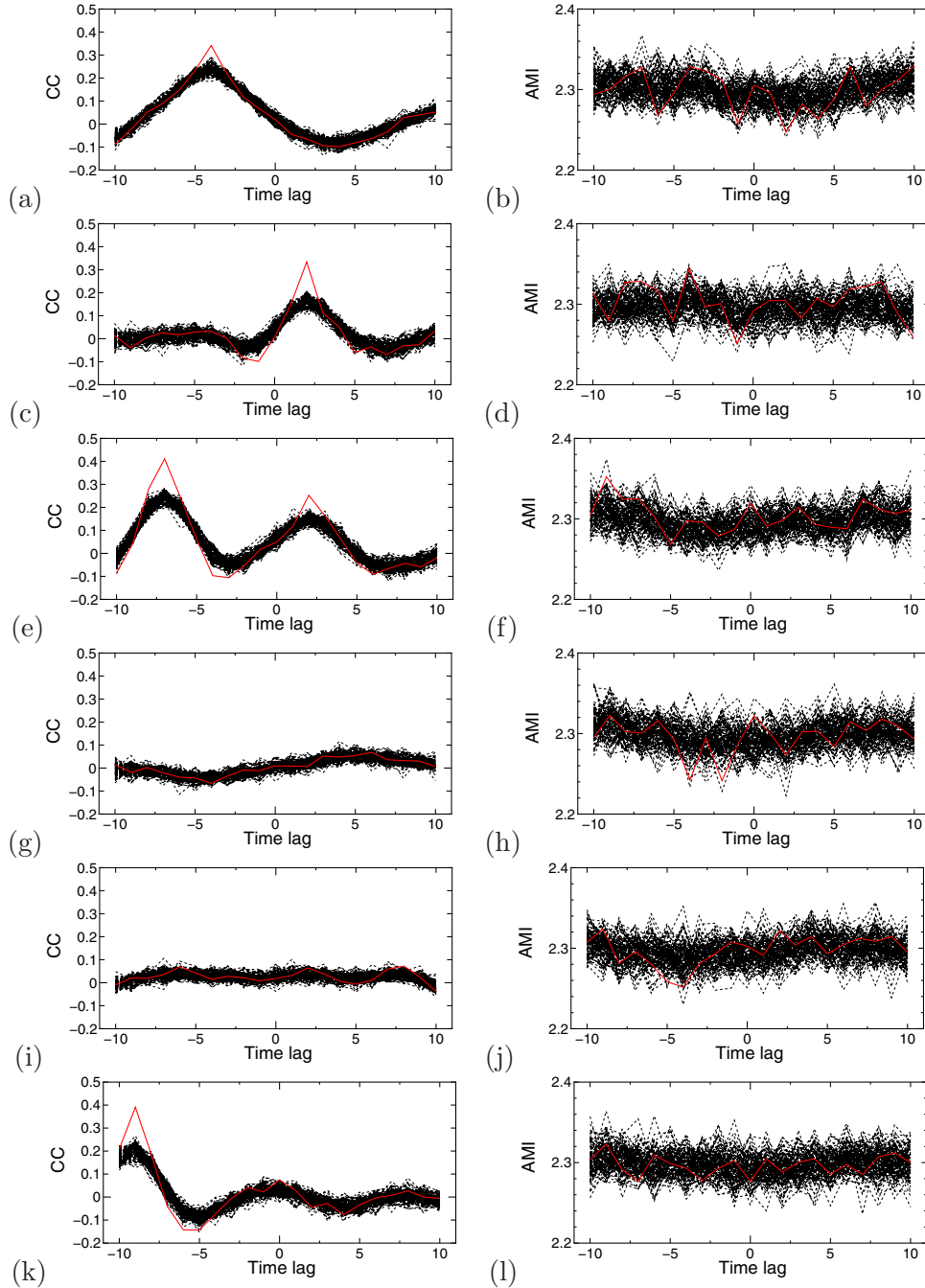|	x_1	x_2	x_3	x_4
x_1	1.0000	–	–	–
x_2	0.3413 (−4)	1.0000	–	–
x_3	0.3337 (2)	0.0688 (6)	1.0000	–
x_4	0.4113 (−7)	0.0725 (8)	0.3906 (−9)	1.0000

FIG. 5. The result of nonlinear system, Eqs. (9)–(12). A plot of (a), (c), (e), (g), (i) and (k) the cross correlation function (CC), and (b), (d), (f), (h), (j) and (l) the average mutual information (AMI), (a) and (b) are result of x_1 and x_2, (c) and (d) are result of x_1 and x_3, (e) and (f) are result of x_1 and x_4, (g) and (h) are result of x_2 and x_3, (i) and (j) are result of x_2 and x_4, and (k) and (l) are result of x_3 and x_4. The solid line is the original data and the dotted lines are the SSS data.

The function $h(x)$ is a static monotonic nonlinear function [32],

$$h(x) = \frac{5.0 \left[\frac{x-a-0.0001}{b-x+0.0001} \right]^{\rho}}{1 + \left[\frac{x-a-0.0001}{b-x+0.0001} \right]^{\rho}}, \tag{13}$$

where $\rho = 3$, $a = -2.0$, and $b = 10.0$. The behaviors of the four time series generated by these models are shown in Fig. 4. The behaviors show irregular fluctuations and it is difficult to know the relationship among the data by visual inspection.

In Eq. (9) there is a product $x_2(t-4)x_4(t-7)$, which is a nonlinear function, and as the data is transformed by a nonlinear function in Eq. (11), $x_3(t)$ is nonlinear data. $x_4(t)$ is no longer linear because $x_1(t-2)$ is included in the model as Eq. (4) shows. The connectivity of the nonlinear system is essentially the same as that of the linear system, Eqs. (1)–(4). That is, the network of the nonlinear system is the same as that shown in Fig. 2.

We estimate the cross correlation function to apply the naive method. All the values are shown in Table II. We need
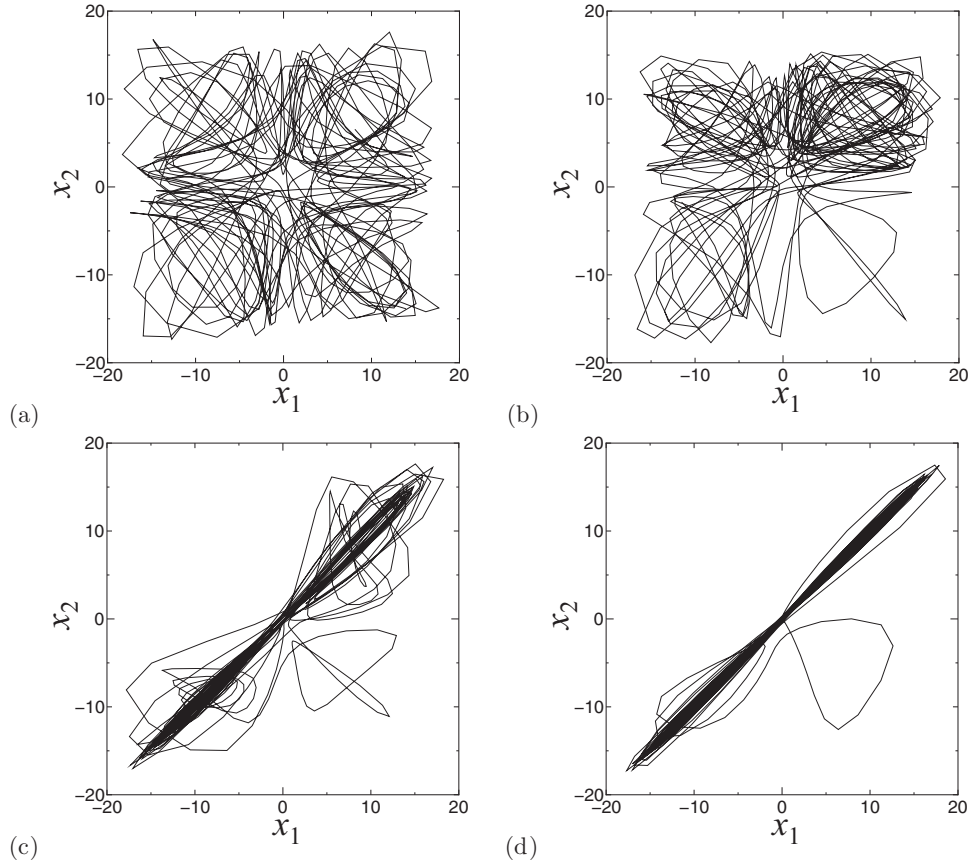
FIG. 6. Phase projections of phase trajectories of x_1 and x_2 of the coupled Lorenz systems, Eq. (14), for different values of coupling: (a) $\gamma = 0.0$ corresponds to the case of no coupling, (b) $\gamma = 1.0$ corresponds to the case of weak coupling, (c) $\gamma = 4.05$ corresponds to the beginning of synchronization of switching, and (d) $\gamma = 6.0$ corresponds to synchronous oscillations of subsystems. The data are calculated using the fourth order Runge-Kutta method with the sampling interval 0.01 and the resampling interval 0.05. The number of data points is 1000 and we use the data to construct the network.

to determine the threshold value to decide whether the link is connected between two components. If we set the value 0.5, as shown in Table II, we cannot connect any link between nodes. This result clearly indicates that only the application of the cross correlation function is not effective.

We apply the SSS method to the data of all possible pairs to verify the connection between two data sets. Figure 5 shows the result. This result indicates that we can discriminate correctly whether there are correlation structures between two

signals. Based on this we can construct the same network as shown in Fig. 2(a).

C. Data from the coupled Lorenz systems

The linear and nonlinear systems used in Secs. IV A and IV B are trivial examples, because the relationships among variables included in the systems are straightforward, as these systems are in actual time. To investigate how the proposed

TABLE III. The largest absolute values of the cross correlation function of all possible pairs between the time lag -10 and 10, where the number in the parentheses is the time lag when the cross correlation function has the largest absolute value. The data are generated by the coupled Lorenz systems with $\gamma = 0.0$, Eq. (14), and the values are estimated using 1000 data points.

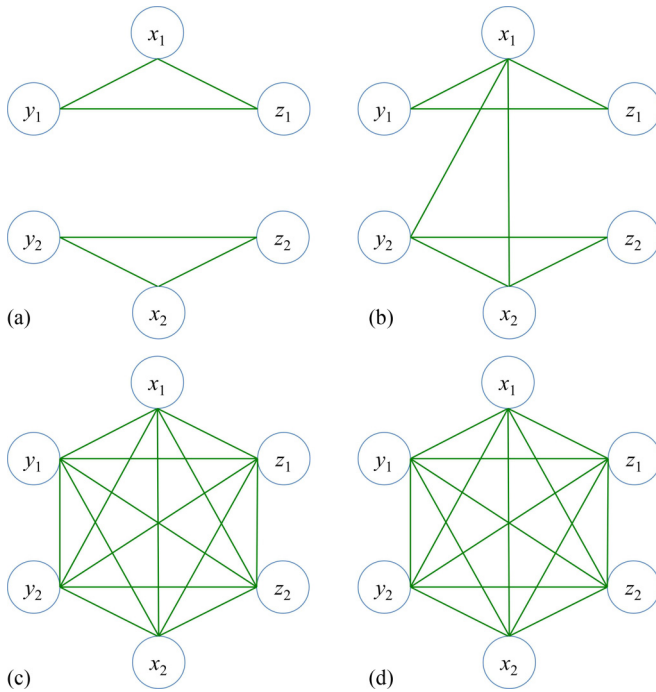|	x_1	y_1	z_1	x_2	y_2	z_2
x_1	1.0000	–	–	–	–	–
y_1	0.9669 (-1)	1.0000	–	–	–	–
z_1	0.0343 (-9)	0.0314 (-8)	1.0000	–	–	–
x_2	0.1238 (-7)	0.1245 (-5)	0.0323 (-10)	1.0000	–	–
y_2	0.1221 (-9)	0.1261 (-7)	0.0320 (8)	0.9664 (-1)	1.0000	–
z_2	0.0876 (-10)	0.0972 (-10)	0.1344 (-5)	0.0445 (10)	0.0146 (-5)	1.0000

FIG. 7. Network of the coupled Lorenz systems using the proposed method: (a) the network when $\gamma = 0.0$, (b) the network when $\gamma = 1.0$, (c) the network when $\gamma = 4.05$, and (d) the network when $\gamma = 6.0$.

method works in uncertain situations, we use the coupled Lorenz systems presented in the form of a differential equation as a less trivial example [21]. The coupled Lorenz systems are

two symmetrically coupled chaotic systems. The equations are given as

$$
\begin{aligned}
\dot{x}_1 &= \sigma(y_1 - x_1) + \gamma(x_2 - x_1), \\
\dot{y}_1 &= r_1 x_1 - x_1 z_1 - y_1, \\
\dot{z}_1 &= x_1 y_1 - b z_1, \\
\dot{x}_2 &= \sigma(y_2 - x_2) + \gamma(x_1 - x_2), \\
\dot{y}_2 &= r_2 x_2 - x_2 z_2 - y_2, \\
\dot{z}_2 &= x_2 y_2 - b z_2, \quad (14)
\end{aligned}
$$

where the parameters are $\sigma = 10, r_1 = 28.8, r_2 = 28, b = 8/3$ and γ is a coupling control parameter [21]. When calculating Eq. (14) we use the fourth order Runge-Kutta method. When the sampling interval of the Runge-Kutta method is small, there is strong collinearity among the data points. In such a case, even if two data channels are independent, when there are similar behaviors or patterns in some parts, the cross correlation function often indicates a large value because of these influences. On the other hand, the smaller the sampling interval the better to obtain the precise trajectories. To avoid this problem we resample the data. We use data using the fourth order Runge-Kutta method with the sampling interval 0.01 and the resampling interval 0.05 [33].

The mutual phase projections of phase trajectories for different values of coupling have been investigated by Anishchenko *et al.* [21]. Figure 6 shows the phase projections of x_1 and x_2.

We first consider when $\gamma = 0.0$ in Eq. (14). In this case the systems are not coupled. From Eq. (14) we expect that there are relationships among x_1, y_1, and z_1, and there are
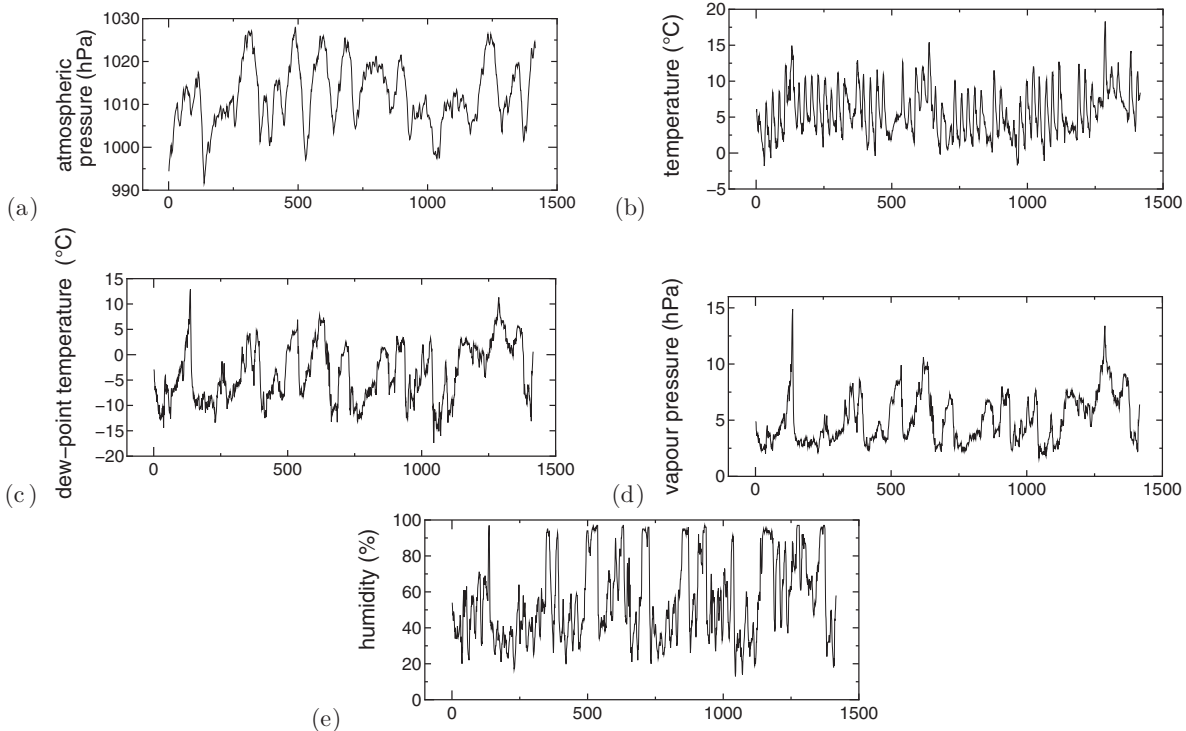

FIG. 8. Hourly meteorological time series in Tokyo, Japan from January 1 to February 28 in 2015 (1416 data points): (a) atmospheric pressure, (b) temperature, (c) dew-point temperature, (d) vapor pressure, and (e) humidity.

also relationships among x_2, y_2, and z_2. We estimate the cross correlation function to apply to the naive method. All the values are shown in Table III. To decide whether the link is connected between two components, we use the threshold value 0.5 as mentioned above. As shown in Table III, we can connect a link between x_1 and y_1 and another link between x_2 and y_2. As the systems are not coupled when $\gamma = 0.0$, it is reasonable that there is no link among some nodes, for example, between x_1 and x_2, x_1 and y_2, and x_1 and z_2. This is as expected. However, there are links we cannot connect among some nodes, although we expect the links, for example, between x_1 and z_1, and y_2 and z_2. We consider that this result is different from our expectation. Hence, this result indicates that only the application of the cross correlation function is not effective.

We apply the proposed method to the data from the coupled Lorenz systems. Figure 7 shows the network of each value of γ shown in Fig. 6. When $\gamma = 0.0$, as this is the case of no coupling, we consider the network shown in Fig. 7(a) is reasonable, because there are links among x_1, y_1, and z_1 and among x_2, y_2, and z_2, and these networks are separated. When $\gamma = 1.0$, this is the case of weak coupling. Figure 7(b) shows that x_1 and x_2 are connected. As x_1 and x_2 are influenced by one another in this case, we consider that this is reasonable. Also, x_1 and y_2 are connected, although y_1 and x_2 are not. This might be because the values of r_1 and r_2 are slightly different. Figures 7(c) and 7(d) show that all nodes are connected. When $\gamma = 4.05$ and $\gamma = 6.0$, as full-scale synchronous oscillations begin, we consider that the results are reasonable.

These results described in Secs. IV A and IV B indicate that we can discriminate correctly whether there are correlation structures between two signals, irrespective of whether data sets are from the linear system and nonlinear system, even if the value of the cross correlation function is small. The results in Sec. IV C indicate that the networks constructed

by the proposed method are the same as our expectation. Therefore, we conclude that the proposed method can verify the connectivity correctly.

V. APPLICATIONS

Based on the results of these computational studies, we apply the proposed method to two experimental systems: (i) hourly meteorological time series data set in Tokyo, Japan and (ii) daily exchange rate data set. As shown in Figs. 8 and 11, each of them shows irregular fluctuations. We use 1416 data points for a meteorological time series data set, and 2255 data points for an exchange rate data set. In all cases we use $A = 1.0$, generate 99 SSS data, and estimate the CC function and the AMI between the time lag -10 and 10.

A. Meteorological data

The meteorological time series data set we use are five different time series: the atmospheric pressure, the temperature, the dew-point temperature, the vapor pressure, and the humidity taken hourly in Tokyo, Japan from January 1 to February 28 in 2015 [34]. The measurement location is $35°41.5''$ north latitude and $139°45.0''$ east longitude. Figure 8 shows the time series.

Figure 9 shows the two results of applying the SSS method to the the meteorological time series. Figures 9(a) and 9(b) show that both the CC and the AMI of the original data fall within the distributions of SSS data. This result indicates that there is no correlation structure between the irregular fluctuations of temperature and vapor pressure. Figure 9(c) shows the CC of the original data fall outside the distributions of SSS data, although Fig. 9(d) shows that the AMI of the original data fall within the distributions. Hence, we consider that there are correlation structures between the irregular
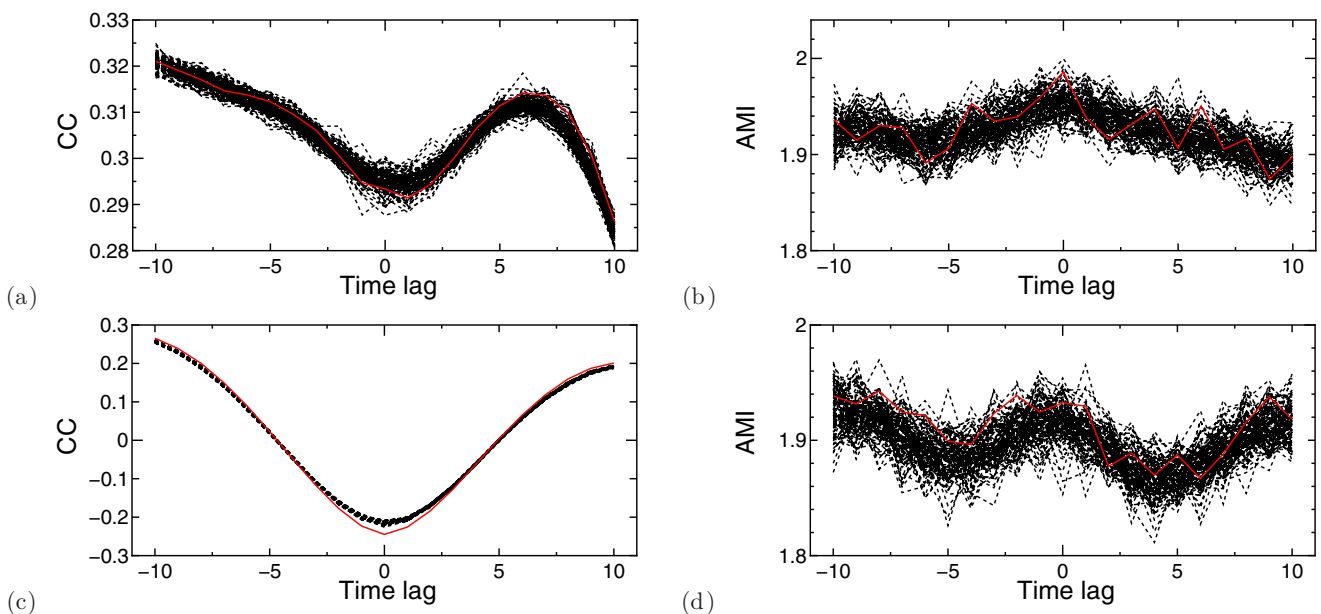

FIG. 9. A plot of (a) and (c) CC, and (b) and (d) AMI for the meteorological time series shown in Fig. 8: (a) and (b) temperature and vapor pressure, and (c) and (d) temperature and humidity.
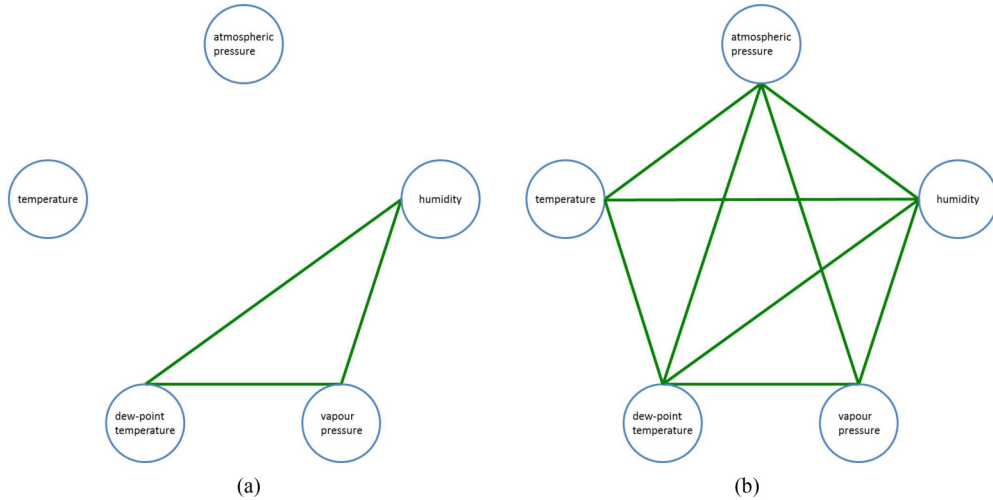
FIG. 10. Network of meteorological time series taken hourly in Tokyo, Japan from January 1 to February 28 in 2015 (1416 data points): (a) the network using the naive method using the CC with the threshold 0.5, and (b) the network using the proposed method.

fluctuations of temperature and humidity. We note that the largest absolute values of the cross correlation function for both cases are not large and these values are smaller than 0.5. This means that temperature, vapor pressure, and humidity are not connected if we apply the naive method.

Figure 10 shows the network of applying the proposed method to the meteorological time series. For comparison we also construct the network using the naive method, where we use the threshold 0.5 for the cross correlation function. Figure 10(a) shows that the network using the naive method has three links. However, Fig. 10(b) shows there are nine links which is more than that on the network using the naive method. From the result, we see that the relationship among meteorological data sets may be more complicated than suggested via the naive approach.

We briefly consider the reciprocal relationship between temperature and humidity. As Fig. 9(c) shows that the behavior of the CC is similar to a downward convex shape, a parabola of a quadratic function which opens upward, we understand

that these are negatively correlated. However, the largest absolute value of the CC is not large as mentioned above. As a result, we would conclude that the temperature and humidity in Tokyo are not correlated. As many roads are paved with asphalt and there are many tall buildings in Tokyo, the environment or situation might be far from nature. Hence, the relationship between the temperature and the humidity in Tokyo might be tenuously tied. However, as the basic physical and meteorological relationship does not change, we consider that the naive method is inadequate.

B. Exchange rate data

The exchange rate data set we use are four different time series: Swiss Franc–US dollar (CHF–USD), European Euro–US dollar (EUR–USD), British Pound–US dollar (GBP–USD), and the Japanese Yen–US dollar (JPY–USD). These currencies are often treated as the major currency. The data set is from January 3, 2006 to December 31, 2014 and we use
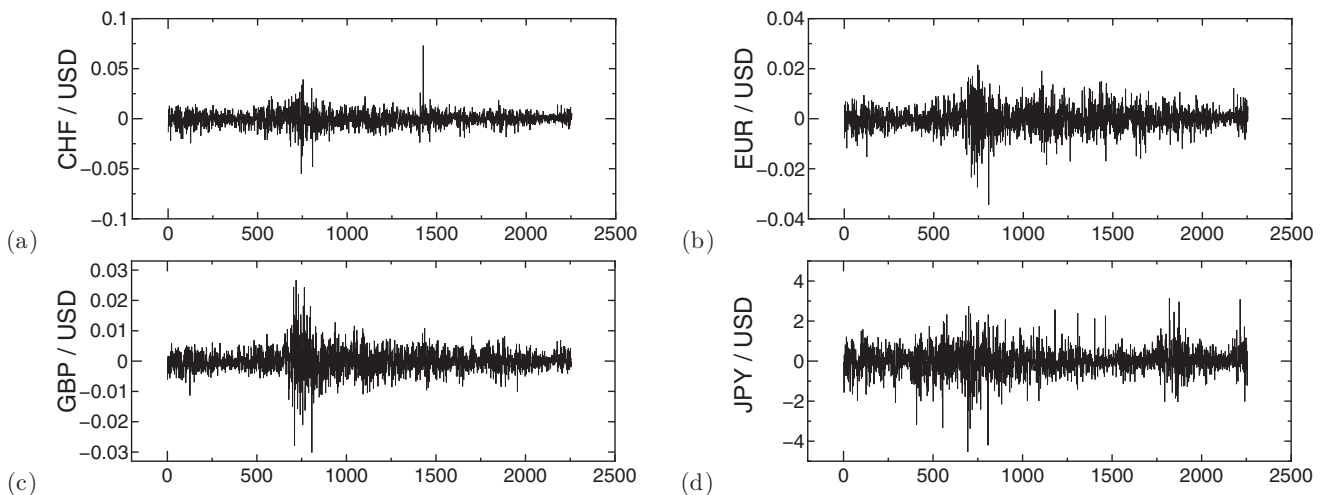

FIG. 11. The first difference data of daily exchange rate set from January 3, 2006 to December 31, 2014 (2255 data points): (a) CHF–USD, (b) EUR–USD, (c) GBP–USD, and (d) JPY–USD.
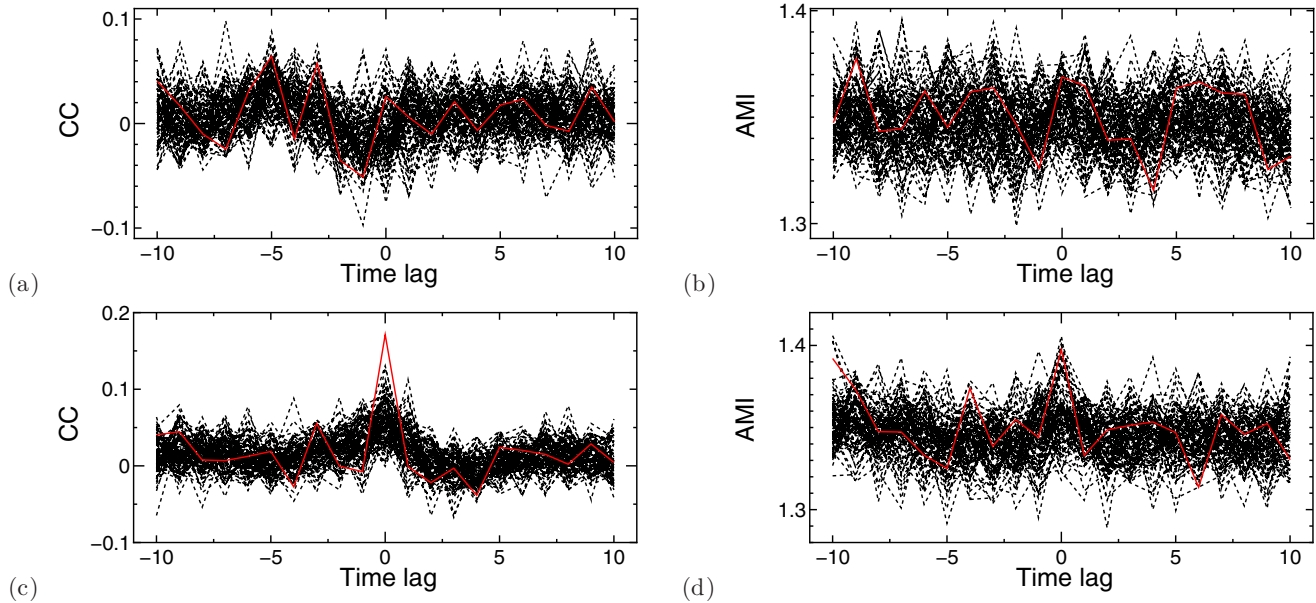
FIG. 12. A plot of (a) and (c) CC, and (b) and (d) AMI for the first difference data of the exchange rate data set shown in Fig. 11: (a) and (b) GBP–USD and JPY–USD, and (c) and (d) EUR–USD and JPY–USD.

the first difference data [35]. Figure 11 shows the four time series. The volatility of these data is not constant and clearly increases and decreases over time.

Figure 12 shows the two results of applying the SSS method to the exchange rate data sets. Figures 12(a) and 12(b) show that both the CC and the AMI of the original data fall within the distributions of SSS data. This result indicates that there is no correlation structure between the irregular fluctuations of GBP–USD and JPY–USD. Figure 12(c) shows the CC of the original data fall outside the distributions of SSS data, although Fig. 12(b) shows that the AMI of the original data fall within the distributions. Hence, we consider that there are correlation structures between the irregular fluctuations of EUR–USD and JPY–USD. We note that the largest absolute values of the cross correlation function for the both the cases are not large and these values are smaller than 0.5. This means that EUR–USD and JPY–USD are not connected if we apply the naive method.

Figure 13 shows the network of applying the naive method and the proposed method to the daily exchange rate data sets. Figure 13(a) shows that the network using the naive method has two links and only JPY–USD has no connection. However, Fig. 13(b) shows there are more links and JPY–USD connects with CHF–USD and EUR–USD.

It should be noted that Figs. 10 and 13 show that there are differences between the networks using the naive method and the proposed method. This indicates that we cannot easily decide whether there are correlation structures between the two signals using the CC of only the original data. Also, this implies that there are more complicated relationships among data sets than we expected.

Price movements of exchange rate are basically determined by the balance of supply and demand for the currency. The main factors are the national context and future prospects about the politics and economics. Hence, price movements of exchange rate can be said to be assessment of an economic

performance of countries. As the United Kingdom (UK), Switzerland, and the Euro area are geopolitically close, the direct economical ties such as exchange people and commercial pursuits are active. The gross domestic product (GDP) of purchasing power parity (PPP) of the Euro area is 13.236 trillion USD, that of UK is 2.569 trillion USD, and that of Switzerland is 473.3 billion USD. That is, the GDP (PPP) of the Euro area is the largest, that of UK is second largest, and that of Switzerland is third largest. From this we consider that the UK and Switzerland have a strong economical influence from the Euro area. As a result, we consider that the behavior of EUR–USD and GBP–USD is similar and that of EUR–USD and CHF–USD is also similar (the largest absolute value of the CC between EUR–USD and GBP–USD is 0.6524 and that between EUR–USD and CHF–USD is 0.7905). However, the UK and Switzerland each have internal affairs and affairs with other countries except the Euro area. Hence, although the UK and Switzerland are influenced by the Euro area, the behavior of GBP–USD and CHF–USD is not similar (the largest absolute value of the CC between GBP–USD and CHF–USD is 0.4914).

On the other hand, when countries are geographically separated, the above mentioned direct economical ties among the countries are generally less active (inactive) than that among geographically close countries. However, we can directly make foreign-currency transactions of various countries at the present day. In this case, as the deals are done based on dealers' circumstances, expectations, and so on, the relationships among the exchange rates might be more complicated than we expect. Switzerland and Japan are geographically very separated. However, Switzerland and Japan have something in common. The common point is that CHF and JPY are regarded as worldwide stable and reliable currencies. As a matter of fact, CHF and JPY buying are often increased in case of emergencies. Hence, it is small wonder that there is some kind of relationship between CHF–USD and JPY–USD.
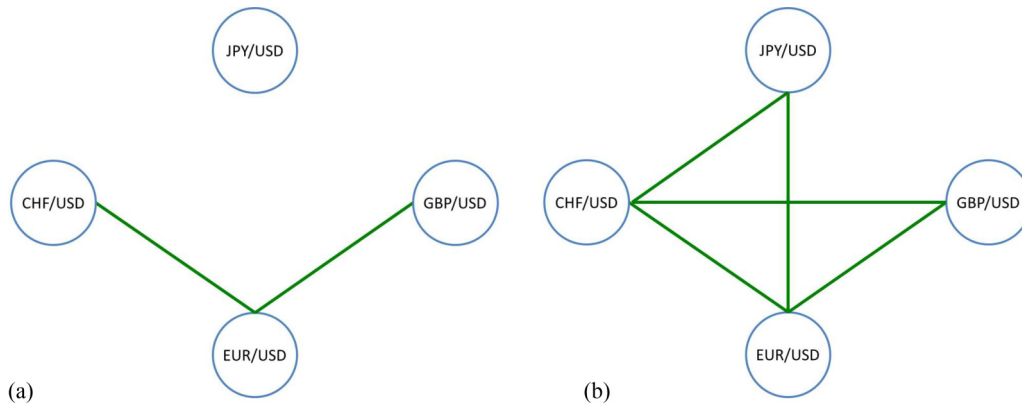
FIG. 13. Network of daily exchange rate data from January 3, 2006 to December 31, 2014 (2255 data points): (a) the network using the naive method using the CC with the threshold 0.5, and (b) the network using the proposed method.

We consider that the network constructed by the proposed method shown in Fig. 13(b) reflects these matters.

VI. CONCLUSION

We have described an algorithm for constructing networks from multivariate linear or nonlinear time series, even when the data are not linearly correlated. To verify the connection between two signals, we apply the previously proposed SSS method, which can investigate correlation structures irrespective of whether the structures are linear or nonlinear. We have demonstrated the application of this algorithm to computational examples using the CC and the AMI as discriminating statistics. Our arguments and computational examples show that this algorithm succeeds in testing correlation structures in irregular fluctuations irrespective of whether the data are linear or nonlinear and constructing the network from a dynamical system perspective.

[1] D. H. Meadows, *Thinking in Systems: A Primer* (Chelsea Green Publishing, London, 2008).

[2] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).

[3] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).

[4] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).

[5] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).

[6] M. Small, D. M. Walker, and C. K. Tse, Phys. Rev. Lett. **99**, 188702 (2007).

[7] M. Small, J. Zhang, and X. Xu, in *Complex Sciences* (Springer, Berlin/Heidelberg, 2009), pp. 2078–2089.

[8] N. Marwan, J. F. Donges, Y. Zou, R. V. Donner, and J. Kurths, Phys. Lett. A **373**, 4246 (2009).

[9] J. Zhang and M. Small, Phys. Rev. Lett. **96**, 238701 (2006).

[10] X. Xu, J. Zhang, and M. Small, Proc. Natl. Acad. Sci. USA **105**, 19601 (2008).

[11] B. Luque, L. Lacasa, F. Ballesteros, and J. Luque, Phys. Rev. E **80**, 046103 (2009).

[12] R. N. Mantegna, Eur. Phys. J. B **11**, 193 (1999).

[13] I. Farkas, H. Jeong, T. Vicsek, A.-L. Barabási, and Z. N. Oltvai, Physica A **318**, 601 (2003).

[14] K. Yamasaki, A. Gozolchiani, and S. Havlin, Phys. Rev. Lett. **100**, 228501 (2008).

[15] A. A. Tsonis and K. L. Swanson, Phys. Rev. Lett. **100**, 228502 (2008).

[16] C. K. Tse, J. Liu, and F. C. M. Lau, J. Empirical Finance **17**, 659 (2010).

[17] M. Nagy, Z. Ákos, D. Biro, and T. Vicsek, Nature (London) **464**, 890 (2010).

[18] D. M. Walker, C. Carmeli, F. J. Pérez-Barbería, M. Small, and E. Pérez-Fernández, Anim. Behav. **79**, 351 (2010).

[19] T. Nakamura, Y. Hirata, and M. Small, Phys. Rev. E **74**, 041114 (2006).

[20] T. Schreiber and A. Schmitz, Physica D **142**, 346 (2000).

[21] V. S. Anishchenko, A. N. Silchenko, and I. A. Khovanov, Phys. Rev. E **57**, 316 (1998).

[22] G. Papadopoulos and D. Kugiumtzis, Physica A **436**, 387 (2015).

[23] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, Physica D **58**, 77 (1992).

[24] H. D. I. Abarbanel, *Analysis of Observed Chaotic Data* (Springer-Verlag, New York, 1996).

[25] J. Theiler and D. Prichard, Physica D **94**, 221 (1996).

[26] The significance level of each test is 0.01. If two statistics are identical (dependent), the significance level for the proposed test is 0.01. If the statistics are independent, the significance level for the test is given by $1.0 - 0.99 \times 0.99 = 0.0199$. Hence, the reality should be somewhere in-between [19].

[27] M. Small, T. Nakamura, and X. Luo, *Nonlinear Phenomena Research Perspectives* (Nova Science Publications, New York, 2007).

[28] T. Nakamura and M. Small, Physica A **377**, 599 (2007).

[29] As the linear and nonlinear systems are toy, there is no special meaning for both the systems. Broadly speaking, we built the systems as follows. We first built the linear system. Variables (terms and time delays) and the coefficients in the linear system were chosen arbitrarily so that the generated multivariate time series did not diverge. Also, we built four models from the data generated by the linear system using 1000 data points and checked that all terms included in the linear system were selected in the best model, because a requisite term is not included in the best model if the term does not work as the variable for the linear system. In such a case we changed the terms or time delays or the coefficient. We repeated these operations until we could build good models. After building the linear system we built the nonlinear system. We added a nonlinear term and a static monotonic nonlinear function to a nonlinear system. The coefficients in the nonlinear system were chosen arbitrarily based on the linear system. Then we checked that the generated multivariate time series for long term did not diverge.

[30] It is possible to justify the value of the threshold using the t test [15]. However, as we can practically set the significant level as one likes (for example, 90%, 95%, 99%, and so on), the threshold value is still arbitrary.

[31] We note that there are cases where the proposed method cannot find the connectivity correctly. In Eq. (1) the coefficient of $x_2(t - 4)$ is 0.3. When we use 0.1 instead of 0.3, we cannot find the relationship between x_1 and x_4. We expect that the proposed method cannot find the relationship when the influence of a variable is not large enough.

[32] P. E. Rapp, C. J. Cellucci, T. A. A. Watanabe, A. M. Albano, and T. I. Schmah, Int. J. Bifurcation Chaos **11**, 983 (2001).

[33] We previously confirmed that the resampling interval 0.05 is enough at least by the following approach. When we apply the SSS method to data without resampling (that is, the data is generated using the Runge-Kutta method with the sampling interval 0.01), even though $\gamma = 0$ [this is a case where Eq. (14) is not coupled], the result indicates that there is some sort of correlation structure between x_1 and x_2 in Eq. (14). However, this result is wrong. When we apply the SSS method to data with the resampling interval 0.05 we could obtain the correct result that x_1 and x_2 are independent.

[34] The data set can be obtained from Japan Meteorological Agency, http://www.jma.go.jp/jma/indexe.html.

[35] The data set can be obtained from PACIFIC Exchange Rate Service, http://fx.sauder.ubc.ca/data.html.