2016 International Symposium on Nonlinear Theory and Its Applications,
NOLTA2016, Yugawara, Japan, November 27th-30th, 2016

On constructing networks from 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
200-1 Monobe-Otsu, Nankoku, Kochi 783-8508 Japan
[3] School of Mathematics and Statistics, The University of Western Australia,
Crawley, WA 6009, Australia
[4] Mineral Resources, CSIRO, Kensington, WA 6151, Australia
Email: * tomo@sim.u-hyogo.ac.jp, † tanizawa@ee.kochi-ct.ac.jp, ‡ michael.small@uwa.edu.au

Abstract—We introduce a method for constructing networks from multivariate nonlinear time series from a deterministic dynamical systems perspective. The method can be applied even when the data exhibit no obvious qualitative similarity: a situation in which the naive method utilising the cross correlation function directly cannot correctly identify connectivity. The method is demonstrated for numerical data sets generated by known systems and applied to several experimental time series.

1. 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 (or underlying system) under study [1]. Elements in the system interact with each other. Once the connectivity has been determined the effect of that connectivity is frequently studied using the concept of complex networks [2].

There are also approaches for constructing networks for multivariate time series [3, 4]. 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" between two signals is with the cross correlation function with a fixed threshold [3, 4]. While this naive approach is expeditious, it is also flawed when one is looking at nonlinear (possibly temporally delayed) interaction in complex systems (in other words, two signals are not similar). We describe the naive approach in detail in Sec. 2.

The most important thing to investigate the relationship between two signals is not similarity but correlation structures from the viewpoint of a deterministic dynamical system. Recently a method to construct such a network for multivariate nonlinear time series has been proposed based on this perspective [5]. To verify the intrinsic (essential) connection between two data sets, the previously proposed small shuffle surrogate (SSS) method is applied in the method, which can investigate correlation structures irrespective of whether the structures are linear or nonlinear [6]. That is, the method constructs networks for multivariate time series, even if there are nonlinearities in these time series.

2. The naive approach to network construction

The most extensively used method to construct networks for multivariate time series can be reduced to the following three steps [3, 4].

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. Although the naive method has been proved to be effective in various cases [3, 4], the range of applicability might be restrictive because "no similarity" is not equivalent to "no correlation" and "no relationship" [6]. Furthermore, there is a possibility that "similarity" might not be equivalent to "relationship." That is, the naive method cannot deal with data appropriately especially when there are nonlinearties. In the next section, we describe an approach to reduce this problem.

3. A different approach to construct networks

The approach of the proposed method is basically the same as the naive method described in Sec. 2. 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 small shuffle surrogate (SSS) method [6], because the SSS method has broad applicability[1] and can examine whether there are correlation structures[2].

3.1. 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 [6]. The SSS method destroys local structures or correlations in irregular fluctuations (short-term variabilities) and preserves the global behaviours 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 $s(t) = x\left[\hat{i}(t)\right]$.

It has been found that choosing $A = 1.0$ is adequate for nearly all purposes [6] — although this parameter choice remains heuristic. Further details of the method and the mechanism are provided in [6]. 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 [6].

3.2. 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 [7].

After the calculation of these statistics, we need to determine whether a null hypothesis (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 [8]. 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 non-parametric significance level is between 0.01 and 0.02 for a one-sided test with two non-independent statistics[3].

4. Numerical Example

We demonstrate the application of our algorithm to one simulated multivariate time series data set, and confirm our theoretical arguments with the several example. For comparison we also apply the naive method to the data sets. In this case, we use $A = 1.0$ for generating SSS data, generate 99 SSS data, and the data is 1000 points with Gaussian observational noise with the mean zero and the standard deviation 0.01.

4.1. Data from a nonlinear system

To investigate whether the proposed method works even if there is nonlinearity, 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:

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

$$x_2(t) = 2.0 + 0.6\, x_2(t-1) - 0.2\, x_2(t-6) + \varepsilon_2(t), \quad (2)$$

$$x_3(t) = h\left[2.2 + 0.2\, x_1(t-2) + 0.3\, x_4(t-9) + \varepsilon_3(t)\right], \quad (3)$$

$$x_4(t) = 1.3 + 0.2\, x_1(t-2) + 0.5\, x_4(t-1) - 0.3\, x_4(t-3) + \varepsilon_4(t), \quad (4)$$

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 function $h(x)$ is a static monotonic nonlinear function [9],

$$h(x) = \frac{5.0\left[\frac{x-a-0.0001}{b-x+0.0001}\right]^{\rho}}{1 + \left[\frac{x-a-0.0001}{b-x+0.0001}\right]^{\rho}}, \quad (5)$$

where $\rho = 3$, $a = -2.0$ and $b = 10.0$. The behaviours of the four time series generated by these models are shown in Fig. 1. The behaviours show irregular fluctuations and it is difficult to know the relationship among the data by visual inspection.

[1] The SSS method can investigate whether there are correlation structures in short-term variabilities among data, irrespective of whether data have similar or different long-term trends.

[2] The term "correlation structures" we use means any structures, irrespective of whether the structures are linear or nonlinear.

[3] 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 [6].

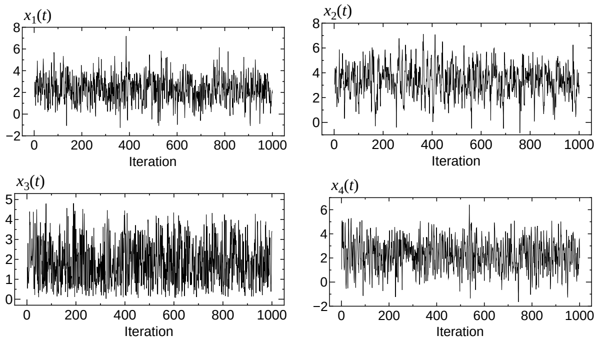

Figure 1: Time series data generated by the nonlinear system, Eqs. (1)–(4). We use the data to construct the network.

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). Eq. (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 this, the connectivity expressions of the nonlinear system become

$$x_1 = f_1(x_2, x_4), \tag{6}$$
$$x_2 = 0, \tag{7}$$
$$x_3 = f_3(x_1, x_4), \tag{8}$$
$$x_4 = f_4(x_1), \tag{9}$$

where f_i stands for the function representing connectivity of the i-th 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 estimate the cross correlation function to apply to the naive method. All the values are shown in Table 1. We need to determine the fixed threshold value to decide whether a link is present between two components. If we set the value 0.5, as shown in Table 1, we cannot connect any link between nodes. 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 apply the SSS method to the data of all possible pairs to verify the connection between two data sets. Figure 3 shows the result. This result indicates that we can discriminate correctly whether there are correlation structures between two signals. Also, other data sets are discriminated correclty. Based on this we can construct the same network as shown in Fig. 2(a).

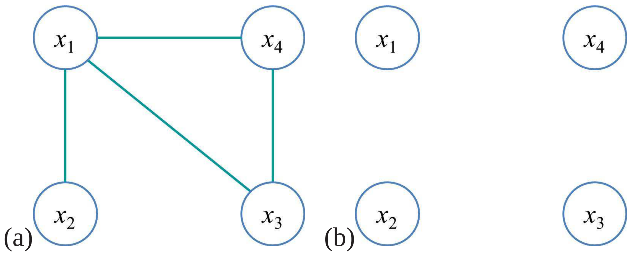

Figure 2: (Colour online) The linkage of network: (a) the connectivity of Eqs. (1)–(4). The same network is obtained when the proposed method is applied to the data shown in Fig. 1. (b) the network when we apply the naive method to the data. As shown in this figure, there is no link among the nodes.

Table 1: 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. (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	0.3413 (-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

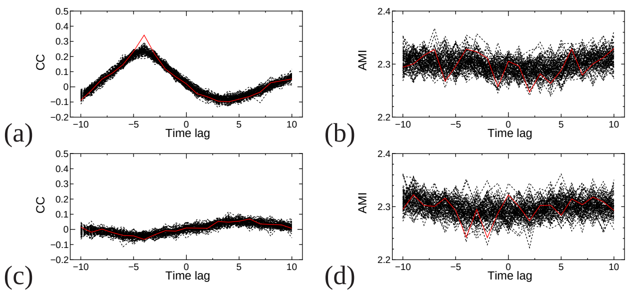

Figure 3: (Colour online) The result of nonlinear system, Eqs. (1)–(4). A plot of (a) and (c) the cross correlation function (CC), and (b) and (d) the average mutual information (AMI), (a) and (b) are result of x_1 and x_2, and (c) and (d) are result of x_2 and x_3, The solid line is the original data and the dotted lines are the SSS data.

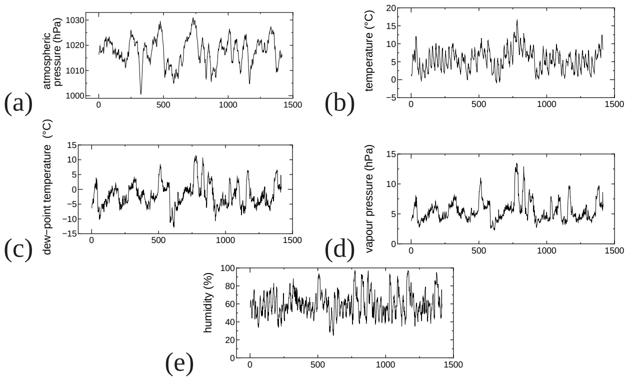

Figure 4: Hourly meteorological time series in Kobe, Japan from 1 January to 28 February in 2013 (1416 data points): (a) atmospheric pressure, (b) temperature, (c) dew-point temperature, (d) vapour pressure, and (e) humidity.

5. Application

Based on the results of the computational studies, we apply the proposed method to one experimental system: hourly meteorological time series data set in Kobe, Japan. The meteorological time series data set are five different time series: the atmospheric pressure, the temperature, the dew-point temperature, the vapour pressure and the humidity taken hourly in Kobe, Japan from 1 January to 28 February in 2013[4]. As shown in Fig. 4, each of them shows irregular fluctuations. We use 1416 data points for a meteorological time series 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 $+10$.

Figure 5 shows networks constructed by the naive method and the proposed method. There are interesting differences between them. The CC functions between the temperature and the dew-point temperature and between the temperature and the vapour pressure are larger than 0.5. Hence, these are connetced as shown in Fig 5(a). However, as the CC and AMI of the original data fall within the distributions of SSS data, these are not connected as shown in Fig 5(b). This might indicate that "similarity" is not equivalent to "relationship."

6. Conclusion

We have introduced an algorithm to construct networks for multivariate time series using the SSS method. The network constructed by the method indicates the intrinsic connectivity of the elements included in the system.

Acknowledgments

T.N would like to acknowledge the partial support of a Grant-in-Aid for Scientific Research (C) (No. 25282094) from the Japan Society for the Promotion of Science (JSPS), and T.T also would like to acknowledge the support of a Grant-in-Aid for Scientific Research (C) (No. 24540419) from JSPS. M.S is funded by the Australian Research Council via a Future Fellowship (FT110100896) and Discovery Project (DP140100203).

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

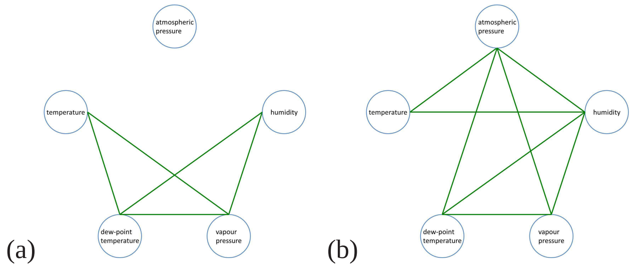

Figure 5: Network of meteorological time series taken hourly in Kobe, Japan shown in Fig. 4: (a) the network using the naive method using the CC with the threshold 0.5, and (b) the network using the proposed method.

References

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

[2] A-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[3] K. Yamasaki, A. Gozolchiani, and S. Havlin. Climate networks around the globe are significantly affected by el niño. *Phys. Rev. Lett*, 100(22):228501, 2008.

[4] M. Nagy, Z. Ákos, D. Biro, and T. Vicsek. Hierarchical group dynamics in pigeon flocks. *Nature (London)*, 464:890–894, 2010.

[5] T. Nakamura, T. Tanizawa, and M. Small. Constructing networks from a dynamical system perspective for multivariate nonlinear time series. *Physical Review E*, 93:032323, March 2016.

[6] T. Nakamura, Y. Hirata, and M. Small. Testing for correlation structures in short-term variabilities with long-term trends of multivariate time series. *Physical Review E*, 74:041114, 2006.

[7] H.D.. Abarbanel. *Analysis of observed Chaotic Data*. Springer-Verlag, New York, 1996.

[8] J. Theiler and D. Prichard. Constrained-realization monte-carlo method for hypothesis testing. *Physica D*, 94:221–235, 1996.

[9] P.E. Rapp, C.J. Cellucci, T.A.A. Watanabe, A.M. Albano, and T.I. Schmah. Surrogate data pathologies and the false-positive rejection of the null hypothesis. *Int. J. Bifurcation Chaos*, 11:983–997, 2001.