

Article

# Inferring Weighted Directed Association Networks from Multivariate Time Series with the Small-Shuffle Symbolic Transfer Entropy Spectrum Method

Yanzhu Hu <sup>1</sup>, Huiyang Zhao <sup>1,2,\*</sup> and Xinbo Ai <sup>1</sup>

<sup>1</sup> Beijing Key Laboratory of Work Safety Intelligent Monitoring, Beijing University of Posts and Telecommunications, Beijing 100876, China; yzhu@263.net (Y.H.); axb@bupt.edu.cn (X.A.)

<sup>2</sup> School of Information Engineering, Xuchang University, Xuchang 461000, China

\* Correspondence: zhyjky@163.com; Tel.: +86-10-6228-3022

Academic Editors: Jinfeng Wang and Kevin H. Knuth

Received: 14 June 2016; Accepted: 2 September 2016; Published: 7 September 2016

**Abstract:** Complex network methodology is very useful for complex system exploration. However, the relationships among variables in complex systems are usually not clear. Therefore, inferring association networks among variables from their observed data has been a popular research topic. We propose a method, named small-shuffle symbolic transfer entropy spectrum (SSSTES), for inferring association networks from multivariate time series. The method can solve four problems for inferring association networks, i.e., strong correlation identification, correlation quantification, direction identification and temporal relation identification. The method can be divided into four layers. The first layer is the so-called data layer. Data input and processing are the things to do in this layer. In the second layer, we symbolize the model data, original data and shuffled data, from the previous layer and calculate circularly transfer entropy with different time lags for each pair of time series variables. Thirdly, we compose transfer entropy spectrums for pairwise time series with the previous layer's output, a list of transfer entropy matrix. We also identify the correlation level between variables in this layer. In the last layer, we build a weighted adjacency matrix, the value of each entry representing the correlation level between pairwise variables, and then get the weighted directed association network. Three sets of numerical simulated data from a linear system, a nonlinear system and a coupled Rossler system are used to show how the proposed approach works. Finally, we apply SSSTES to a real industrial system and get a better result than with two other methods.

**Keywords:** network inference; multivariate time series; surrogate data method; transfer entropy

## 1. Introduction

### 1.1. Problem Statement

Association networks are found in many domains, such as networks of citation patterns across scientific articles [1–3], networks of actors co-starring in movies [4–6], networks of regulatory influence among genes [7,8], and networks of functional connectivity between regions of the brain [9,10]. The rules defining edges in association networks are not the same. In general, if the relationships among nodes are explicit, we can define a rule for their connectivity and establish the network easily. However, the relationships among components are unknown in many real complex systems, so association network inference has become a popular research topic. Many complex systems belong to the industrial field and the datasets obtained from these complex systems are multivariate time series. Therefore, we aim at studying association network inference from multivariate time series and also attempt to deal appropriately with the problems of the edges' direction and weight in the network.

## 1.2. Related Works

Association network inference has been a research topic for several years. We will review some methods that have been proposed so far to address the undetermined relationships among variables. The most classical approach is based on correlation. For instance, Guo et al. [7] incorporated the distance correlation into inferring gene regulatory networks from the gene expression data without any underlying distribution assumptions. Maucher et al. [11] used Pearson correlation as an elementary correlation measure to detect regulatory dependencies in a gene regulatory network. The association networks generated by basic correlation approaches usually include many indirect relationships which need to be detected and removed to increase the power of the network inference approach. Therefore, a major challenge in inferring association networks is the identification of direct relationships between variables. The classical approach to detect indirect relationships is based on partial correlations, which imposes the control of one gene on the relationship of others. Han and Zhu [12] proposed a method based on the matrix of thresholding partial correlation coefficients (MTPCC) for network inference from expression profiles. The corresponding undirected dependency graph (UDG) was obtained as a model of the regulatory network of *S. cerevisiae*. Yuan et al. [8] proposed a directed partial correlation (DPC) method as an efficient and effective solution to regulatory network inference. It combines the efficiency of partial correlation for setting up network topology by testing conditional independence, and the concept of Granger causality to assess topology change with induced interruptions. Wang et al. [13] focused on gene group interactions and inferred these interactions using appropriate partial correlations between genes, that is, the conditional dependencies between genes after removing the influences of a set of other functionally related genes.

Moreover, Gaussian Graphical Models also perform well for inferring association networks in specific experimental datasets. Schäfer and Strimmer [14] introduced a framework for small-sample inference of graphical models from gene expression data to detect conditionally dependent genes. Huynh-Thu et al. [15] proposed an algorithm using tree-based ensemble methods like random forests or extra-trees for the inference of GRNs (Genetic Regulatory Networks) that was the best performer in the DREAM4 In Silico Multifactorial Challenge.

Some approaches to infer association networks rely on information theory-based similarity measures. Margolin et al. [16] described a computational protocol for the ARACNE algorithm, an information-theory method for identifying transcriptional interactions between gene products using microarray expression profile data. Faith et al. [17] developed and applied the context likelihood of relatedness (CLR) algorithm, also used mutual information as a metric of similarity between the expression profiles of two genes. Zoppoli et al. [18] proposed a method called TimeDelay-ARACNE. It tries to extract dependencies between two genes at different time delays, providing a measure of these dependencies in terms of mutual information. TimeDelay-ARACNE can infer small local networks of time regulated gene-gene interactions detecting their versus and also discovering cyclic interactions when only a medium-small number of measurements are available. Villaverde et al. [19] reviewed some of the existing information theory methodologies for network inference, and clarified their differences.

In addition, approaches rooted in Bayesian Networks (BN) employ probabilistic graphical models in order to infer causal relationships between variables. Aliferis et al. [20] presented an algorithmic framework for learning local causal structure around target variables of interest in the form of direct causes/effects and Markov blankets applicable to very large data sets with relatively small samples. The selected feature sets can be used for causal discovery and classification. Dondelinger et al. [21] introduced a novel information sharing scheme to infer gene regulatory networks from multiple sources of gene expression data. They illustrate and test this method on a set of synthetic data, using three different measures to quantify the network reconstruction accuracy. As a review paper, Lian et al. [22] first discussed the evolution of molecular biology research from reductionism to holism. This is followed by a brief insight on various computational and statistical methods used in GRN inference before focusing on reviewing the current development and applications of DBN-based methods.

Granger causality (GC) is also a very popular tool for association networks inference. It can assess the presence of directional association between two time series of a multivariate data set. GC was introduced originally by Wiener [23], and later formalized by Granger [24] in terms of linear vector autoregressive (VAR) modeling of multivariate stochastic processes. Tilghman and Rosenbluth [25] presented Granger Causality as a method for inferring communications links among a collection of wireless transmitters from externally measurable features. The link inference method was applicable to inferring the link topology of broad classes of wireless networks, regardless of the nature of the Medium Access Control (MAC) protocol used. Cecchi et al. [9] presented a scalable method, based on the Granger causality analysis of multivariate linear models, to compute the structure of causal links over large scale dynamical systems that achieves high efficiency in discovering actual functional connections. The method was proved well to deal with autoregressive models of more than 10,000 variables. Schiatti et al. [26] compared the GC with a novel measure, termed extended GC (eGC), able to capture instantaneous causal relationships. The practical estimation of eGC worked with a two-step procedure, first detecting the existence of zero-lag correlations, and then assigning them to one of the two possible causal directions using pairwise measures of non-Gaussianity. Montalto et al. [27] introduced a new Granger causality measure called neural networks Granger causality. Instead of fitting predefined models, (linear ones in the original proposal by Granger) they trained a neural network to estimate the target using only the past states that can better explain the target series, by using the non-uniform embedding technique.

Of course, there are many more methods for association networks inference and we have not mentioned above, such as neural network [28], SparCC (Sparse Correlations for Compositional data) [29], S-estimator [30,31], Maximal Information Coefficient (MIC) [32], Local Similarity Analysis (LSA) [33,34], Transfer Entropy [35–37], and so on. They all showed some excellent performance through experiment and observation.

Until now, we have not seen a description that can deal with any association network inference problem. Although any of the abovementioned researches have its advantages proved by different styles, it is not always suitable for one's particular network inference problem. Because each strategy applies different assumptions, they each have different strengths and limitations and highlight complementary aspects of the network. Some of these popular tools are non-directional, e.g., correlation or partial correlation, mutual information measures and Bayesian Networks, thus these measures cannot satisfy one's directed association networks inference study [36]. Granger causality has acquired preeminent status in the study of interactions and is able to detect asymmetry in the interaction. However, its limitation is that the model must be appropriately matched to the underlying dynamics of the examined system, otherwise model misspecification may lead to spurious causalities [37]. Some of the proposed methods cannot detect indirect relationships, such as basic correlation, mutual information and Bayesian Networks. Some of the proposed methods mainly deal with linear problem, e.g., Pearson correlation and Spearman correlation but are not appropriate for nonlinear problems.

### 1.3. Primary Contribution of This Work

To address the issues mentioned above, we will propose an approach called small-shuffle symbolic transfer entropy spectrum (SSSTES). This work addresses five challenges:

- (1) Time series being non-stationary: the probabilities are estimated from observations of a single instance over a long time series. It is very important that the time series is statistically stationary over the period of interest, which can be a practical problem with transfer entropy calculations [38]. In most cases the time series from real industrial complex systems are non-stationary.
- (2) Time series being continuous: it is problematic to calculate the transfer entropy on continuous-valued time series. Thus, here we will resort to a solution.
- (3) Strong relationships identification: in general, when carrying out correlation analysis, we are more interested in strong correlations than weak correlations. Because the relationships among these variables are unknown, strong correlations are more convincing but weak correlations have

a greater probability of misidentification and this may bring serious consequences. We don't take the indirect correlation into account in the whole paper.

- (4) The direction and quantity of influence: the causality relation identification is crucial for network prediction and evolution. It is difficult to detect the directional influence that one variable exerts on another between two variables.
- (5) Temporal relation identification: we attempt to detect the specific temporal relation-based time lags, namely the function relation of time.

In the next section, we will propose a method of inferring association networks from multivariate time series. The emphasis is on how to solve the five challenges mentioned above. Section 3 will apply the proposed method to two numerical examples whose coupled relationships of their components are clear and the values are time-varying. We summarize the results of this paper and identify out some topics for further study in Section 4.

## 2. Methods

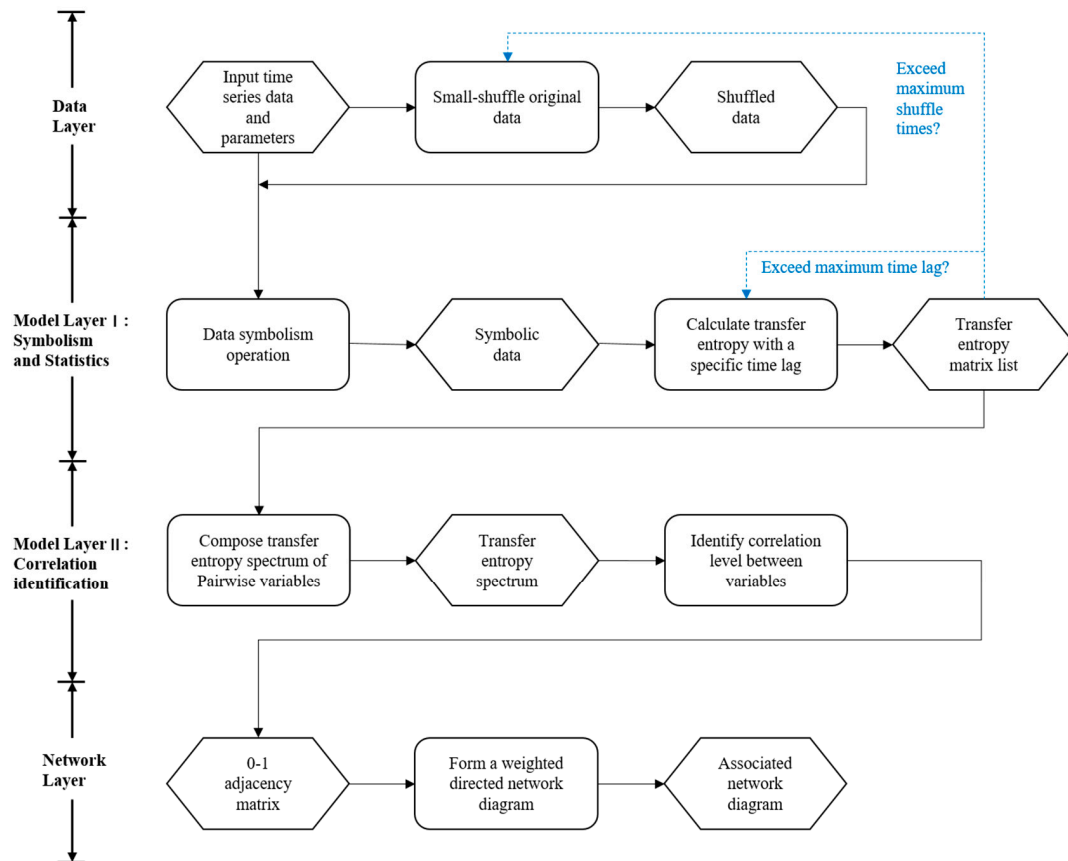
In this section, we will explain the proposed approach in detail. First, we will show you an integrated framework of the approach, and then carry out a detailed description around the framework.

### 2.1. Main Principle

The approach designed for association network inference takes exploration and application into account thus minimizing human intervention when modeling. Therefore, the approach starts with inputting data and ends with outputting a network inferred from multivariate time series. The modelling process is transparent for users. The main principle of the proposed approach is shown in Figure 1.

The integrated framework has four layers. The first layer, the so-called Data Layer, is the interface interaction with users. One thing to do in this layer is to input the original multivariate time series and modelling parameters, the other thing to do is to shuffle the original data several times with a surrogate data method. What to do next are data operation and modelling the approach. We can divide the complicated modelling work into two layers. The things to do in Model I Layer are time series symbolism and transfer entropy calculation. We will review some theories have been put forward and explain how to synthesize them for our purpose. The output of this layer is a complicated list of transfer entropy matrices. The thing to do in Model II Layer is strong correlation identification with an approach called transfer entropy spectrum. The output of this layer is a 0–1 matrix. In the last layer, we will get a weighted directed network. The start node of an arrowed edge represents a driven variable and the end node represents its corresponding variable. The weight of an edge quantifies the correlation between two nodes, i.e., time series variables.

As shown in Figure 1, there are six key processing operations, represented by rounded rectangles, to accomplish association networks inference. Thus, we will introduce the six steps one by one in the rest of this section.



**Figure 1.** Transfer entropy-based framework to infer association networks from multivariate time series. The black solid arrowed line in the flow diagram represents the determined sequential process and the blue dashed arrowed line, along with a Boolean condition, represents potential process. When the value of condition expression is false, the corresponding process will be carried out. Each rounded rectangle represents a key processing operations using a specific method and each hexagon represents a staged result.

### 2.2. Small-Shuffle Surrogate Data Method

The technique of surrogate data analysis is a randomization test method [39]. Given time series data, surrogate time series are constructed consistent with the original data and some null hypothesis. The random-shuffle surrogate (RSS) method proposed in [39] can test whether data can be fully described by independent and identically distributed random variables. As summarized in [38,39], the limitation of the RSS method is that it destroys any correlation structure in the data. That is, not only the short-term relationships but also the long-trend relationships between two variables are destroyed. The RSS method assumes global stationarity and performs a pairwise linear decoupling between channels, but in many typical examples the individual channels are also influenced by other nonstationary variations, so we prefer to use the small-shuffled surrogate (SSS) method proposed in [40–42].

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 scale. The steps using SSS method are described 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) Shuffle the index of  $x(t)$ :

$$i'(t) = i(t) + A \times g(t) \tag{1}$$

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:

$$s(t) = x[\hat{i}(t)] \quad (2)$$

Parameter  $A$  reflects the extent of shuffling data. A higher value of parameter  $A$  results in more difference between the surrogate data and the original data. On the contrary, the smaller the value of  $A$ , the less the difference. The parameter  $A$  is input at the beginning of the method and its empirical value is 1.0.

### 2.3. Time Series Symbolization

The time series symbolization technique was introduced with the concept of permutation entropy [43,44]. This technique has helped many other researches on time series achieve new progress and brought us some new techniques, e.g., permutation entropy [43] and symbolic transfer entropy (STE) [44]. It is helpful to deal with the problem of continuous and non-linear time series. The principle of time series symbolization is described as follows.

For original multivariate time series, let two time series  $V_1, V_2$ , be  $\{v_{1,t}\}, \{v_{2,t}\}$  respectively,  $t = 1, 2, \dots, k$ . The embedding parameters in order to form the reconstructed vector of the time series  $V_1$  are the embedding dimension  $m_1$  and the time delay  $\tau_1$ . Accordingly,  $m_2$  and  $\tau_2$  are the embedding parameters defined for  $V_2$ . The reconstructed vector of  $V_1$  is defined as:

$$v_{1,t} = (v_{1,t}, v_{1,t-\tau_1}, \dots, v_{1,t-(m_1-1)\tau_1})', \quad (3)$$

where  $t = 1, 2, \dots, k'$  and  $k' = k - \max((m_1 - 1)\tau_1, (m_2 - 1)\tau_2)$ .

For each vector  $v_{1,t}$ , the ranks of its components assign a rank-point  $\hat{v}_{1,t} = [r_{1,t}, r_{2,t}, L, r_{m_1,t}]$  where  $r_{j,t} \in \{1, 2, \dots, m_1\}$  for  $j = 1, 2, \dots, m_1$ .  $\hat{v}_{2,t}$  is defined accordingly.

### 2.4. Symbolic Transfer Entropy Calculation with Different Time Lags

Symbolic transfer entropy means that our transfer entropy calculation is based on the symbolic time series data presented in Section 2.3. Symbolic transfer entropy is defined as follows [44]:

$$STE_{v_2 \rightarrow v_1} = \sum p(\hat{v}_{1,t+\tau}, \hat{v}_{1,t}, \hat{v}_{2,t}) \log \frac{p(\hat{v}_{1,t+\tau} | \hat{v}_{1,t}, \hat{v}_{2,t})}{p(\hat{v}_{1,t+\tau} | \hat{v}_{1,t})}, \quad (4)$$

where  $\tau$  is the time delay,  $p(\hat{v}_{1,t+\tau}, \hat{v}_{1,t}, \hat{v}_{2,t})$ ,  $p(\hat{v}_{1,t+\tau} | \hat{v}_{1,t}, \hat{v}_{2,t})$  and  $p(\hat{v}_{1,t+\tau} | \hat{v}_{1,t})$  are the joint and conditional distributions estimated on the rank vectors as relative frequencies, respectively.

Symbolic transfer entropy uses a convenient rank transform to find an estimate of the transfer entropy on continuous data without the need for kernel density estimation. Since slow drifts do not have a direct effect on the ranks, it still works well for non-stationary time series [36]. Due to the fact the time delay is underdetermined, the symbolic transfer entropy is calculated  $n$  times for each pair of time series. This process is described using Algorithm 1 below.



---

**Algorithm 1: Symbolic Transfer Entropy Calculation with Different Time Lags**


---

**Input:** *STS*, symbolic time series

*tm*, maximum time delay

**Output:** *STEML*, a list of symbolic transfer entropy matrix

**Method:**

```

for (t = 1; t <= tm; t++) {
    colNum = column number of STS
    for (i = 1; i <= colNum; i++) {
        for (j = 1; j <= colNum; j++) {
            if (j ≠ i) {
                STS(j) = the column j of STS
                STS(i) = the column i of STS
                STE_matrix [i, j] = call_STE_Function(STS(j), STS(i), t)
            }
        }
    }
    Element t of STEML = STE_matrix
}

```

**Return** *STEML*

---

We first use Algorithm 1 to get a list of symbolic transfer entropy matrices on original time series. Then we shuffle the original data several times which has been specified at the beginning of our method. We repeat the Algorithm 1 on each shuffled data accordingly.

### 2.5. Symbolic Transfer Entropy Spectrum Composition

The Symbolic Transfer Entropy Spectrum (STES) is defined as follows.

The symbolic transfer entropy spectrum between time series *Y* and *X* is composed of their many symbolic transfer entropy curves drawn in a rectangular coordinate system. The horizontal axis represents different time delays and the vertical axis represents transfer entropy. One of the transfer entropy curves comes from the original data and other curves come from shuffled data.

Let  $L_{Y \rightarrow X}^o$  be the transfer entropy curve of the original data, and  $L_{Y \rightarrow X}^s$  be the transfer entropy curve of the shuffled data, then the symbolic transfer entropy spectrum between *Y* and *X* can be denoted as follows:

$$\text{STES}_{Y \rightarrow X} = \{L_{Y \rightarrow X}^o, \{L_{Y \rightarrow X}^s\}\} \quad (5)$$

In order to form the transfer entropy spectrum, we must understand the structure of the output in Section 2.4. The output is a complicated list of transfer entropy matrices. For each data point, original data or shuffled data, a list of transfer entropy matrices with different delays is returned after applying Algorithm 1. Thus, for all data, the returned result of the last step is a list of transfer entropy matrix lists. The parameters input at the beginning of the method are maximum time delay *tm* and shuffling times *sm*. Let *tm* = 10, *sm* = 99, then the output of the last step is a list of 100 elements and each element is a list of 10 transfer entropy matrices. Moreover, each entry of the transfer entropy matrix reflects the correlation strength of a pair of time series. Thus, according to the definition of transfer entropy spectrum, we first split the output of Section 2.5 into pieces and then form the transfer entropy spectrum in a certain way.

### 2.6. Strong Correlation Identification

The target of the proposed method in this paper is strong correlation identification and is not all correlations among multivariate time series. The scenario for this method is that we don't know the

relationships in the complex system. We pay more attention to the precision of correlation identification but not the recall, because the misidentification of relationships among variables may have serious consequences for our data analysis.

Our decision whether a strong correlation exists or not between two variables is made by the characteristic of the transfer entropy spectrum. This characteristic is based on the theory of hypothesis testing which is often used in surrogate data methods [31,36,39,42]. Discriminating statistics are necessary for surrogate data hypothesis testing. The cross correlation and average mutual information were selected as discriminating statistics in [41,42], and partial symbolic transfer entropy in [36]. In this paper, we consider transfer entropy as the discriminating statistics. The surrogate data method also needs a null hypothesis. Applying a statistical hypothesis test can result in two outcomes, i.e., the null hypothesis is rejected or not. There are two type of errors when using the hypothesis testing. If the null hypothesis is rejected and it is true, this is called type I error; if we fail to reject the null hypothesis when it is in fact false, this is called type II error. The null hypothesis in our proposed method is that there is no short-term correlation structure between the data or that the irregular fluctuations are independent. In the symbolic transfer entropy spectrum, if the symbolic transfer entropy of the original data falls outside the distribution of the SSS data and there exists an outlier point which value is greater than any other points' value, we can reject the null hypothesis. As a result, we consider that there is a short-term correlation structure between the data and this correlation is a strong correlation. Otherwise, we accept the null hypothesis and consider that there is not a strong correlation between the data. The output of this step is an adjacency matrix and its entry  $a_{ij}$  is denoted as follows:

$$a_{ij} = \begin{cases} 1, \exists STE_{i \rightarrow j}^o(t) > \max(STE_{i \rightarrow j}^s(t)) \\ 0, \text{others} \end{cases} \quad (6)$$

where  $t \in (1, 2, \dots, tm)$ ,  $s \in (1, 2, \dots, sm)$ ,  $STE_{i \rightarrow j}^o(t)$  is the symbolic transfer entropy from variable  $i$  to variable  $j$  with a time delay  $t$  and  $STE_{i \rightarrow j}^s$  is the symbolic transfer entropy with all different time delays from variable  $i$  to variable  $j$ .

### 2.7. Association Network Inference

The association network inferred from multivariate time series can be denoted as  $G = (V, E)$ . Here  $V = \{v_1, v_2, \dots, v_n\}$  is the set of vertices, i.e., time series variables, and  $E$  is the set of edges, i.e., the strong correlations, identified in the Section 2.6, between each pair of vertices in  $V$ .

From the 0–1 adjacency matrix from the last step, we have determined the direction of the network. In this step, we assign a weight to the edges in  $E$ . The selected measure for the weight is the corresponding maximum symbolic transfer entropy of original data calculated in Section 2.4 and the Equation (6) is transformed as follows:

$$a_{ij}' = \begin{cases} \max(STE_{i \rightarrow j}^o(t)), & a_{ij} = 1 \\ 0, & a_{ij} = 0 \end{cases} \quad (7)$$

where  $i$  is the driven variable, and  $j$  is the response variable. Finally, we can plot the association network based on the weighted adjacency matrix denoted as Equation (7) and carry out deep network analysis.

## 3. Results

In this section, we demonstrate the application of the proposed method to simulated time series data from two types of complex system, i.e., a linear system and a nonlinear system. The relationships among the variables in these two examples are clear and therefore we can assess our method by some measures.

In all the following cases, the parameters for modelling with SSSTES method are shuffling amplitude  $A = 1.0$ , the dimension of the symbolic time series  $m = 2$ , maximum time delay  $tm = 10$ ,



maximum shuffling times  $sm = 99$ , time points  $t = 1, 2, \dots, 1000$ . These parameters are input in the Data Layer shown in Figure 1.

### 3.1. Numerical Example from Linear System

First, we apply our method to a linear system which has four time series variables, i.e.,  $x_1(t)$ ,  $x_2(t)$ ,  $x_3(t)$ ,  $x_4(t)$ . The relationships among these variables are modelled by the following expressions [42]:

$$x_1(t) = 1.3 + 0.4x_1(t-1) - 0.2x_1(t-3) + 0.3x_2(t-4) + 0.2x_4(t-7) + r_1(t), \quad (8)$$

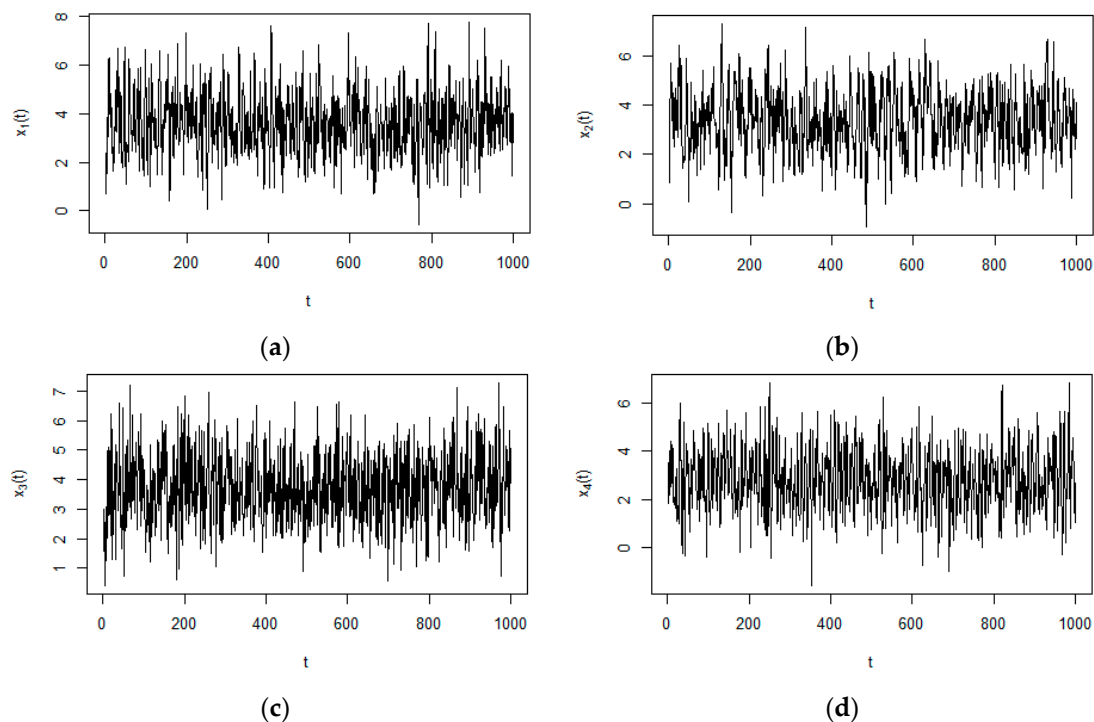
$$x_2(t) = 2.0 + 0.6x_2(t-1) - 0.2x_2(t-6) + r_2(t), \quad (9)$$

$$x_3(t) = 2.2 + 0.2x_1(t-2) + 0.3x_4(t-9) + r_3(t), \quad (10)$$

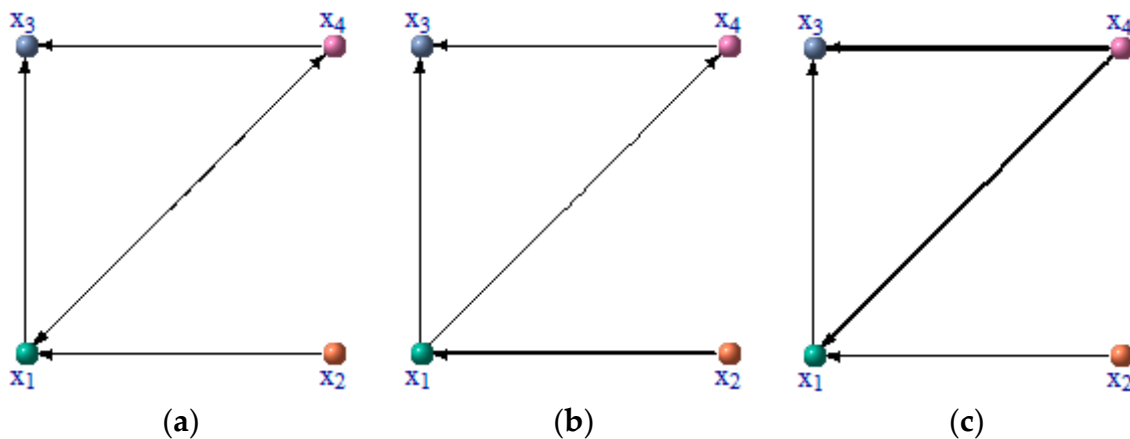
$$x_4(t) = 1.5 + 0.2x_1(t-2) + 0.5x_4(t-1) - 0.3x_4(t-3) + r_4(t), \quad (11)$$

where  $r_i(t)$  ( $i = 1, 2, 3, 4$ ) are random noise, independent and identically distributed Gaussian random variables with mean zero and standard deviation 1.0. These four time series are shown in Figure 2.

It is difficult for us to find the relationships among the four time series variables from Figure 2. Their fluctuations seem to be irregular and not have any obvious trend, but they have linear relationships in reality. If the variable  $y$  is a linear combination of variables  $x_1, x_2, \dots, x_n$ , we say  $y$  is a response variable and  $x_1, x_2, \dots, x_n$  are the drive variables. In the network, we denote the drive-response relationship between  $y$  and  $x_1$  as a arrowed edge from  $x_1$  to  $y$ . Therefore, the corresponding network of the above linear system is shown in Figure 3a.



**Figure 2.** Linear multivariate time series generated by Equations (8)–(11). (a) Time series of  $x_1$ ; (b) Time series of  $x_2$ ; (c) Time series of  $x_3$ ; (d) Time series of  $x_4$ .



**Figure 3.** The original network and inferred networks. (a) the original association network constructed from Equations (8)–(11); (b) the inferred association network in Section 3.1; (c) the inferred association network in Section 3.2.

As shown in Figure 3a, variable  $x_1$  is driven by two other variables  $x_2, x_4$ , variable  $x_3$  is driven by variables  $x_1, x_4$  and  $x_4$  is driven by  $x_1$ . However,  $x_2$  is not driven by any other variables and it is only a driven variable of  $x_1$ . It is noted that there are autocorrelations in Equations (8)–(11) but we do not show the autocorrelations in Figure 3a. In this paper, we focus on the relationships among different variables but are not concerned with the autocorrelations.

After generating the simulated data by Equations (8)–(11) in the Data Layer shown in Figure 1, what we should do is modelling with the proposed method SSSTES. This process has been described in detail in Sections 2.3–2.6. The shuffled data used in the modelling process is generated with the method described in Section 2.2. One output of the Model Layer is the symbolic transfer entropy spectra shown in Figure 4. Since the STE values are rather small, they are multiplied by 100 for ease of plotting. There are twelve pairs of relationships among four time series and they are all shown in Figure 4. Per line we have two pairs of relationships. One is the relationship from time series  $x$  to  $y$  and the other is the reverse relationship. For example, the left plot in first line is the symbolic transfer entropy spectrum from time series  $x_1$  to time series  $x_2$  and the right one is the symbolic transfer entropy spectrum from time series  $x_2$  to time series  $x_1$ . Other symbolic transfer entropy spectra are all similarly shown in Figure 4.

Then, we need to identify the strong relationships in Figure 4. The method to identify the strong relationship between a pair of time series is described in Section 2.6. With this method we identify the strong relationships  $v_{21}, v_{13}, v_{14}, v_{43}$ . As a result, we conclude that time series  $x_1$  is influenced by time series  $x_2$  and this conclusion is consistent with Equation (8). By contrast with Equations (8)–(11), we find that other three identified strong relationships  $v_{13}, v_{14}, v_{43}$  are also correct.

The other output of the Model Layer is a 0–1 adjacency matrix, corresponding to the symbolic transfer entropy spectrum. For instance, the resulted adjacency matrix from Figure 4 is described by Equation (12):

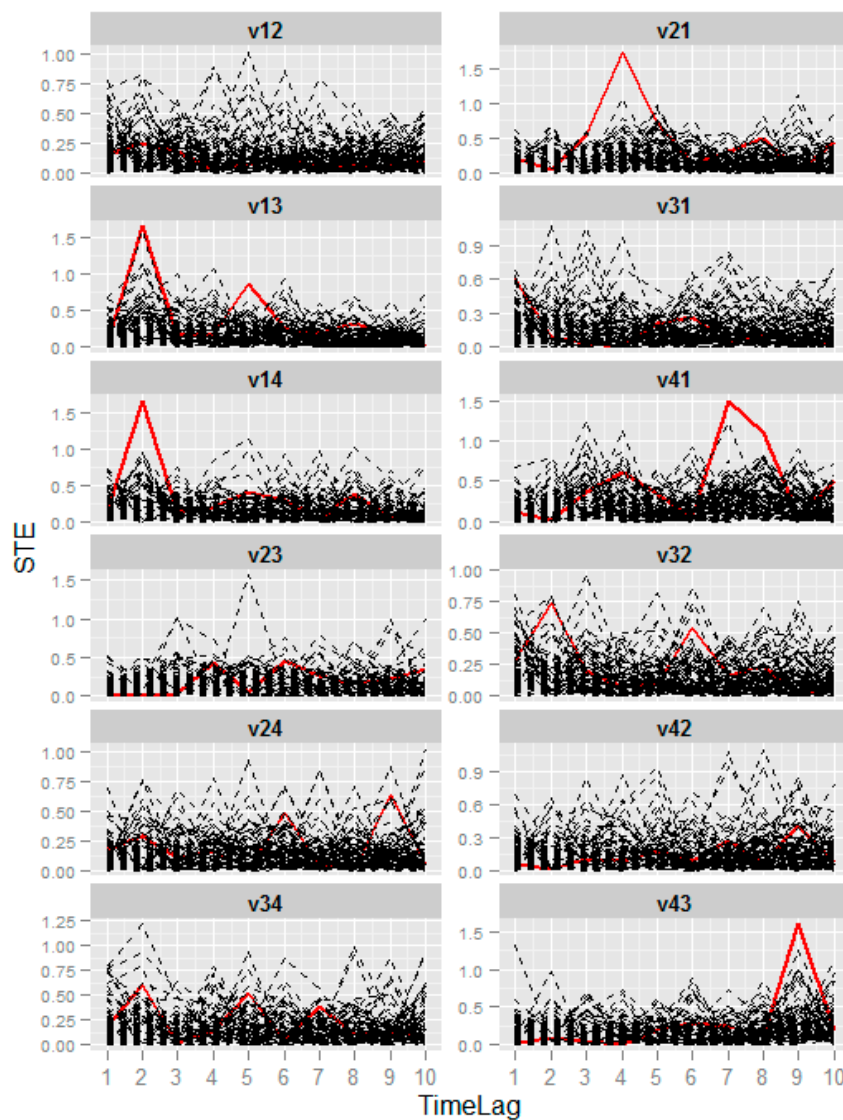
$$C = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad (12)$$

Finally, we infer a weighted directed association network in the last layer. From Equation (12), we can get a directed network and then we should quantify the correlation strength between those pairs of relationships that have been identified out above. Therefore, we introduce a correlation measure into adjacency matrix  $C$  and get a new weighted adjacency matrix  $C'$ , whose entries is described as

Equation (7). The selected measure is the maximum symbolic transfer entropy with different time lags of original data. Then, we get the weighted adjacency matrix  $C'$  as follows:

$$C' = \begin{pmatrix} 0 & 0 & 1.66 & 1.65 \\ 1.74 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1.60 & 0 \end{pmatrix} \tag{13}$$

The association network corresponding to the matrix  $C'$  is shown in Figure 3b. In Figure 3b, each time series is mapped as a node, and each arrowed edge stands for a drive-response relationship, and we associate each edge with a weight value, i.e., the max symbolic transfer entropy value, which is mapped as the width of the lines. As we see, the relationship from  $x_4$  to  $x_3$  is the strongest one. In Figure 3, the original network (a) has five directed edges and the inferred network; (b) has four edges. By comparison, we find that the four edges of inferred network all exist in the original network, thus we get a higher precision.



**Figure 4.** Symbolic transfer entropy spectrums of linear system. Plot  $v_{12}$  is the symbolic transfer entropy spectrums between time series 1 and 2. Plots  $v_{21}, v_{13}, v_{31}, v_{14}, v_{41}, v_{23}, v_{32}, v_{24}, v_{42}, v_{34}, v_{43}$  are the symbolic transfer entropy spectrums between other time series pairs respectively.

Next, we assess the correlation identification of the proposed method using two indicators, i.e., precision and sensitivity (or recall, true positive rate) [45,46]. Precision is defined as Equation (14) and sensitivity is defined as Equation (15):

$$P = \frac{TP}{TP + FP'} \quad (14)$$

$$S = \frac{TP}{TP + FN}. \quad (15)$$

Here,  $TP$  is the numbers of edges which are in the intersections between the original edge set and the inferred edge set,  $FP'$  is the number of edges which are in the inferred edge set but not in the original edge set and  $FN$  is the number of edges which are not in the inferred edge set but are in the original edge set. In order to test whether the model is sensitive to the system noise, we generate ten groups of data using Equations (8)–(11) and then apply the proposed method to these data. As a result, we get ten precision values and sensitivity values and their average values shown in Table 1. From Table 1, the average precision of our model is higher to 0.92 and the average sensitivity achieve to 0.74 although it is inferior to precision.

**Table 1.** The model assessment on 10 groups of data generated by Equations (8)–(11). The values of Precision, Sensitivity and precision of time lags ( $PTL$ ) in the table are rounded to two decimals.

ID	Precision	Sensitivity	$PTL$
1	1.00	0.80	1.00
2	1.00	0.60	1.00
3	1.00	0.60	1.00
4	0.80	0.80	0.75
5	0.60	0.60	1.00
6	1.00	0.80	1.00
7	1.00	0.80	1.00
8	0.80	0.80	1.00
9	1.00	0.80	1.00
10	1.00	0.80	1.00
Average	0.92	0.74	0.98

Next, we discuss the temporal relation identification of the proposed method. Please note that the following discussion is based on those edges inferred correctly. The time lag we assign to two correlation variables is the time point when STE of original data achieve the maximum value. Based on this definition, we define a measure, i.e., the precision of time lags ( $PTL$ ), to assess the temporal relation identification of the proposed method. It is defined as Equation (16):

$$PTL = \frac{TPL}{TPL + FPL} \quad (16)$$

Here,  $TPL$  is the correct number of temporal relation identification in those edges which have been identified correctly,  $FPL$  is the error number of temporal relation identification which have been identified correctly. The results of  $PTL$  are shown in Table 1. We get a higher  $PTL$  of 0.98.

It is proved that the proposed method is good at inferring association networks from linear multivariate time series. In Section 3.2, we will validate our method using a nonlinear system.

### 3.2. Numerical Example from Nonlinear System

In this section, we validate whether the proposed method work well for a nonlinear system. The simulated data is generated by Equations (17)–(20) [42]:

$$x_1(t) = 1.3 + 0.2x_1(t-1) - 0.1x_1(t-3) + 0.1x_2(t-4) x_4(t-7) + r_1(t), \quad (17)$$

$$x_2(t) = 2.0 + 0.6x_2(t-1) - 0.2x_2(t-6) + r_2(t), \quad (18)$$

$$x_3(t) = h[2.2 + 0.2x_1(t-2) + 0.3x_4(t-9) + r_3(t)], \quad (19)$$

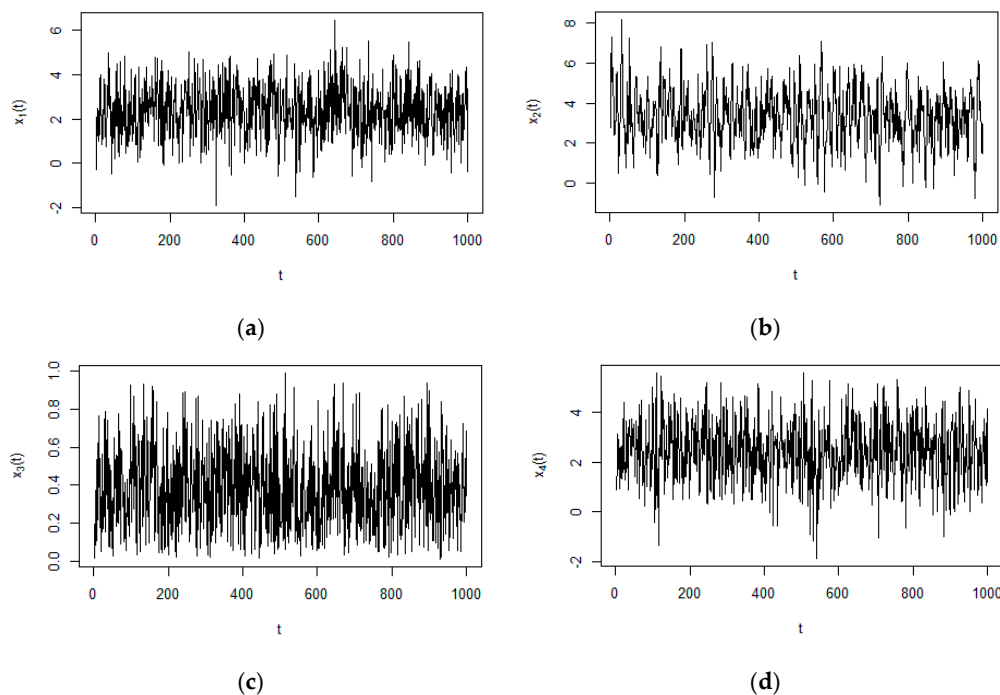
$$x_4(t) = 1.5 + 0.2x_1(t-2) + 0.5x_4(t-1) - 0.3x_4(t-3) + r_4(t). \quad (20)$$

Here,  $r_i(t)$  ( $i = 1, 2, 3, 4$ ) are random noise, independent and identically distributed Gaussian random variables with mean zero and standard deviation 1.0. Two changes are made to get a nonlinear system compared with Equations (8)–(11). One change is made with  $x_1(t)$ . In Equation (17), there is a product  $x_2(t-4) x_4(t-7)$ , which is a nonlinear function. In addition, in Equation (20), there is a product  $0.2x_1(t-2)$  and this makes  $x_4(t)$  also nonlinear. The other change can be found in Equation (19). The function  $h(\cdot)$  is a static monotonic nonlinear function and it is denoted as follows [47]:

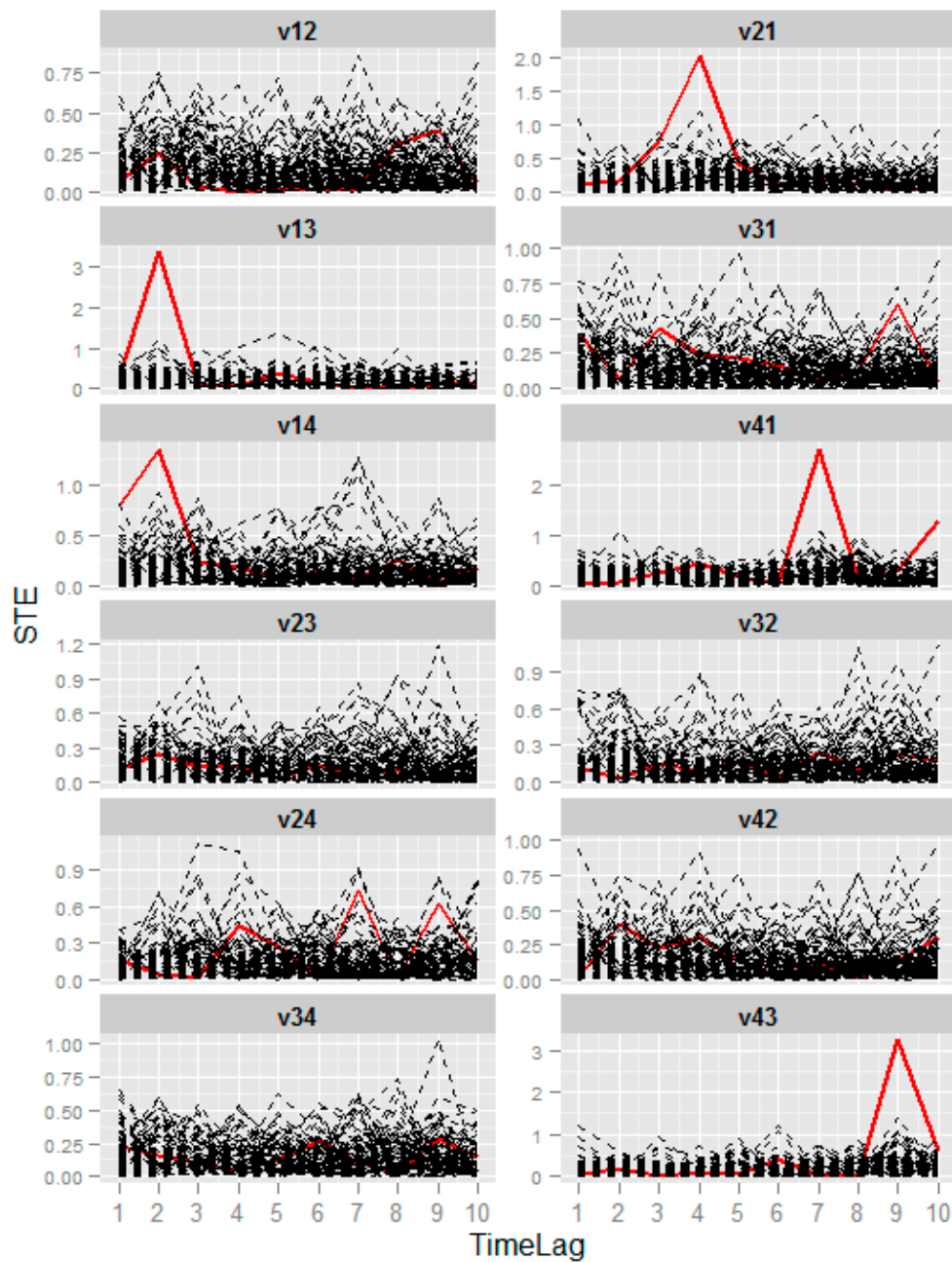
$$h(x) = \frac{\left[\frac{x-a-0.0001}{b-x+0.0001}\right]^\rho}{1 + \left[\frac{x-a-0.0001}{b-x+0.0001}\right]^\rho} \quad (21)$$

Here,  $\rho = 3$ ,  $a = -2$ ,  $b = 10$ . The time series generated by Equations (17)–(20) are shown in Figure 5. The original network of this nonlinear system is the same as is Figure 3a. We apply the proposed method to this nonlinear system and the process is the same as that described in Section 3.1. The resulting symbolic transfer entropy spectrum is shown in Figure 6. From this figure, we conclude that time series  $x_1$  is influenced by time series  $x_2$  and  $x_4$ . Time series  $x_3$  is influenced by time series  $x_1$  and  $x_4$ . These identified relationships are correct by contrast to Equations (17)–(20). We can denote the output of symbolic transfer entropy spectrum as an adjacency matrix Equation (22).

$$C = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix} \quad (22)$$



**Figure 5.** Nonlinear multivariate time series generated by Equations (17)–(20). (a) Time series of  $x_1$ ; (b) Time series of  $x_2$ ; (c) Time series of  $x_3$ ; (d) Time series of  $x_4$ .



**Figure 6.** Symbolic transfer entropy spectra of a nonlinear system. Plot  $v_{12}$  is the symbolic transfer entropy spectrums between time series 1 and 2. Plots  $v_{21}, v_{13}, v_{31}, v_{14}, v_{41}, v_{23}, v_{32}, v_{24}, v_{42}, v_{34}, v_{43}$  are the symbolic transfer entropy spectrums between other time series pairs, respectively.

This is a 0–1 adjacency matrix. We aim to get a weighted directed network, so we assign a weight to each edge following the method described in Section 2.7. Then, we get the weighted adjacency matrix which is denoted as Equation (23):

$$C' = \begin{pmatrix} 0 & 0 & 3.36 & 0 \\ 2.03 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 2.72 & 0 & 3.23 & 0 \end{pmatrix} \quad (23)$$



From this matrix, we get the association network which is shown in Figure 3c. The inferred network has four edges and they are all contained in the original network which is shown in Figure 3a. Therefore, we consider that the proposed method works well for nonlinear systems.

To end of this section, we also assess the performance of the proposed method when it is applied to the nonlinear system. The indicators are still precision, sensitivity [45,46] and *PTL*, described in Section 3.1. The results measured on ten groups of data are shown in Table 2. From Table 2, we see that the average precision of our model is higher to 0.90, the average sensitivity achieved is 0.72 and the precision of time lag identification is 0.98.

**Table 2.** The model assessment on 10 groups of data generated by Equations (17)–(20). The values of precision, sensitivity and *PTL* in the table are rounded to two decimals.

ID	Precision	Sensitivity	<i>PTL</i>
1	0.75	0.60	1.00
2	1.00	0.80	1.00
3	0.80	0.80	1.00
4	1.00	0.80	1.00
5	1.00	0.80	1.00
6	1.00	0.40	1.00
7	1.00	0.80	1.00
8	0.67	0.80	0.75
9	0.80	0.80	1.00
10	1.00	0.60	1.00
Average	0.90	0.72	0.98

### 3.3. Numerical Example from the Coupled Rossler Systems

The linear system and nonlinear system, illustrated in Sections 3.1 and 3.2, are common systems, in which the relationships among time series variables are determined. In this section, we validate whether the proposed method work well for the coupled chaotic system, in which the relationships between components are uncertain and complicated. The coupled Rossler systems are two symmetrically coupled chaotic systems. The equations are given as follows:

$$\dot{x}_1 = -(y_1 + z_1), \quad (24)$$

$$\dot{y}_1 = x_1 + a_1 * y_1 + c_1 * (y_2 - y_1), \quad (25)$$

$$\dot{z}_1 = b_1 + z_1 * (x_1 - w), \quad (26)$$

$$\dot{x}_2 = -(y_2 + z_2), \quad (27)$$

$$\dot{y}_2 = x_2 + a_2 * y_2 + c_2 * (y_1 - y_2), \quad (28)$$

$$\dot{z}_2 = b_2 + z_2 * (x_2 - w), \quad (29)$$

where the parameters are  $a_1 = 0.2$ ,  $a_2 = 0.4$ ,  $b_1 = b_2 = 0.2$ ,  $w_1 = w_2 = 5.7$ ,  $c_1 = c_2 = 0.2$  and  $c_1, c_2$  are coupling control parameters. In this coupled chaotic system, we generate data using the fourth order Runge-Kutta method with the sampling interval 0.01.

This coupled Rossler system has six components and their time series are shown in Figure 7. The relationships among these six time series are shown in Figure 8a. From Equation (24), we know  $x_1$  is driven by  $y_1$  and  $z_1$ . From Equation (25), we know  $y_1$  is driven by  $y_2$ . From Equation (26), we know  $z_1$  is driven by  $x_1$ . In the same way, we can get the other relationships in Figure 8a from Equations (27)–(29). It is emphasized that we don't take the indirect relationships into account, not only for this example, but also for the other examples. After applying the proposed method to the coupled Rossler system, we get the inferred association network shown in Figure 8b. There are 10 edges

in Figure 8a and there are eight edges in Figure 8b. By comparison, the precision is 0.875 and the sensitivity is 0.7 according to their definitions in Section 3.1.

Next, we also assess the performance of the proposed method when it is applied to the coupled Rossler system. For the time series shown in Figure 7, the experiments are carried out ten times. By comparison and calculation, ten groups of results are obtained, shown in Table 3. The assessment indicators are precision and sensitivity. From Table 3, we see that the average precision of our model is higher to 0.84 and the average sensitivity is 0.62. It is thus proved that our model method is also effective for chaotic time series.

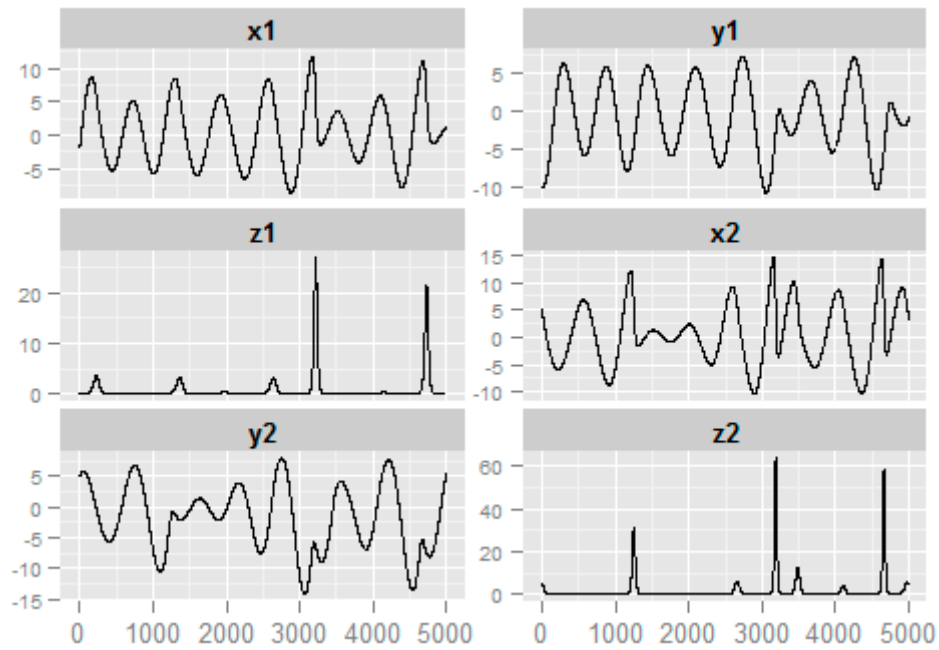


Figure 7. The multivariate time series in the coupled Rossler system.

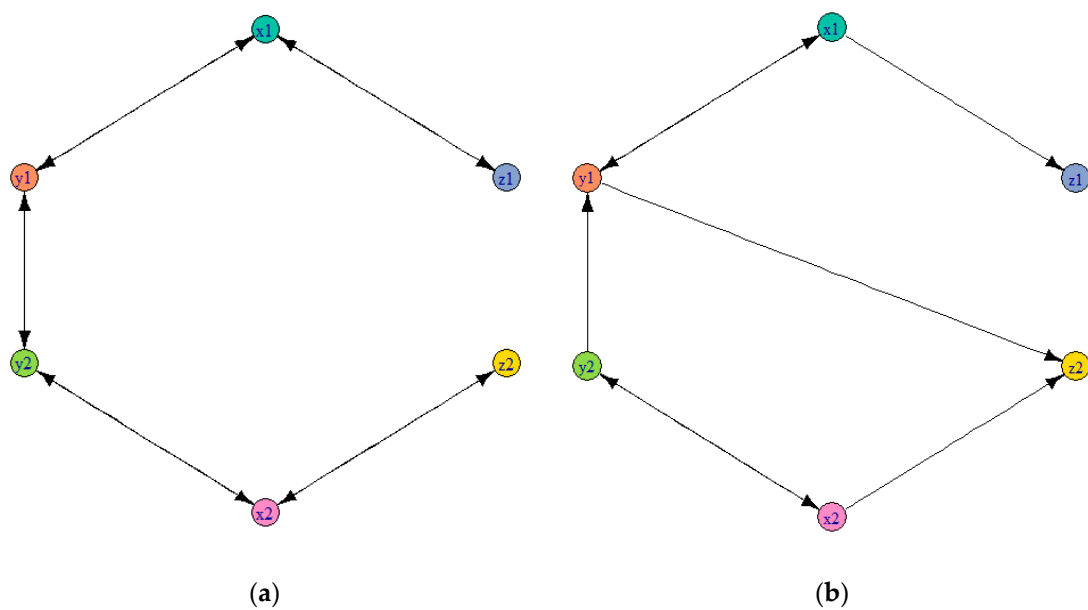


Figure 8. The original network and inferred network from the coupled Rossler systems. (a) The original association network constructed from Equations (24)–(29); (b) the inferred association network.

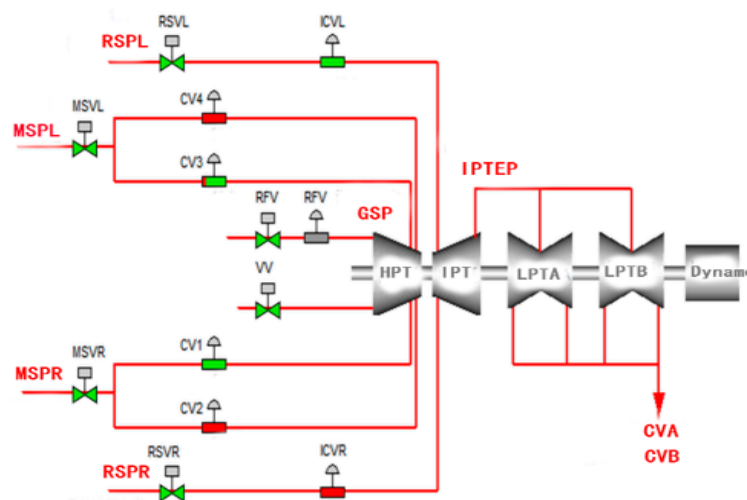
**Table 3.** The model assessment on coupled chaotic data generated by Equations (24)–(29). The values of precision and sensitivity in the table are rounded to two decimals.

ID	Precision	Sensitivity
1	1.00	0.60
2	0.75	0.60
3	0.75	0.60
4	0.86	0.60
5	0.75	0.60
6	0.88	0.70
7	0.75	0.60
8	0.88	0.70
9	1.00	0.60
10	0.75	0.60
Average	0.84	0.62

### 3.4. Application

In this section, we apply the proposed method to a real industrial system which is an important part in the whole power generation system, i.e., the steam turbine system. The basic principle of a steam turbine system is shown in Figure 9. The eight monitored parameters are main steam pressure left (MSPL), main steam pressure right (MSPR), reheat steam pressure left (RSPL), reheat steam pressure right (RSPR), governing stage pressure (GSP), IP turb exh pressure (IPTEP), condenser vacuum A (CVA) and condenser vacuum B (CVB), respectively. The font color of these parameters is red. In addition, the main components of turbine are also shown in Figure 9, i.e., high pressure turb (HPT), intermediate pressure turb (IPT), low pressure turb A (LPTA), low pressure turb B (LPTB) and dynamo. The other symbols are switches or valves.

The data set was obtained from the distributed control system (DCS) of a power plant belonging to China Huadian Corporation, which is one of the five state-owned sole proprietorship power generation corporations. We use the eight monitoring parameters mentioned above and 5000 records for analysis. The time series of these monitored parameters are shown in Figure 10. Figure 11 shows the original network and inferred network from the time series. The original network is drawn according to the work principle of the steam turbine shown in Figure 9.



**Figure 9.** The work principle of steam turbine system. MSPL, MSPR, GSP, RSPL, RSPR, IPTEP, CVA, CVB are the monitored parameters for analysis.

Based on the experiments from simulated numerical examples in Sections 3.1–3.3, we apply the proposed method to the multivariate time series from the DCS of the power plant. The inferred association network is shown in Figure 11b. In comparison with Figure 11a, we can see that there are nine edges in the inferred network and six edges exist in both the original network and inferred network. Therefore, the precision is 0.67 and the sensitivity is 0.75. The result in application is not as good as the result obtained in Sections 3.1 and 3.2. The error relationships are MSPL-IPTEP, MSPR-IPTEP and GSP-IPTEP respectively. However, this doesn't mean that there are no correlations between them. Indeed, we can conclude from the original network that these three pairs of correlations are all right. The reason why we consider them as error relationships is that they are all indirect correlations, but we only consider the direct correlations as correct. Therefore, although the precision is a bit low, it still performs well in this application.

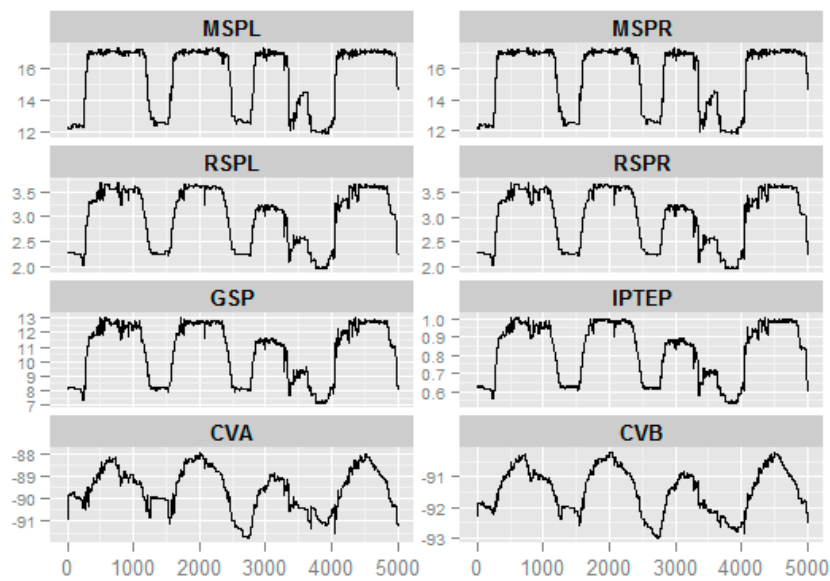


Figure 10. The time series of eight monitored parameters from the power system DCS.

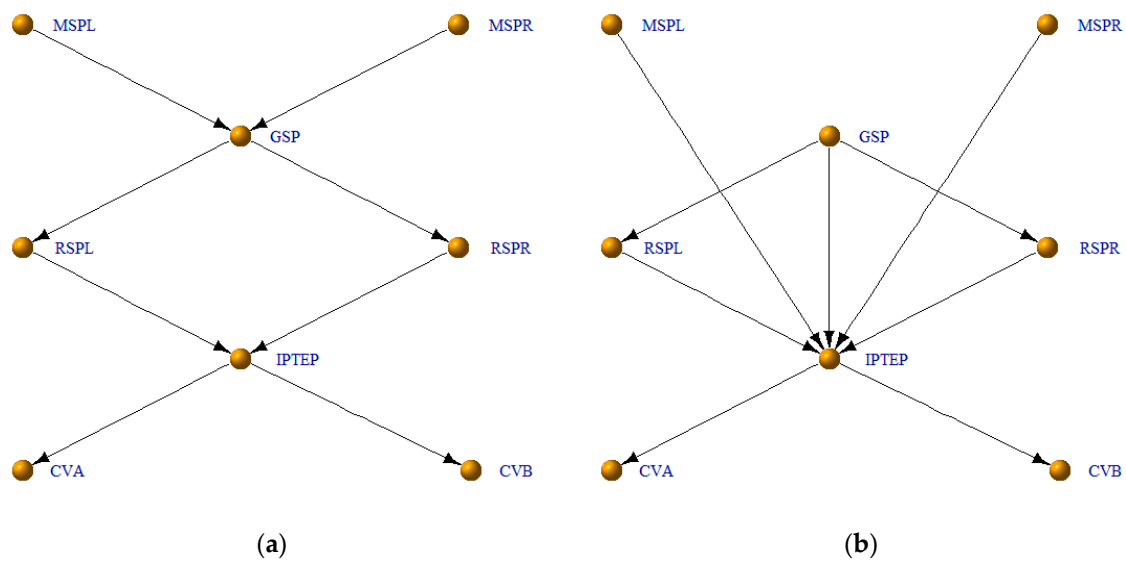


Figure 11. The original network and inferred network from the DCS time series. (a) The original association network constructed from the DCS time series; (b) The inferred association network from the DCS time series.

Next, we will make a comparison between SSSTES and two other methods, i.e., random-shuffle symbolic transfer entropy (RSSTE) [36] and Granger Causality (GC) [24,48]. We apply these three methods to the power data set, respectively, and get the results shown in Table 4. For the two surrogate data methods, i.e., SSSTES and RSSTE, the shuffling times is 99 and the values of precision and sensitivity are the average values. The average precision of our proposed SSSTES method is 0.67 and the average sensitivity is 0.75. The average precision of RSSTE is 0.27 and the average sensitivity is 0.38. The precision of GC is 0.17 and the sensitivity is 1.00. By comparison, it is proved that our proposed method is better than the other two methods in this application. Although the sensitivity of GC is the highest, the cost is the lowest precision. It means that GC identifies all the correlations but causes a high rate of error identification. However, SSSTES is a moderate method, which maintains a high precision and meanwhile the sensitivity in an acceptable range.

**Table 4.** The algorithm assessment of three different methods. The values of precision and sensitivity in the table are rounded to two decimals.

ID	Method	Precision	Sensitivity
1	SSSTES	0.67	0.75
2	RSSTE	0.27	0.38
3	GC	0.17	1.00

#### 4. Conclusions

In order to infer a weighted directed association network from multivariate time series, we have proposed a method named Small-Shuffle Symbolic Transfer Entropy Spectrum (SSSTES) which synthesizes the Symbolic Transfer Entropy (STE) and Small-Shuffle Surrogate (SSS) methods. We first proposed the framework of the method. It is composed of four layers, i.e., Data Layer, Network Layer and two Model Layers. Then we described the six main process of SSSTES from Section 2.2 to Section 2.7. The most complicated and important processes are STES composition and strong correlation identification. After the method description, we applied the proposed method to numerically simulate linear, nonlinear and coupled chaotic systems. We used three indicators, i.e., precision, sensitivity and PTL, to assess the proposed method. As a result, the method shows good performance in the linear system, nonlinear system, and coupled chaotic system. Finally, we applied the proposed method to an empirical investigation on a stream turbine system in a power plant. As a comparison, we also used another two methods, i.e., RSSTE and GC, which can deal with directional correlation. The results show that our proposed method is better than the other two methods. In general, the method performs well on both simulated numerical examples and a real complex system. It can not only identify the strong correlations, but also find out the time delay between pairwise time series when the temporal correlation is determined.

Although it is illustrated that the proposed method is good at inferring association networks from multivariate time series, there are still some topics that are worth studying in the future. First, in this paper, it is considered that the misidentification of relationships may have serious consequences, thus we aim to achieve a strong correlation identification and ignore the proportion of identified relationships among all relationships existing in the complex system. Therefore, we will attempt to improve the sensitivity of SSSTES. Second, although the method deals with many issues, such as correlation identification, correlation quantification, direction identification, indirect correlation and temporal relation identification, some of their abilities are not good enough and influence the result, e.g., the sensitivity in Section 3.3 and the precision in Section 3.4. We should make further efforts on these topics. Nevertheless, the proposed method still can serve as a heuristic tool for inferring association networks from multivariate time series so as studying the system deeply with complex network knowledge.

**Acknowledgments:** We thank all challenge participants for their invaluable contribution. We thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper. This paper is supported by the National Natural Science Foundation of China (61503034), by the Beijing Science and Technology Plan (D151100004715001), by the Key Scientific Research Project of Henan Province Universities (15B520031) and by the Xuchang Science and Technology Plan (1502098).

**Author Contributions:** Yanzhu Hu provided the thought of the method; Huiyang Zhao designed the algorithm and performed the experiments; Xinbo Ai wrote the paper. All authors have read and approved the final manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- De Solla Price, D.J. Networks of scientific papers. *Science* **1965**, *149*, 510–515. [[CrossRef](#)]
- Evans, T.S.; Hopkins, N.; Kaube, B.S. Universality of performance indicators based on citation and reference counts. *Scientometrics* **2012**, *93*, 473–495. [[CrossRef](#)]
- Goldberg, S.R.; Anthony, H.; Evans, T.S. Modelling citation networks. *Scientometrics* **2015**, *105*, 1577–1604. [[CrossRef](#)]
- Watts, D.J.; Strogatz, S.H. Collective dynamics of ‘small-world’ networks. *Nature* **1998**, *393*, 440–442. [[CrossRef](#)] [[PubMed](#)]
- Barabási, A.-L.; Albert, R. Emergence of scaling in random networks. *Science* **1999**, *286*, 509–512. [[PubMed](#)]
- Fernández-Rosales, I.Y.; Liebovitch, L.S.; Guzmán-Vargas, L. The dynamic consequences of cooperation and competition in small-world networks. *PLoS ONE* **2015**, *10*, e0126234. [[CrossRef](#)] [[PubMed](#)]
- Guo, X.; Zhang, Y.; Hu, W.; Tan, H.; Wang, X. Inferring nonlinear gene regulatory networks from gene expression data based on distance correlation. *PLoS ONE* **2014**, *9*, e87446. [[CrossRef](#)] [[PubMed](#)]
- Yuan, Y.; Li, C.-T.; Windram, O. Directed partial correlation: Inferring large-scale gene regulatory network through induced topology disruptions. *PLoS ONE* **2011**, *6*, e16835. [[CrossRef](#)] [[PubMed](#)]
- Cecchi, G.A.; Garg, R.; Rao, A.R. Inferring brain dynamics using granger causality on fmri data. In Proceedings of the 5th IEEE International Symposium on Biomedical Imaging, Paris, France, 14–17 May 2008. [[CrossRef](#)]
- Deng, L.; Sun, J.; Cheng, L.; Tong, S. Characterizing dynamic local functional connectivity in the human brain. *Sci. Rep.* **2016**, *6*, 26976. [[CrossRef](#)] [[PubMed](#)]
- Maucher, M.; Kracher, B.; Kühl, M.; Kestler, H.A. Inferring Boolean network structure via correlation. *Bioinformatics* **2011**, *27*, 1529–1536. [[CrossRef](#)] [[PubMed](#)]
- Han, L.; Zhu, J. Using matrix of thresholding partial correlation coefficients to infer regulatory network. *Biosystems* **2008**, *91*, 158–165. [[CrossRef](#)] [[PubMed](#)]
- Wang, Y.X.R.; Jiang, K.; Feldman, L.J.; Bickel, P.J.; Huang, H. Inferring gene-gene interactions and functional modules using sparse canonical correlation analysis. *Ann. Appl. Stat.* **2014**, *9*, 300–323. [[CrossRef](#)]
- Schäfer, J.; Strimmer, K. An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics* **2005**, *21*, 754–764. [[CrossRef](#)] [[PubMed](#)]
- Huynh-Thu, V.A.; Irrthum, A.; Wehenkel, L.; Geurts, P. Inferring regulatory networks from expression data using tree-based methods. *PLoS ONE* **2010**, *5*, e12776. [[CrossRef](#)] [[PubMed](#)]
- Margolin, A.A.; Wang, K.; Lim, W.K.; Kustagi, M.; Nemenman, I.; Califano, A. Reverse engineering cellular networks. *Nat. Protoc.* **2006**, *1*, 662–671. [[CrossRef](#)] [[PubMed](#)]
- Faith, J.J.; Hayete, B.; Thaden, J.T.; Mogno, I.; Wierzbowski, J.; Cottarel, G.; Kasif, S.; Collins, J.J.; Gardner, T.S. Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. *PLoS Biol.* **2007**, *5*. [[CrossRef](#)] [[PubMed](#)]
- Zoppoli, P.; Morganello, S.; Ceccarelli, M. TimeDelay-ARACNE: Reverse engineering of gene networks from time-course data by an information theoretic approach. *BMC Bioinform.* **2010**, *11*. [[CrossRef](#)] [[PubMed](#)]
- Villaverde, A.; Ross, J.; Banga, J. Reverse engineering cellular networks with information theoretic methods. *Cells* **2013**, *2*, 306–329. [[CrossRef](#)] [[PubMed](#)]
- Aliferis, C.F.; Statnikov, A.; Tsamardinos, I.; Mani, S.; Koutsoukos, X.D. Local causal and markov blanket induction for causal discovery and feature selection for classification part I: Algorithms and empirical evaluation. *J. Mach. Learn. Res.* **2010**, *11*, 171–234.



21. Dondelinger, F.; Husmeier, D.; Lèbre, S. Dynamic bayesian networks in molecular plant science: Inferring gene regulatory networks from multiple gene expression time series. *Euphytica* **2012**, *183*, 361–377. [[CrossRef](#)]
22. En Chai, L.; Saberi Mohamad, M.; Deris, S.; Khim Chong, C.; Choon, Y.W.; Omatu, S. Current development and review of dynamic bayesian network-based methods for inferring gene regulatory networks from gene expression data. *Curr. Bioinform.* **2014**, *9*, 531–539. [[CrossRef](#)]
23. Wiener, N.; Wiener, N. The theory of prediction. *Mod. Math. Eng.* **1956**, *1*, 125–139.
24. Granger, C.W.J. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **1969**, *37*, 424–438. [[CrossRef](#)]
25. Tilghman, P.; Rosenbluth, D. Inferring wireless communications links and network topology from externals using granger causality. In *MILCOM 2013—2013 IEEE Military Communications Conference*; IEEE: Piscataway, NJ, USA, 2013; pp. 1284–1289.
26. Schiatti, L.; Nollo, G.; Rossato, G.; Faes, L. Extended granger causality: A new tool to identify the structure of physiological networks. *Physiol. Meas.* **2015**, *36*. [[CrossRef](#)] [[PubMed](#)]
27. Montalto, A.; Stramaglia, S.; Faes, L.; Tessitore, G.; Prevede, R.; Marinazzo, D. Neural networks with non-uniform embedding and explicit validation phase to assess granger causality. *Neural Netw.* **2015**, *71*, 159–171. [[CrossRef](#)] [[PubMed](#)]
28. Mahdevar, G.; Nowzaridalini, A.; Sadeghi, M. Inferring gene correlation networks from transcription factor binding sites. *Genes Genet. Syst.* **2013**, *88*, 301–309. [[CrossRef](#)] [[PubMed](#)]
29. Friedman, J.; Alm, E.J. Inferring correlation networks from genomic survey data. *PLoS Comput. Biol.* **2012**, *8*, e1002687. [[CrossRef](#)] [[PubMed](#)]
30. Carmeli, C.; Knyazeva, M.G.; Innocenti, G.M.; de Feo, O. Assessment of EEG synchronization based on state-space analysis. *Neuroimage* **2005**, *25*, 339–354. [[CrossRef](#)] [[PubMed](#)]
31. Walker, D.M.; Carmeli, C.; Pérez-Barbería, F.J.; Small, M.; Pérez-Fernández, E. Inferring networks from multivariate symbolic time series to unravel behavioural interactions among animals. *Anim. Behav.* **2010**, *79*, 351–359. [[CrossRef](#)]
32. Reshef, D.N.; Reshef, Y.A.; Finucane, H.K.; Grossman, S.R.; McVean, G.; Turnbaugh, P.J.; Lander, E.S.; Mitzenmacher, M.; Sabeti, P.C. Detecting novel associations in large data sets. *Science* **2011**, *334*, 1518–1524. [[CrossRef](#)] [[PubMed](#)]
33. Xia, L.C.; Ai, D.; Cram, J.; Fuhrman, J.A.; Sun, F. Efficient statistical significance approximation for local similarity analysis of high-throughput time series data. *Bioinformatics* **2013**, *29*, 230–237. [[CrossRef](#)] [[PubMed](#)]
34. Ruan, Q.; Dutta, D.; Schwalbach, M.S.; Steele, J.A.; Fuhrman, J.A.; Sun, F. Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. *Bioinformatics* **2006**, *22*, 2532–2538. [[CrossRef](#)] [[PubMed](#)]
35. Montalto, A.; Faes, L.; Marinazzo, D. Mute: A matlab toolbox to compare established and novel estimators of the multivariate transfer entropy. *PLoS ONE* **2014**, *9*. [[CrossRef](#)] [[PubMed](#)]
36. Ai, X. Inferring a drive-response network from time series of topological measures in complex networks with transfer entropy. *Entropy* **2014**, *16*, 5753–5776. [[CrossRef](#)]
37. Papan, A.; Kyrtsov, C.; Kugiumtzis, D.; Diks, C. *Partial Symbolic Transfer Entropy*; University of Amsterdam: Amsterdam, The Netherlands, 2013; pp. 13–16.
38. Thorniley, J. An improved transfer entropy method for establishing causal effects in synchronizing oscillators. In *ECAL 2011*; MIT Press: Cambridge, MA, USA, 2011; pp. 797–804.
39. Theiler, J.; Eubank, S.; Longtin, A.; Galdrikian, B.; Doyne Farmer, J. Testing for nonlinearity in time series: The method of surrogate data. *Phys. D Nonlinear Phenom.* **1992**, *58*, 77–94. [[CrossRef](#)]
40. Small, M. *Applied Nonlinear Time Series Analysis: Applications in Physics, Physiology and Finance*; World Scientific: Singapore, 2005; p. 260.
41. Nakamura, T.; Hirata, Y.; Small, M. Testing for correlation structures in short-term variabilities with long-term trends of multivariate time series. *Phys. Rev. E* **2006**, *74*, 041114. [[CrossRef](#)] [[PubMed](#)]
42. Nakamura, T.; Tanizawa, T.; Small, M. Constructing networks from a dynamical system perspective for multivariate nonlinear time series. *Phys. Rev. E* **2016**, *93*, 032323. [[CrossRef](#)] [[PubMed](#)]
43. Bandt, C.; Pompe, B. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.* **2002**, *88*, 174102. [[CrossRef](#)] [[PubMed](#)]
44. Staniek, M.; Lehnertz, K. Symbolic transfer entropy. *Phys. Rev. Lett.* **2008**, *100*, 158101. [[CrossRef](#)] [[PubMed](#)]

45. Ma, H.; Aihara, K.; Chen, L. Detecting causality from nonlinear dynamics with short-term time series. *Sci. Rep.* **2014**, *4*. [[CrossRef](#)] [[PubMed](#)]
46. Hill, S.M.; Heiser, L.M.; Cokelaer, T.; Unger, M.; Nesser, N.K.; Carlin, D.E.; Zhang, Y.; Sokolov, A.; Paull, E.O.; Wong, C.K.; et al. Inferring causal molecular networks: Empirical assessment through a community-based effort. *Nat. Methods* **2016**, *13*, 310–318. [[CrossRef](#)] [[PubMed](#)]
47. Rapp, P.E.; Cellucci, C.J.; Watanabe, T.A.A.; Albano, A.M.; Schmah, T.I. Surrogate data pathologies and the false-positive rejection of the null hypothesis. *Int. J. Bifurc. Chaos* **2001**, *11*, 983–997. [[CrossRef](#)]
48. Sims, C.A. Money, income, and causality. *Am. Econ. Rev.* **1972**, *62*, 540–552.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).