





Article

Detecting Causality in Multivariate Time Series via Non-Uniform Embedding

Ziyu Jia ¹ , Youfang Lin ¹, Zehui Jiao ¹ , Yan Ma ²  and Jing Wang ^{1,*} 

¹ School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China; ziyujia@bjtu.edu.cn (Z.J.); yflin@bjtu.edu.cn (Y.L.); zhjiao@bjtu.edu.cn (Z.J.)

² Division of Interdisciplinary Medicine and Biotechnology, Department of Medicine, Beth Israel Deaconess Medical Center/Harvard Medical School, Boston, MA 02215, USA; dr.yan.ma@gmail.com

* Correspondence: wj@bjtu.edu.cn

Received: 11 November 2019; Accepted: 13 December 2019; Published: 16 December 2019



Abstract: Causal analysis based on non-uniform embedding schemes is an important way to detect the underlying interactions between dynamic systems. However, there are still some obstacles to estimating high-dimensional conditional mutual information and forming optimal mixed embedding vector in traditional non-uniform embedding schemes. In this study, we present a new non-uniform embedding method framed in information theory to detect causality for multivariate time series, named LM-PMIME, which integrates the low-dimensional approximation of conditional mutual information and the mixed search strategy for the construction of the mixed embedding vector. We apply the proposed method to simulations of linear stochastic, nonlinear stochastic, and chaotic systems, demonstrating its superiority over partial conditional mutual information from mixed embedding (PMIME) method. Moreover, the proposed method works well for multivariate time series with weak coupling strengths, especially for chaotic systems. In the actual application, we show its applicability to epilepsy multichannel electrocorticographic recordings.

Keywords: causal analysis; non-uniform embedding; multivariate time series; conditional mutual information

1. Introduction

In recent years, various time series analysis methods have been proposed to detect interactions between complex systems [1–3]. The study of causality, in particular, has attracted wide attention of researchers. There are two classic methods in the time series causal analysis: Granger causality [4] and transfer entropy [5,6]. Both methods are based on time series prediction for causal analysis. In addition, the relationship between Granger causality and transfer entropy is demonstrated [7]: the two methods are equivalent under Gaussian assumptions. Furthermore, Hlavackova–Schindler [8] extends the equivalence of the two causality methods for generalized Gaussian processes which satisfy some additional conditions on probability density distributions.

With the development of multivariate state space reconstruction, different embedding schemes [9–13] are used in Granger causality and transfer entropy. The common idea of those embedding schemes is to reconstruct the past of the whole system represented by all variables with reference to the target variable, in order to form a mixed embedding vector containing the most significant past variables to explain the target variable. Non-uniform embedding schemes are proposed to solve the problems of redundancy and arbitrariness in uniform embedding schemes [11,14]. Vlachos et al. propose a causality measure for bivariate variables based on the mixed embedding scheme: the conditional mutual information from mixed embedding (MIME) [15]. Kugiumtzis et al. extend the MIME method to multivariable time series and form the partial MIME (PMIME) method [16].

The PMIME method successfully solves the problem of detecting direct causality in multivariable time series. In addition, it is gradually applied to complex systems such as physiology [17,18] and finance [19,20].

Although the causal analysis using non-uniform embedding has practical advantages, there are still some key shortcomings that need to be overcome. One shortcoming is the curse of dimensionality, which makes the estimation of mutual information inaccurate as the dimension of the embedded space increases [21–24]. Another shortcoming is related to the mixed embedding vector. The greedy strategy uses a sequential forward approach to select the lagged variables and finally form the mixed embedding vector [11,15,16]. That is to say, the lagged variables that have been embedded will not be changed in the mixed embedding vector. As the iteration increases, more lagged variables are selected until the final mixed embedding vector is formed. Therefore, the inaccuracy of the initial embedding will have a large impact on the results. The above shortcomings will be highlighted when there are multivariate time series of weak causal coupling strengths in practical applications.

In order to overcome the above shortcomings, we propose a new non-uniform embedding method named LM-PMIME for multivariate time series based on the low-dimensional approximation of conditional mutual information (CMI) and the mixed search strategy. The main contribution of the proposed method is to reduce the dimension of the embedded space by replacing the original estimate with a low-dimensional approximation of CMI. In addition, a mixed strategy, which has taken the place of the greedy strategy, was adopted as an embedded strategy to optimize the initial embedding. The proposed method works well for multivariate time series with weak coupling strengths.

The rest of the paper is organized as follows. In Section 2, we propose the multivariate non-uniform embedding in accordance with the low dimensional approximation of CMI and a mixed search strategy. In Section 3, we perform the simulation experiments in order to verify the effectiveness of the LM-PMIME method. In Section 4, by analyzing the electrocorticographic (ECoG) recordings from an epileptic patient, the applicability of the proposed method to actual data is shown. Finally, a summary is presented in Section 5.

2. Method

In this section, we first introduce the traditional PMIME method. Then we expound a low dimensional approximation of CMI and a mixed search strategy. Finally, we present the LM-PMIME method for multivariable non-uniform embedding.

2.1. PMIME Method

The PMIME method, a generalization of the MIME method for bivariate variables [15], is developed by Kugiumtzis et al. [16] to detect the directional coupling in multivariable time series. Let K variables $X, Y, Z_1, \dots, Z_{K-2}$ constitute an overall dynamical system $\{x_t, y_t, z_{1,t}, \dots, z_{K-2,t}\}_{t=1}^n$. Suppose that the driving subsystem is X and the target subsystem is Y . In other words, the current value of variable Y is affected by the past of variable X . $Z = \{Z_1, \dots, Z_{K-2}\}$ represent the remaining subsystems.

We estimate the causal effect of X on Y conditioned by $Z = \{Z_1, \dots, Z_{K-2}\}$. It is necessary to form a set of variables representing the past of the subsystems. The lags of X, Y and Z are sought within a range given by a maximum lag for each variable, e.g., L_x for X and L_y for Y . W_t is defined as the set of all lagged variables at time t , containing the parts $x_t, x_{t-1}, \dots, x_{t-L_x}$ of X and the same for Y and Z . It is usually assumed that the maximum lag L for all variables is the same ($L_x = L_y = L_z$). The larger the value of L , the more lagged variables are included in W_t . The key step of the PMIME method is to form the mixed embedding vector $\mathbf{v}_t \in W_t$ using non-uniform embedding. The stopping criterion and greedy selection are applied to the process of embedding. The detailed method is described below as follows:

1. An empty embedding vector $\mathbf{v}_t^0 = \emptyset$ is initialized.

2. At the first iteration $k = 1$, the embedding vector $w_t^1 \in W_t$ is selected most related to y_t :

$$w_t^1 = \arg \max_{w \in W_t} I(y_t; w), \quad (1)$$

where $I(\cdot)$ represents mutual information. Mutual information is estimated by the k-nearest neighbors (k-NNs) method. Then we have $\mathbf{v}_t^1 = [w_t^1]$. At the same time, w_t^1 is removed from W_t .

3. At the iteration $k > 1$, the mixed embedding vector is augmented by the component w_t^k of W_t , giving most information about y_t additionally to the information already contained in $\mathbf{v}_t^{k-1} = [w_t^1, \dots, w_t^{k-1}]$. w_t^k will be selected by a standard through calculating the maximum value of the conditional mutual information, $w_t^k = \arg \max_{w \in W_t} I(y_t; w | \mathbf{v}_t^{k-1})$, i.e., at the iteration $k = 2$, $w_t^2 = \arg \max_{w \in W_t} I(y_t; w | \mathbf{v}_t^1)$, where the CMI is calculated by the k-NNs estimator, and the mixed embedding vector is $\mathbf{v}_t^2 = [w_t^1, w_t^2]$. By using greedy forward method, each w_t^k will be embedded in the already embedded vector \mathbf{v}_t^{k-1} until the process stops. The termination criterion is quantified as:

$$I(y_t; \mathbf{v}_t^{k-1}) / I(y_t; \mathbf{v}_t^k) > A, \quad (2)$$

where the threshold $A < 1$ and the general value of A is 0.95 or 0.97 in [15,16]. That is, the additional information of w_t^k selected at the iteration k is not large enough. The embedding process will stop and we have the mixed embedding vector $\mathbf{v}_t = \mathbf{v}_t^{k-1}$. Any combination of the lagged variables $X, Y, Z_1, \dots, Z_{K-2}$ may be included in \mathbf{v}_t .

4. To calculate the causality strength of X on Y conditioned by the other variables in Z , the index is defined as

$$R_{X \rightarrow Y|Z} = \frac{I(y_t; \mathbf{v}_t^X | \mathbf{v}_t^Y, \mathbf{v}_t^Z)}{I(y_t; \mathbf{v}_t)}, \quad (3)$$

where \mathbf{v}_t^X represents the component of X in \mathbf{v}_t . It is the same with \mathbf{v}_t^Y and \mathbf{v}_t^Z . The causal effect of X to Y depends on the components of X in \mathbf{v}_t^X .

2.2. The Proposed Method

2.2.1. Low Dimensional Approximation of CMI

As the dimension of the mixed embedding vector increases, the estimation of CMI becomes less reliable. Because of an increasing volume of state space, the estimation of entropy rates progressively decreases towards zero [15]. Therefore, in order to overcome the problems caused by computing high-dimensional CMI, the low-dimensional approximation of CMI is a better alternative. The low-dimensional approximation can improve the accuracy of conditional mutual information estimation and reduce the computational cost.

The low-dimensional approximation of CMI is studied by researchers in the feature selection [25–31]. Brown et al. [23] emphasize that lots of feature selection heuristics are all approximate iterative maximizers of the conditional likelihood. Consequently, the methods are summarized as a parameterized general standard:

$$I(w; y_t) - \beta \sum_{w_i \in \mathbf{v}_t} I(w; w_i) + \gamma \sum_{w_j \in \mathbf{v}_t} I(w; w_j | y_t), \quad (4)$$

where the difference between different standards depends on the parameters (β and γ). For example, the JMI standard [26] is obtained with $\gamma = \beta = 1/|\mathbf{v}_t|$. β and γ are different in standards such as MRMR standard [28], and CIFE standard [29]. Recent studies illustrate that the higher-order feature interactions are considered to optimize feature selection standard. Therefore, we need to consider the second-order interactions between the features compared to Equation (4), such as $I(w; w_j | w_i)$ [24].

$$I(w; y_t) - \beta \sum_{w_i \in \mathbf{v}_t} I(w; w_i) + \gamma \sum_{w_i \in \mathbf{v}_t} I(w; w_i | y_t) - \delta \sum_{w_i \in \mathbf{v}_t} \sum_{w_j \in \mathbf{v}_t; i \neq j} I(w; w_j | w_i), \quad (5)$$

where $\beta = \gamma = 1/|\mathbf{v}_t|$ and $\delta = 1/|\mathbf{v}_t|(|\mathbf{v}_t| - 1)$. Using Equation (5), the original high-dimensional MI based standard is decomposed into a group of low-dimensional MI quantities. We apply this low-dimensional approximation to the selection of lagged variables.

2.2.2. Mixed Search Strategy

An applicable search strategy is important for building a mixed embedding vector. Because the greedy search strategy has high computational efficiency and good practicability, it has become the preferred strategy for embedding. However, the greedy strategy uses a sequential forward approach to select lagged variables, which rely heavily on the initial embedded vector. That is to say, the initial embedded vector is not accurate and the subsequent selection will get worse.

To solve the above problem, we present a mixed strategy to avoid inaccuracies in the initial embedding. The mixed strategy consists of two strategies: the traversal strategy and the greedy strategy. The application of the strategy is determined by defining a strategy adjustment factor m . Assuming that a number of iterations is k , the traversal strategy is applied when $1 < k \leq m$. For example, when using the traversal strategy, it is necessary to calculate the possible combinations of all lagged variables before determining the mixed embedding vector of the current step. That is to say, we need to calculate C_{K*L}^k combinations in total, and then select the combination of the largest conditional mutual information as the mixed embedding vector of the current step. The greedy strategy is applied when $k > m$. This strategy is the same as the one used by the PMIME method.

2.2.3. LM-PMIME Method

We propose the LM-PMIME method to detect the directional coupling in multivariable time series according to the low-dimensional approximation of CMI and the mixed search strategy. In the LM-PMIME method, the mixed strategy determines the way to select lagged variables. Meanwhile, whether the variable will be embedded depends on the low dimensional approximation of CMI. Figure 1 illustrates the flow of the LM-PMIME method.

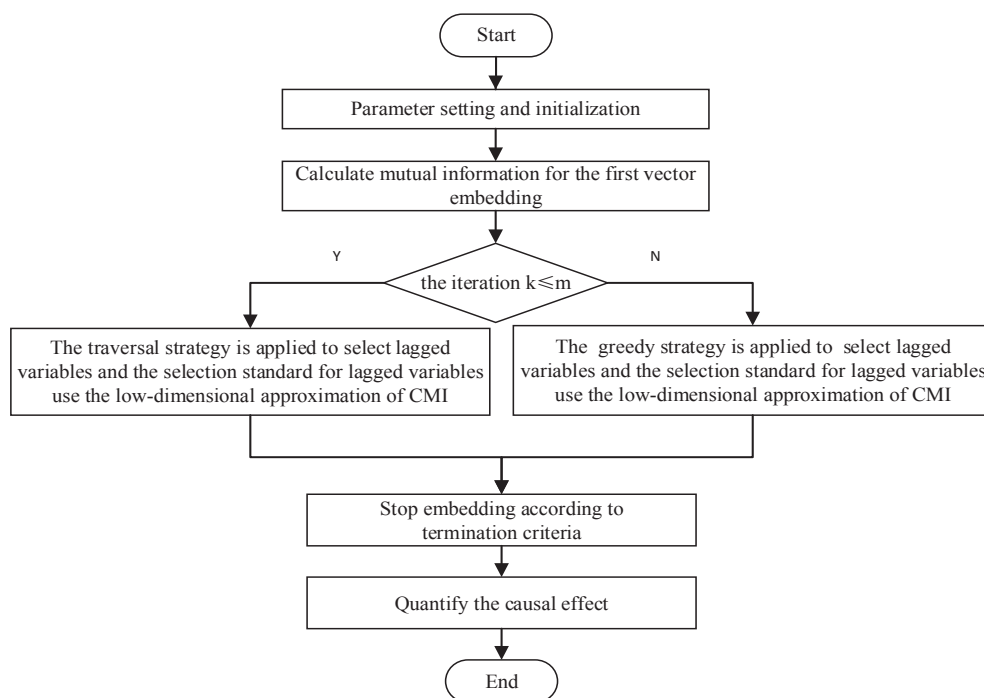


Figure 1. The flowchart of the low-dimensional approximation of CMI and the mixed search strategy(LM)-partial conditional mutual information from mixed embedding (PMIME) method.

The detailed LM-PMIME method is as follows:

1. Initialize an empty embedding vector $\mathbf{v}_t^0 = \emptyset$.
2. At the first iteration $k = 1$, the embedding vector $w_t^1 \in W_t$ is selected most related to y_t :

$$w_t^1 = \arg \max_{w \in W_t} I(y_t; w) \tag{6}$$

Then we have $\mathbf{v}_t^1 = [w_t^1]$.

3. At the iteration $1 < k \leq m$, w_t^k will be selected by a standard through calculating the maximum value of the low dimensional approximation of CMI.

$$w_t^k = \arg \max_{w \in W_t} I(w; y_t) - \beta \sum_{w_i \in \mathbf{v}_t^{k-1}} I(w; w_i) + \gamma \sum_{w_i \in \mathbf{v}_t^{k-1}} I(w; w_i | y_t) - \delta \sum_{w_i \in \mathbf{v}_t^{k-1}} \sum_{w_j \in \mathbf{v}_t^{k-1}, i \neq j} I(w; w_j | w_i), \tag{7}$$

where $\beta = \gamma = 1/|\mathbf{v}_t|$ and $\delta = 1/|\mathbf{v}_t|(|\mathbf{v}_t| - 1)$. The traversal strategy is applied to select \mathbf{v}_t^k , i.e., at the iteration $k = 4$ and $m = 5$, \mathbf{v}_t^4 needs to be selected. First, clear the already embedded vector \mathbf{v}_t^3 and calculate C_{K*L}^4 combinations in total. Then select the combination of largest conditional mutual information as \mathbf{v}_t^4 of the current step. Finally, $k = k + 1$.

4. At the iteration $k > m$, greedy strategy is used. Each w_t^k will be embedded in the already embedded vector \mathbf{v}_t^{k-1} until the process stops. The standard of low dimensional approximation is still used before stopping.
5. The termination criterion is quantified as:

$$I(y_t; \mathbf{v}_t^{k-1}) / I(y_t; \mathbf{v}_t^k) > A, \tag{8}$$

where the threshold $A < 1$ and threshold A near 1, e.g., $A = 0.95$, allows the inclusion of a new component in the mixed embedding vector even if the augmented vector explains very little of the information on y_t that was not explained at the previous step. Although the statistical significance threshold α is widely used [32,33], the accuracy of causal analysis is lower than the threshold A and the amount of calculation is large [15,16]. Therefore, we employ the threshold A for non-uniform embedding. The general value of A is 0.95 or 0.97 in [15,16]. That is, the additional information of w_t^k selected at the iteration k is not large enough. The embedding process will stop and we have the mixed embedding vector $\mathbf{v}_t = \mathbf{v}_t^{k-1}$. In addition, any combination of the lagged variables $X, Y, Z_1, \dots, Z_{K-2}$ may be included in \mathbf{v}_t .

6. To calculate the causality strength of X on Y conditioned by the other variables in Z , the index is defined as:

$$R_{X \rightarrow Y|Z} = \frac{I(y_t; \mathbf{v}_t^x | \mathbf{v}_t^y, \mathbf{v}_t^z)}{I(y_t; \mathbf{v}_t)}, \tag{9}$$

where \mathbf{v}_t^x represents the component of X in \mathbf{v}_t . It is the same with \mathbf{v}_t^y and \mathbf{v}_t^z . The causality strength of X to Y depends on the components of X in \mathbf{v}_t^x . The presence of components of X in the mixed embedding vector shows that X has effect on Y and then the derived causality strength $R_{X \rightarrow Y|Z}$ is positive, whereas the absence shows no effect and causality strength $R_{X \rightarrow Y|Z}$ is exactly zero. In addition, the $R_{X \rightarrow Y|Z}$ is considered significant if it is positive in the PMIME method and proposed method.

3. Simulation Study

To evaluate the effectiveness of the LM-PMIME, a series of synthetic time series from linear stochastic, nonlinear stochastic and chaotic systems are applied to compare the effectiveness of the traditional PMIME method. The effects of data length and coupling strength are considered in the numerical experiment. To show the performance of the low-dimensional approximation of CMI, the M-PMIME method only with the improvement of mixed strategy is added for comparison.

In the above models, the accuracy of the estimated mutual information is vital for embedding vector selection [28]. The two most common methods for estimating mutual information are the histogram and kernel methods. The former one is time efficient but not highly accurate [34]. The latter one has higher accuracy but comes with huge computational pressure [35]. We applied the k-nearest neighbors (k-NNs) method to estimate mutual information because the k-NNs estimator is suitable for high-dimensional data [36].

We calculate all methods on 100 realizations from each system to assess statistically the evaluation indicators (sensitivity, specificity, and F1 score). The connections between variables are classified as coupled directions and uncoupled directions to compute the confusion matrix.

3.1. Linear Multivariate Stochastic Process

The linear vector autoregressive (VAR) process is composed of order 4 in 5 time series (model 1 in [37]).

$$\begin{cases} x_{1,t} = 0.4x_{1,t-1} - 0.5x_{1,t-2} + 0.4x_{5,t-1} + e_{1,t} \\ x_{2,t} = 0.4x_{2,t-1} - 0.3x_{1,t-4} + 0.4x_{5,t-2} + e_{2,t} \\ x_{3,t} = 0.5x_{3,t-1} - 0.7x_{3,t-2} - 0.3x_{5,t-3} + e_{3,t} \\ x_{4,t} = 0.8x_{4,t-3} + 0.4x_{1,t-2} + 0.3x_{2,t-2} + e_{4,t} \\ x_{5,t} = 0.7x_{5,t-1} - 0.5x_{5,t-2} - 0.4x_{4,t-1} + e_{5,t} \end{cases} \quad (10)$$

where $e_{i,t}$, $i = 1, \dots, 5$, are Gaussian noise. $X_1 \rightarrow X_2$, $X_1 \rightarrow X_4$, $X_2 \rightarrow X_4$, $X_4 \rightarrow X_5$, $X_5 \rightarrow X_1$, $X_5 \rightarrow X_2$, and $X_5 \rightarrow X_3$ are the true causal connections in this process.

We use $A = 0.97$ and $L = 6$, which matches the larger lag for the three methods in the process. In addition, the LM-PMIME method and the M-PMIME method use the parameter $m = 2$. The results from the linear VAR process with the time series length of 512 are shown in Figure 2. The direction of the causal effect is from row to column in the matrix representation, e.g., the causal connection $X_1 \rightarrow X_2$ is represented as (1, 2) in the matrix representation. Hence, true causal relationships in the process are (1, 2), (1, 4), (2, 4), (4, 5), (5, 1), (5, 2) and (5, 3) at the matrix elements. The mean values of coupling measured by the three methods are positive and high on these matrix elements. It is proved that the three methods have good sensitivity to true couplings. However, Figure 2 shows that there are lots of false positives in the traditional method using high-dimensional CMI. In contrast, the LM-PMIME method has better results than the other two methods, because the method reduces false positives. The sensitivity, specificity, and F1 scores are obtained from 100 realizations of linear VAR process with varying length of time series. The values of the specific indicators are listed in Table 1. The F1 score of the LM-PMIME method has better results on the linear VAR process with different time series lengths. Furthermore, the F1 score calculated by the LM-PMIME method increases as the time series length increases. These better results are likely due to the great improvement of specificity by the proposed method. At the same time, the F1 score reflects that the PMIME method and M-PMIME method have achieved similar results in the linear VAR process. It shows that the mixed strategy does not work in this process. However, the following experiments show that the mixed strategy works well on the chaotic system.

3.2. Nonlinear Multivariate Stochastic Process

The nonlinear VAR process is of order 1 in three variables NLVAR₃(1) (model 7 in [38]).

$$\begin{cases} x_{1,t} = 3.4x_{1,t-1}(1 - x_{1,t-1}^2)e^{-x_{1,t-1}^2} + 0.4e_{1,t} \\ x_{2,t} = 3.4x_{2,t-1}(1 - x_{2,t-1}^2)e^{-x_{2,t-1}^2} + 0.5x_{1,t-1}x_{2,t-1} + 0.4e_{2,t} \\ x_{3,t} = 3.4x_{3,t-1}(1 - x_{3,t-1}^2)e^{-x_{3,t-1}^2} + 0.3x_{2,t-1} + 0.5x_{1,t-1}^2 + 0.4e_{3,t} \end{cases} \quad (11)$$

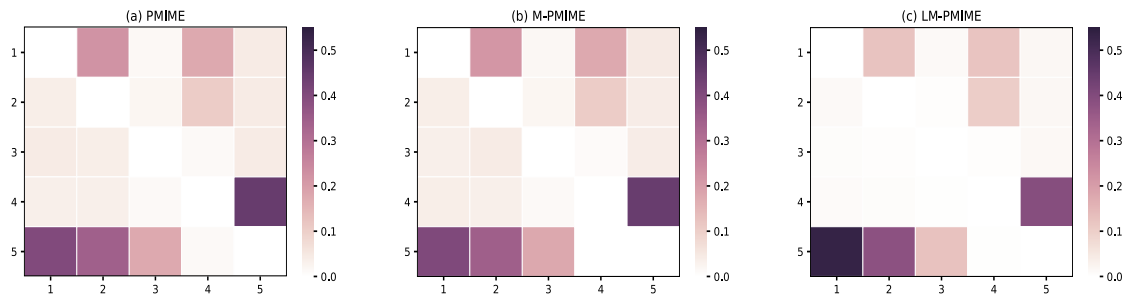


Figure 2. Matrix representation of causality for the linear vector autoregressive (VAR) process. Retrieved by (a) traditional PMIME method, (b) mixed search strategy (M)-PMIME method, (c) and LM-PMIME method with k-nearest neighbors (k-NNs) estimator. The length of the time series is 512. $m = 2$ is used for the M-PMIME method and the LM-PMIME method. The remaining parameters of the three methods are the same ($L = 6, A = 0.97$). Color maps for the mean values of coupling measurements are obtained from 100 realizations of the linear VAR process. The direction of causal influence is from row to column in the matrix. The true causal connections in this linear VAR process are at the matrix elements (1, 2), (1, 4), (2, 4), (4, 5), (5, 1), (5, 2) and (5, 3).

Table 1. Evaluation indicators are obtained from 100 realizations of linear VAR process with varying time series length for the three different methods. $A = 0.97$ and $L = 6$ are the parameters common to the three methods. In addition, the LM-PMIME method and the M-PMIME method use the parameter $m = 2$.

	Sensitivity	Specificity	F1 Score
$n = 256$			
PMIME	0.988	0.492	0.600
M-PMIME	0.989	0.481	0.596
LM-PMIME	0.797	0.741	0.647
$n = 512$			
PMIME	1.000	0.567	0.643
M-PMIME	0.994	0.727	0.645
LM-PMIME	0.855	0.763	0.693
$n = 1024$			
PMIME	1.000	0.697	0.719
M-PMIME	0.940	0.729	0.713
LM-PMIME	0.877	0.807	0.739

The true causal connections in NLVAR₃(1) are $X_1 \rightarrow X_2, X_1 \rightarrow X_3, X_2 \rightarrow X_3$. The results obtained from 100 realizations of the nonlinear VAR process are shown in Figure 3 for $n = 512, A = 0.97, L = 6$. The strategy adjustment factor $m = 3$ determines the application of the strategies for the LM-PMIME method and M-PMIME method. The true causal connections in NLVAR₃(1) are represented at the matrix elements (1,2), (1,3), and (2,3). For all methods, the mean values of coupling measurements on these matrix elements are positive and high. It turns out that all methods have good sensitivity to true couplings. But there are many false positives in the traditional methods using high-dimensional CMI. Hence, the LM-PMIME method significantly outperforms the others. The evaluation indicators are obtained from NLVAR₃(1) by increasing the time series length. The values of the specific indicators are listed in Table 2. The F1 score of LM-PMIME method has better results on NLVAR₃(1) with different time series lengths. In addition, the F1 score will increase as the time series length increases. The low-dimensional approximation of CMI can greatly improve specificity, although mixed strategy does not work in NLVAR₃(1).

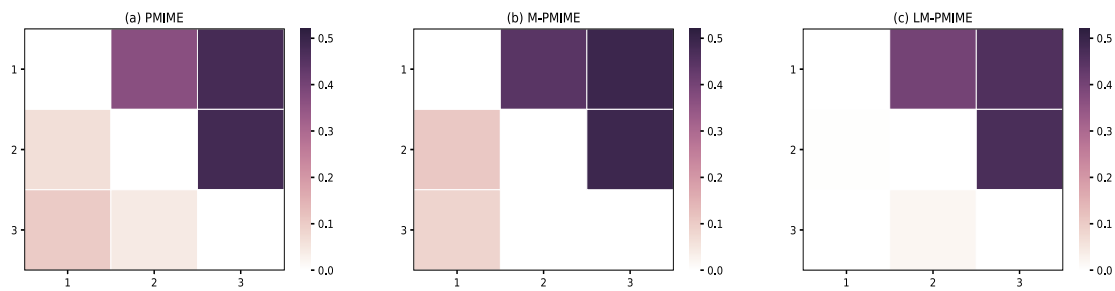


Figure 3. Matrix representation of causality for $NLVAR_3(1)$. Retrieved by (a) traditional PMIME method, (b) M-PMIME method, (c) and LM-PMIME method with k-NNs estimator. The time series length is 512. $m = 3$ is used for the M-PMIME method and the LM-PMIME method. The remaining parameters of the three methods are the same ($L = 6, A = 0.97$). Color maps for the mean values of coupling measurements are obtained from 100 realizations of $NLVAR_3(1)$. The true causal connections in $NLVAR_3(1)$ are at the matrix elements (1,2), (1,3), (2,3).

Table 2. Evaluation indicators are obtained from 100 realizations of $NLVAR_3(1)$ with varying time series length for the three different methods. $A = 0.97$ and $L = 6$ are the parameters common to the three methods. In addition, the LM-PMIME method and the M-PMIME method use the parameter $m = 3$.

	Sensitivity	Specificity	F1 Score
		$n = 256$	
PMIME	0.973	0.650	0.737
M-PMIME	0.976	0.615	0.712
LM-PMIME	0.860	0.844	0.792
		$n = 512$	
PMIME	1.000	0.681	0.758
M-PMIME	1.000	0.662	0.748
LM-PMIME	0.950	0.887	0.873
		$n = 1024$	
PMIME	1.000	0.860	0.877
M-PMIME	1.000	0.790	0.827
LM-PMIME	0.989	0.892	0.896

3.3. Coupled Henon Maps

The system of K coupled chaotic Henon maps (model 6 in [16]) defined as

$$\begin{cases} x_{1,t} = 1.4 - x_{1,t-1}^2 + 0.3x_{1,t-2} \\ x_{i,t} = 1.4 - (Cx_{i-1,t-1} + (1 - C)x_{i,t-1})^2 + 0.3x_{i,t-2} \end{cases} \quad \text{for } i = 2, \dots, K \quad (12)$$

$X_{i-1} \rightarrow X_i$, where $i = 2, \dots, K$, are the true causal connections in the K coupled chaotic Henon maps. The results from 100 realizations of the coupled Henon maps with the coupling strength $C = 0.1$ are shown in Figure 4 for $n = 1024, K = 6, A = 0.95, L = 5, m = 2$. In addition to this, the results of only changing the coupling strength $C = 0.3$ are shown in Figure 5. The true causal connections in the coupled Henon maps are at the matrix elements $(i - 1, i)$, where $i = 2, \dots, 6$. There is almost no false positive for all methods. However, Figures 4 and 5 illustrate that the proposed methods have better performance than the traditional method when there are true causal connections. All methods will detect stronger causal connections as the coupling strength C of the system increases. The evaluation indicators are obtained from coupled Henon maps with the variables K from 3 to 9. The values of the specific indicators are listed in Tables 3 and 4. The results show that the F1 score of the LM-PMIME method is higher than the others when the coupling strength is low. Although the F1 score may be affected by the number of variables K in the simulation experiments, the F1 score for the LM-PMIME method is above 0.9. The LM-PMIME method and the M-PMIME method greatly improve

the specificity, especially the former method. It is proved that both low-dimensional approximation of CMI and the mixed strategies play an important role in coupled Henon maps when the coupling strength is low.

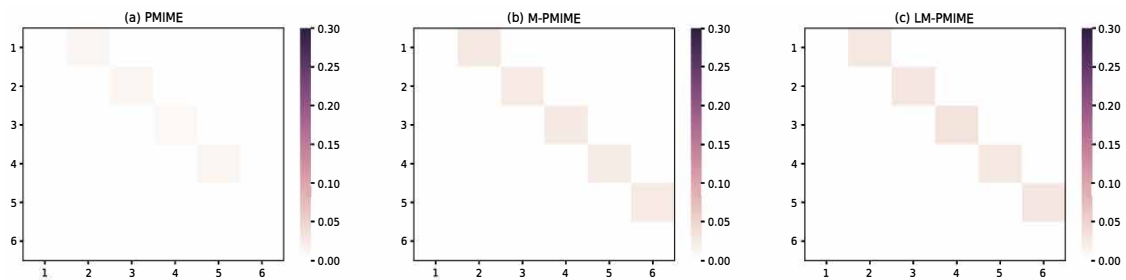


Figure 4. Matrix representation of causality for $K = 6$ variables of the coupled Henon maps ($C = 0.1$). Retrieved by (a) traditional PMIME method, (b) M-PMIME method, (c) and LM-PMIME method with k-NNs estimator. The time series length is 1024. $m = 2$ is used for the M-PMIME method and the LM-PMIME method. The remaining parameters of the three methods are the same ($L = 5, A = 0.95$). Color maps for the mean values of coupling measurements are obtained from 100 realizations of the coupled Henon maps. The true causal connections in the coupled Henon maps are at the matrix elements $(i - 1, i)$, where $i = 2, \dots, 6$.

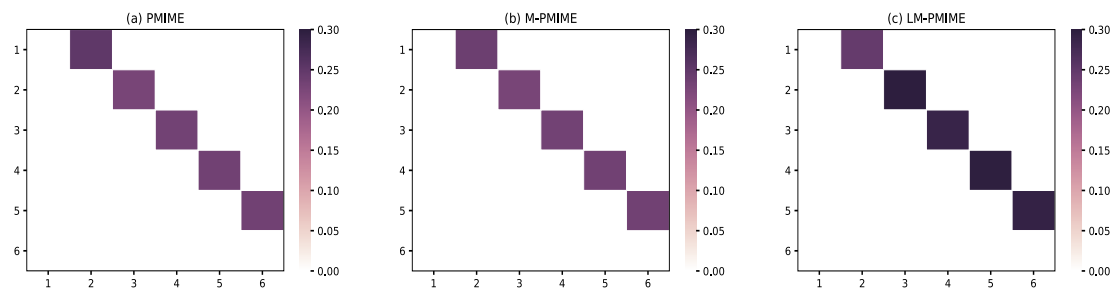


Figure 5. Matrix representation of causality for $K = 6$ variables of the coupled Henon maps ($C = 0.3$). Retrieved by (a) traditional PMIME method, (b) M-PMIME method, (c) and LM-PMIME method with k-NNs estimator. The time series length is 1024. $m = 2$ is used for the M-PMIME method and the LM-PMIME method. The remaining parameters of the three methods are the same ($L = 5, A = 0.95$). Color maps for the mean values of coupling measurements are obtained from 100 realizations of the coupled Henon maps. The true causal connections in the coupled Henon maps are at the matrix elements $(i - 1, i)$, where $i = 2, \dots, 6$.

Table 3. Evaluation indicators are obtained from 100 realizations of K variables of the Henon maps ($C = 0.1$) for the three different methods. $A = 0.95$ and $L = 5$ are the parameters common to the three methods. In addition, the LM-PMIME method and the M-PMIME method use the parameter $m = 2$.

	Sensitivity	Specificity	F1 Score
		$K = 3$	
PMIME	0.175	1.000	0.297
M-PMIME	0.715	1.000	0.834
LM-PMIME	0.945	1.000	0.972
		$K = 6$	
PMIME	0.217	1.000	0.357
M-PMIME	0.674	1.000	0.806
LM-PMIME	0.926	0.998	0.950
		$K = 9$	
PMIME	0.204	1.000	0.338
M-PMIME	0.700	1.000	0.824
LM-PMIME	0.895	0.998	0.904

Table 4. Evaluation indicators are obtained from 100 realizations of K variables of the Henon maps ($C = 0.3$) for the three different methods. $A = 0.95$ and $L = 5$ are the parameters common to the three methods. In addition, the LM-PMIME method and the M-PMIME method use the parameter $m = 2$.

	Sensitivity	Specificity	F1 Score
		$K = 3$	
PMIME	1.000	1.000	1.000
M-PMIME	1.000	1.000	1.000
LM-PMIME	1.000	1.000	1.000
		$K = 6$	
PMIME	1.000	1.000	1.000
M-PMIME	1.000	1.000	1.000
LM-PMIME	1.000	1.000	1.000
		$K = 9$	
PMIME	1.000	1.000	1.000
M-PMIME	1.000	1.000	1.000
LM-PMIME	1.000	1.000	1.000

3.4. Coupled Lorenz System

The chaotic system of three coupled identical Lorenz oscillators (model 5 in [16]) is defined as

$$\begin{cases} \dot{x}_1 = -10x_1 + 10y_1, & \dot{x}_i = -10x_i + 10y_i + C(x_{i-1} - x_i), \\ \dot{y}_1 = -x_1z_1 + 28x_1 - y_1, & \dot{y}_i = -x_iz_i + 28x_i - y_i, \\ \dot{z}_1 = x_1y_1 - \frac{8}{3}z_1, & \dot{z}_i = x_iy_i - \frac{8}{3}z_i, \end{cases} \quad (13)$$

where $i = 2, 3$. The differential equations by the explicit Runge-Kutta (4,5) method are solved in MATLAB. The true causal connections in the three coupled Lorenz oscillators are $X_{i-1} \rightarrow X_i$, where $i = 2, 3$.

The results from 100 realizations of the three coupled Lorenz oscillators with the coupling strength $C = 3$ are shown in Figure 6, for $n = 512$, $A = 0.95$, $L = 5$, $m = 3$. In addition, The evaluation indicators are listed in Table 5. The values of the specific indicators are obtained from the Lorenz oscillators with varying time series length and the remaining parameters are the same. The F1 scores of the proposed methods are much higher than the traditional PMIME method. The M-PMIME method performs best when the time series is short. That is to say, the mixed strategy plays a role in improving the F1 score. However, the F1 score of the LM-PMIME method is the highest as the time series length increases.

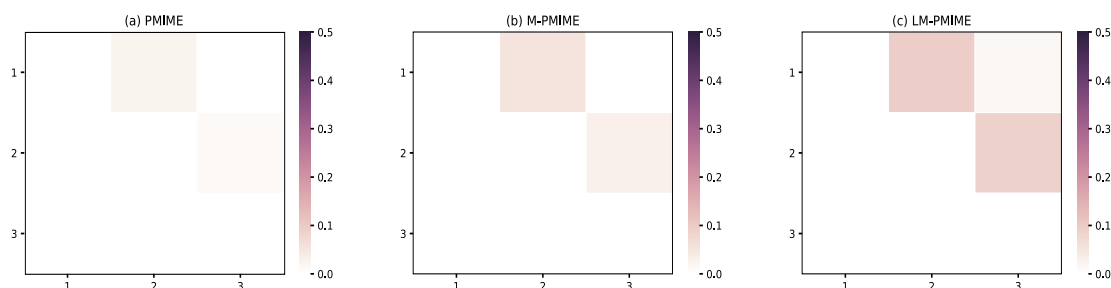


Figure 6. Matrix representation of causality for the three coupled Lorenz oscillators. Retrieved by (a) traditional PMIME method, (b) M-PMIME method, (c) and LM-PMIME method with k-NNs estimator. The length of the time series is 512 with coupling strength $C = 3$. $m = 3$ is used for the M-PMIME method and the LM-PMIME method. The remaining parameters of the three methods are the same ($L = 5$, $A = 0.95$). Color maps for the mean values of coupling measurements are obtained from 100 realizations of the three coupled Lorenz oscillators. The true causal connections in the three coupled Lorenz oscillators are at the matrix elements $(i - 1, i)$, where $i = 2, 3$.

Table 5. Evaluation indicators are obtained from 100 realizations of the three coupled Lorenz oscillators ($C = 3$) with varying time series length for the three different methods. $A = 0.95$ and $L = 5$ are the parameters common to the three methods. In addition, the LM-PMIME method and the M-PMIME method use the parameter $m = 3$.

	Sensitivity	Specificity	F1 Score
		$n = 256$	
PMIME	0.225	0.997	0.364
M-PMIME	0.660	0.863	0.617
LM-PMIME	0.805	0.665	0.541
		$n = 512$	
PMIME	0.185	1.000	0.312
M-PMIME	0.640	0.913	0.658
LM-PMIME	0.875	0.743	0.631
		$n = 1024$	
PMIME	0.175	1.000	0.297
M-PMIME	0.670	0.909	0.674
LM-PMIME	0.970	0.756	0.687

The evaluation indicators are obtained from 100 realizations of the Lorenz oscillators with coupling strength C from 1 to 5 for the three different methods. The time series length is 512 and $A = 0.95$, $L = 5$, $m = 3$. The values of the indicators are listed in Table 6. The results show that the LM-PMIME method performs best when the C is low, such as $C = 1$. Although the F1 score of the traditional PMIME method increases as the coupling strength C increases, it is still much worse than the proposed methods. Figure 7 is the matrix representation of causality for the three coupled Lorenz oscillators with coupling strength $C = 5$. The true causal connections are $(i - 1, i)$ in the matrix elements, where $i = 2, 3$. Only for the LM-PMIME method, the mean values of coupling measurements on these matrix elements are positive and high.

Table 6. Evaluation indicators are obtained from 100 realizations of the three coupled Lorenz oscillators ($n = 512$) with coupling strength C from 1 to 5 for the three different methods. $A = 0.95$ and $L = 5$ are the parameters common to the three methods. In addition, the LM-PMIME method and the M-PMIME method use the parameter $m = 3$.

	Sensitivity	Specificity	F1 Score
		$C = 1$	
PMIME	0.000	1.000	0.000
M-PMIME	0.155	0.926	0.221
LM-PMIME	0.375	0.830	0.381
		$C = 2$	
PMIME	0.075	1.000	0.141
M-PMIME	0.565	0.893	0.583
LM-PMIME	0.825	0.765	0.623
		$C = 3$	
PMIME	0.185	1.000	0.312
M-PMIME	0.640	0.913	0.658
LM-PMIME	0.875	0.743	0.631
		$C = 4$	
PMIME	0.260	1.000	0.413
M-PMIME	0.740	0.892	0.698
LM-PMIME	0.920	0.710	0.627
		$C = 5$	
PMIME	0.320	0.997	0.481
M-PMIME	0.725	0.873	0.660
LM-PMIME	0.960	0.731	0.661

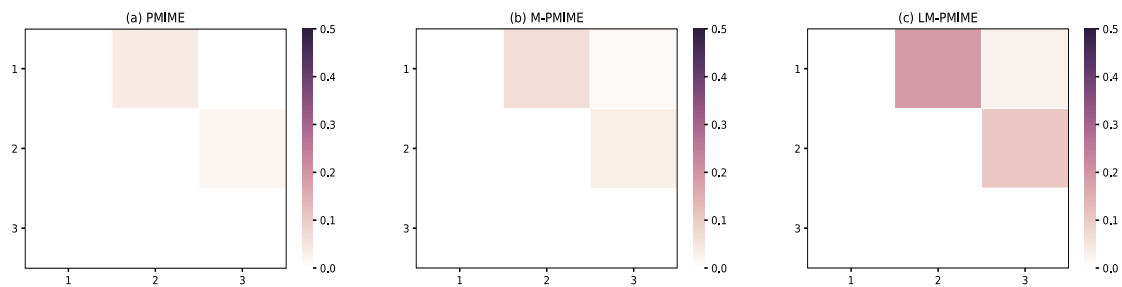


Figure 7. Matrix representation of causality for the three coupled Lorenz oscillators. Retrieved by (a) traditional PMIME method, (b) M-PMIME method, (c) and LM-PMIME method with k-NNs estimator. The length of the time series is 512 with coupling strength $C = 5$. $m = 3$ is used for the M-PMIME method and the LM-PMIME method. The remaining parameters of the three methods are the same ($L = 5$, $A = 0.95$). Color maps for the mean values of coupling measurements are obtained from 100 realizations of the three coupled Lorenz oscillators. The direction of causal influence is from row to column in the matrix. The true causal connections in the three coupled Lorenz oscillators are at the matrix elements $(i - 1, i)$, where $i = 2, 3$.

4. Application to Epilepsy ECoG Signals

In the study, we show the applicability of the LM-PMIME method to actual electrocorticographic (ECoG) to explore key contacts of the human subject with intractable epilepsy. A public dataset from a 39-year-old woman with medically refractory complex partial seizures is used. The dataset contains 8 seizure epochs and eight pre-seizure epochs. Each epoch contains 76-time series obtained from the eight-by-eight electrode grid and two depth electrodes with six contacts each. In addition, the duration of each epoch is 10 s and the length of each time series is 4000 (see [39] for more details). The data is recorded at 400 Hz, which is downsampled to 100 Hz.

We use PMIME method and LM-PMIME method to analyze the seizure data and the pre-seizure data, respectively. To assess the causal matrices of different physiological states estimated by each method, we compute the average causal strengths (the mean values of the coupling measurements over all epochs in the same physiological state) as shown in Figure 8. The brighter the colors are, the more significant causal connections are. As a result, it is obvious from the causal matrices of LM-PMIME method that contact 73 has more impact on the other contacts, highlighting that it is the key contact in the pre-seizure data (see Figure 8b). Figure 8 illustrates the difference of total numbers of significant connections between the seizure state and the pre-seizure state. The proposed method highlights the key contact 50 (see Figure 9b) and these discovered key contacts are consistent with many researchers [39–41].

The LM-PMIME method gives an obvious causal driver located at the contact 73 from the second depth electrode strip in the pre-seizure data shown in Figure 8. Therefore, the contact 73 may be associated with seizures. Although not yet clinically observable, it has been suggested that the second depth electrode primarily affects cortical activity in [40,41]. In addition, the proposed method successfully identifies a key contact from the data: contact 50, which exhibits the most significant change in the betweenness centrality. The contact is considered the primary target of therapeutic intervention in [38], because contacts with statistically significant increases in betweenness centrality may lead to seizures.

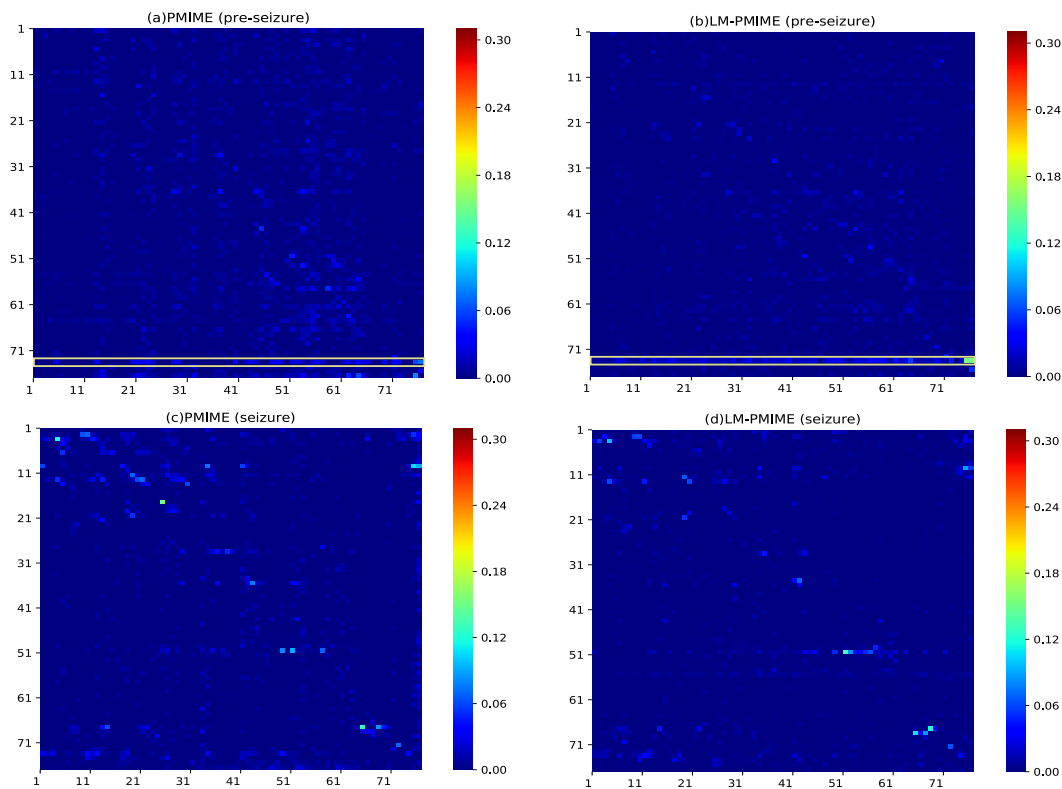


Figure 8. Results for multivariate electrocorticographic (ECoG) data. Matrices of causalities reflect the pre-seizure state (**top**) and the seizure state (**bottom**) estimated by the PMIME method and the LM-PMIME method. The causal strengths are averaged (the mean values of the coupling measurements over all epochs in the same physiological state). Contacts 1 to 64 belong to an eight-by-eight electrode grid, and contacts 65 to 76 belong to two depth electrodes. The direction of causal influence is from row to column in the matrices. The brighter colors indicate more significant values. The key contact is marked by a rectangular box. The parameter $A = 0.95$ and $m = 2$ are set for the different methods.

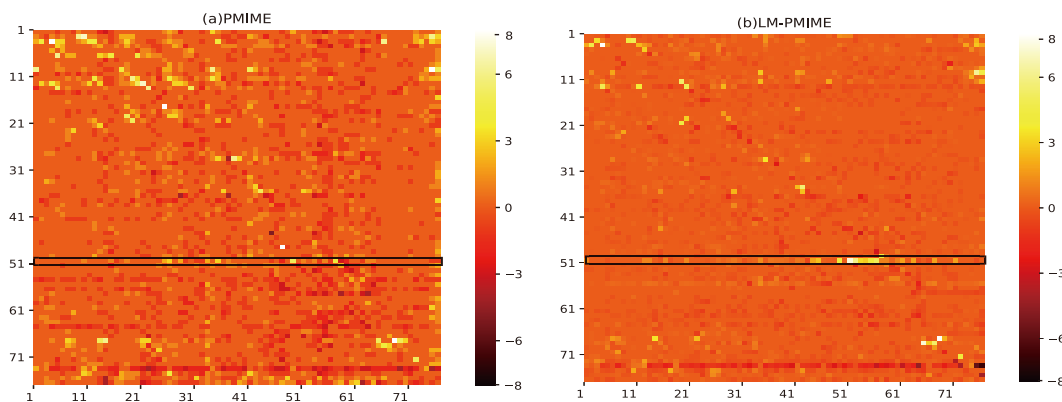


Figure 9. Results for multivariate ECoG data. Matrices reflect the difference of total numbers of significant connections between the seizure state and the pre-seizure state (seizure minus pre-seizure). The numbers are respectively summed from 8 seizure epochs and eight pre-seizure epochs. Contacts 1 to 64 belong to an eight-by-eight electrode grid, and contacts 65 to 76 belong to two depth electrodes. The brighter colors indicate more significant values. The key contact is marked by a rectangular box. The parameter $A = 0.95$ and $m = 2$ are set for the different methods.

5. Discussion and Conclusions

In this study, we show effective modifications for the well-known non-uniform embedding method: PMIME, which quantifies causality by means of information theoretic measures.

The advantage of the non-uniform embedding compared with uniform embedding is that it can reduce the dimension of the state space by selecting the relevant variables which contribute the most to explain the target variable. Therefore, it has been proved that the non-uniform embedding process is more flexible for state-space reconstruction [11,15,42]. However, there are still some obstacles to estimating high-dimensional CMI and constructing an optimal mixed embedding vector in the traditional non-uniform embedding methods. The proposed LM-PMIME method overcomes the above shortcomings of traditional methods. The major contribution of the proposed LM-PMIME method, which is based on the low-dimensional approximation of conditional mutual information and the mixed search strategy, is that improves the traditional non-uniform embedding methods. The curse of dimensionality is avoided by replacing the original estimate with a low-dimensional approximation of CMI. In addition, a mixed strategy instead of the greedy strategy is used as an embedded strategy to solve the problem of initial embedding inaccuracy. Hence, the mixed embedding vector becomes more parsimonious by maximizing the correlation with the target variable and minimizing the redundancy between the selected variables. In order to form the optimal mixed embedded vector, there are also other propositions. For example, in [22] a preselection scheme for subsets of causal predictors is used to search an optimal subset and detect the synergetic variables. In addition, many researchers adopt the OCE algorithm [43] or the PCMCI [30] algorithm to estimate the causal graphs. Different from these preselection methods, the LM-PMIME method relies on both the low-dimensional approximation and the mixed search strategy to improve the conditions. In all simulation systems, the LM-PMIME method performs better than the traditional methods according to the F1 score. Because of the complexity of chaotic systems, true causality is often difficult to detect. However, the LM-PMIME method significantly improves the sensitivity in chaotic systems. In the remaining simulation systems, the LM-PMIME method reduces false positives and increases the specificity. The experiments also adopt the comparison method M-PMIME, which improves the search strategy without using low-dimensional approximation. By the M-PMIME method, it can be found that the mixed search strategy works well in chaotic systems, especially the systems with low coupling strengths. In addition, the low-dimensional approximation of CMI plays an important role in linear and nonlinear systems. Therefore, we combine both the low-dimensional approximation of CMI and the mixed search strategy to form a new non-uniform embedding method LM-PMIME for multivariate time series. Although the LM-PMIME method works better than the traditional method, there are limitations to the proposed method. It still does not exceed the efficiency of the standard uniform embedding methods under the non-uniform embedding framework. In addition, the strategy adjustment factor m is larger, the traversal strategy is used more times in iterations. The main role of the traversal strategy is to correct the initial embedding inaccuracy, which is often time-consuming. Therefore, the recommended value is $m < 5$ and the simulation results generally take $m = 2$ or $m = 3$ to achieve better results. This is because a smaller value of m can reduce the computation time and it is enough to optimize the initial embedding.

Generally speaking, the proposed low dimensional approximation of CMI and mixed search strategy improve the non-uniform embedding process, which is also applicable to other causal analysis methods based on non-uniform embedding [32]. In this study, we present the LM-PMIME method to detect directional coupling for multivariate time series. The effectiveness and applicability of the LM-PMIME method are demonstrated by a large number of experiments. The LM-PMIME method works well for multivariate time series with weak coupling strengths. In addition, the proposed LM-PMIME method, a causal analysis method, has great potential to be adopted in other applications, e.g., prediction of dynamic systems. We will further study the non-uniform embedding method and extend its applications.

Author Contributions: All authors have contributed substantially to this work: Conceptualization, Z.J. (Ziyu Jia) and J.W.; Data curation, Z.J. (Ziyu Jia) and Y.M.; Formal analysis, Z.J. (Ziyu Jia), Z.J. (Zehui Jiao), Y.M. and J.W.; Funding acquisition, Y.L. and J.W.; Investigation, Z.J. (Ziyu Jia) and J.W.; Methodology, Z.J. (Ziyu Jia) and J.W.; Project administration, Y.L. and J.W.; Resources, Y.L. and J.W.; Software, Z.J. (Ziyu Jia); Supervision, J.W.;

Validation, Z.J. (Ziyu Jia) and J.W.; Visualization, Z.J. (Ziyu Jia) and Z.J. (Zehui Jiao); Writing—original draft, Z.J. (Ziyu Jia); Writing—review and editing, Y.L., Z.J. (Zehui Jiao), Y.M. and J.W.

Funding: Financial support by the National Natural Science Foundation of China (61603029) and the Fundamental Research Funds for the Central Universities (2018RC002) are gratefully acknowledged.

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

References

1. Paluš, M.; Krakovská, A.; Jakubík, J.; Chvosteková, M. Causality, dynamical systems and the arrow of time. *Chaos Interdiscip. J. Nonlinear Sci.* **2018**, *28*, 075307. [[CrossRef](#)] [[PubMed](#)]
2. Krakovská, A.; Jakubík, J.; Chvosteková, M.; Coufal, D.; Jajcay, N.; Paluš, M. Comparison of six methods for the detection of causality in a bivariate time series. *Phys. Rev. E* **2018**, *97*, 042207. [[CrossRef](#)] [[PubMed](#)]
3. Paluš, M.; Komárek, V.; Hrnčíř, Z.; Štěrbová, K. Synchronization as adjustment of information rates: Detection from bivariate time series. *Phys. Rev. E* **2001**, *63*, 046211. [[CrossRef](#)] [[PubMed](#)]
4. Granger, C.W.J. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **1969**, 424–438. [[CrossRef](#)]
5. Marko, H. The bidirectional communication theory—a generalization of information theory. *IEEE Trans. Commun.* **1973**, *21*, 1345–1351. [[CrossRef](#)]
6. Schreiber, T. Measuring information transfer. *Phys. Rev. Lett.* **2000**, *85*, 461. [[CrossRef](#)]
7. Barnett, L.; Barrett, A.B.; Seth, A.K. Granger causality and transfer entropy are equivalent for Gaussian variables. *Phys. Rev. Lett.* **2009**, *103*, 238701. [[CrossRef](#)]
8. Schindlerova, K. Equivalence of Granger Causality and Transfer Entropy: A Generalization. *Appl. Mathem. Sci.* **2011**, *5*, 3637–3648.
9. Mao, X.; Shang, P. Transfer entropy between multivariate time series. *Commun. Nonlinear Sci. Numer. Simul.* **2017**, *47*, 338–347. [[CrossRef](#)]
10. Montalto, A.; Stramaglia, S.; Faes, L.; Tessitore, G.; Prevete, R.; Marinazzo, D. Neural networks with non-uniform embedding and explicit validation phase to assess Granger causality. *Neural Netw.* **2015**, *71*, 159–171. [[CrossRef](#)]
11. Faes, L.; Nollo, G.; Porta, A. Information-based detection of nonlinear Granger causality in multivariate processes via a nonuniform embedding technique. *Phys. Rev. E* **2011**, *83*, 051112. [[CrossRef](#)] [[PubMed](#)]
12. Runge, J. Causal network reconstruction from time series: From theoretical assumptions to practical estimation. *Chaos Interdiscip. J. Nonlinear Sci.* **2018**, *28*, 075310. [[CrossRef](#)] [[PubMed](#)]
13. Zhao, X.; Sun, Y.; Li, X.; Shang, P. Multiscale transfer entropy: Measuring information transfer on multiple time scales. *Commun. Nonlinear Sci. Numer. Simul.* **2018**, *62*, 202–212. [[CrossRef](#)]
14. Sysoev, I.V.; Ponomarenko, V.I.; Pikovsky, A. Reconstruction of coupling architecture of neural field networks from vector time series. *Commun. Nonlinear Sci. Numer. Simul.* **2018**, *57*, 342–351. [[CrossRef](#)]
15. Vlachos, I.; Kugiumtzis, D. Nonuniform state-space reconstruction and coupling detection. *Phys. Rev. E* **2010**, *82*, 016207. [[CrossRef](#)]
16. Kugiumtzis, D. Direct-coupling information measure from nonuniform embedding. *Phys. Rev. E* **2013**, *87*, 062918. [[CrossRef](#)]
17. Kugiumtzis, D.; Koutlis, C.; Tsimpiris, A.; Kimiskidis, V.K. Dynamics of epileptiform discharges induced by transcranial magnetic stimulation in genetic generalized epilepsy. *Int. J. Neural Syst.* **2017**, *27*, 1750037. [[CrossRef](#)]
18. Kugiumtzis, D.; Kimiskidis, V.K. Direct causal networks for the study of transcranial magnetic stimulation effects on focal epileptiform discharges. *Int. J. Neural Syst.* **2015**, *25*, 1550006. [[CrossRef](#)]
19. Papan, A.; Kyrtsov, C.; Kugiumtzis, D.; Diks, C. Financial networks based on Granger causality: A case study. *Physica A* **2017**, *482*, 65–73. [[CrossRef](#)]
20. Papan, A.; Kyrtsov, C.; Kugiumtzis, D.; Diks, C. Assessment of resampling methods for causality testing: A note on the US inflation behavior. *PLoS ONE* **2017**, *12*, e0180852. [[CrossRef](#)]
21. Runge, J.; Heitzig, J.; Petoukhov, V.; Kurths, J. Escaping the curse of dimensionality in estimating multivariate transfer entropy. *Phys. Rev. Lett.* **2012**, *108*, 258701. [[CrossRef](#)] [[PubMed](#)]
22. Runge, J.; Donner, R.V.; Kurths, J. Optimal model-free prediction from multivariate time series. *Phys. Rev. E* **2015**, *91*, 052909. [[CrossRef](#)] [[PubMed](#)]

23. Brown, G.; Pocock, A.; Zhao, M.J.; Luján, M. Conditional likelihood maximisation: A unifying framework for information theoretic feature selection. *J. Mach. Learn. Res.* **2012**, *13*, 27–66.
24. Vinh, N.X.; Zhou, S.; Chan, J.; Bailey, J. Can high-order dependencies improve mutual information based feature selection? *Pattern Recognit.* **2016**, *53*, 46–58. [[CrossRef](#)]
25. Battiti, R. Using mutual information for selecting features in supervised neural net learning. *IEEE Trans. Neural Netw. Learn. Syst.* **1994**, *5*, 537–550. [[CrossRef](#)]
26. Yang, H.H.; Moody, J. Data visualization and feature selection: New algorithms for nongaussian data. *NIPS* **2000**, 687–693.
27. Fleuret, F. Fast binary feature selection with conditional mutual information. *J. Mach. Learn. Res.* **2004**, *5*, 1531–1555.
28. Peng, H.; Long, F.; Ding, C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226–1238. [[CrossRef](#)]
29. Lin, D.; Tang, X. Conditional infomax learning: An integrated framework for feature extraction and fusion. *ECCV* **2006**, 68–82.
30. Runge, J.; Nowack, P.; Kretschmer, M.; Flaxman, S.; Sejdinovic, D. Detecting causal associations in large nonlinear time series datasets. *arXiv* **2017**, arXiv:1702.07007.
31. Meyer, P.E.; Bontempi, G. On the use of variable complementarity for feature selection in cancer classification. *ECCV* **2006**, 91–102.
32. 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*, e109462. [[CrossRef](#)] [[PubMed](#)]
33. Faes, L.; Kugiumtzis, D.; Nollo, G.; Jurysta, F.; Marinazzo, D. Estimating the decomposition of predictive information in multivariate systems. *Phys. Rev. E* **2015**, *91*, 032904. [[CrossRef](#)]
34. Hacine-Gharbi, A.; Ravier, P.; Harba, R.; Mohamadi, T. Low bias histogram-based estimation of mutual information for feature selection. *Pattern Recognit. Lett.* **2012**, *33*, 1302–1308. [[CrossRef](#)]
35. Kwak, N.; Choi, C.H. Input feature selection by mutual information based on Parzen window. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *12*, 1667–1671. [[CrossRef](#)]
36. Kraskov, A.; Stögbauer, H.; Grassberger, P. Estimating mutual information. *Phys. Rev. E* **2004**, *69*, 066138. [[CrossRef](#)]
37. Schelter, B.; Winterhalder, M.; Hellwig, B.; Guschlbauer, B.; Lüking, C.H.; Timmer, J. Direct or indirect? Graphical models for neural oscillators. *J. Physiol. Paris* **2006**, *99*, 37–46. [[CrossRef](#)]
38. Gourévitch, B.; Le Bouquin-Jeannès, R.; Faucon, G. Linear and nonlinear causality between signals: Methods, examples and neurophysiological applications. *Biol. Cybern.* **2006**, *95*, 349–369. [[CrossRef](#)]
39. Kramer, M.A.; Kolaczyk, E.D.; Kirsch, H.E. Emergent network topology at seizure onset in humans. *Epilepsy Res.* **2008**, *79*, 173–186. [[CrossRef](#)]
40. Marinazzo, D.; Pellicoro, M.; Stramaglia, S. Kernel method for nonlinear Granger causality. *Phys. Rev. Lett.* **2008**, *100*, 144103. [[CrossRef](#)]
41. Faes, L.; Marinazzo, D.; Stramaglia, S. Multiscale information decomposition: Exact computation for multivariate Gaussian processes. *Entropy* **2017**, *19*, 408. [[CrossRef](#)]
42. Porta, A.; Faes, L. Wiener–Granger causality in network physiology with applications to cardiovascular control and neuroscience. *Proc. IEEE* **2016**, *104*, 282–309. [[CrossRef](#)]
43. Sun, J.; Taylor, D.; Bollt, E.M. Causal network inference by optimal causation entropy. *SIAM J. Appl. Dyn. Syst.* **2015**, *14*, 73–106. [[CrossRef](#)]

