MDPI

Article

Causalized Convergent Cross Mapping and Its Implementation in Causality Analysis

Boxin Sun [1], Jinxian Deng [1], Norman Scheel [2], David C. Zhu [3], Jian Ren [1], Rong Zhang [4,5] and Tongtong Li [1,6,*,†]

1   Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI 48824, USA; sunboxin@msu.edu (B.S.); dengjinx@msu.edu (J.D.); renjian@msu.edu (J.R.)
2   Department of Radiology, Michigan State University, East Lansing, MI 48824, USA; scheelno@msu.edu
3   Department of Radiology, Albert Einstein College of Medicine, Bronx, NY 10461, USA; zhuda@msu.edu
4   Institute for Exercise and Environmental Medicine, Texas Health Presbyterian Hospital Dallas, Dallas, TX 75231, USA; rongzhang@texashealth.org
5   Department of Neurology and Internal Medicine, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
6   Michigan Alzheimer's Disease Research Center, Ann Arbor, MI 48109, USA
*   Correspondence: tongli@msu.edu; Tel.: +1-517-355-5230
†   Current address: 2120 Engineering Building, 428 S. Shaw Lane, East Lansing, MI 48824-1226, USA.

Abstract: Rooted in dynamic systems theory, convergent cross mapping (CCM) has attracted increased attention recently due to its capability in detecting linear and nonlinear causal coupling in both random and deterministic settings. One limitation with CCM is that it uses both past and future values to predict the current value, which is inconsistent with the widely accepted definition of causality, where it is assumed that the future values of one process cannot influence the past of another. To overcome this obstacle, in our previous research, we introduced the concept of causalized convergent cross mapping (cCCM), where future values are no longer used to predict the current value. In this paper, we focus on the implementation of cCCM in causality analysis. More specifically, we demonstrate the effectiveness of cCCM in identifying both linear and nonlinear causal coupling in various settings through a large number of examples, including Gaussian random variables with additive noise, sinusoidal waveforms, autoregressive models, stochastic processes with a dominant spectral component embedded in noise, deterministic chaotic maps, and systems with memory, as well as experimental fMRI data. In particular, we analyze the impact of shadow manifold construction on the performance of cCCM and provide detailed guidelines on how to configure the key parameters of cCCM in different applications. Overall, our analysis indicates that cCCM is a promising and easy-to-implement tool for causality analysis in a wide spectrum of applications.

Keywords: causality; causalized convergent cross mapping; directed information

1. Introduction

Causality analysis aims to find the relationship between causes and effects by exploring the directional influence of one variable on the other, and it has been a central topic in science, economy, climate, and many other fields [1–9]. Compared with correlation, which reflects the mutual dependence between two variables, causality analysis may provide additional information since two time series with low correlation may have strong unidirectional or bi-directional causal coupling between them. Some representative examples can be found in [9].

The first practical causal analysis framework is Granger Causality (GC), which was proposed by Granger in 1969 [10]. GC is a statistical approach that relies on a multi-step linear prediction model and aims to determine whether the values of one time series are useful in predicting the future values of the other. As a well-known technique, the validity

and computational simplicity of GC have been widely recognized [11–14]. At the same time, it has also been noticed that when there exists instantaneous and/or strong nonlinear interactions between two regions, GC analysis may lead to invalid results [9,15]. Moreover, GC may not be able to detect the causation in deterministic settings [10,16].

In 1990, directed information (DI)—the first causality detection tool based on information theory—was proposed by Massey [17] when studying discrete memoryless communication channels with feedback. DI measures the directed information flowing from one sequence to the other. As an information-theoretic framework, a major advantage of DI is that it is a universal method that does not rely on any model assumptions of the signals and is not limited by linearity or separability [18,19]. In refs. [9,18], the performance of DI in causality analysis was demonstrated using both simulated data and experimental fMRI data. It was found that DI is capable of detecting both linear and non-linear causal relationships. However, it was also noticed that the direct evaluation of DI relies heavily on probability estimation and tends to be sensitive to data length as well as the step size used in the quantization process [9].

In 2012, convergent cross mapping (CCM), a new causality model based on state space reconstruction was proposed by Sugihara et al. [16], and it was demonstrated that CCM could serve as an effective tool in addressing non-separable systems and identifying weakly coupled variables under deterministic settings, which may not be covered by GC. Since then, CCM has attracted considerable attention from the research community in many different fields [20–28].

Recall that causality aims to determine whether the current and past values of one time series are useful in predicting the future values of another in addition to its own past values. In CCM, however, both the past and future values are utilized to reconstruct the current value [9]. As a result, the causality defined by CCM is inconsistent with the original, widely accepted definition of causality where the key assumption is that the future values of one process cannot influence the past of the other.

Motivated by this observation, in [9], we introduced the concept of causalized convergent cross mapping (cCCM). More specifically, if only the current and historical values of X and the past values of Y are used to predict the current value $Y(t)$, and vice versa, then CCM is converted to causalized CCM. We further proved the approximate equivalence of DI and cCCM under stationary ergodic Gaussian random processes [9].

This study is a continued work of our previous research [9] and is focused on the implementation perspective of cCCM in causality detection. More specifically, in this study, we aimed to further investigate the effectiveness of cCCM in identifying both linear and nonlinear causal coupling in various settings through a large number of examples, including Gaussian random variables with additive noise, sinusoidal waveforms, autoregressive models, stochastic processes with a dominant spectral component embedded in noise, deterministic chaotic maps, and systems with memory, as well as experimental functional Magnetic Resonance Imaging (fMRI) data. In particular, we analyze the impact of shadow manifold construction on the performance of cCCM and provide detailed guidelines on how to configure the key parameters of cCCM (especially the shadow manifold dimension and time lag) in different applications. Moreover, we examine the noise effect in cCCM and show that, in general, reliable causality detection can be achieved when the signal-to-noise ratio (SNR) is larger than 15 dB. Overall, our analysis indicates that CCM is a promising and easy-to-implement tool for causality analysis in a wide spectrum of applications.

The rest of the paper is organized as follows. In Section 2, we briefly revisit the original CCM, the causalized CCM (cCCM), the conditional equivalence between cCCM and DI, and the extension of bivariate cCCM to multivariate cCCM. In Section 3, we present the major results of the study, where we discuss the impact of noise on the performance of cCCM, evaluate the effectiveness of cCCM in causality analysis through numerous examples, provide detailed guidelines on the configuration of cCCM, and compare the performances of bivariate and multivariate cCCM and GC through both simulation examples and ex-

perimental fMRI data. Finally, we present the conclusions drawn from this research and provide related discussions in Section 4.

2. A Revisit of Causalized Convergent Cross Mapping

In this section, we first briefly revisit convergent cross mapping (CCM) [16] and introduce the concept of causalized CCM (cCCM) [9]. We then present the conditional equivalence of cCCM and the directed information framework [9] and introduce the extension of bivariate cCCM to multivariate cCCM.

Convergent cross mapping (CCM). The CCM algorithm is based on state-space reconstruction. Consider two dynamically coupled variables X and Y that share the same attractor manifold $\mathbf{M}$. Let $\mathbf{X}^n = [X_1, X_2, \cdots, X_n]$ and $\mathbf{Y}^n = [Y_1, Y_2, \cdots, Y_n]$ be the time series consisting of samples of X and Y, respectively. The CCM causality analysis framework can be summarized as follows:

- **Step 1:** Construct the shadow manifolds with respect to $\mathbf{X}^n$ and $\mathbf{Y}^n$.

$$\mathbf{M}_x = \{\mathbf{x}_t \mid \mathbf{x}_t = [X_t, X_{t-\tau}, \cdots, X_{t-(E-1)\tau}], \ t = (E-1)\tau + 1, \cdots, n\}, \quad (1)$$

$$\mathbf{M}_y = \{\mathbf{y}_t \mid \mathbf{y}_t = [Y_t, Y_{t-\tau}, \cdots, Y_{t-(E-1)\tau}], \ t = (E-1)\tau + 1, \cdots, n\}. \quad (2)$$

- **Step 2:** For each vector $\mathbf{x}_t$, find its $E + 1$ nearest neighbors and denote the time indices (from closest to farthest) of the $E + 1$ nearest neighbors of $\mathbf{x}_t$ by $t_1, \cdots, t_{E+1}$.
- **Step 3:** If the two signals X and Y are dynamically coupled, then the nearest neighbors of $\mathbf{x}_t$ in $\mathbf{M}_x$ would be mapped to the nearby points of Y_t on manifold $\mathbf{M}$. The estimated Y_t based on $\mathbf{M}_x$, or say the cross mapping from X to Y, is defined as

$$\hat{Y}_t | \mathbf{M}_x = \sum_{i=1}^{E+1} w_i Y_{t_i} \quad (3)$$

where

$$w_i = \frac{u_i}{\sum\limits_{j=1}^{E+1} u_j}, \ \text{with} \ u_i = exp\{-\frac{d(\mathbf{x}_t, \mathbf{x}_{t_i})}{d(\mathbf{x}_t, \mathbf{x}_{t_1})}\},$$

where d denotes the Euclidean distance between two vectors. Please note that for every i, ω_i is a function of t. The cross mapping from Y to X can be defined in a similar way. As n increases, it is expected that $\hat{X}_t | \mathbf{M}_y$ and $\hat{Y}_t | \mathbf{M}_x$ would converge to X_t and Y_t, respectively.

- **Step 4:** The cross mapping correlations are defined as

$$\rho_{CCM}(X \rightarrow Y) = \rho(\mathbf{Y}^n, \hat{\mathbf{Y}}^n) \ \text{and} \ \rho_{CCM}(Y \rightarrow X) = \rho(\mathbf{X}^n, \hat{\mathbf{X}}^n) \quad (4)$$

where ρ denotes the Pearson correlation.

- **Step 5:** If $\rho_{CCM}(X \rightarrow Y) > \rho_{CCM}(Y \rightarrow X)$ and converges faster than $\rho_{CCM}(Y \rightarrow X)$, then we say that the causal effect of X on Y is stronger than that in the reverse.

Geometric illustration of convergent cross mapping. Here, we provide the geometric illustration of convergent cross mapping from the shadow manifold $\mathbf{M}_x$ to the shadow manifold $\mathbf{M}_y$ under both strong and weak causal coupling.

Figure 1a corresponds to the situation when there is a strong causal relationship from X to Y, and Figure 1b illustrates the case when there is only a weak causation. For illustration purpose, the dimension of the shadow manifold was chosen to be $E = 2$, the neighborhood of $\mathbf{x}_t$ is represented using a the simplex consisting of three nearest neighbors, and the neighborhood of $\mathbf{y}_t$ is represented in the same way.
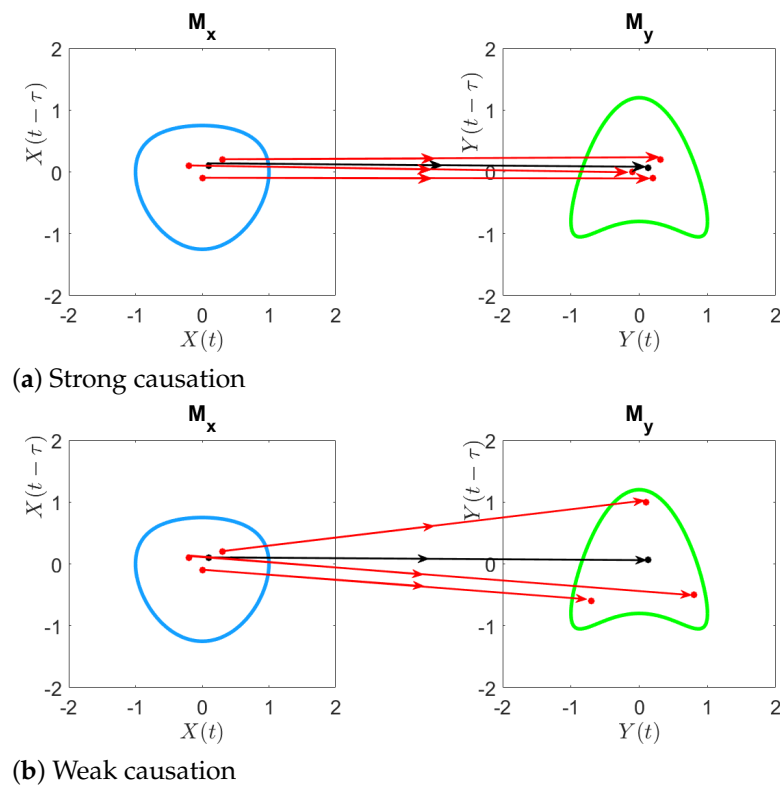
(**a**) Strong causation

(**b**) Weak causation

Figure 1. Geometric illustration of the cross mapping from $\mathbf{M}_x$ to $\mathbf{M}_y$. (**a**) When strong causation exists from X to Y, the nearest neighbors of $\mathbf{x}_t$ are mapped to the nearest neighbors of $\mathbf{y}_t$. (**b**) When there is only weak causation from X to Y, the nearest neighbors of $\mathbf{x}_t$ are no longer mapped to the nearest neighbors of $\mathbf{y}_t$.

Causalized convergent cross mapping (cCCM). Note that in CCM, both the past and future values are used in data reconstruction, which is inconsistent with the original definition of causality where it is assumed that the future values of one process cannot impact the past of another. For this reason, we propose the concept of causalized convergent cross mapping (cCCM).

More specifically, in CCM, if we limit the search of all the nearest neighbors in $\mathbf{M}_x$ to $t_i < t$, i.e., we only use the current and previous values of X and the past values of Y to predict the current value Y_t, operating in the same way for the other direction, and then we obtained causalized CCM. That being said, **Step 2** in cCCM now becomes

- **Step 2 for cCCM:** For each vector $\mathbf{x}_t$, find its $E+1$ nearest neighbors in $\mathbf{M}_x$ with an index smaller than t and denote the time indices (from closest to farthest) of the $E+1$ nearest neighbors of $\mathbf{x}_t$ by $t_1, ... t_{E+1}$. Note that for $i = 1, 2, \cdots, E+1$, we now have $t_i < t$.

Then, we follow Steps 3–5 above, and denote the corresponding causalized cross mapping correlation, or the cCCM causation, as ρ_{cCCM}.

Conditional equivalence between cCCM and directed information. As an information-theoretic causality model, directed information (DI) measures the information flow from one time series to the other. DI plays a central role in causality analysis for two reasons. First, it is a universal method that does not have any modeling constraints on the sequences to be evaluated [29,30]. Second, DI serves as the pivot that links existing causality models GC [10,18], transfer entropy (TE) [9,31,32], and dynamic causal modeling (DCM) [33,34] through conditional equivalence between them.

In [9], we proved the conditional equivalence between cCCM and DI under Gaussian variables and used DI as a bridge to connect cCCM to other representative tools of causality analysis. More specifically, we showed that if (i) X and Y are dynamically coupled, zero-

mean Gaussian random variables and their joint distribution is bivariate Gaussian, and (ii) $\mathbf{X}^n, \mathbf{Y}^n$ are stationary ergodic random processes; then, when n is sufficiently large,

$$\bar{I}_n(X \to Y) \approx -\frac{1}{2} \log(1 - \rho_{cCCM}^2(X \to Y)), \tag{5}$$

where $\bar{I}_n(X \to Y)$ denotes the average DI from X to Y, measured in bits per sample. The conditional equivalence of DI and cCCM under Gaussian random variables was demonstrated in [9] using experimental fMRI data.

This result also connects cCCM to other representative causality analysis frameworks in the family—GC, TE (Transfer Entropy, 2000 [31]), and DCM (Dynamic causal modelling, 2003 [33])—through the conditional equivalence between them under Gaussian random variables [9,12].

It is worth pointing out that the simulation-based analysis in [9] suggested that cCCM is often more robust in causality detection than DI. This is mainly because the DI calculation is based on probability estimation, which is sensitive to the step size used in the quantization process[35]. cCCM, on the other hand, gets around this obstacle through geometric cross mapping between the corresponding shadow manifolds, at the cost of a higher computational complexity. More specifically, cCCM relies on a K-nearest neighbor search and has a computational complexity of $O(n^2)$ in the sequence length n, but the computational complexity of DI is only $O(n)$.

Extension of bivariate cCCM to multivariate cCCM. Bivariate cCCM can be extended to multivariate conditional cCCM [9] based on a multivariate KNN search, which takes a similar approach as in the multivariate KNN predictability approaches [36–39].

Let $\mathbf{\Omega} = \{X_1, \cdots, X_L\}$ denote the set of dynamically coupled random variables that share the same attractor manifold. As shown in [9], the multivariate conditional cCCM from $X_j \to X_i$ with respect to $\mathbf{\Omega} \smallsetminus \{X_i, X_j\}$ (i.e., all the remaining random variables in $\mathbf{\Omega}$) is defined using the causality ratio as

$$\text{cCCM}(X_j \to X_i | \mathbf{\Omega} \smallsetminus \{X_i, X_j\}) = \frac{\text{Var}(\mathbf{e}_i^n | \mathbf{\Omega} \smallsetminus \{X_j\}) - \text{Var}(\mathbf{e}_i^n | \mathbf{\Omega})}{\text{Var}(\mathbf{e}_i^n | \mathbf{\Omega} \smallsetminus \{X_j\})},$$

where $\mathbf{e}_i^n | \mathbf{\Omega} \smallsetminus \{X_j\}$ denotes the estimation error vector based on $\mathbf{\Omega} \smallsetminus \{X_j\}$, and $\mathbf{e}_i^n | \mathbf{\Omega}$ is the estimation error vector based on the whole $\mathbf{\Omega}$. The definition can be adjusted by modifying $\mathbf{\Omega}$ to reflect the conditional cCCM with respect to either an individual random variable or a group of random variables.

3. Results

3.1. The Impact of Estimation Error in cCCM

Note that CCM and cCCM are based on data reconstruction, and the reconstructed data converge to the true data as the data length goes to infinity when there exists causal coupling between the random variables under consideration. Here, we consider the impact of estimation error in cCCM.

As an example, we consider $\rho_{cCCM}(X \to Y) = \rho(\mathbf{Y}^n, \hat{\mathbf{Y}}^n) \approx \rho(Y, \hat{Y})$. Note that

$$\hat{Y}_t | \mathbf{M}_x = \sum_{i=1}^{E+1} w_i Y_{t_i} \tag{6}$$

where $t_i < t$, and

$$w_i = \frac{u_i}{\sum_{j=1}^{E+1} u_j}, \quad with \ u_i = exp\{-\frac{d(\mathbf{x}_t, \mathbf{x}_{t_i})}{d(\mathbf{x}_t, \mathbf{x}_{t_1})}\}.$$

When there exists estimation error, we can model $\hat{Y}$ as

$$\hat{Y} = Y + n_e \tag{7}$$

where n_e denotes the estimation error, which is independent of Y. In assuming n_e is of zero-mean and variance σ_e^2, it can be shown that (please refer to the Supplementary file of [9])

$$\rho(Y, \hat{Y}) = \frac{\sigma_Y}{\sqrt{\sigma_Y^2 + \sigma_e^2}}. \tag{8}$$

where σ_Y^2 denotes the variance of Y. This result implies that the cCCM value $\rho_{\text{cCCM}}(X \to Y) \approx \rho(Y, \hat{Y})$ decreases as the estimation error power increases.

In the following, using the noise-free case as the benchmark, we examine the noise effect on cCCM through simulation examples, including Gaussian random variable and its signed and squared versions (**Examples 1** and **2**, respectively), as well as sinusoidal waveforms (**Example 3**). As shown in Table 1, when we increase the SNR from 0 dB to 20 dB, the cCCM value of the noisy signal gradually converges to the noise-free result. More specifically, our results suggest that reliable causality detection can be achieved when the SNR is larger than 15 dB.

Table 1. Impact of estimation error (n_1 and n_2 are AWGN noise generated independently of X and Y, respectively.)

Examples	Direction	ρ_{cCCM} under Different SNR Values					
		0 dB	5 dB	10 dB	15 dB	20 dB	Noise Free
1. $X_0 = \text{randn}(1000, 1)$, $X = X_0 + n_1$, $Y = \text{sgn}(X_0) + n_2$,	$X \to Y$	0.1896	0.4691	0.6951	0.7819	0.8162	0.8215
	$Y \to X$	0.2598	0.5522	0.6858	0.7133	0.7190	0.7807
	Difference	-0.0701	-0.0830	0.0093	0.0686	0.0972	0.0408
2. $X_0 = \text{randn}(1000, 1)$, $X = X_0 + n_1$, $Y = X_0^2 + n_2$,	$X \to Y$	0.1335	0.4332	0.6946	0.8141	0.8460	0.8639
	$Y \to X$	0.0033	0.0197	0.0510	0.1028	0.0473	0.0290
	Difference	0.1302	0.4134	0.6435	0.7113	0.7987	0.8349
3. $X(t) = \sin(t) + n_1$, $Y(t) = \cos(t) + n_2$. $t = 0 : 0.01 : 2\pi$	$X \to Y$	0.1960	0.3234	0.4728	0.6708	0.8917	0.9999
	$Y \to X$	0.3281	0.4981	0.6513	0.7761	0.9080	0.9999
	Difference	-0.1321	-0.1748	-0.1785	-0.1053	-0.0163	0

Here, randn(1000,1) returns a 1000-by-1 matrix of normally distributed random numbers.

The performance of cCCM is not only affected by noise but also closely related to the selection of E and τ. For **Examples 1** and **2** in Table 1, we chose $E = 5$ and $\tau = 1$. For **Example 3**, we used $E = 5$ and $\tau = 5$. Here, a larger τ is used mainly because $X(t)$ and $Y(t)$ are significantly over-sampled in **Example 3**. More discussion on the choice of shadow manifold parameters can be found in Section 3.2.

3.2. Causality Detection Using cCCM and the Choice of Shadow Manifold Parameters

In this section, we illustrate the performance of cCCM (together with CCM) in causality detection through simulation examples, including autoregressive models, stochastic processes with a dominant spectral component embedded in noise, deterministic chaotic maps, and systems with memory. As will be seen, these examples show that CCM and cCCM are sensitive to changes in coupling strength. It can be observed that CCM tends to result in larger causation values than cCCM; this is expected since CCM uses both the past and future values of X to predict the current value of Y (and vice versa), while cCCM only uses both the past values of X to predict the current value of Y (and vice versa).

We will also discuss the choice of key parameters—the dimension of shadow manifold E and the time lag τ—in the cCCM algorithm and the impact of these parameters on the detection of causal relationships. According to Takens' theorem [40] and Whitney's embedding theorem [41,42], the "magic number" is $E = 2d + 1$, and often less [16], where d is the dimension of the attractor **M** shared by X and Y. Another parameter, the time lag τ, is generally chosen as $\tau = 1$. When the signal is over-sampled, $\tau > 1$ can also be used.

It should be noted that for an accurate assessment of the causation, the sampling rate should always be chosen to be larger than the Nyquist rate. Otherwise, the causal relationship identified by cCCM may be invalid since the under-sampled sequences cannot capture the total information in the original signals.

3.2.1. Examples on Autoregressive Models

Example 4:

Let X and Y be random processes given by

$$X(t+1) = 0.5X(t) + 0.05Y(t) + n_1(t),$$
$$Y(t+1) = 0.65X(t) + 0.08Y(t) + n_2(t),$$

where $n_1, n_2 \sim N(0, 0.05^2)$, n_1 and n_2 are independent, $t = [0, 1, 2, \cdots, 2047]$, and $X(0) = Y(0) = 1.5$. We chose $E = 5$ and $\tau = 1$, and then the cCCM and CCM values between these two time series are

$$\rho_{\text{cCCM}}(X \to Y) = 0.5067, \ \rho_{\text{cCCM}}(Y \to X) = 0.2210,$$
$$\rho_{\text{CCM}}(X \to Y) = 0.5165, \ \rho_{\text{CCM}}(Y \to X) = 0.2294.$$

The convergence of CCM and cCCM with respect to the data length is shown in Figure 2.

Example 5:

Let X and Y be random processes given by

$$X(t+1) = 0.6X(t) + 0.3Y(t) + n_1(t),$$
$$Y(t+1) = 0.02X(t) + 0.8Y(t) + n_2(t),$$

where $n_1, n_2 \sim N(0, 0.05^2)$, n_1 and n_2 are independent, $t = [0, 1, \cdots, 2047]$, and $X(0) = Y(0) = 1.5$. Then, the cCCM and CCM values between these two time series are

$$\rho_{\text{cCCM}}(X \to Y) = 0.3599, \ \rho_{\text{cCCM}}(Y \to X) = 0.5589$$
$$\rho_{\text{CCM}}(X \to Y) = 0.4140, \ \rho_{\text{CCM}}(Y \to X) = 0.6222$$

The convergence of CCM and cCCM with respect to the data length is shown in Figure 2.
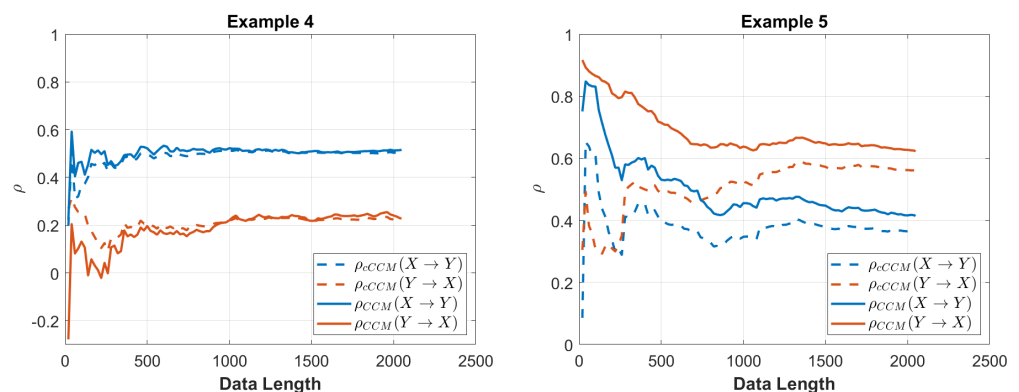

Figure 2. Performance of cCCM and CCM versus the data length for Examples 4 and 5.

3.2.2. Examples on Stochastic Processes with a Dominant Spectral Component

Example 6:

Let X and Y be two stochastic processes given by

$$X(t) = 0.1 \sin(5\pi t) + 0.6 \sin(20\pi t) + n_1(t)$$
$$Y(t) = 0.6 \sin(20\pi t) + n_2(t)$$

where n_1, n_2 are independent AWGN noise with SNR $= 10$ dB, and $t = 0 : 0.005 : 2$ (here, 0.005 is the step size). Then, the cCCM and CCM values between these two time series are

$$\rho_{\text{cCCM}}(X \to Y) = 0.9314, \ \rho_{\text{cCCM}}(Y \to X) = 0.9175$$
$$\rho_{\text{CCM}}(X \to Y) = 0.9362, \ \rho_{\text{CCM}}(Y \to X) = 0.9242$$

The convergence of CCM and cCCM with respect to the data length is shown in Figure 3.

Example 7:

Let X and Y be two stochastic processes given by

$$X(t) = 0.6 \sin(5\pi t) + 0.1 \sin(20\pi t) + n_1(t)$$
$$Y(t) = 0.1 \sin(20\pi t) + n_2(t)$$

where n_1 and n_2 are independent AWGN noise with SNR$= 10$dB, and $t = 0 : 0.005 : 2$. Then, the cCCM and CCM values between these two time series are

$$\rho_{\text{cCCM}}(X \to Y) = 0.7108, \ \rho_{\text{cCCM}}(Y \to X) = 0.0616.$$
$$\rho_{\text{CCM}}(X \to Y) = 0.7657, \ \rho_{\text{CCM}}(Y \to X) = 0.0517.$$

The convergence of CCM and cCCM with respect to the data length is shown in Figure 3.
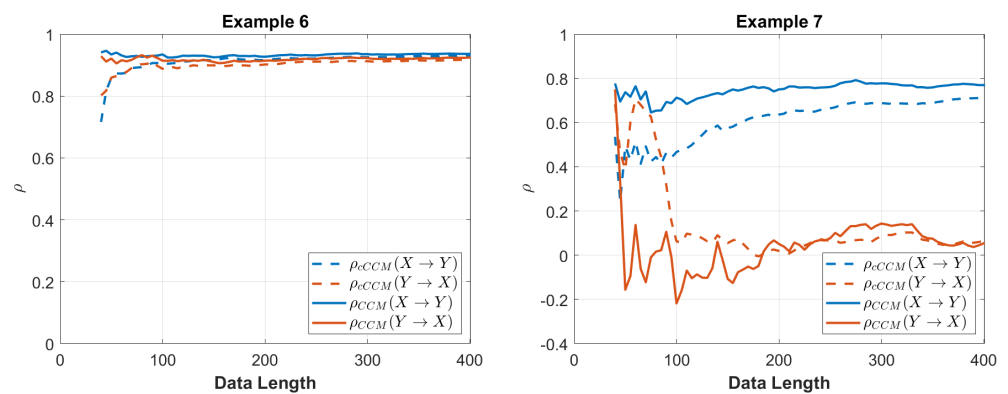
Figure 3. Performance of cCCM and CCM versus the data length for Examples 6 and 7.

We selected $\tau = 5$ in **Examples 6** and **7** to reduce the impact of noise; please refer to Section 3.2.5 for more details.

3.2.3. Examples on Deterministic Chaotic Maps

Example 8:

Let X and Y be two stochastic processes given by

$$X(t+1) = X(t)[3.8 - 3.8X(t)],$$
$$Y(t+1) = Y(t)[3.2 - 3.2Y(t) - 0.1X(t)],$$

where $t = [0, 1, \cdots, 2047]$, $X(0) = 0.7$, and $Y(0) = 0.1$. Then, the cCCM and CCM values between these two time series are

$$\rho_{cCCM}(X \to Y) = 0.2164, \ \rho_{cCCM}(Y \to X) = 0.8923.$$
$$\rho_{CCM}(X \to Y) = 0.1679, \ \rho_{CCM}(Y \to X) = 0.9705.$$

The convergence of CCM and cCCM with respect to the data length is shown in Figure 4.

Example 9:

Let X and Y be two stochastic processes given by

$$X(t+1) = X(t)[3.8 - 3.8X(t) - 0.1Y(t)],$$
$$Y(t+1) = Y(t)[3.2 - 3.2Y(t) - 0.1X(t)],$$

where $t = [0, 2047]$, $X(0) = 0.7$, and $Y(0) = 0.1$. Then, the cCCM and CCM values between these two time series are

$$\rho_{cCCM}(X \to Y) = 0.8693, \ \rho_{cCCM}(Y \to X) = 0.9122.$$
$$\rho_{CCM}(X \to Y) = 0.9704, \ \rho_{CCM}(Y \to X) = 0.9717.$$

The convergence of CCM and cCCM with respect to the data length is shown in Figure 4.
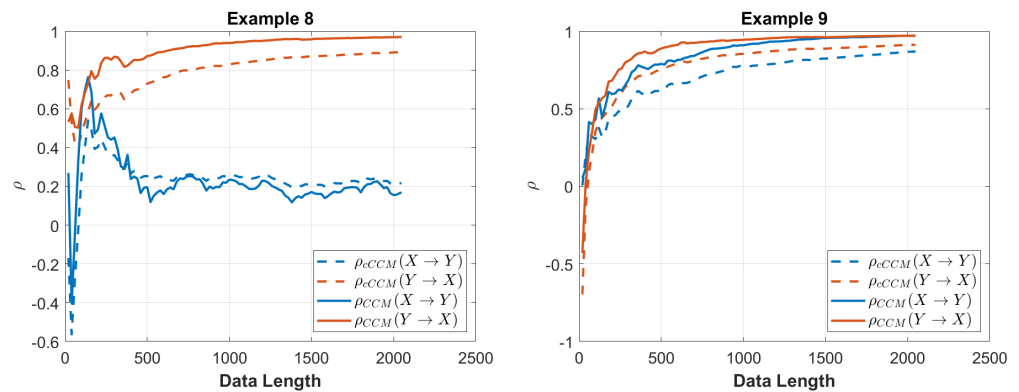
Figure 4. Performance of cCCM and CCM versus the data length for Examples 8 and 9.

3.2.4. Examples on Systems with Memory

In this subsection, we examined the causal relationship in systems with memory (Examples 10–14) using CCM and cCCM under different choices of E and τ.

Example 10:

Consider a system with memory

$$X = \text{randn}(1024, 1),$$
$$Y(t) = 0.2X(t-1) + 0.85X(t-4).$$

Here, (i) the MATLAB command "randn(1024,1)" returns an 1024-by-1 matrix of normally distributed random numbers; (ii) $t = [0, 1, \cdots, 1023]$ and $X(t) = 0$ while $t < 0$. The results corresponding to different E or τ values are displayed in Table 2.

Table 2. Results for **Example 10**.

Values for E and τ	cCCM	CCM
$E = 3, \tau = 1$	$\rho_{\text{cCCM}}(X \to Y) : 0.0198$ $\rho_{\text{cCCM}}(Y \to X) : -0.0600$ $\text{MSE}(\mathbf{X}^n, \hat{\mathbf{X}}^n) : 1.2416$ $\text{MSE}(\mathbf{Y}^n, \hat{\mathbf{Y}}^n) : 0.9031$	$\rho_{\text{CCM}}(X \to Y) : 0.0495$ $\rho_{\text{CCM}}(Y \to X) : 0.0087$ $\text{MSE}(\mathbf{X}^n, \hat{\mathbf{X}}^n) : 1.1896$ $\text{MSE}(\mathbf{Y}^n, \hat{\mathbf{Y}}^n) : 0.8711$
$E = 3, \tau = 2$	$\rho_{\text{cCCM}}(X \to Y) : 0.9395$ $\rho_{\text{cCCM}}(Y \to X) : 0.0079$ $\text{MSE}(\mathbf{X}^n, \hat{\mathbf{X}}^n) : 1.1876$ $\text{MSE}(\mathbf{Y}^n, \hat{\mathbf{Y}}^n) : 0.0846$	$\rho_{\text{CCM}}(X \to Y) : 0.9535$ $\rho_{\text{CCM}}(Y \to X) : 0.0630$ $\text{MSE}(\mathbf{X}^n, \hat{\mathbf{X}}^n) : 1.1472$ $\text{MSE}(\mathbf{Y}^n, \hat{\mathbf{Y}}^n) : 0.0632$
$E = 5, \tau = 1$	$\rho_{\text{cCCM}}(X \to Y) : 0.9522$ $\rho_{\text{cCCM}}(Y \to X) : -0.0210$ $\text{MSE}(\mathbf{X}^n, \hat{\mathbf{X}}^n) : 1.0941$ $\text{MSE}(\mathbf{Y}^n, \hat{\mathbf{Y}}^n) : 0.0868$	$\rho_{\text{CCM}}(X \to Y) : 0.9776$ $\rho_{\text{CCM}}(Y \to X) : -0.0225$ $\text{MSE}(\mathbf{X}^n, \hat{\mathbf{X}}^n) : 1.0871$ $\text{MSE}(\mathbf{Y}^n, \hat{\mathbf{Y}}^n) : 0.0458$

We can see that the causation from $X \to Y$ cannot be fully captured when $E = 3$ and $\tau = 1$.

Example 11:

Consider

$$X = \text{randn}(1024, 1),$$
$$Y(t) = 0.85X(t - 1) + 0.85X(t - 4).$$

where $t = [0, 1, \cdots, 1023]$, and $X(t) = 0$ while $t < 0$. Then, for different E or τ values, the results are displayed in Table 3.

Table 3. Results for **Example 11**.

Values for E and τ	cCCM	CCM
$E = 3, \tau = 1$	$\rho_{\text{cCCM}}(X \to Y) : 0.5451$ $\rho_{\text{cCCM}}(Y \to X) : -0.0521$ $\text{MSE}(\mathbf{X}^n, \hat{\mathbf{X}}^n) : 1.2205$ $\text{MSE}(\mathbf{Y}^n, \hat{\mathbf{Y}}^n) : 0.9445$	$\rho_{\text{CCM}}(X \to Y) : 0.5833$ $\rho_{\text{CCM}}(Y \to X) : -0.0616$ $\text{MSE}(\mathbf{X}^n, \hat{\mathbf{X}}^n) : 1.2204$ $\text{MSE}(\mathbf{Y}^n, \hat{\mathbf{Y}}^n) : 0.8887$
$E = 3, \tau = 2$	$\rho_{\text{cCCM}}(X \to Y) : 0.5651$ $\rho_{\text{cCCM}}(Y \to X) : 0.0382$ $\text{MSE}(\mathbf{X}^n, \hat{\mathbf{X}}^n) : 1.1485$ $\text{MSE}(\mathbf{Y}^n, \hat{\mathbf{Y}}^n) : 0.8996$	$\rho_{\text{CCM}}(X \to Y) : 0.5818$ $\rho_{\text{CCM}}(Y \to X) : 0.0750$ $\text{MSE}(\mathbf{X}^n, \hat{\mathbf{X}}^n) : 1.1060$ $\text{MSE}(\mathbf{Y}^n, \hat{\mathbf{Y}}^n) : 0.8884$
$E = 5, \tau = 1$	$\rho_{\text{cCCM}}(X \to Y) : 0.9496$ $\rho_{\text{cCCM}}(Y \to X) : -0.0113$ $\text{MSE}(\mathbf{X}^n, \hat{\mathbf{X}}^n) : 1.0797$ $\text{MSE}(\mathbf{Y}^n, \hat{\mathbf{Y}}^n) : 0.1742$	$\rho_{\text{CCM}}(X \to Y) : 0.9762$ $\rho_{\text{CCM}}(Y \to X) : 0.0123$ $\text{MSE}(\mathbf{X}^n, \hat{\mathbf{X}}^n) : 1.0597$ $\text{MSE}(\mathbf{Y}^n, \hat{\mathbf{Y}}^n) : 0.0906$

We can see that the causation from $X \to Y$ cannot be fully captured when $E = 3$ and $\tau = 1, 2$.

Example 12:

Consider a system with different dominant delays from **Example 11**:

$$X = \text{randn}(1024, 1),$$
$$Y(t) = 0.85X(t - 2) + 0.85X(t - 4).$$

where $t = [0, 1, \cdots, 1023]$, and $X(t) = 0$ while $t < 0$. Then, for different E or τ values, the results are displayed in Table 4.

Table 4. The results of **Example 12**.

Values for E and τ	cCCM	CCM
$E = 3, \tau = 1$	$\rho_{\text{cCCM}}(X \rightarrow Y) : 0.5373$ $\rho_{\text{cCCM}}(Y \rightarrow X) : -0.0490$ $\text{MSE}(\boldsymbol{X}^n, \hat{\boldsymbol{X}}^n) : 1.2765$ $\text{MSE}(\boldsymbol{Y}^n, \hat{\boldsymbol{Y}}^n) : 0.9430$	$\rho_{\text{CCM}}(X \rightarrow Y) : 0.5819$ $\rho_{\text{CCM}}(Y \rightarrow X) : -0.0151$ $\text{MSE}(\boldsymbol{X}^n, \hat{\boldsymbol{X}}^n) : 1.2074$ $\text{MSE}(\boldsymbol{Y}^n, \hat{\boldsymbol{Y}}^n) : 0.8876$
$E = 3, \tau = 2$	$\rho_{\text{cCCM}}(X \rightarrow Y) : 0.9696$ $\rho_{\text{cCCM}}(Y \rightarrow X) : -0.0184$ $\text{MSE}(\boldsymbol{X}^n, \hat{\boldsymbol{X}}^n) : 1.2204$ $\text{MSE}(\boldsymbol{Y}^n, \hat{\boldsymbol{Y}}^n) : 0.0854$	$\rho_{\text{CCM}}(X \rightarrow Y) : 0.9910$ $\rho_{\text{CCM}}(Y \rightarrow X) : -0.0416$ $\text{MSE}(\boldsymbol{X}^n, \hat{\boldsymbol{X}}^n) : 1.2342$ $\text{MSE}(\boldsymbol{Y}^n, \hat{\boldsymbol{Y}}^n) : 0.0266$
$E = 5, \tau = 1$	$\rho_{\text{cCCM}}(X \rightarrow Y) : 0.9471$ $\rho_{\text{cCCM}}(Y \rightarrow X) : -0.0052$ $\text{MSE}(\boldsymbol{X}^n, \hat{\boldsymbol{X}}^n) : 1.0647$ $\text{MSE}(\boldsymbol{Y}^n, \hat{\boldsymbol{Y}}^n) : 0.1802$	$\rho_{\text{CCM}}(X \rightarrow Y) : 0.9762$ $\rho_{\text{CCM}}(Y \rightarrow X) : 0.0088$ $\text{MSE}(\boldsymbol{X}^n, \hat{\boldsymbol{X}}^n) : 1.0573$ $\text{MSE}(\boldsymbol{Y}^n, \hat{\boldsymbol{Y}}^n) : 0.0913$

We can see that the causation from $X \rightarrow Y$ cannot be fully captured when $E = 3$ and $\tau = 1$.

Example 13:

Consider

$$X = \text{randn}(1024, 1),$$
$$Y(t) = 0.8X(t - 1) + 0.8X(t - 4) + 0.6X(t - 5)$$

where $t = [0, 1, \cdots, 1023]$, and $X(t) = 0$ while $t < 0$. Then, for different E or τ values, the results are displayed in Table 5.

Table 5. The results of **Example 13**.

Values for E and τ	cCCM	CCM
$E = 5, \tau = 1$	$\rho_{\text{cCCM}}(X \rightarrow Y) : 0.7900$ $\rho_{\text{cCCM}}(Y \rightarrow X) : -0.0054$ $\text{MSE}(\boldsymbol{X}^n, \hat{\boldsymbol{X}}^n) : 1.0799$ $\text{MSE}(\boldsymbol{Y}^n, \hat{\boldsymbol{Y}}^n) : 0.5641$	$\rho_{\text{CCM}}(X \rightarrow Y) : 0.8242$ $\rho_{\text{CCM}}(Y \rightarrow X) : -0.0016$ $\text{MSE}(\boldsymbol{X}^n, \hat{\boldsymbol{X}}^n) : 1.0736$ $\text{MSE}(\boldsymbol{Y}^n, \hat{\boldsymbol{Y}}^n) : 0.4796$
$E = 3, \tau = 2$	$\rho_{\text{cCCM}}(X \rightarrow Y) : 0.4700$ $\rho_{\text{cCCM}}(Y \rightarrow X) : -0.0197$ $\text{MSE}(\boldsymbol{X}^n, \hat{\boldsymbol{X}}^n) : 1.1890$ $\text{MSE}(\boldsymbol{Y}^n, \hat{\boldsymbol{Y}}^n) : 1.1990$	$\rho_{\text{CCM}}(X \rightarrow Y) : 0.4972$ $\rho_{\text{CCM}}(Y \rightarrow X) : -0.0194$ $\text{MSE}(\boldsymbol{X}^n, \hat{\boldsymbol{X}}^n) : 1.1829$ $\text{MSE}(\boldsymbol{Y}^n, \hat{\boldsymbol{Y}}^n) : 1.1764$
$E = 6, \tau = 1$	$\rho_{\text{cCCM}}(X \rightarrow Y) : 0.9388$ $\rho_{\text{cCCM}}(Y \rightarrow X) : -0.0300$ $\text{MSE}(\boldsymbol{X}^n, \hat{\boldsymbol{X}}^n) : 1.0425$ $\text{MSE}(\boldsymbol{Y}^n, \hat{\boldsymbol{Y}}^n) : 0.2551$	$\rho_{\text{CCM}}(X \rightarrow Y) : 0.9694$ $\rho_{\text{CCM}}(Y \rightarrow X) : -0.0194$ $\text{MSE}(\boldsymbol{X}^n, \hat{\boldsymbol{X}}^n) : 1.0419$ $\text{MSE}(\boldsymbol{Y}^n, \hat{\boldsymbol{Y}}^n) : 0.1462$

From this example, we can see the following: (i) when $E = 5, \tau = 1$, we have $\mathbf{x}(t) = [X(t), X(t-1), \cdots, X(t-4)]$, and the causation corresponding to item $0.6X(t-5)$ cannot be captured; (ii) when $E = 3$ and $\tau = 2$, we have $\mathbf{x}(t) = [X(t), X(t-2), X(t-4)]$, and the causation corresponding to items $0.8X(t-1)$ and $0.6X(t-5)$ cannot be captured; and (iii) when $E = 6$ and $\tau = 1$, we have $\mathbf{x}(t) = [X(t), X(t-1), \cdots, X(t-5)]$, and the causation corresponding to all the items can be captured.

Now, if we consider the time-delayed causality, in which $X(t)$ remains the same and $Y_1(t) = Y(t+1)$, then this is equivalent to considering the causality from $X_1(t) = X(t-1)$ to $Y(t)$. In this case, as shown in Table 6, even when $E = 5$ and $\tau = 1$, we have $\mathbf{x}_1(t) = [X(t-1), X(t-2), \cdots, X(t-5)]$, and the causation corresponding to all the items can be captured.

Table 6. The results of time-delayed causality analysis in **Example 13**.

Values for E and τ	cCCM	CCM
$E = 5, \tau = 1$	$\rho_{\text{cCCM}}(X \to Y_1) : 0.9490$ $\rho_{\text{cCCM}}(Y_1 \to X) : 0.5947$ $\text{MSE}(\mathbf{X}^n, \hat{\mathbf{X}}^n) : 0.5841$ $\text{MSE}(\mathbf{Y}_1^n, \hat{\mathbf{Y}}_1^n) : 0.2032$	$\rho_{\text{CCM}}(X \to Y_1) : 0.9753$ $\rho_{\text{CCM}}(Y_1 \to X) : 0.6380$ $\text{MSE}(\mathbf{X}^n, \hat{\mathbf{X}}^n) : 0.5362$ $\text{MSE}(\mathbf{Y}_1^n, \hat{\mathbf{Y}}_1^n) : 0.1061$
$E = 3, \tau = 2$	$\rho_{\text{cCCM}}(X \to Y_1) : 0.6621$ $\rho_{\text{cCCM}}(Y_1 \to X) : 0.5309$ $\text{MSE}(\mathbf{X}^n, \hat{\mathbf{X}}^n) : 0.6814$ $\text{MSE}(\mathbf{Y}_1^n, \hat{\mathbf{Y}}_1^n) : 0.8317$	$\rho_{\text{CCM}}(X \to Y_1) : 0.6854$ $\rho_{\text{CCM}}(Y_1 \to X) : 0.5683$ $\text{MSE}(\mathbf{X}^n, \hat{\mathbf{X}}^n) : 0.6404$ $\text{MSE}(\mathbf{Y}_1^n, \hat{\mathbf{Y}}_1^n) : 0.7992$
$E = 6, \tau = 1$	$\rho_{\text{cCCM}}(X \to Y_1) : 0.9421$ $\rho_{\text{cCCM}}(Y_1 \to X) : 0.5861$ $\text{MSE}(\mathbf{X}^n, \hat{\mathbf{X}}^n) : 0.5874$ $\text{MSE}(\mathbf{Y}_1^n, \hat{\mathbf{Y}}_1^n) : 0.2555$	$\rho_{\text{CCM}}(X \to Y_1) : 0.9705$ $\rho_{\text{CCM}}(Y_1 \to X) : 0.6405$ $\text{MSE}(\mathbf{X}^n, \hat{\mathbf{X}}^n) : 0.5332$ $\text{MSE}(\mathbf{Y}_1^n, \hat{\mathbf{Y}}_1^n) : 0.1450$

Example 14:

Consider

$$X = \text{randn}(1024, 1),$$
$$Y(t) = 0.8X(t - 4) + 0.6X(t - 5)$$

where $t = [0, 1, \cdots, 1023]$, and $X(t) = 0$ while $t < 0$. Then, for different E or τ values, the results are displayed in Table 7.

Table 7. The results of **Example 14**.

Values for E and τ	cCCM	CCM
$E = 5, \tau = 1$	$\rho_{\text{cCCM}}(X \to Y) : 0.6947$ $\rho_{\text{cCCM}}(Y \to X) : -0.0250$ $\text{MSE}(\mathbf{X}^n, \hat{\mathbf{X}}^n) : 1.0910$ $\text{MSE}(\mathbf{Y}^n, \hat{\mathbf{Y}}^n) : 0.4670$	$\rho_{\text{CCM}}(X \to Y) : 0.7286$ $\rho_{\text{CCM}}(Y \to X) : -0.0024$ $\text{MSE}(\mathbf{X}^n, \hat{\mathbf{X}}^n) : 1.0849$ $\text{MSE}(\mathbf{Y}^n, \hat{\mathbf{Y}}^n) : 0.4287$
$E = 3, \tau = 2$	$\rho_{\text{cCCM}}(X \to Y) : 0.7019$ $\rho_{\text{cCCM}}(Y \to X) : -0.0327$ $\text{MSE}(\mathbf{X}^n, \hat{\mathbf{X}}^n) : 1.1890$ $\text{MSE}(\mathbf{Y}^n, \hat{\mathbf{Y}}^n) : 1.2170$	$\rho_{\text{CCM}}(X \to Y) : 0.7309$ $\rho_{\text{CCM}}(Y \to X) : 0.0119$ $\text{MSE}(\mathbf{X}^n, \hat{\mathbf{X}}^n) : 1.1829$ $\text{MSE}(\mathbf{Y}^n, \hat{\mathbf{Y}}^n) : 1.1760$
$E = 6, \tau = 1$	$\rho_{\text{cCCM}}(X \to Y) : 0.9479$ $\rho_{\text{cCCM}}(Y \to X) : -0.0300$ $\text{MSE}(\mathbf{X}^n, \hat{\mathbf{X}}^n) : 1.0425$ $\text{MSE}(\mathbf{Y}^n, \hat{\mathbf{Y}}^n) : 0.2551$	$\rho_{\text{CCM}}(X \to Y) : 0.9718$ $\rho_{\text{CCM}}(Y \to X) : -0.0194$ $\text{MSE}(\mathbf{X}^n, \hat{\mathbf{X}}^n) : 1.0419$ $\text{MSE}(\mathbf{Y}^n, \hat{\mathbf{Y}}^n) : 0.1462$

In this example, both $E = 5, \tau = 1$ and $E = 3, \tau = 2$ can only capture the causation corresponding to $0.8X(t - 4)$, and $E = 6$ and $\tau = 1$ can capture the overall causation accurately.

Now, if we consider the time-delayed causality, in which $X(t)$ remains the same and $Y_3(t) = Y(t + 3)$, then this is equivalent to considering the causality from $X_3(t) = X(t - 3)$ to $Y(t)$. In this case, as shown in Table 8, $E = 5$ and $\tau = 1$ work even better than $E = 6$ and $\tau = 1$ since $E = 5$ leads to a manifold with a lower dimension and, hence, a higher nearest neighbor density.

Table 8. The results of time-delayed causality analysis in **Example 14**.

Values for E and τ	cCCM	CCM
$E = 5, \tau = 1$	$\rho_{\text{cCCM}}(X \rightarrow Y_3) : 0.9543$ $\rho_{\text{cCCM}}(Y_3 \rightarrow X) : -0.0378$ $\text{MSE}(\boldsymbol{X}^n, \hat{\boldsymbol{X}}^n) : 1.0968$ $\text{MSE}(\boldsymbol{Y}_3^n, \hat{\boldsymbol{Y}}_3^n) : 0.1210$	$\rho_{\text{CCM}}(X \rightarrow Y_3) : 0.9782$ $\rho_{\text{CCM}}(Y_3 \rightarrow X) : -0.0095$ $\text{MSE}(\boldsymbol{X}^n, \hat{\boldsymbol{X}}^n) : 1.0736$ $\text{MSE}(\boldsymbol{Y}_3^n, \hat{\boldsymbol{Y}}^n) : 0.0636$
$E = 3, \tau = 2$	$\rho_{\text{cCCM}}(X \rightarrow Y_3) : 0.4445$ $\rho_{\text{cCCM}}(Y \rightarrow X) : -0.0606$ $\text{MSE}(\boldsymbol{X}^n, \hat{\boldsymbol{X}}^n) : 1.2451$ $\text{MSE}(\boldsymbol{Y}^n, \hat{\boldsymbol{Y}}_3^n) : 0.7748$	$\rho_{\text{CCM}}(X \rightarrow Y_3) : 0.4568$ $\rho_{\text{CCM}}(Y_3 \rightarrow X) : -0.0611$ $\text{MSE}(\boldsymbol{X}^n, \hat{\boldsymbol{X}}^n) : 1.2306$ $\text{MSE}(\boldsymbol{Y}_3^n, \hat{\boldsymbol{Y}}_3^n) : 0.7743$
$E = 6, \tau = 1$	$\rho_{\text{cCCM}}(X \rightarrow Y_3) : 0.9424$ $\rho_{\text{cCCM}}(Y_3 \rightarrow X) : -0.0293$ $\text{MSE}(\boldsymbol{X}^n, \hat{\boldsymbol{X}}^n) : 1.0545$ $\text{MSE}(\boldsymbol{Y}_3^n, \hat{\boldsymbol{Y}}_3^n) : 0.1590$	$\rho_{\text{CCM}}(X \rightarrow Y_3) : 0.9703$ $\rho_{\text{CCM}}(Y_3 \rightarrow X) : -0.0173$ $\text{MSE}(\boldsymbol{X}^n, \hat{\boldsymbol{X}}^n) : 1.0432$ $\text{MSE}(\boldsymbol{Y}_3^n, \hat{\boldsymbol{Y}}_3^n) : 0.0927$

From Examples 10–14, it can be seen that in systems with memory, the selection of the shadow manifold dimension E and the signal lag τ largely rely on the positions of the dominant delays in the channel impulse response.

It can be seen that in systems with memory, for the accurate evaluation of CCM and cCCM causality, the following conditions need to be satisfied:

(a) $E \cdot \tau > d_{d,max}$, where $d_{d,max}$ denotes the largest dominant delay.
(b) For each t, the shadow manifold constructing vector $x(t) = [X(t), X(t - \tau), \cdots, X(t - (E - 1)\tau)]$ should contain all the samples corresponding to the dominant delays.

It is also observed that if the conditions above are not satisfied, time-delayed cCCM from $X(t - \tau)$ to $Y(t)$ might still capture the causation accurately if the instantaneous information exchange between $X(t)$ and $Y(t)$ is not significant. More specifically, if we consider a linear time-invariant (LTI) system $Y(t) = X(t) * h(t) = \sum_{l=0}^{L} h(l)X(t - l)$, where $h(t)$ denotes the channel impulse response, when $h(0)$ is negligibly small, we say that there is no significant instantaneous information exchange between $X(t)$ and $Y(t)$.

In the following two examples, we compare the performance between cCCM and Granger causality (GC) for systems with memory.

Example 15:

Consider a system with memory:

$$X = \text{randn}(1024, 1), \tag{9}$$
$$Y(t) = 0.8X(t) + 0.2X(t - 1) + 0.2X(t - 2) + 0.2X(t - 5) + n(t), \tag{10}$$

where $t = [0, 1, \cdots, 1023]$, and $X(t) = 0$ while $t < 0$. Here, we assume that $n(t) \sim N(0, \sigma^2)$ is independent of X. We then compare the performances of GC and CCM under different noise powers, and the results are shown in Table 9.

Table 9. Results for **Example 15**.

Noise	$n(t) \sim N(0, \sigma^2)$				
	$\sigma^2 = 0$	$\sigma^2 = 10^{-6}$	$\sigma^2 = 10^{-2}$	$\sigma^2 = 10^{-1}$	$\sigma^2 = 4$
SNR(dB)	∞ (noise free)	52.26 dB	12.26 dB	2.26 dB	-7.74 dB
$GC(X \rightarrow Y)$	6.078×10^{-5}	6.383×10^{-4}	0.0065	0.0345	0.0277
$GC(Y \rightarrow X)$	6.862×10^{-5}	6.589×10^{-4}	0.0019	0.0040	0.0035
$\rho_{\text{cCCM}}(X \rightarrow Y)$	0.9169	0.9168	0.9070	0.8314	0.1897
$\rho_{\text{cCCM}}(Y \rightarrow X)$	0.9024	0.9023	0.8873	0.8043	0.1413

As can be seen, as long as the signal-to-noise ratio (SNR) is not too small, cCCM can capture the strong bidirectional causality between $X(t)$ and $Y(t)$, but GC cannot. This is mainly because cCCM takes the instantaneous information exchange between $X(t)$ and $Y(t)$ into consideration, but GC does not. That is, when there exists instantaneous information exchange between $X(t)$ and $Y(t)$, GC may fail to capture the causal coupling between $X(t)$ and $Y(t)$.

It is also observed that the ρ_{cCCM} value decreases as the noise power increases, which is consistent with our analysis in Section 3.1. When $\sigma^2 = 4$ and SNR $= -7.74$ dB, both cCCM and GC can no longer deliver valid results due to the strong noise effect.

Recall that the most commonly used method in Granger causality [10–12] analysis is to compare the following two prediction errors e_i and $\tilde{e}_i$:

$$Y_i = \sum_{j=1}^{K} a_j Y_{i-j} + e_i$$

$$Y_i = \sum_{j=1}^{K} b_j Y_{i-j} + \sum_{j=1}^{L} c_j X_{i-j} + \tilde{e}_i,$$

And the Granger causality is defined to be the log-likelihood ratio

$$GC(X \to Y) = \ln \frac{|\text{cov}(\mathbf{e})|}{|\text{cov}(\tilde{\mathbf{e}})|}, \tag{11}$$

where $\mathbf{e} = [e_1, e_2, \ldots, e_n]^T$, $\tilde{\mathbf{e}} = [\tilde{e}_1, \tilde{e}_2, \ldots, \tilde{e}_n]^T$, and $|\text{cov}(\cdot)|$ stands for the determinant of the covariance matrix.

Our results in Table 9 and the definition of GC suggest that the small fluctuations in the GC values as the noise variance increases from 10^{-6} to 4 are more likely to reflect the impact of the noise rather than the detection of the causality.

Example 16:

Consider

$$X = \text{randn}(1024, 1),$$
$$Y(t) = 0.8X(t-1) + 0.2X(t-2) + 0.2X(t-5) + n(t),$$

where $t = [0, 1, \cdots, 1023]$, $X(t) = 0$ while $t < 0$, and $n(t) \sim N(0, 10^{-6})$ is an independently generated Gaussian noise. Then, the Granger causality between X and Y is

$$GC(X \to Y) = 11.3901,$$
$$GC(Y \to X) = 0.0002.$$

and the causality detected by cCCM is

$$\rho_{\text{cCCM}}(X \to Y) = 0.9138,$$
$$\rho_{\text{cCCM}}(Y \to X) = 0.0128.$$

In this example, there is no instantaneous information exchange between $X(t)$ and $Y(t)$, and both GC and cCCM detect the strong unidirectional causality from X to Y and deliver consistent results.

3.2.5. Additional Examples on the Selection of the Dimension of the Shadow Manifold E and Time Lag τ

In this subsection, we illustrate the impact of E and τ on the performance of cCCM through some additional simulation examples, including single-tone time series embedded in noise (**Example 13**) and Gaussian stochastic process (**Example 14**). As in **Example 3**, it was found that a large $E \cdot \tau$ value may help enhance the performance of cCCM under

noise. However, it is also noticed that if E is too large, cCCM may no longer deliver valid results, as the excessively high dimension of the shadow manifold significantly reduces the density of the nearest neighbors, leading to inaccurate state-space reconstruction and causality evaluation.

Example 17:

This is a revisit of Example 3, with additional discussions on the selection of E and τ and different sampling instants. Consider the following noisy single-tone time series:

$$X(t) = \sin(t) + n_1(t),$$
$$Y(t) = \cos(t) + n_2(t),$$

where $t = 0 : 0.01\pi : 2\pi$, and $n_1(t)$ and $n_2(t)$ are independent AWGN noises with SNRs varying in 0, 5, 10, 15, and 20 dB, or equal to 0 for all t in the noise-free case. By changing the values for E and τ, we are able to observe different noise effects. The simulation results for $E = 5, \tau = 1$ and $E = 5, \tau = 5$ are shown in Table 10 below.

Table 10. Performance of cCCM under additive white Guassian noise with different E and τ values.

Values for E and τ	Direction	0 dB	5 dB	10 dB	15 dB	20 dB	Noise Free
$E = 5, \tau = 1$	$X \to Y$	0.2378	0.2410	0.3497	0.5156	0.6273	0.9566
	$Y \to X$	0.2476	0.4446	0.6194	0.7267	0.8447	0.9945
$E = 5, \tau = 5$	$X \to Y$	0.3497	0.6116	0.8278	0.9432	0.9799	0.9985
	$Y \to X$	0.5570	0.7566	0.8893	0.9693	0.9877	0.9991

As can be seen, as we increase the length of the data span $E \cdot \tau$, the noise effect is reduced. In particular, compared with $E = 5$ and $\tau = 1$, when we choose $E = 5$ and $\tau = 5$, a much better noise immunity is achieved, since $E \cdot \tau$ is sufficiently long.

Note that increasing τ leads to the downsampling of the time series, and increasing E expands the dimension of the shadow manifolds. An over-increase in E or τ might downgrade the performance of cCCM. From Example 14, it can be seen that if E is much larger than $2d + 1$, cCCM may deliver inaccurate results.

Example 18:

Consider

$$X = \text{randn}(1024, 1), \quad Y = |X|,$$

In this example, there is a strong unidirectional causality from X to Y, but very weak causation in the inverse direction. Choose $\tau = 1$. From Figure 5, we can see that as E increases, the cCCM value keeps on decreasing and reduces to 0.2 when $E = 50$, which no longer reflects the strong unidirectional causality from X to Y.
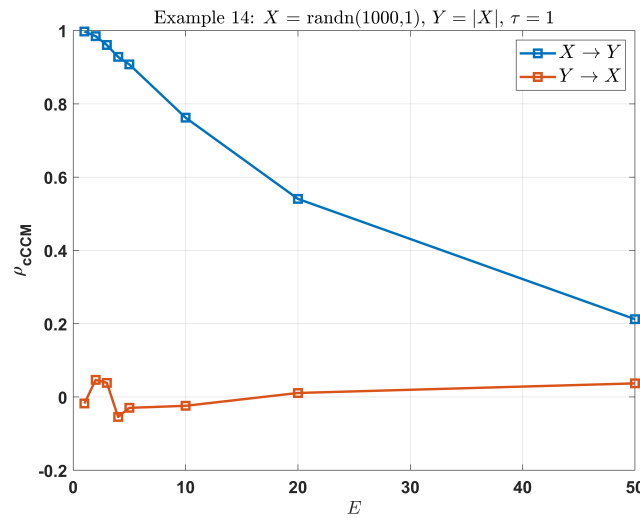
Figure 5. cCCM results for **Example 18** ($X = randn(1024, 1)$, $Y = |X|$): an excessively large E may downgrade the performance of cCCM; here, $\tau = 1$.

3.2.6. Examples of the Impact of Sampling Frequency on cCCM

In this subsection, we show that for the accurate assessment of causation, signals under consideration should be sampled with a sampling frequency higher than the Nyquist rate.

Example 19:

Consider

$$X = \sin(k\pi t),$$
$$Y = \cos(k\pi t),$$

where $t = 0 : 0.005 : 4$, and $k = 150, 200, 400$. From Figure 6, it can be seen that if the sampling frequency is higher than the Nyquist rate, then strong bidirectional causal coupling can be detected between X and Y. On the other hand, if the sampling frequency is lower than the Nyquist rate, then the resulted cCCM value is no longer valid.
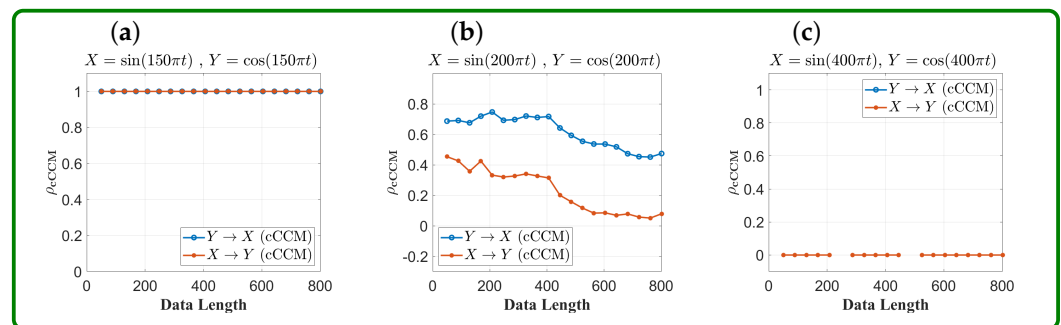
Figure 6. Impact of sampling frequency on cCCM convergence speed: an illustration using sinusoidal waveforms with different frequencies. **(a)** f_0 = 75 Hz; **(b)** f_0 = 100 Hz; **(c)** f_0 = 200 Hz. Here, f_0 denotes the frequency of the corresponding sinusoidal waveform. The sampling time sequence was chosen as $t = 0 : 0.005 : 4$; that is, sampling frequency f_s = 200 Hz. As can be seen, cCCM works well when the sampling rate is above the Nyquist rate, as shown in **(a)** but may or may not deliver meaningful results when the sampling frequency is below or equal to the Nyquist rate, as shown in **(b,c)**.

3.2.7. Examples on Data Repetition in Causality Analysis

The following examples illustrate that even if X and Y are two independent signals that are not causally coupled, a causal pattern can be enforced in the concatenated time series through data repetition.

Example 20:

Let $X = \text{randn}(1000, 1)$ and $Y = \text{randn}(1000, 1)$ be two independent normally distributed time series. We have

$$\rho_{\text{cCCM}}(X \to Y) = 0.0198, \quad \rho_{\text{cCCM}}(Y \to X) = -0.0326.$$

That is, X and Y are not causally coupled. Consider

$$\tilde{X} = [X; X; X] \quad \tilde{Y} = [Y; Y; Y],$$

Then, we have

$$\rho_{\text{cCCM}}(\tilde{X} \to \tilde{Y}) = 0.7860, \quad \rho_{\text{cCCM}}(\tilde{Y} \to \tilde{X}) = 0.7810.$$

As can be seen, data concatenation results in strong causality that does not exist in the original X and Y.

3.2.8. An Example of Multivariate Conditional cCCM

Example 21:

Let $X_0 = \text{randn}(1024,1)$, $Y_0 = \text{randn}(1024,1)$, and $Z = \text{randn}(1024,1)$ be independent and normally distributed Gaussian random variables. Consider

$$X = 0.7X_0 + 10Z,$$
$$Y = 0.4Y_0 + 12Z.$$

Then, the bivariate cCCM between X and Y is

$$\rho_{\text{cCCM}}(X \to Y) = 0.9574, \quad \rho_{\text{cCCM}}(Y \to X) = 0.9601,$$

which provides a delusion that there exists strong bidirectional causality between X and Y. On the other hand, the multivariate cCCM between X and Y conditioning on Z is

$$\text{cCCM}(X \to Y|Z) = 0.0298, \quad \text{cCCM}(Y \to X|Z) = 0.0306,$$

which accurately reflect the independent relationship between X and Y. From this example, it can be seen that conditional cCCM can help inspect the dependence among the random variables under consideration and may deliver more accurate results in the causality evaluation.

3.3. Application of cCCM for Brain Causality Analysis Using Experimental fMRI Data

In this study, we applied both bivariate and multivariate cCCM for a causality analysis of the brain network using experimental fMRI data and compared the results with those of GC [10,43].

We considered an fMRI dataset where fourteen right-handed healthy college students (7 males and 7 females, 23.4 ± 4.2 years of age) from Michigan State University volunteered to participate in a task-driven fMRI-based study. For each subject, fMRI datasets were collected on a visual stimulation condition with a scene–object fMRI paradigm, where each volume of images was acquired 192 times (8 min) while a subject was presented with 12 blocks of visual stimulation after an initial 10 s resting period. In a predefined randomized order, the scenery pictures were presented in six blocks, and the object pictures were presented in another six blocks. All pictures were unique. In each block, ten pictures were presented continuously for 25 s (2.5 s for each picture), followed with a 15 s baseline condition (a white screen with a black fixation cross at the center). The subject needed to press their right index finger once when the screen was switched from the baseline to the picture condition. More details on fMRI data acquisition and preprocessing can be found in [44].

Region of Interest (**ROI**) selection: we selected 10 ROI regions, including the left primary visual cortex (LV1), left parahippocampal place area (LPPA), left sensory motor cortex (LSMC), left parahippocampal white matter (LPWM), left retrosplenial cortex (LRSC), right primary visual cortex (RV1), right parahippocampal place area (RPPA), right sensory motor cortex (RSMC), right frontal white matter (RFWM), and right retrosplenial cortex (RRSC).

3.3.1. Results for Bivariate and Multivariate cCCM

Note that the total length of the fMRI BOLD time series under visual stimulation condition was $n = 192$, with the sampling period being 2.5 s. In the literature, it was reported that increasing the sampling rate of the fMRI signal can improve the robustness of the causality analysis [45]. Here, we first interpolate the fMRI sequence by a factor of 2 using the spline interpolation command in MATLAB and then conducted causality analysis for all the possible unidirectional regional pairs using both bivariate and multivariate cCCM.

The causality analysis results based on bivariate cCCM (averaged over all 14 subjects) are shown in Figure 7. Our results suggest the presence of unidirectional causality from LV1 → LSMC, RV1 → LSMC, LV1 → LPWM, LV1 → RFWN, and LPPA → LPWM.

The results corresponding to multivariate conditional cCCM with respect to individual brain regions are shown in Figure 8. As can be seen, RV1 has the most significant impact on the conditional causality from LV1 → LSMC, $LV1 → LPWM$, and $LV1 → RFWM$. This implies that RV1 has the highest inter-region dependence with LV1. For the same reason, LV1 has the most significant impact on the conditional causality from RV1 → LSMC. That is, multivariate conditional cCCM with respect to individual regions can detect unidirectional causality and also reflect the impact of interdependence between the ROIs on the conditional causality.
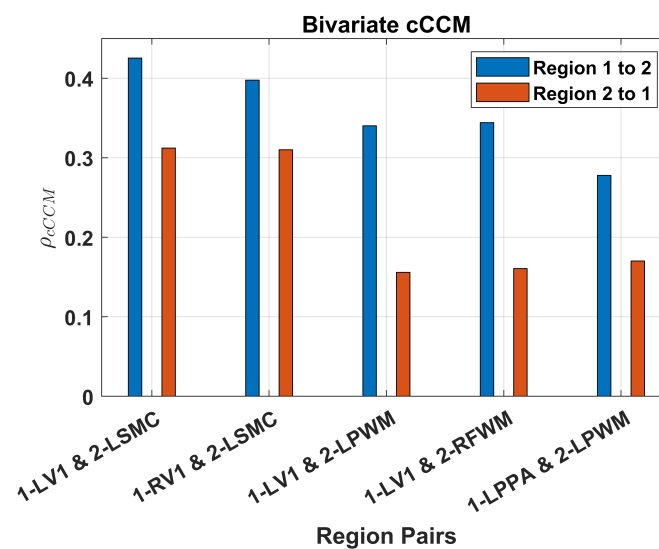
Figure 7. FMRI-based causality analysis using bivariate cCCM. Unidirectional causality was detected in the brain network under a visual simulation condition with a scene–object fMRI paradigm.
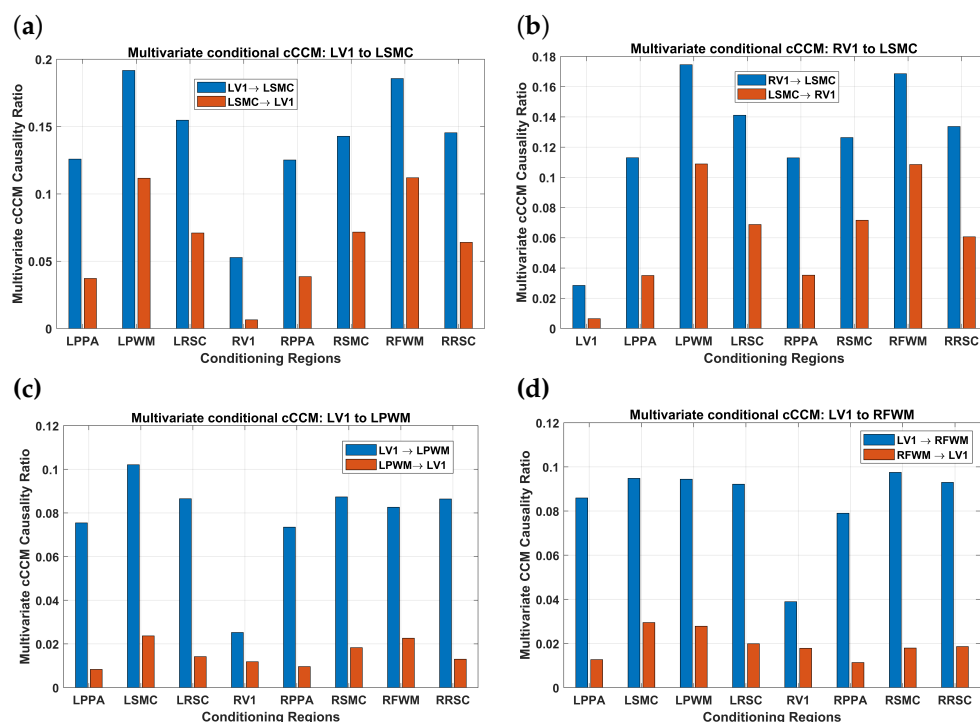
Figure 8. FMRI-based causality analysis using multivariate conditional cCCM with respect to individual regions. (**a**) $LV1 \rightarrow LSMC$, (**b**) $RV1 \rightarrow LSMC$, (**c**) $LV1 \rightarrow LPWM$, (**d**) $LV1 \rightarrow RFWM$. The results indicate that multivariate cCCM (with respect to individual regions) can detect unidirectional causality and also reflect the impact of interdependence between the ROIs on the conditional causality. More specifically, it can be seen that due to the dependence between the brain regions, multivariate conditional CCM values are much smaller than the bivariate cCCM values. In particular, RV1 has the most significant impact on the conditional causality from $LV1 \rightarrow$ LSMC, $LV1 \rightarrow LPWM$, and $LV1 \rightarrow RFWM$. This implies that RV1 has the highest dependence with LV1. For the same reason, LV1 has the the most significant impact on the conditional causality from RV1 $\rightarrow$ LSMC.

3.3.2. Results for Bivariate and Multivariate Granger Causality (GC)

For comparison purposes, we analyzed the brain network causality using both bivariate and multivariate GC [12] with the same fMRI dataset.

From Figures 7 and 9, it can be seen that bivariate GC delivers similar results as cCCM except for the causal coupling from $LV1 \rightarrow LSMC$. More specifically, cCCM shows that there exists unidirectional causality from $LV1 \rightarrow LSMC$, while GC shows that there exists bidirectional causality between $LV1$ and $LSMC$ but no significant unidirectional causality. In ref. [9], the DI-based causality analysis also verified the presence of unidirectional causality from $LV1 \rightarrow LSMC$ for the same dataset, which is consistent with the results of cCCM. These results suggest that for this fMRI dataset, cCCM tends to deliver a more accurate causality evaluation than GC.
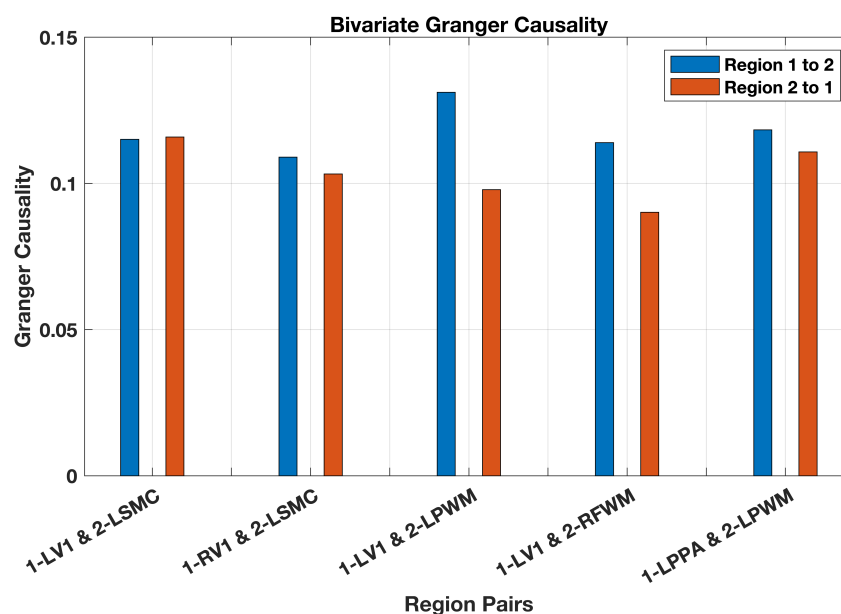
Figure 9. FMRI-based causality analysis using bivariate GC. The results of GC are largely consistent with those of bivariate cCCM except from $LV1 \rightarrow LSMC$. This may be because (i) cCCM takes the instantaneous information exchange between $X(t)$ and $Y(t)$ into consideration, but GC does not; and (ii) cCCM can capture both linear and nonlinear causal causal coupling, and GC may have difficulty in detecting nonlinear causality. That is, when there exists instantaneous information exchange and/or a nonlinear causal relationship between $X(t)$ and $Y(t)$, GC may fail to capture the underlying causal coupling accurately.

A natural question arises: how should we explain the difference between cCCM and GC for the causality analysis here? Since based on the central limit theorem, fMRI signals can be modeled as Gaussian random variables for which cCCM and GC are conditionally equivalent. The underlying argument is that the equivalence between cCCM and GC under Gaussian random variables is subject to two conditions: (i) both $X(t)$ and $Y(t)$ follow the linear auto-regression model; and (ii) there is no significant instantaneous information exchange between $X(t)$ and $Y(t)$. More specifically, cCCM takes the instantaneous information exchange between $X(t)$ and $Y(t)$ into consideration, but GC does not. For this reason, when there exists instantaneous information exchange between $X(t)$ and $Y(t)$, GC may fail to capture the causal coupling between $X(t)$ and $Y(t)$, but cCCM succeeds. This is demonstrated through simulations in Example 15. In addition, cCCM can capture both linear and nonlinear causal causal coupling, but GC may have difficulty in detecting nonlinear causality. For these reasons, cCCM might be a more robust causality analysis tool than GC.

In the multivariate case, the theoretical relationship between cCCM and GC is not clear yet. In comparing Figures 8 and 10, it can be seen that the results of multivariate cCCM and GC are largely consistent for $LV1 \rightarrow LSMC$ and $LV1 \rightarrow RFWM$. However, they deliver very different results for the conditional causality from $RV1 \rightarrow LSMC$ and $LV1 \rightarrow LPWM$ with respect to other individual regions. In particular, for these two region pairs, the results of multivariate cCCM with respect to other individual regions are consistent with their bivariate counterparts and also reflect the impact of inter-region dependence on the conditional causality. However, the corresponding results of multivariate GC with respect to other individual regions vary significantly with the region under consideration, and 50% or more are no longer consistent with those of the bivariate GC.
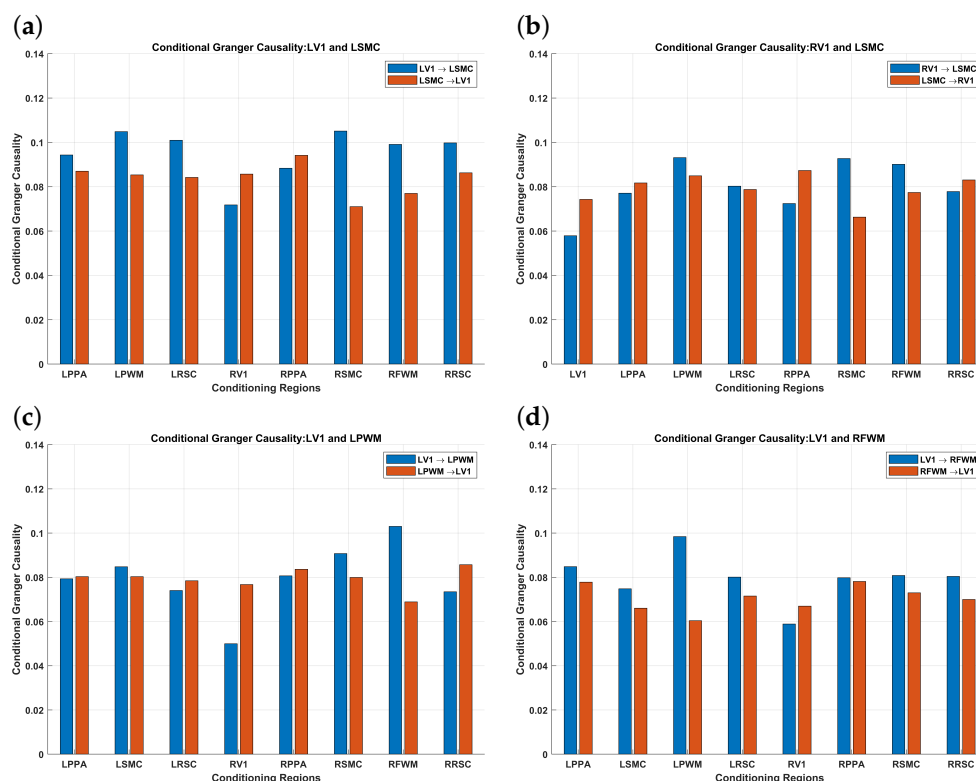
Figure 10. FMRI-based causality analysis using multivariate conditional GC with respect to individual regions. (**a**) $LV1 \rightarrow LSMC$, (**b**) $RV1 \rightarrow LSMC$, (**c**) $LV1 \rightarrow LPWM$, (**d**) $LV1 \rightarrow RFWM$. It can be seen that the results of multivariate cCCM and GC are largely consistent for (**a**,**d**). However, for (**b**,**c**), multivariate cCCM and GC deliver very different results. In particular, for the conditional causality from $RV1 \rightarrow LSMC$ and $LV1 \rightarrow LPWM$, the results of multivariate cCCM with respect to other individual regions are consistent with their bivariate counterparts and also reflect the impact of inter-region dependence on the conditional causality. However, the corresponding results of multivariate GC with respect to other individual regions vary significantly with the region under consideration, and 50% or more are no longer consistent with those of the bivariate GC.

Further theoretical analysis is needed on the theoretical relationship between conditional GC and multivariate cCCM, as well as the relationship between DI and the recent minimum entropy framework [46] in both bivariate and multivariate scenarios.

4. Conclusions and Discussion

In this paper, we revisited the definition of original CCM, identified the gap between CCM and the traditional definition of causality, presented causalized CCM (cCCM), and discussed the conditional equivalence of cCCM and directed information and the extension of bivariate cCCM to multivariate cCCM. We then evaluated the effectiveness of cCCM in the detection of causality through a large number of examples including Gaussian random variables with additive noise, sinusoidal waveforms, autoregressive models, stochastic processes with a dominant spectral component embedded in noise, deterministic chaotic maps, and systems with memory, as well as experimental fMRI data. We also examined the impact of noise on the performance of cCCM, and our results suggest that, in general, reliable results can be achieved when SNR > 15 dB. In particular, we provided detailed discussions on the choice of the dimension of the shadow manifolds E and the time lag τ and the impact of these parameters on the detection of causal relationships using cCCM. Finally, we applied both bivariate and multivariate cCCM for the causality analysis of the brain network using experimental fMRI data and compared the results with those of GC.

Based on the conditional equivalence of cCCM and DI [9], we can see that cCCM provides an alternative way to evaluate the directed information transfer between stationary

ergodic Gaussian random variables. Compared with DI, which relies heavily on probability estimation and tends to be sensitive to data length and quantization step size, cCCM, on the other hand, gets around this problem through geometric cross mapping between the manifolds involved.

However, the advantage of cross-mapping-based causality detection techniques comes with prices. The major limitation with CCM and cCCM is that they are based on the K-nearest neighbor (KNN) search algorithm and hence have a high computation complexity $O(n^2)$, where n is the data length. The convergence speeds of CCM and cCCM also vary with the signals under applications and need to be taken into consideration in causality analysis, especially in dynamic systems where the causal relationships are time-variant. It is worthy to point out that when combined with the sliding window approach [47,48], cCCM can be used to evaluate time-varying causality in dynamic networks such as brain networks [49].

Overall, both our theoretical [9] and numerical analysis demonstrated that cCCM is a promising and easy-to-implement tool for causality detection in a wide spectrum of applications. In this paper, we showed that appropriate choices of E, τ, and the sampling frequency are critical for cCCM-based causality analysis and provided detailed guidelines on the configuration of cCCM. We wish that this paper can serve as a helpful reference on the implementation of cCCM for causality detection in different applications.

Abbreviations

The following abbreviations are used in this manuscript:

CCM	Convergent cross mapping;
cCCM	Causalized convergent cross mapping;
GC	Granger causality;
DI	Directed information;
fMRI	Functional magnetic resonance imaging;
SNR	Signal-to-noise ratio;

DMN	Default mode network;
BOLD	Blood-oxygen-level-dependent;
KNN	K-nearest neighbor;
TE	Transfer entropy;
DCM	Dynamic causal modeling;
AWGN	Additive white Gaussian noise;
MSE	Mean square error;
CR	Causality ratio;
ROI	Region of interest;
LV1	Left primary visual cortex;
LPPA	Left parahippocampal place area;
LSMC	Left sensory motor cortex;
LPWM	Left parahippocampal white matter;
LRSC	Left retrosplenial cortex;
RV1	Right primary visual cortex;
RPPA	Right parahippocampal place area;
RSMC	Right sensory motor cortex;
RFWM	Right frontal white matter;
RRSC	Right retrosplenial cortex.

References

1. Hua, J.C.; Jin Kim, E.; He, F. Information Geometry Theoretic Measures for Characterizing Neural Information Processing from Simulated EEG Signals. *Entropy* **2024**, *26*, 213. [CrossRef]
2. Ma, Y.; Qian, J.; Gu, Q.; Yi, W.; Yan, W.; Yuan, J.; Wang, J. Network Analysis of Depression Using Magnetoencephalogram Based on Polynomial Kernel Granger Causality. *Entropy* **2023**, *25*, 1330. [CrossRef] [PubMed]
3. Wang, Q.; Yao, W.; Bai, D.; Yi, W.; Yan, W.; Wang, J. Schizophrenia MEG Network Analysis Based on Kernel Granger Causality. *Entropy* **2023**, *25*, 1006. [CrossRef]
4. Stokes, P.A.; Purdon, P.L. A study of problems encountered in Granger causality analysis from a neuroscience perspective. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E7063–E7072. [CrossRef] [PubMed]
5. Zhang, X.; Mormino, E.C.; Sun, N.; Sperling, R.A.; Sabuncu, M.R.; Yeo, B.T.T.; Weiner, M.W.; Aisen, P.; Weiner, M.; Aisen, P.; et al. Bayesian model reveals latent atrophy factors with dissociable cognitive trajectories in Alzheimer's disease. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E6535–E6544. [CrossRef]
6. Hillebrandt, H.; Friston, K.J.; Blakemore, S.J. Effective connectivity during animacy perception—Dynamic causal modelling of Human Connectome Project data. *Sci. Rep.* **2014**, *4*, 6240. [CrossRef]
7. Marinescu, I.E.; Lawlor, P.N.; Kording, K.P. Quasi-experimental causality in neuroscience and behavioural research. *Nat. Hum. Behav.* **2018**, *2*, 891–898. [CrossRef]
8. Deshpande, G.; Santhanam, P.; Hu, X. Instantaneous and causal connectivity in resting state brain networks derived from functional MRI data. *NeuroImage* **2011**, *54*, 1043–1052. [CrossRef]
9. Deng, J.; Sun, B.; Scheel, N.; Renli, A.B.; Zhu, D.C.; Zhu, D.; Ren, J.; Li, T.; Zhang, R. Causalized convergent cross-mapping and its approximate equivalence with directed information in causality analysis. *PNAS Nexus* **2023**, *3*, 422. [CrossRef]
10. Granger, C.W.J. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica* **1969**, *37*, 424. [CrossRef]
11. Granger, C.W.J.; Newbold, P. *Forecasting Economic Time Series*; Elsevier: Amsterdam, The Netherlands, 1977; p. 225. [CrossRef]
12. Barnett, L.; Seth, A.K. The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference. *J. Neurosci. Methods* **2014**, *223*, 50–68. [CrossRef]
13. Mannino, M.; Bressler, S.L. Foundational perspectives on causality in large-scale brain networks. *Phys. Life Rev.* **2015**, *15*, 107–123. [CrossRef] [PubMed]
14. Seth, A.K.; Chorley, P.; Barnett, L.C. Granger causality analysis of fMRI BOLD signals is invariant to hemodynamic convolution but not downsampling. *NeuroImage* **2013**, *65*, 540–555. [CrossRef]
15. David, O.; Guillemain, I.; Saillet, S.; Reyt, S.; Deransart, C.; Segebarth, C.; Depaulis, A. Identifying neural drivers with functional MRI: an electrophysiological validation. *PLoS Biol.* **2008**, *6*, 2683–2697. [CrossRef]
16. Sugihara, G.; May, R.; Ye, H.; hao Hsieh, C.; Deyle, E.; Fogarty, M.; Munch, S. Detecting Causality in Complex Ecosystems. *Science* **2012**, *338*, 496–500. [CrossRef] [PubMed]
17. Massey, J. Causality, feedback, and directed information. In Proceedings of the the International Symposium on Information Theory and Its Applications, Waikiki, HI, USA, 11 November 1990; pp. 303–305.
18. Wang, Z.; Alahmadi, A.; Zhu, D.C.; Li, T. Causality Analysis of fMRI Data Based on the Directed Information Theory Framework. *IEEE Trans. Biomed. Eng.* **2016**, *63*, 1002–1015. [CrossRef]
19. Amblard, P.O.; Michel, O.J.J. On directed information theory and Granger causality graphs. *J. Comput. Neurosci.* **2011**, *30*, 7–16. [CrossRef]

20. Tsonis, A.A.; Deyle, E.R.; May, R.M.; Sugihara, G.; Swanson, K.; Verbeten, J.D.; Wang, G. Dynamical evidence for causality between galactic cosmic rays and interannual variation in global temperature. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 3253–3256. [CrossRef] [PubMed]

21. Deyle, E.R.; Maher, M.C.; Hernandez, R.D.; Basu, S.; Sugihara, G. Global environmental drivers of influenza. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 13081–13086. [CrossRef]

22. Liu, O.R.; Gaines, S.D. Environmental context dependency in species interactions. *Proc. Natl. Acad. Sci. USA* **2022**, *119*. [CrossRef]

23. Chang, C.W.; Miki, T.; Ye, H.; Souissi, S.; Adrian, R.; Anneville, O.; Agasild, H.; Ban, S.; Be'eri-Shlevin, Y.; Chiang, Y.R.; et al. Causal networks of phytoplankton diversity and biomass are modulated by environmental context. *Nat. Commun.* **2022**, *13*, 1140. [CrossRef]

24. Chen, D.; Sun, X.; Cheke, R.A. Inferring a Causal Relationship between Environmental Factors and Respiratory Infections Using Convergent Cross-Mapping. *Entropy* **2023**, *25*, 807. [CrossRef]

25. Wang, J.Y.; Kuo, T.C.; Hsieh, C.H. Causal effects of population dynamics and environmental changes on spatial variability of marine fishes. *Nat. Commun.* **2020**, *11*, 2635. [CrossRef]

26. McCracken, J.M.; Weigel, R.S. Convergent cross-mapping and pairwise asymmetric inference. *Phys. Rev. E* **2014**, *90*, 062903. [CrossRef] [PubMed]

27. Breston, L.; Leonardis, E.J.; Quinn, L.K.; Tolston, M.; Wiles, J.; Chiba, A.A. Convergent cross sorting for estimating dynamic coupling. *Sci. Rep.* **2021**, *11*, 20374. [CrossRef]

28. Wismüller, A.; Abidin, A.Z.; D'Souza, A.M.; Wang, X.; Hobbs, S.K.; Leistritz, L.; Nagarajan, M.B. Nonlinear functional connectivity network recovery in the human brain with mutual connectivity analysis (MCA): Convergent cross-mapping and non-metric clustering. *Proc. SPIE Int. Soc. Opt. Eng.* **2015**, *3*, 94170M. [CrossRef]

29. Permuter, H.H.; Kim, Y.H.; Weissman, T. Interpretations of Directed Information in Portfolio Theory, Data Compression, and Hypothesis Testing. *IEEE Trans. Inf. Theory* **2011**, *57*, 3248–3259. [CrossRef]

30. Soltani, N.; Goldsmith, A. Inferring neural connectivity via measured delay in directed information estimates. In Proceedings of the 2013 IEEE International Symposium on Information Theory, Istanbul, Turkey, 7–12 July 2013; pp. 2503–2507. [CrossRef]

31. Schreiber, T. Measuring information transfer. *Phys. Rev. Lett.* **2000**, *85*, 461–464. [CrossRef]

32. 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] [PubMed]

33. Friston, K.; Harrison, L.; Penny, W. Dynamic causal modelling. *NeuroImage* **2003**, *19*, 1273–1302. [CrossRef]

34. Wang, Z.; Liang, Y.; Zhu, D.C.; Li, T. The Relationship of Discrete DCM and Directed Information in fMRI-Based Causality Analysis. *IEEE Trans. Mol. Biol. Multi-Scale Commun.* **2018**, *4*, 3–13. [CrossRef]

35. Ridderinkhof, K.R.; Ullsperger, M.; Crone, E.A.; Nieuwenhuis, S. The Role of the Medial Frontal Cortex in Cognitive Control. *Science* **2004**, *306*, 443–447. [CrossRef] [PubMed]

36. Porta, A.; Faes, L.; Bari, V.; Marchi, A.; Bassani, T.; Nollo, G.; Perseguini, N.M.; Milan, J.; Minatel, V.; Borghi-Silva, A.; et al. Effect of Age on Complexity and Causality of the Cardiovascular Control: Comparison between Model-Based and Model-Free Approaches. *PLoS ONE* **2014**, *9*, e89463. [CrossRef] [PubMed]

37. Porta, A.; Faes, L. Wiener–Granger Causality in Network Physiology With Applications to Cardiovascular Control and Neuroscience. *Proc. IEEE* **2016**, *104*, 282–309. [CrossRef]

38. Porta, A.; Bari, V.; Gelpi, F.; Cairo, B.; Maria, B.D.; Tonon, D.; Rossato, G.; Faes, L. On the Different Abilities of Cross-Sample Entropy and K-Nearest-Neighbor Cross-Unpredictability in Assessing Dynamic Cardiorespiratory and Cerebrovascular Interactions. *Entropy* **2023**, *25*, 599. [CrossRef] [PubMed]

39. Abarbanel, H.D.I.; Carroll, T.A.; Pecora, L.M.; Sidorowich, J.J.; Tsimring, L.S. Predicting physical variables in time-delay embedding. *Phys. Rev. E* **1994**, *49*, 1840–1853. [CrossRef] [PubMed]

40. Takens, F. Detecting strange attractors in turbulence. In *Lecture Notes in Mathematics*; Springer: Berlin/Heidelberg, Germany, 1981; pp. 366–381. [CrossRef]

41. Packard, N.H.; Crutchfield, J.P.; Farmer, J.D.; Shaw, R.S. Geometry from a Time Series. *Phys. Rev. Lett.* **1980**, *45*, 712–716. [CrossRef]

42. Whitney, H.; Eells, J.; Toledo, D. *Collected Papers of Hassler Whitney (Contemporary Mathematicians)*; Birkhäuser: Basel, Switzerland, 1992.

43. Geweke, J.F. Measures of Conditional Linear Dependence and Feedback Between Time Series. *J. Am. Stat. Assoc.* **1984**, *79*, 907. [CrossRef]

44. Zhu, D.C.; Majumdar, S. Integration of resting-state FMRI and diffusion-weighted MRI connectivity analyses of the human brain: limitations and improvement. *J. Neuroimag. Off. J. Am. Soc. Neuroimag.* **2014**, *24*, 176–186. [CrossRef]

45. Lin, F.H.; Ahveninen, J.; Raij, T.; Witzel, T.; Chu, Y.H.; Jääskeläinen, I.P.; Tsai, K.W.K.; Kuo, W.J.; Belliveau, J.W. Increasing fMRI Sampling Rate Improves Granger Causality Estimates. *PLoS ONE* **2014**, *9*, e100319. [CrossRef]

46. Ning, L. An information-theoretic framework for conditional causality analysis of brain networks. *Netw. Neurosci.* **2024**, *3*, 1–38. [CrossRef]

47. Allen, E.A.; Damaraju, E.; Plis, S.M.; Erhardt, E.B.; Eichele, T.; Calhoun, V.D. Tracking Whole-Brain Connectivity Dynamics in the Resting State. *Cereb. Cortex* **2014**, *24*, 663–676. [CrossRef] [PubMed]

48. Schumacher, J.; Peraza, L.R.; Firbank, M.; Thomas, A.J.; Kaiser, M.; Gallagher, P.; O'Brien, J.T.; Blamire, A.M.; Taylor, J.P. Dynamic functional connectivity changes in dementia with Lewy bodies and Alzheimer's disease. *Neuroimag. Clin.* **2019**, *22*, 101812. [CrossRef] [PubMed]

49. Deng, J.; Sun, B.; Kavcic, V.; Liu, M.; Giordani, B.; Li, T. Novel methodology for detection and prediction of mild cognitive impairment using resting-state EEG. *Alzheimer's Dement.* **2024**, *20*, 411. [CrossRef] [PubMed]