Approximate Equivalence of Causalized Convergent Cross Mapping and Directed Information under Stationary Ergodic Gaussian Random Processes

Tongtong Li[1,2] Jinxian Deng[1] Boxin Sun[1] Jian Ren[1]
[1] Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI, USA
[2] Michigan Alzheimer's Disease Research Center, Ann Arbor, MI, 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. In this paper, we revisit the concept of causalized convergent cross mapping and prove its conditional equivalence with directed information under stationary ergodic Gaussian random processes. The theoretical result is demonstrated through simulation examples.**

***Index Terms*—Causality, causalized convergent cross mapping, directed information**

I. 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]–[3]. 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 [3].

The first practical causal analysis framework is Granger Causality (GC), which was proposed by Granger in 1969 [4]. 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 [5], [6]. 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 [3], [7]. Moreover, GC may not be able to detect the causation in deterministic settings [4], [8].

In 1990, directed information (DI)—the first causality detection tool based on information theory—was proposed by Massey [9] 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 [10], [11]. In [3], [10], 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 [3].

In 2012, convergent cross mapping (CCM), a new causality model based on state space reconstruction was proposed by Sugihara et al. [8], 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 [12]. 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 [3]. 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 [3], 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. In this paper, we first revisit CCM, cCCM, and DI, and then demonstrate the conditional equivalence of DI and cCCM under stationary ergodic Gaussian random processes [3] through both theoretical analysis and simulation examples.

II. A Revisit to Causalized Convergent Cross-Mapping and Directed Information

In this section, we briefly revisit the concepts of convergent cross mapping (CCM) [8] and causalized CCM (cCCM) [3].

Convergent Cross-Mapping (CCM) Consider two dynamically coupled variables X and Y which 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. Recall that the convergent cross mapping (CCM) algorithm can be summarized as:

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

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

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

- *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 $t_1, ...t_{E+1}$.
- *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 $\mathbf{M}_y$. The estimated Y_t based on $\mathbf{M}_x$, or say the cross mapping from X to Y, is defined as:

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

where

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

here d denotes the Euclidean distance between two vectors. 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

$$\rho_{\text{CCM}}(X \to Y) = \rho(\mathbf{Y}^n, \hat{\mathbf{Y}}^n)$$
$$\rho_{\text{CCM}}(Y \to X) = \rho(\mathbf{X}^n, \hat{\mathbf{X}}^n) \tag{4}$$

where ρ denotes the Pearson correlation.

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

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}.

Directed Information (DI) The DI from $\mathbf{X}^n$ to $\mathbf{Y}^n$ is defined as [9]:

$$I(\mathbf{X}^n \to \mathbf{Y}^n) = \sum_{i=1}^{n} [H(Y_i|\mathbf{Y}^{i-1}) - H(Y_i|\mathbf{Y}^{i-1}, \mathbf{X}^i)] = \sum_{i=1}^{n} I(\mathbf{X}^i; Y_i|\mathbf{Y}^{i-1})$$

The average DI from X to Y, measured in bits per sample is defined as

$$\bar{I}_n(X \to Y) = \frac{I(\mathbf{X}^n \to \mathbf{Y}^n)}{n}$$

III. The Approximate Equivalence between cCCM and DI for Gaussian Variables

In this section, we first revisit the result of Gel'fand [13] where he established the closed-form relationship of mutual information and Pearson correlation for Gaussian variables. We then show the approximate equivalence of cCCM and DI following Takens' theorem [14] and the result of Gel'fand [13] as well as the Shannon–McMillan–Breiman theorem [15].

A. Closed-Form Relationship between Pearson Correlation and Mutual Information

Mutual information (MI) is an information-theoretic metric that measures both linear and nonlinear dependence between two random variables. Let X and Y be two random variables, then the mutual information between them is defined as

$$I(X;Y) = H(X) - H(X|Y), \tag{5}$$

where $H(X)$ is the entropy of X and $H(X|Y)$ is the conditional entropy. Let Ω_x and Ω_y denote the sample space of X and Y, respectively; let $P_Y(y) = Pr\{Y = y\}$, and $P_{X|Y}(x|y) = Pr\{X = x|Y = y\}$, then

$$H(X|Y) = -\sum_{y \in \Omega_y} P_Y(y) \sum_{x \in \Omega_x} P_{X|Y}(x|y) \log P_{X|Y}(x|y) \tag{6}$$

$$= -\sum_{y \in \Omega_y} \sum_{x \in \Omega_x} P_{X,Y}(x,y) \log \left[\frac{P_{X,Y}(x,y)}{P_Y(y)}\right], \tag{7}$$

where $P_{X,Y}(X,Y) = Pr\{X = x, Y = y\}$.

In general, Pearson correlation represents the linear dependence between two random variables, while mutual information characterizes both linear and non-linear dependence between them. However, closed-form relationship between Pearson correlation and mutual information can be derived under certain conditions. More specifically, when X and

Y follow normal distributions and their joint distribution is bivariate normal, that is $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$, $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$ and

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \Sigma \right), \qquad \Sigma = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$$

then it can be proved that (Gel'fand,1957 [13])

$$I(X;Y) = -\frac{1}{2}\log(1-\rho^2), \tag{8}$$

where ρ is the ensemble average representation of the Pearson correlation between X and Y. In fact,

$$H(X) = \frac{1}{2}\log(2\pi e\sigma_x^2), \qquad H(Y) = \frac{1}{2}\log(2\pi e\sigma_y^2), \tag{9}$$

$$\begin{aligned} H(X,Y) &= \frac{1}{2}\log[(2\pi e)^2|\Sigma|] \\ &= \frac{1}{2}\log[(2\pi e)^2(1-\rho^2)\sigma_x^2\sigma_y^2]. \end{aligned} \tag{10}$$

Equation (8) then follows from equations (5),(9) and (10).

The result above was extended to vector case in 2012 [16], [17] by Arellano-Valle et al.

B. Approximate Equivalence between Causalized-CCM and DI

Theorem 1. *Let X, Y be two dynamically coupled zero-mean Gaussian random variables which are also bivariate Gaussian and share the same attractor manifold $\boldsymbol{M}$. $\boldsymbol{X}^n = \{X_1, \cdots, X_n\}$, $\boldsymbol{Y}^n = \{Y_1, \cdots, Y_n\}$ are the time series contain the samples of X, Y, respectively, and are stationary ergodic random processes. Define*

$$\bar{I}_n(X \to Y) = \frac{I(\boldsymbol{X}^n \to \boldsymbol{Y}^n)}{n}, \quad \bar{I}_n(Y;\hat{Y}) = \frac{I(\boldsymbol{Y}^n;\hat{\boldsymbol{Y}}^n)}{n}. \tag{11}$$

Then

$$\lim_{n\to\infty} \bar{I}_n(X \to Y) = \lim_{n\to\infty} \bar{I}_n(Y;\hat{Y}), \tag{12}$$

and when n is sufficiently large,

$$\bar{I}_n(X \to Y) \approx \bar{I}_n(Y;\hat{Y}) \approx -\frac{1}{2}\log(1-\rho^2_{cCCM}(X \to Y)), \tag{13}$$

where $\rho_{cCCM}(X \to Y) = \rho\left(\mathbf{Y}^n, \hat{\mathbf{Y}}^n\right)$. The result also holds in the reverse direction.

Now, we can prove the approximate equivalence of causalized CCM and DI under Gaussian variables in two steps.

Step 1: Show that if the two signals X and Y are dynamically coupled, then

$$\lim_{n\to\infty} \bar{I}_n(X \to Y) = \lim_{n\to\infty} \bar{I}_n(Y;\hat{Y}) \tag{14}$$

where

$$\bar{I}_n(X \to Y) = \frac{I(\mathbf{X}^n \to \mathbf{Y}^n)}{n}, \qquad \bar{I}_n(Y;\hat{Y}) = \frac{I(\mathbf{Y}^n;\hat{\mathbf{Y}}^n)}{n},$$

and

$$\hat{Y}(t)|\mathbf{M}_x = \sum_{i=1}^{E+1} w_i Y(t_i), \qquad t = 1, \cdots, n, \tag{15}$$

is the estimated value of $Y(t)$ in causalized-CCM, here $t_i < t$ and

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

In fact, recall that the directed information from time series $\mathbf{X}^n$ to $\mathbf{Y}^n$ is defined as

$$\begin{aligned} I(\mathbf{X}^n \to \mathbf{Y}^n) &= \sum_{i=1}^{n}[H(Y_i|\mathbf{Y}^{i-1}) - H(Y_i|\mathbf{Y}^{i-1}, \mathbf{X}^i)] \\ &= \sum_{i=1}^{n} I(\mathbf{X}^i; Y_i|\mathbf{Y}^{i-1}) \end{aligned}$$

and the average DI from X to Y, measured in bits per sample, is defined as

$$\bar{I}_n(X \to Y) = \frac{I(\mathbf{X}^n \to \mathbf{Y}^n)}{n}. \tag{16}$$

The mutual information between $\mathbf{Y}^n$ and $\hat{\mathbf{Y}}^n$ is given by

$$\begin{aligned} I(\mathbf{Y}^n;\hat{\mathbf{Y}}^n) &= H(\mathbf{Y}^n) - H(\mathbf{Y}^n|\hat{\mathbf{Y}}^n) \\ &= \sum_{i=1}^{n}[H(Y_i|\mathbf{Y}^{i-1}) - H(Y_i|\mathbf{Y}^{i-1}, \hat{\mathbf{Y}}^n)] \\ &= \sum_{i=1}^{n}[H(Y_i|\mathbf{Y}^{i-1}) - H(Y_i|\mathbf{Y}^{i-1}, \mathbf{X}^i)] \\ &+ \sum_{i=1}^{n}[H(Y_i|\mathbf{Y}^{i-1}, \mathbf{X}^i) - H(Y_i|\mathbf{Y}^{i-1}, \hat{\mathbf{Y}}^i)] \\ &+ \sum_{i=1}^{n}[H(Y_i|\mathbf{Y}^{i-1}, \hat{\mathbf{Y}}^i) - H(Y_i|\mathbf{Y}^{i-1}, \hat{\mathbf{Y}}^n)] \end{aligned} \tag{17}$$

Define $a_i = H(Y_i|\mathbf{Y}^{i-1}, \mathbf{X}^i) - H(Y_i|\mathbf{Y}^{i-1}, \hat{\mathbf{Y}}^i)$ and $b_i = H(Y_i|\mathbf{Y}^{i-1}, \hat{\mathbf{Y}}^i) - H(Y_i|\mathbf{Y}^{i-1}, \hat{\mathbf{Y}}^n)$, then it follows that

$$I(\mathbf{Y}^n;\hat{\mathbf{Y}}^n) = I(\mathbf{X}^n \to \mathbf{Y}^n) + \sum_{i=1}^{n} a_i + \sum_{i=1}^{n} b_i \tag{18}$$

Model Y as

$$Y_i = \hat{Y}_i + e_i, \tag{19}$$

where $\hat{Y}_i = \hat{Y}(i)$ is defined in equation (15). Without loss of generality, we can assume that e_i is independent of $\hat{Y}_i$, and has zero-mean and variance $\sigma^2_{e_i}$. Following the argument in Takens' theorem [14], as $i \to \infty$, the shadow manifolds $\mathbf{M}_x$ and $\mathbf{M}_y$ become denser and the neighborhood shrinks such that $\hat{Y}_i \to Y_i$. It then follows that $\lim_{i\to\infty} \sigma^2_{e_i} = 0$. Note that

$$0 \le H(Y_i|\mathbf{Y}^{i-1}, \mathbf{X}^i) \le H(e_i) \le \frac{1}{2}\log(2\pi e\sigma^2_{e_i}). \tag{20}$$

It then follows that

$$H(Y_i|\mathbf{Y}^{i-1}, \mathbf{X}^i) \to 0 \quad \text{as} \quad i \to \infty. \tag{21}$$

Similarly,

$$H(Y_i|\mathbf{Y}^{i-1},\hat{\mathbf{Y}}^{i})\to 0,\qquad H(Y_i|\mathbf{Y}^{i-1},\hat{\mathbf{Y}}^{n})\to 0\ \text{ as }\ i\to\infty. \tag{22}$$

Following the definition of a_i and b_i, we have

$$a_i\to 0,\quad b_i\to 0\ \text{ as }\ i\to\infty. \tag{23}$$

Define

$$\bar{I}_n(Y;\hat{Y})=\frac{I(\mathbf{Y}^n;\hat{\mathbf{Y}}^n)}{n}, \tag{24}$$

it then follows from equation (18)that

$$\frac{I(\mathbf{Y}^n;\hat{\mathbf{Y}}^n)}{n}=\frac{I(\mathbf{X}^n\to\mathbf{Y}^n)}{n}+\frac{1}{n}\sum_{i=1}^{n}a_i+\frac{1}{n}\sum_{i=1}^{n}b_i \tag{25}$$

Note that $a_i\to 0,\quad b_i\to 0\ \text{ as }\ i\to\infty$, it follows that

$$\frac{1}{n}\sum_{i=1}^{n}a_i\to 0,\qquad \frac{1}{n}\sum_{i=1}^{n}b_i\to 0\ \text{ as }\ n\to\infty \tag{26}$$

That is, $\lim_{n\to\infty}\bar{I}_n(X\to Y)=\lim_{n\to\infty}\bar{I}_n(Y;\hat{Y})$. This implies that when n is sufficiently large,

$$\bar{I}_n(X\to Y)\approx\bar{I}_n(Y;\hat{Y}). \tag{27}$$

Step 2: Show that: if Y and $\hat{Y}$ are Gaussian random variables and their joint distribution is bivariate Gaussian, then when n is sufficiently large,

$$\bar{I}_n(X\to Y)\approx\bar{I}_n(Y;\hat{Y})\approx-\frac{1}{2}\log(1-\rho^2_{\text{cCCM}}(X\to Y)), \tag{28}$$

where $\rho_{\text{cCCM}}(X\to Y)=\rho(\mathbf{Y}^n,\hat{\mathbf{Y}}^n)$.

This can be obtained from the closed-form relationship between Pearson correlation and mutual information. Let $I(Y;\hat{Y})$ and $\rho(Y,\hat{Y})$ denote mutual information and Pearson correlation between Y and $\hat{Y}$, respectively. If Y and $\hat{Y}$ are Gaussian random variables and their joint distribution is bivariate Gaussian, based on Gel'fand's result in equation (8), we have

$$I(Y;\hat{Y})=-\frac{1}{2}\log(1-\rho^2(Y,\hat{Y})). \tag{29}$$

Recall that the mutual information rate $\bar{I}_n(Y;\hat{Y})=\frac{I(\mathbf{Y}^n;\hat{\mathbf{Y}}^n)}{n}$ represents the average mutual information between $\mathbf{Y}^n$ and $\hat{\mathbf{Y}}^n$, measured in bits per sample, and each sample in $\mathbf{Y}^n$ can be regarded as a message emitted from the memoryless random source Y, and each sample in $\hat{\mathbf{Y}}^n$ is emitted from $\hat{Y}$. On the other hand, $I(Y;\hat{Y})$, is the average information (virtually) "transmitted" between Y and $\hat{Y}$, measured in bits per sample. That is, they have the same physical meaning. As $n\to\infty$, both $\mathbf{Y}^n$ and $\hat{\mathbf{Y}}^n$ are stationary ergodic random processes, following the Shannon–McMillan–Breiman theorem [15], we have

$$\bar{I}_n(Y;\hat{Y})\to I(Y;\hat{Y}),\quad\text{as}\quad n\to\infty. \tag{30}$$

Similarly,

$$\rho(\mathbf{Y}^n,\hat{\mathbf{Y}}^n)\to\rho(Y,\hat{Y}),\quad\text{as}\quad n\to\infty, \tag{31}$$

we can then get equation (28) from equations (27), (29)-(31). The result in the reverse direction can be proved in a similar manner.

IV. Simulation results

The approximate equivalence of cCCM and DI is demonstrated using resting-state fMRI data, which are often modeled as Gaussian random variables [20]. For fMRI, we investigate the baseline data of 30 subjects from the risk reduction for Alzheimer's disease (rrAD) trial [19], where 18 common regions of interest of the Default Mode Network (DMN) were extracted and sorted in descending order by their connection strength to the isthmus of the posterior cingulate cortex seed region time course. The fMRI data of each brain region is regarded as a dynamic manifold with a deterministic attractor but perturbed by random afferent input and noise. Here, the total length of the BOLD (blood-oxygen-level-dependent) signal is 284 samples, with a sampling period of 2.5s.

The relationship between the estimated DI and cCCM causation is illustrated in Figure 1. Figure 1 (A) and 1 (B) plot the DI and cCCM values between all the $18\times 17=306$ region pairs in the DMN for all the 30 subjects. That is, each figure has $18\times 17\times 30=9180$ points. We can see that there is a log-relationship between them—DI can be represented using cCCM and vice versa. It should be noted though, due to the finite data size, the quantization error in the digitization process of DI calculation [19], and the noise in the fMRI data, the estimated DI and ρ_{cCCM} satisfy the following relationship, which is a linear transformation of equation (28),

$$\hat{I}_n(X\to Y)\approx a[-1/2\log_2(1-\rho^2_{\text{cCCM}}(X\to Y))]+b \tag{32}$$

where $a=0.7945, b=0.2578$ in this case.

We also conduct causality analysis of the DMN using DI and cCCM for a randomly selected participant from the rrAD trial. The results is presented in Figure 2. As can be seen, both bidirectional and unidirectional causality can be observed in the DMN of the selected participant.

V. Conclusions and Discussions

In this paper, we revisited the definition of CCM, identified the gap between CCM and the traditional definition of causality, presented causalized CCM (cCCM), and demonstrated the conditional equivalence of cCCM and directed information through both theoretical analysis and simulation examples. More details on implementation of cCCM for causality detection can be found in [21], where we demonstrated the effectiveness of cCCM in the detection of causality under different scenarios through a large number of examples, and provided detailed discussions on the configuration of the cCCM algorithm.

Acknowledgement

This research was partially supported by National Science Foundations under awards 2032709 and 1919154.

References

[1] P. A. Stokes and P. L. Purdon, "A study of problems encountered in granger causality analysis from a neuroscience perspective," *Proceedings of the National Academy of Sciences*, vol. 114, 8 2017.

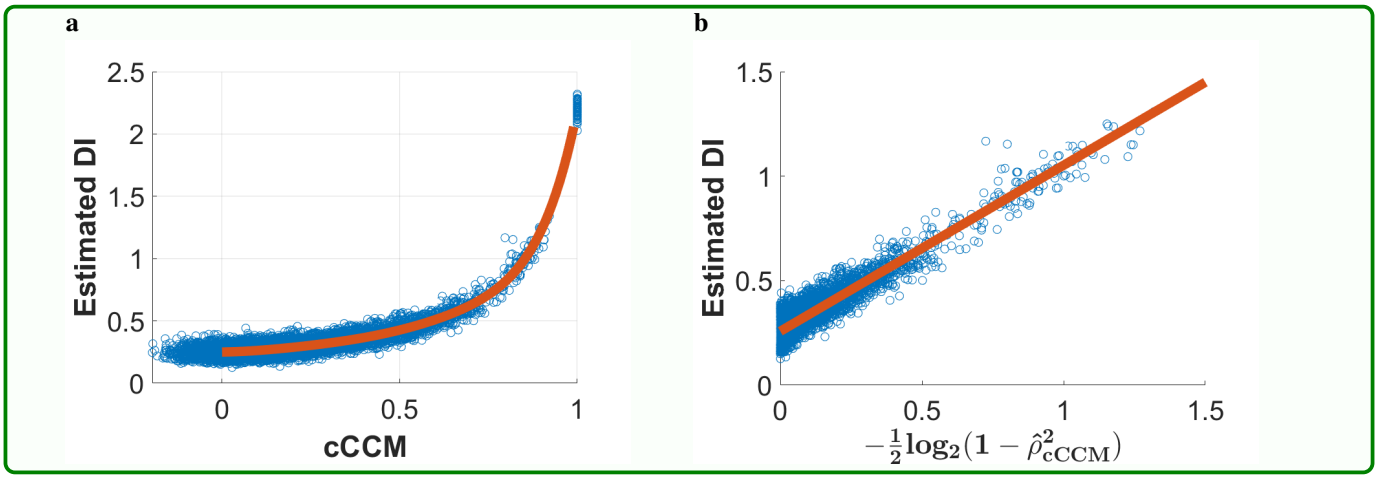

Figure 1. **Demonstration of the approximate equivalence of cCCM and DI using resting-state fMRI data of 30 subjects from the rrAD trials** [18], [19]. **a**) The approximate log-relationship between estimated DI and cCCM. **b**) The relationship between the directly estimated DI and cCCM-predicted DI.

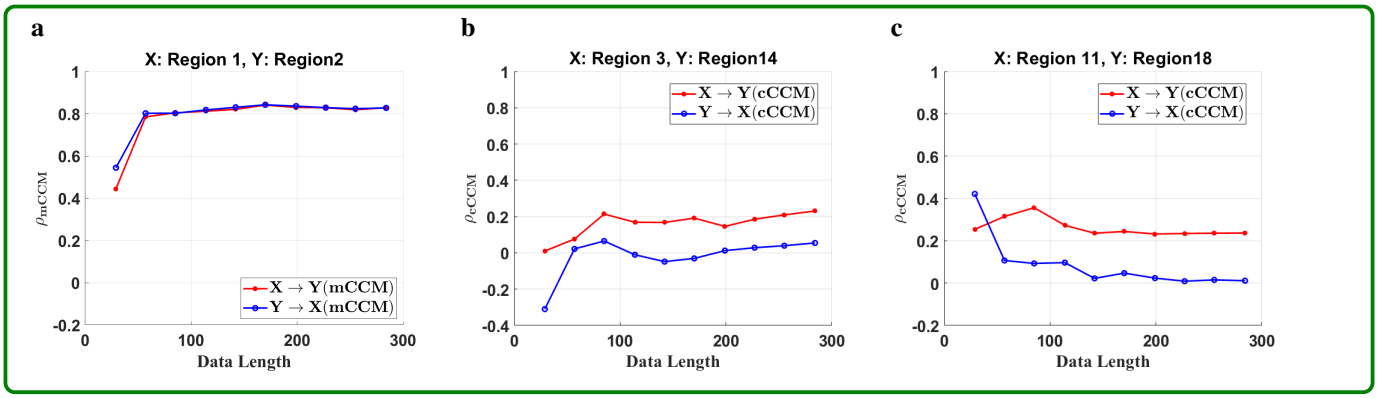

Figure 2. **Causality analysis for a single participant**. **a**)-**c**), Illustration of cCCM causation versus fMRI data length for different region pairs. Region names: 1-Posterior Cingulum right, 2-Posterior Cingulum left, 3-Precuneus / Angular Gyrus right, 11-Medial pre-frontal Thalamus right, 14-Hippocampus Subiculum / Presubiculum left, 18-BA11 Inferior Frontal Gyrus left.

[2] H. Hillebrandt, K. J. Friston, and S.-J. Blakemore, "Effective connectivity during animacy perception – dynamic causal modelling of human connectome project data," *Scientific Reports*, vol. 4, p. 6240, 9 2014.

[3] J. Deng, B. Sun *et al.*, "Causalized convergent cross-mapping and its approximate equivalence with directed information in causality analysis," *PNAS Nexus*, vol. 3, 2024.

[4] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, p. 424, 8 1969.

[5] C. Granger and P. Newbold, *Forecasting Economic Time Series*. Elsevier, 1977.

[6] L. Barnett and A. K. Seth, "The mvgc multivariate granger causality toolbox: A new approach to granger-causal inference," *Journal of Neuroscience Methods*, vol. 223, pp. 50–68, 2 2014.

[7] O. David, I. Guillemain *et al.*, "Identifying neural drivers with functional mri: an electrophysiological validation." *PLoS biology*, vol. 6, pp. 2683–97, 12 2008.

[8] G. Sugihara, R. May *et al.*, "Detecting causality in complex ecosystems," *Science*, vol. 338, pp. 496–500, 10 2012.

[9] J. Massey, "Causality, feedback, and directed information," in *The International Symposium on Information Theory and Its Applications*, 11 1990, pp. 303–305.

[10] Z. Wang, A. Alahmadi *et al.*, "Causality analysis of fmri data based on the directed information theory framework," *IEEE Transactions on Biomedical Engineering*, vol. 63, pp. 1002–1015, 5 2016.

[11] P.-O. Amblard and O. J. J. Michel, "On directed information theory and granger causality graphs," *Journal of Computational Neuroscience*, vol. 30, pp. 7–16, 2 2011.

[12] E. R. Deyle, M. C. Maher *et al.*, "Global environmental drivers of influenza," *Proceedings of the National Academy of Sciences*, vol. 113, pp. 13 081–13 086, 11 2016.

[13] I. Gel'fand and Y. A, "Calculation of amount of information about a random function contained in another such function," in *American Mathematical Society translations*, 12 1957, pp. 199–246.

[14] F. Takens, *Detecting strange attractors in turbulence*. SpringerVerlag, 1981, pp. 366–381.

[15] P. H. Algoet and T. M. Cover, "A sandwich proof of the shannon-mcmillan-breiman theorem," *The Annals of Probability*, vol. 16, 4 1988.

[16] R. B. Arellano-Valle, J. E. Contreras-Reyes, and M. G. Genton, "Shannon entropy and mutual information for multivariate skew-elliptical distributions," *Scandinavian Journal of Statistics*, vol. 40, pp. 42–62, 3 2013.

[17] A. Komaee, "Mutual information rate between stationary gaussian processes," *Results in Applied Mathematics*, vol. 7, p. 100107, 8 2020.

[18] N. Scheel, J. N. Keller *et al.*, "Evaluation of noise regression techniques in resting-state fmri studies using data of 434 older adults," *Frontiers in Neuroscience*, vol. 16, 10 2022.

[19] ClinicalTrials.gov, "Risk reduction for alzheimer's disease (rrad) (2016-2022)," 2 2022.

[20] K. J. Friston, A. P. Holmes *et al.*, "Statistical parametric maps in functional imaging: A general linear approach," *Human Brain Mapping*, vol. 2, pp. 189–210, 1 1994.

[21] B. Sun, J. Deng *et al.*, "Causalized convergent cross mapping and its implementation in causality analysis," *Entropy*, vol. 26, p. 539, 6 2024.