

RESEARCH ARTICLE

A study for multiscale information transfer measures based on conditional mutual information

Xiaogeng Wan ^{*}, Lanxi Xu

Department of Mathematics, College of Science, Beijing University of Chemical Technology, Beijing, China

* wxgbj88@sina.com



Abstract

As the big data science develops, efficient methods are demanded for various data analysis. Granger causality provides the prime model for quantifying causal interactions. However, this theoretic model does not meet the requirement for real-world data analysis, because real-world time series are diverse whose models are usually unknown. Therefore, model-free measures such as information transfer measures are strongly desired. Here, we propose the multi-scale extension of conditional mutual information measures using MORLET wavelet, which are named the WM and WPM. The proposed measures are computational efficient and interpret information transfer by multi-scales. We use both synthetic data and real-world examples to demonstrate the efficiency of the new methods. The results of the new methods are robust and reliable. Via the simulation studies, we found the new methods outperform the wavelet extension of transfer entropy (WTE) in both computational efficiency and accuracy. The features and properties of the proposed measures are also discussed.

OPEN ACCESS

Citation: Wan X, Xu L (2018) A study for multiscale information transfer measures based on conditional mutual information. PLoS ONE 13(12): e0208423. <https://doi.org/10.1371/journal.pone.0208423>

Editor: Lawrence M. Ward, University of British Columbia, CANADA

Received: June 15, 2018

Accepted: November 17, 2018

Published: December 6, 2018

Copyright: © 2018 Wan, Xu. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: X. Wan is supported by the Fundamental Research Funds for the Central Universities (ZY1823) and funding from the Faculty of Sciences at Beijing University of Chemical Technology (YY1707).

Competing interests: The authors declare no conflict of interest of any nature or kind in any product, service and/or company that could be

1 Introduction

As big data science developments, practical time series methods are demanded to study the complexity and dynamics of the data. Real-world data are time series usually obtained by experiments or observations whose models are diverse and the data are often nonlinear or non-stationary [1–2], e.g. the EEG time series measured from experiments [2–5] and financial data observed from real-world markets [6]. Therefore, efficient method is necessary to study the dynamics of these complex systems.

Various directed methods have been developed for studying the directed interaction between time series. The most classic causality measure is the Granger causality (GC) [7–8], it is a prime model for causality measures which uses significance tests to detect the directed dependency of one time series on another time series [7–8]. However, GC has many limitations, it is bivariate time domain method suits for only linear models [7–8]. Other similar methods have been derived to cover the limitation of GC [9–12]. For instance, the conditional Granger causality is a multivariate method that can detect direct interactions between time series [10], the frequency domain GC is derived for frequency domain data analysis [10–11], and the nonlinear GC can be applied to nonlinear data analysis [12]. More advanced measures

construed as influencing the position presented in, or the review of, the manuscript.

are such as the Partial Direct Coherence (PDC) [13–17] and Directed Transfer Function (DTF) [18–19], which are sophisticated frequency domain measures. However, both PDC and DTF are linear measures whose validity rely on the linear autoregressive or moving average model fitting [13–19].

Since real world time series are often nonlinear and sometimes nonstationary, the data models are usually varied and unknown, therefore linear causal measures are farfetched to analyze the real-world time series. To analyze the dynamics and interactions between real-world time series and also theoretic models, information transfer measures are preferred [1, 12, 20–30]. Information transfer measures are used to detect the directed information transfer between coupled time series which can be used to study the direction of interactions of complex networks. However, the information transfer is a different notion to the causal effect [31–34]. In [33], J. T. Lizier and M. Prokopenko have used the study of the transfer entropy [27] and information flow to differentiate the concepts of information transfer and causal effect [33]. N. Ay and D. Polani have introduced in [31] the notion of causal independence which allows for defining a measure for the strength of a causal effect. In their work, they call this notion the information flow which is compared with the transfer entropy. A relevant work is presented by Wibral et. al. in [34], in which an extension method from transfer entropy is proposed that account for delayed source-target interactions, while crucially retaining the conditioning on the embedded target state at the immediately previous time step [34]. This new extension method proposed by Wibral et. al. in [34] is proved as the only relevant option in keeping with the Wiener's principle of causality. To clarify the notional causality which is different from notion of information transfer, a clear and systematic literature of causality, including theories and causal models is presented in the work by J. Pearl in [32].

Transfer entropy (TE) is a fundamental information transfer measure proposed by T. Schreiber [27], it is a directed information transfer method that evaluates the bivariate information transfer between coupled time series. Due to the model-free nature of information methods, information transfer measures such as transfer entropy (TE) are preferred in many studies to analyze the interactions for varied models. To suffice the needs of different analysis, many other information transfer methods are derived from the transfer entropy. For instance, the symbolic transfer entropy (STE) [28] and the Partialized Transfer Entropy (PTE) [35] are derived to improve TE for particular applications. However, these transfer entropy measures are computationally redundant, in that they use uniform embeddings in their expressions, which leads to high computational complexity and redundancy in their computations, because variables of no significant contribution to the information transfer detection are also included in the computation [22, 29].

To solve this problem, non-uniform state-space embedding methods such as MIME (conditional mutual information from mixed embeddings) [29] and partial MIME (PMIME, a direct version of MIME) [22] are developed to reduce the computational redundancies. Both MIME and PMIME use a progressive scheme of a maximum criterion and a stopping criterion to select significant contributive components from uniform state-space embedding vector to form refined embedding vectors for information transfer detection [22, 29]. In consequence, both methods are computational efficient and have wide applications to various data analysis [22, 26, 29, 35–38].

Most of these measures require data stationarity [7–8, 13–19]. Real world data such as financial and biological time series are not seriously stationary. Wavelet is an ideal tool for non-stationary data analysis who presents good solutions to time and frequency allocations and outperforms the short-time Fourier transforms [24, 39–43]. In [24], MORLET wavelet [40] is introduced to TE to cover non-stationary and discontinuous data analysis [24]. Since TE is primitive and computationally redundant, we are inspired to use MORLET wavelet to

extend MIME and PMIME into computational efficient multi-scale measures that cover the deficiency of WTE, and the new extensions are expected to be useful in various data analysis particular the real-world data analysis.

In this paper, we introduce the MORLET wavelet extension to MIME and PMIME and study their efficiency in application to both model data and real-world time series. The paper is organized as follows. In the Introduction section, we review the background of this study. In the section of Materials and methods, we describe the formulative wavelet extension of MIME and PMIME. In the Results section, four synthetic models (theoretic maps) and two real-world examples (EEG and financial data) are used to demonstrate the efficiency of the proposed extensions, where all simulation studies are compared to the wavelet-extension of TE (WTE). In the Discussion section, application and features of the new methods are discussed. The final conclusion of this study is drawn in the Conclusion section.

2 Materials and methods

In this section, we introduce the details of the wavelet extension of MIME and PMIME. Here, we refer the two wavelet extensions as WM and WPM, respectively.

2.1 MORLET wavelet extension of MIME (WM)

Conditional mutual information from mixed embeddings (MIME) [29] is a time domain non-linear information transfer measure whose wavelet extension is described as follows.

Assume X and Y are two arbitrary time series of length N, L_X and L_Y denote the maximum time lag for X and Y. To compute the WM for $Y \rightarrow X$, a mother function $\psi(\eta) = \pi^{-1/4} e^{i\omega_0\eta} e^{-\eta^2/2}$ [2, 24, 39–43] is used to convert the X and Y time series into MORLET wavelet coefficients [24]

$$V_{s_i, \tau_X} = \frac{1}{\sqrt{s_i}} \sum_{t=1}^N X_t \psi^* \left(\frac{t - \tau_X}{s_i} \right) \tag{1}$$

$$W_{s_i, \tau_Y} = \frac{1}{\sqrt{s_i}} \sum_{t=1}^N Y_t \psi^* \left(\frac{t - \tau_Y}{s_i} \right) \tag{2}$$

where $\omega_0 \in [5,6]$ is the normalized frequency, time lags τ_X ($1 \leq \tau_X \leq L_X$) and τ_Y ($1 \leq \tau_Y \leq L_Y$) are the translation parameters used to localize the wavelet, s_i ($1 \leq i \leq m$, m is the total number of time scales) is the time scale that determines the width and resolution of the wavelet, * represents the complex conjugation [24]. This wavelet setting is the same to the WTE [24].

The WM are computed for every time scale s_i ($1 \leq i \leq m$). For each time scale s_i ($1 \leq i \leq m$), a future embedding vector of X of time horizon T [29] ($T \in \mathbb{N}^+$) is defined as

$$V_F(s_i) = (V_{s_i, \tau+1}, V_{s_i, \tau+2}, \dots, V_{s_i, \tau+T}); \tag{3}$$

A collective set of candidate components is also defined at the time scale:

$$B(s_i) = (V_{s_i, \tau}, V_{s_i, \tau-1}, \dots, V_{s_i, \tau-L_X}, W_{s_i, \tau}, W_{s_i, \tau-1}, \dots, W_{s_i, \tau-Y}); \tag{4}$$

The same progressive scheme of MIME [29] is used for each time scale s_i . The progressive scheme starts with an empty vector $\mathbf{b}_0(s_i) = \emptyset$. In the first iterative cycle, WM goes through B (s_i) to find the element x' that satisfies the maximum criterion [29]:

$$I : I(x'; V_F(s_i)) = \max_{x' \in B(s_i)} I(x; V_F(s_i)), \quad x' \in B(s_i). \tag{5}$$

where $I(x; V_F(s_i))$ is the mutual information rate between the x and $V_F(s_i)$. The element x' that

satisfies the maximum criterion is selected to join $\mathbf{b}_0(s_i)$ that forms $\mathbf{b}_1(s_i) = (x')$, x' is then removed from $B(s_i)$ and obtains $B_{k-1}(s_i) = B(s_i) \setminus \{x'\}$ [29].

At a k -th iterative cycle, WM seeks the element x' in the remaining set $B_{k-1}(s_i)$ (obtained from the $k-1$ -th iterative cycle) that satisfies the maximum criterion [29]

$$I : I(x'; V_F(s_i) | \mathbf{b}_{k-1}(s_i)) = \max_{x \in B_{k-1}(s_i)} I(x; V_F(s_i) | \mathbf{b}_{k-1}(s_i)), \quad x' \in B_{k-1}(s_i), \quad (6)$$

and moves the element x' from $B_{k-1}(s_i)$ to $\mathbf{b}_{k-1}(s_i)$ to obtain the enlarged embedding vector $\mathbf{b}_k(s_i) = (x', \mathbf{b}_{k-1}(s_i))$ and $B_k(s_i) = B_{k-1}(s_i) \setminus \{x'\}$.

The progressive scheme stops at a $k+1$ -th iterative cycle and uses $\mathbf{b}_k(s_i)$ as the final selected embedding vector, if the following stopping criterion is satisfied [29]:

$$\frac{I(\mathbf{b}_k(s_i); V_F(s_i))}{I(\mathbf{b}_{k+1}(s_i); V_F(s_i))} > A, \quad (7)$$

Here, A is the significance threshold (between 0 and 1) controls the inclusion of embedding components [29]. This stopping criterion ensures contributive components to be included while prevents useless components from being added. The progressive scheme stops if no significant information can be given when including new component is included [29, 35–38].

The WM for time scale s_i is evaluated by [29]:

$$WM_{Y \rightarrow X}(s_i) = 1 - \frac{I(V_F(s_i); \mathbf{b}_k^V(s_i))}{I(V_F(s_i); \mathbf{b}_k(s_i))} = \frac{I(V_F(s_i); \mathbf{b}_k^W(s_i) | \mathbf{b}_k^V(s_i))}{I(V_F(s_i); \mathbf{b}_k(s_i))} \quad (8)$$

We note that this WM information transfer between coupled time series is evaluated at the same time scale s_i ($i = 1, 2, \dots, 64$), which means that the WM does not evaluate the cross scales information transfers. This is limitation of this method. The evaluation of cross-scale information transfers will be our research of the next stage.

2.2 MORLET wavelet extension PMIME (WPM)

WPM is the multi-variate version of WM, which inferences only the direct interactions. Without the loss of generality, assume X, Y and Z are three arbitrary time series of length N , L_X, L_Y, L_Z are the maximum time lags for the three time series. WPM first converts the X, Y, Z time series into MORLET wavelet coefficients [24]:

$$V_{s_i, \tau_X} = \frac{1}{\sqrt{s_i}} \sum_{t=1}^N X_t \Psi^* \left(\frac{t - \tau_X}{s_i} \right) \quad (9)$$

$$W_{s_i, \tau_Y} = \frac{1}{\sqrt{s_i}} \sum_{t=1}^N Y_t \Psi^* \left(\frac{t - \tau_Y}{s_i} \right) \quad (10)$$

$$U_{s_i, \tau_Z} = \frac{1}{\sqrt{s_i}} \sum_{t=1}^N Z_t \Psi^* \left(\frac{t - \tau_Z}{s_i} \right) \quad (11)$$

where the mother function and all the other wavelet parameters are the same to WM.

WPM values are computed for every time scale. To compute the WPM of $Y \rightarrow X$, a future embedding vector [22] of time horizon T is defined:

$$V_F(s_i) = (V_{s_i, \tau+1}, V_{s_i, \tau+2}, \dots, V_{s_i, \tau+T}). \quad (12)$$

Different from WM, the collective set of candidate components are multivariate that are defined by all time series in the system [22]:

$$B(s_i) = (V_{s_i,\tau}, V_{s_i,\tau-1}, \dots, V_{s_i,\tau-L_X}, W_{s_i,\tau}, W_{s_i,\tau-1}, \dots, W_{s_i,\tau-Y}, U_{s_i,\tau}, U_{s_i,\tau-1}, \dots, U_{s_i,\tau-Z}); \quad (13)$$

The initial selected embedding vector is again an empty vector $\mathbf{b}_0(s_i) = \emptyset$.

WPM follows the same progressive scheme and the maximum criterion to WM with the only difference of the collective set of candidate components. WPM selects candidate components from all rather than two variables in the system which contribute to the inference of direct information transfers [22].

If the progressive scheme stops at a $k+1$ -th iterative cycle and uses $\mathbf{b}_k(s_i)$ as the final selected embedding vector. The WPM for $Y \rightarrow X$ at time scale s_i is given by

$$WPM_{Y \rightarrow X}(s_i) = \frac{I(V_F(s_i); b_k^W(s_i) | b_k^V(s_i), b_k^U(s_i))}{I(V_F(s_i); b_k(s_i))} \quad (14)$$

where $b_k^V(s_i)$, $b_k^W(s_i)$ and $b_k^U(s_i)$ are the X, Y and Z components of $\mathbf{b}_k(s_i)$ [22, 29], respectively. Also, we should note that the WPM evaluates information transfer between wavelet coefficients of the same scales.

2.3 Bias correction by surrogate data

We use time-shifted surrogate [22,24,29,44–49,50] to test the significance of the results. Take WM as an example. Let $\{V_{s_i,\tau_X}\}$ and $\{W_{s_i,\tau_Y}\}$ to denote the MORLET wavelet coefficients for the two arbitrary time series X and Y, the $WM_{X \rightarrow Y}(s_i)$ indicates the WM information transfer from X to Y evaluated at time scale s_i . We fix $\{V_{s_i,\tau_X}\}$ and permute the temporal indices of $\{W_{s_i,\tau_Y}\}$ randomly [22,24,29,50] to obtain the surrogate of $\{W_{s_i,\tau_Y}\}$. Next, we apply the WM method on the original series of $\{V_{s_i,\tau_X}\}$ and the surrogate time series of $\{W_{s_i,\tau_Y}\}$, the results are denoted as $WM_{X \rightarrow Y}(s_i, q)$, where q is the index for the surrogates of $\{W_{s_i,\tau_Y}\}$. Thus, the bias corrected WM for $X \rightarrow Y$ is given by [24]

$$WM_{C,X \rightarrow Y}(s_i) = WM_{X \rightarrow Y}(s_i) - \max_q \{WM_{X \rightarrow Y}(s_i, q)\} \quad (15)$$

In the following context, we use $WM_{C,X \rightarrow Y}(s_i)$ to denote the bias corrected WM information transfer from X to Y [22,24,29,50]. The bias corrected WM for the inverse direction, and the bias corrected WPM and WTE are similarly defined. In all simulations, we use $q = 10$ [22,24,29,50].

2.4 Contrast information transfer

To obtain the dominance of interaction between coupled time series, we compute the contrast information transfer between paired variables [24]. For example, to analyze the dominance of interaction between two time series X and Y, we compute the contrast information transfer between $X \rightarrow Y$ and $Y \rightarrow X$ for each time scale s_i : $\Omega_{WM,X \rightarrow Y}(s_i) = WM_{C,X \rightarrow Y}(s_i) - WM_{C,Y \rightarrow X}(s_i)$. If $\Omega_{WM,X \rightarrow Y}(s_i) > 0$, the dominant information transfer is detected for $X \rightarrow Y$; and vice versa, if $\Omega_{WM,Y \rightarrow X}(s_i) > 0$, the dominant information transfer is detected for $Y \rightarrow X$. The contrast information transfer for WPM and WTE are similarly defined.

3 Results

In this section, we use six examples, including both synthetic data and real-world time series, to demonstrate the efficiency of WM and WPM. In these examples, various types of interactions are displayed, and all simulation studies are compared with the method of WTE.

3.1 Synthetic data

Synthetic data are examples of model time series generated by equations. These examples include the Henon maps, linear autoregressive models, and Lorenz systems [22, 29, 50–53], which are indeterministic systems that are frequently used for time series analysis [22, 29, 50, 53]. Here, we use four synthetic examples to demonstrate the efficiency of the proposed information transfer measures.

3.1.1 Cosine map. The cosine map consists of two unidirectionally coupled first order autoregressive processes where one of the processes contains a cosine [24]:

$$\begin{aligned} x_{t+1} &= 0.7x_t + 0.7 \cos(0.3t) + n_t^{(x)}(0, \sigma^2), \\ y_{t+1} &= 0.7y_t + n_t^{(y)}(0, \sigma^2) + ex_t, \end{aligned} \tag{16}$$

where $n_t^{(x)}, n_t^{(y)}$ are independent zero mean Gaussian random processes with variance $\sigma^2 = 2$, and $e \in [0, 1]$ is the coupling strength controls the linear interaction from X_t to Y_t . The data of this example can be found in [S1 Dataset](#).

The initial data are randomly generated from normal distribution with zero mean and unit variance. Each data series contains 5×10^4 data points. For comparison purpose, we use the same MORLET parameters for WM and WPM as recommended for WTE [24, 40]: $r = 0.125$, $\omega_0 = 6, s_0 = 0.5, V = 10$, and $n = 64$.

To analyze the effect of stopping criteria, we compute the WM and WPM as functions of the stopping criterion A . The stopping criterion A represents the proportionality of the conditional mutual information between the current and the previous iterative cycles. In general applications [22, 29, 46], A is a value close to but no greater than 1 [22, 29]. Larger values of A representing looser criteria, while smaller values of A imply more rigid criteria. In the MIME and PMIME studies, the usual choice of the stopping criteria is $A = 0.95$ for MIME and $A = 0.97$ for PMIME. These choices of A value are obtained by various simulation studies, $A = 0.95$ and 0.97 are appropriate A values that not only allow useful lagged values to be detected but also prevent false positiveness from being included [22, 29, 46]. To study the impact of the stopping criteria A on evaluation of the information transfer, we alter the criterion $A = 0.91, 0.93, 0.95, 0.97, 0.99$, and present the WM and WPM results (directional inference) in [Table 1](#). Since the cosine map has only two processes, the multivariate measure WPM has same results to that of the bivariate measure WM. We can see from [Table 1](#), when the coupling strength $0.1 \leq e \leq 1$, both WM and WPM indicate the correct information transfer from X to Y . When $e = 0$, the coupling disappears and no flow of information is detected.

An example of the contrast WM results between X and Y (coupling strength $e = 0.5$) at different scales are shown in [Fig 1](#). In this figure, we can see that different stopping criteria A presents similar results of the contrast WM between X and Y . However, the stopping criterion theoretically becomes looser when A increases. In later Henon map analysis, we will see that $A = 0.95$ and $A = 0.97$ are good choices, but $A = 0.95$ is a bit rigid than $A = 0.97$ in the

Table 1. WM information transfer for the cosine map with different stopping criteria and coupling strengths.

Coupling strength	A = 0.91	A = 0.93	A = 0.95	A = 0.97	A = 0.99
e = 0	-	-	-	-	-
0.1 ≤ e ≤ 1	X->Y	X->Y	X->Y	X->Y	X->Y

This table shows the directional inference of the cosine map by using WM at different stopping criteria A and coupling strength e . Since the cosine map is bivariate, the WM and WPM have the same results on this example.

<https://doi.org/10.1371/journal.pone.0208423.t001>

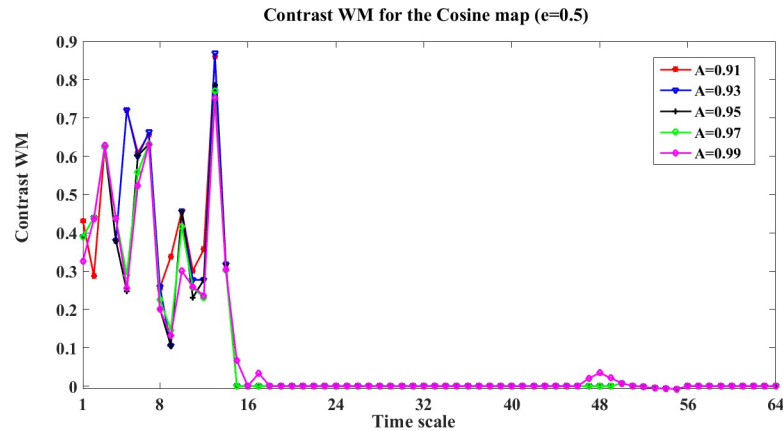


Fig 1. The contrast WM for the Cosine map at different stopping criteria A (e = 0.5). This figure shows the contrast WM $\Omega_{WM,X \rightarrow Y}(s_i, e) = WM_{C,X \rightarrow Y}(s_i, e) - WM_{C,Y \rightarrow X}(s_i, e)$ ($i = 1, 2, \dots, 64, e = 0, 0.1, 0.2, \dots, 1$) for the cosine map (coupling $e = 0.5$). The curves with different colors represent the contrast WM at different stopping criteria ($A = 0.91, 0.93, 0.95, 0.97, 0.99$).

<https://doi.org/10.1371/journal.pone.0208423.g001>

directional inference of indirect interactions, therefore we use $A = 0.97$ for both WM and WPM in all simulation studies.

In this example, we use the stopping criterion $A = 0.97$ and the referenced embedding parameters [22, 29, 36–38]: $T = 1$ (time horizon, prediction time step) and $L_{max} = 5$ (the maximum time lag). The WM, WPM and WTE information transfer values are filtered with the surrogate bias correction. Moreover, the contrast information transfer values between X and Y are computed for WM, WPM and WTE which are shown in Figs 2–4, respectively.

In these figures, the contrast information transfer values are plotted against time scales and coupling strength. We can see that both all three measures inference the correct linear interaction from $X \rightarrow Y$, which are supported by the non-negative surfaces of $\Omega_{WM,X \rightarrow Y}$ (Fig 2), $\Omega_{WPM,X \rightarrow Y}$ (Fig 3), and $\Omega_{WTE,X \rightarrow Y}$ (Fig 4).

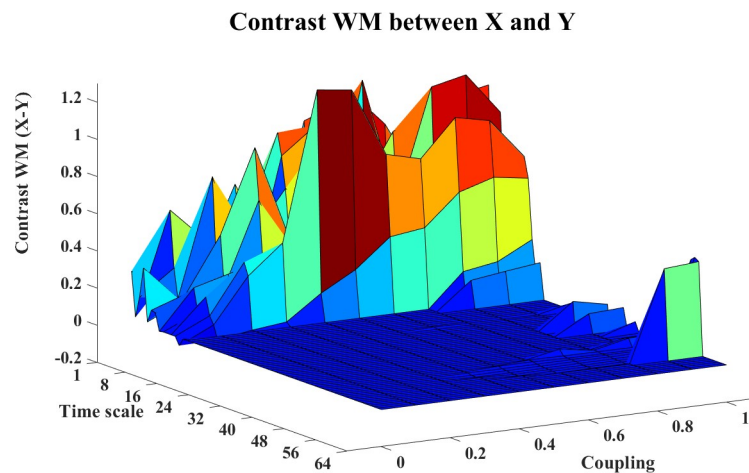


Fig 2. The contrast WM information transfer values for the cosine map. This figure shows the 3D surface of the contrast WM $\Omega_{WM,X \rightarrow Y}(s_i, e) = WM_{C,X \rightarrow Y}(s_i, e) - WM_{C,Y \rightarrow X}(s_i, e)$ ($i = 1, 2, \dots, 64, e = 0, 0.1, 0.2, \dots, 1$) for the cosine map. This surface represents the values of the contrast information transfer against different time scale s_i and coupling strength e . The non-negative surface implicates the directed influence from $X \rightarrow Y$.

<https://doi.org/10.1371/journal.pone.0208423.g002>

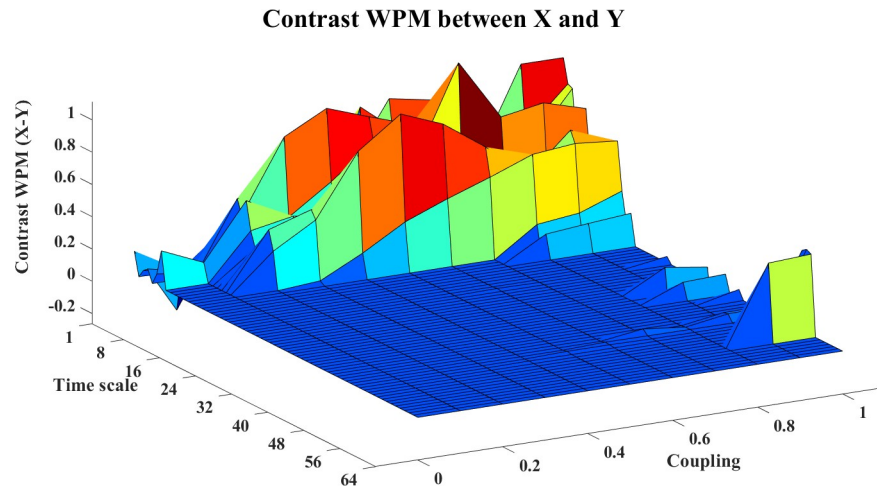


Fig 3. The contrast WPM information transfer values for the cosine map. This figure shows the 3D surface of the contrast WPM $\Omega_{WPM,X \rightarrow Y}(s_i, e) = WPM_{C,X \rightarrow Y}(s_i, e) - WPM_{C,Y \rightarrow X}(s_i, e)$ ($i = 1, 2, \dots, 64, e = 0, 0.1, 0.2, \dots, 1$) for the cosine map. The surface represents the values of the corrected information transfer against time scale s_i and coupling strength e . The non-negative surface implies the directed information transfer from $X \rightarrow Y$.

<https://doi.org/10.1371/journal.pone.0208423.g003>

In these figures, we can see a ‘cosine’ shape (a slowly increasing ridge or a ‘tail’ rather than well-localized in scale) of the surfaces, which nicely reflects the ‘cosine’ influence on the dynamics of the coupled system [24].

To see the information transfer in each independent direction, we plot the bias corrected information transfer values in Fig 5.

3.1.2 Four unidirectionally coupled Henon maps. The four unidirectionally coupled Henon maps are defined by the equations [22,24–25,29,36–38,51–52]:

$$\begin{aligned}
 X_{1,n+1} &= 1.4 - x_{1,n}^2 + 0.3X_{1,n-1}, \\
 X_{2,n+1} &= 1.4 - 0.3X_{1,n}X_{2,n} - 0.7X_{2,n}^2 + 0.3X_{2,n-1}, \\
 X_{3,n+1} &= 1.4 - 0.3X_{2,n}X_{3,n} - 0.7X_{3,n}^2 + 0.3X_{3,n-1}, \\
 X_{4,n+1} &= 1.4 - 0.3X_{3,n}X_{4,n} - 0.7X_{4,n}^2 + 0.3X_{4,n-1},
 \end{aligned}
 \tag{17}$$

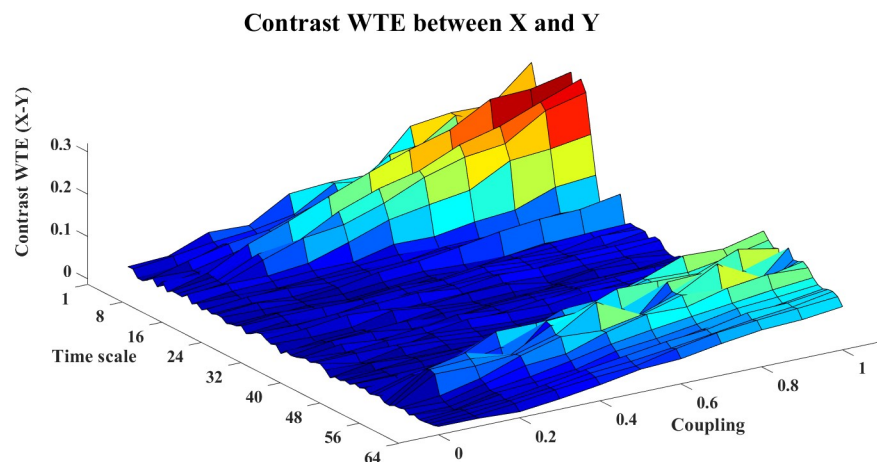


Fig 4. The contrast WTE information transfer values for the cosine map. In this figure, the 3D surface presents the contrast WTE $\Omega_{WTE,X \rightarrow Y}(s_i, e) = WTE_{C,X \rightarrow Y}(s_i, e) - WTE_{C,Y \rightarrow X}(s_i, e)$ ($i = 1, 2, \dots, 64, e = 0, 0.1, 0.2, \dots, 1$) for the cosine map at different time scale s_i and coupling strength e . The non-negative surface indicates the directed information transfer from $X \rightarrow Y$.

<https://doi.org/10.1371/journal.pone.0208423.g004>

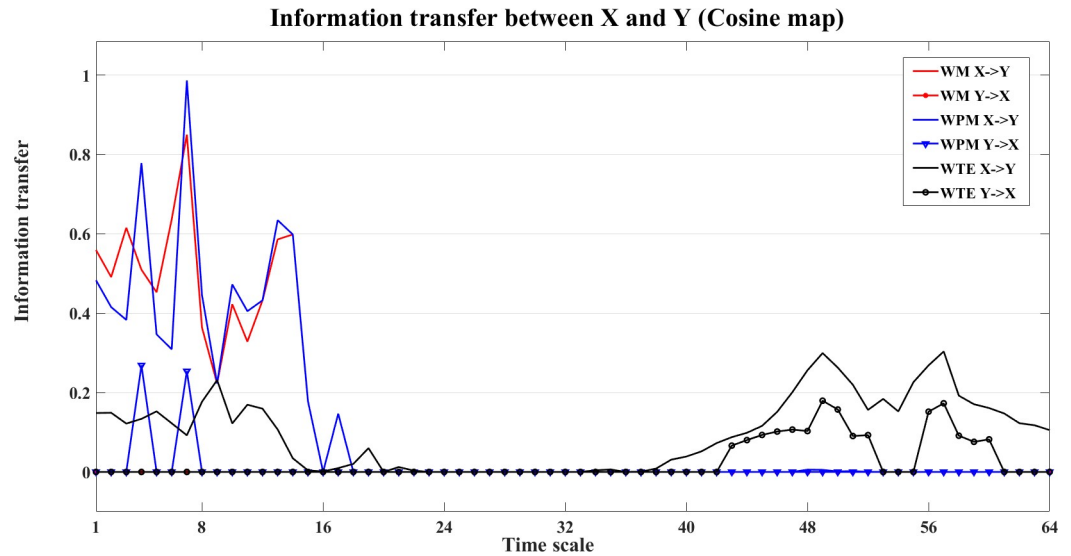


Fig 5. Information transfer between X and Y. In this figure, we present the WM (red), WPM (blue) and WTE (black) information transfer values of cosine map (coupling strength $e = 0.8$) in each independent direction. We can see that all measures present dominant information transfer values for $X \rightarrow Y$.

<https://doi.org/10.1371/journal.pone.0208423.g005>

the direct nonlinear directed interactions are from $X_{i-1} \rightarrow X_i, i = 2, 3, 4$. The wavelet parameters used are $r = 0.125, \omega_0 = 6, s_0 = 0.2, V = 8$, and $n = 64$ [24]. The other WM and WPM parameters are the same to the previous example. The data of this example can be found in S2 Dataset.

These Henon maps have more than two variables, the WPM results are different from that of the WM. To study the influence of the stopping criteria on the multivariate example, we make the following analysis on the direct interaction between X_2 and X_3 , and also the indirect interaction between X_1 and X_3 . Results of the directional inference of different stopping criteria A is shown in Table 2. In this table, both WM and WPM detect all correct interactions between the unidirectionally coupled Henon maps. When the stopping criteria is small ($A \leq 0.95$), WM detects only direct interactions, when $A \geq 0.97$, WM also detects the indirect interactions from $X_1 \rightarrow X_3, X_2 \rightarrow X_4$. WPM is a direct measure, so it detects only the direct interactions for all the different stopping criteria.

To analyze the influence of the stopping criteria, we take X_1 and X_3 as an example. Fig 6 shows the contrast WM between X_1 and X_3 with different stopping criteria and time scales. In this figure, the indirect interaction from $X_1 \rightarrow X_3$ can only be detected when $A \geq 0.97$, and when $A = 0.99$, WM presents false positiveness for $X_3 \rightarrow X_1$. Because WPM is a direct measure, it gives all zeros for the information transfer between X_1 and X_3 for all stopping criteria.

Table 2. Directional inference of WM and WPM with different stopping criteria.

Stopping criteria	WM	WPM
$A = 0.91$	$X_1 \rightarrow X_2, X_2 \rightarrow X_3, X_3 \rightarrow X_4$	$X_1 \rightarrow X_2, X_2 \rightarrow X_3, X_3 \rightarrow X_4$
$A = 0.93$	$X_1 \rightarrow X_2, X_2 \rightarrow X_3, X_3 \rightarrow X_4$	$X_1 \rightarrow X_2, X_2 \rightarrow X_3, X_3 \rightarrow X_4$
$A = 0.95$	$X_1 \rightarrow X_2, X_2 \rightarrow X_3, X_3 \rightarrow X_4$	$X_1 \rightarrow X_2, X_2 \rightarrow X_3, X_3 \rightarrow X_4$
$A = 0.97$	$X_1 \rightarrow X_2, X_2 \rightarrow X_3, X_3 \rightarrow X_4, X_1 \rightarrow X_3, X_2 \rightarrow X_4$	$X_1 \rightarrow X_2, X_2 \rightarrow X_3, X_3 \rightarrow X_4$
$A = 0.99$	$X_1 \rightarrow X_2, X_2 \rightarrow X_3, X_3 \rightarrow X_4, X_1 \rightarrow X_3, X_2 \rightarrow X_4$	$X_1 \rightarrow X_2, X_2 \rightarrow X_3, X_3 \rightarrow X_4$

This table shows the directional inference of WM and WPM between the four unidirectionally coupled Henon maps with different stopping criteria A . We can see that WPM detects only the direct interactions, while WM also detects indirect direction when $A \geq 0.97$.

<https://doi.org/10.1371/journal.pone.0208423.t002>

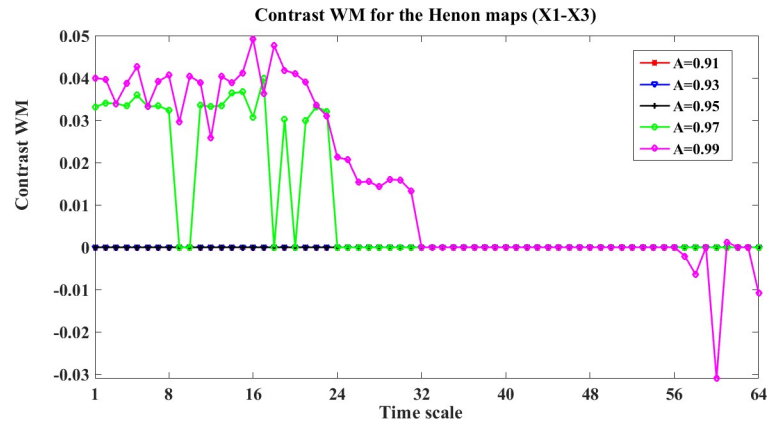


Fig 6. The contrast WM for the four unidirectionally coupled Henon maps with different stopping criteria (X1-X3). This figure shows the line plots of the contrast WM (contrast values between X1 and X3) $\Omega_{WM,X1 \rightarrow X3}(s_i, e) = WM_{C,X1 \rightarrow X3}(s_i, e) - WM_{C,X3 \rightarrow X1}(s_i, e)$ ($i = 1, 2, \dots, 64$) for the Henon maps between X2 and X3. This curves with different colors represent the contrast WM results obtained by different stopping criteria.

<https://doi.org/10.1371/journal.pone.0208423.g006>

Similarly, we analyze the criteria effect on the direct interactions. We take the direct interaction X2->X3 as an example. The WM and WPM information transfer between X2 and X3 with different stopping criteria A are shown in Figs 7 and 8. We can see from these figures, the different stopping criteria presents similar values of the contrast results for both WM and WPM, and the trends of the contrast WM and the contrast WPM are similar. This is because this interaction from X2->X3 is direct, and the WM and WPM may present similar results.

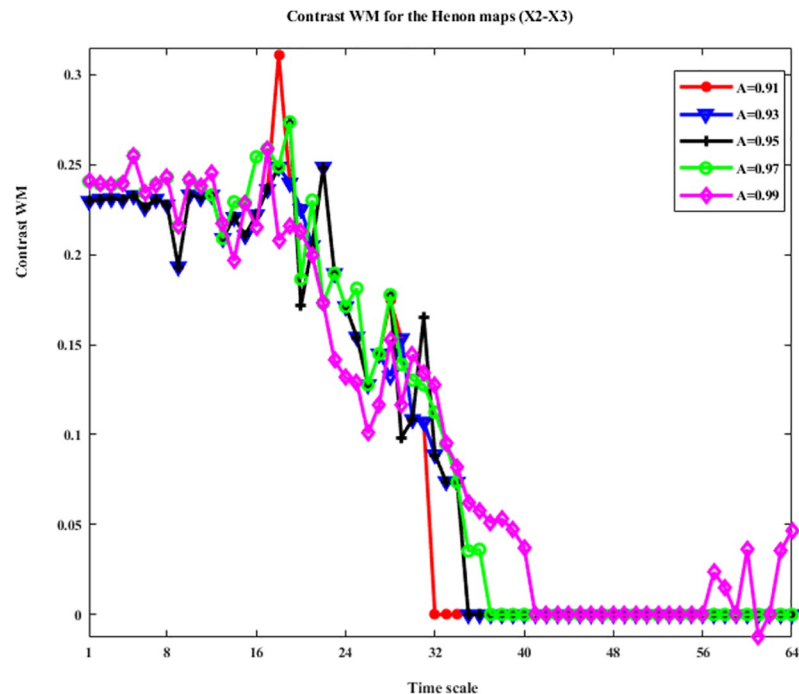


Fig 7. The contrast WM for the four unidirectionally coupled Henon maps with different stopping criteria (X1-X3). This figure shows the contrast WM between X1 and X3: $\Omega_{WM,X1 \rightarrow X3}(s_i, e) = WM_{C,X1 \rightarrow X3}(s_i, e) - WM_{C,X3 \rightarrow X1}(s_i, e)$ ($i = 1, 2, \dots, 64$) with different stopping criteria and time scales. This curves with different colors represent the contrast WM obtained by different stopping criteria.

<https://doi.org/10.1371/journal.pone.0208423.g007>

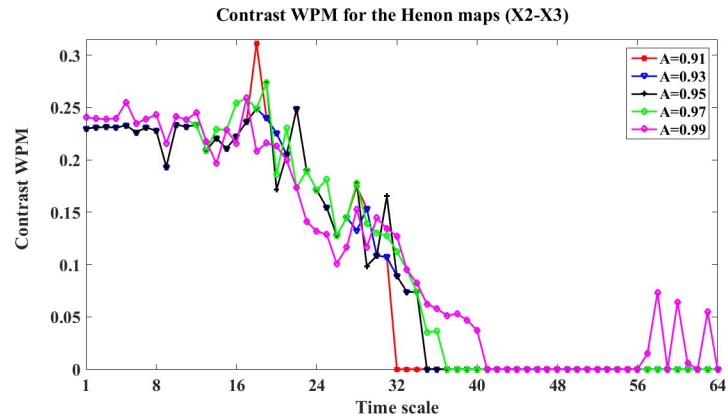


Fig 8. The contrast WPM for the four unidirectionally coupled Henon maps with different stopping criteria (X1-X3). This figure shows the contrast WM between X1 and X3: $\Omega_{WPM,X1 \rightarrow X3}(s_i, e) = WPM_{C,X1 \rightarrow X3}(s_i, e) - WPM_{C,X3 \rightarrow X1}(s_i, e)$ ($i = 1, 2, \dots, 64$) with different stopping criteria and time scales. This curves with different colors represent the contrast WPM obtained by different stopping criteria.

<https://doi.org/10.1371/journal.pone.0208423.g008>

Also, we note that in the contrast results of WM, when $A = 0.99$ the contrast WM presents negative values which implies a false direction from $X3 \rightarrow X2$. This maybe because the A is large and the stopping criteria become too loose that false positive is detected in this situation.

For the overall situation for the directional inference, the average information transfer values (over time scales) for the Henon maps are shown in Fig 9. In this figure, the average information transfer values are plotted by color-matrices, the correspondence between the colors and the information transfer values is shown in the color-bar. In the color-matrices, the directional inference of each lattice is from the row channel to the column channel, e.g. the (1,2)-th lattice in the color-matrix represents the average information transfer for $X1 \rightarrow X2$. We can see that WM indicates all the correct directions of interactions, while WPM inferences only the direct interactions. In this study, WTE only indicates clear interaction from $X1 \rightarrow X2$,

Average information transfer for the four unidirectionally coupled Henon maps

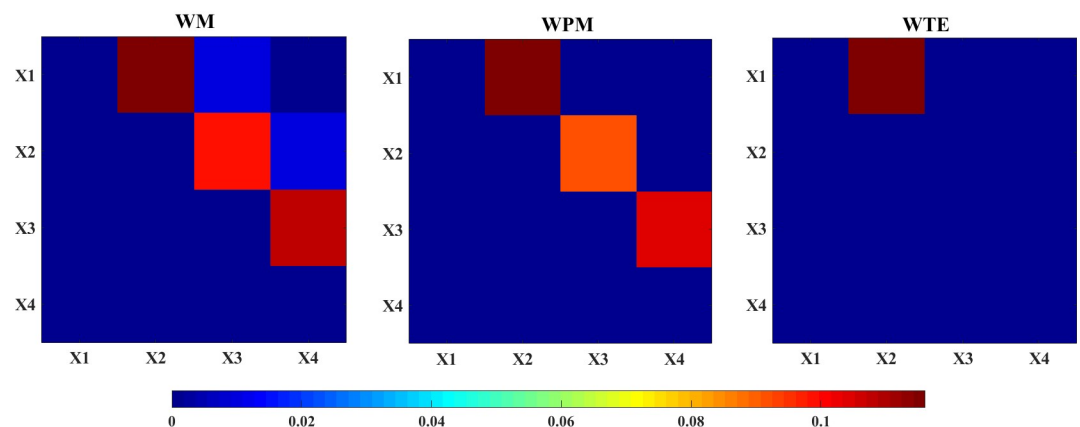


Fig 9. The color-map for the average information transfer between the four unidirectionally coupled Henon maps. The three color-graphs separately show the average WM (left), WPM (middle) and WTE (right) information transfer for the four unidirectionally coupled Henon maps. The direction of each lattice is read from the row channel to the column channel. In this figure, WM indicates all correct interactions from $X_i \rightarrow X_j$ where $i < j$, WPM indicates only the direct directions $X_i \rightarrow X_{i+1}$, $i = 1, 2, 3$, while WTE indicates only $X1 \rightarrow X2$ and other directions are failed to be detected.

<https://doi.org/10.1371/journal.pone.0208423.g009>

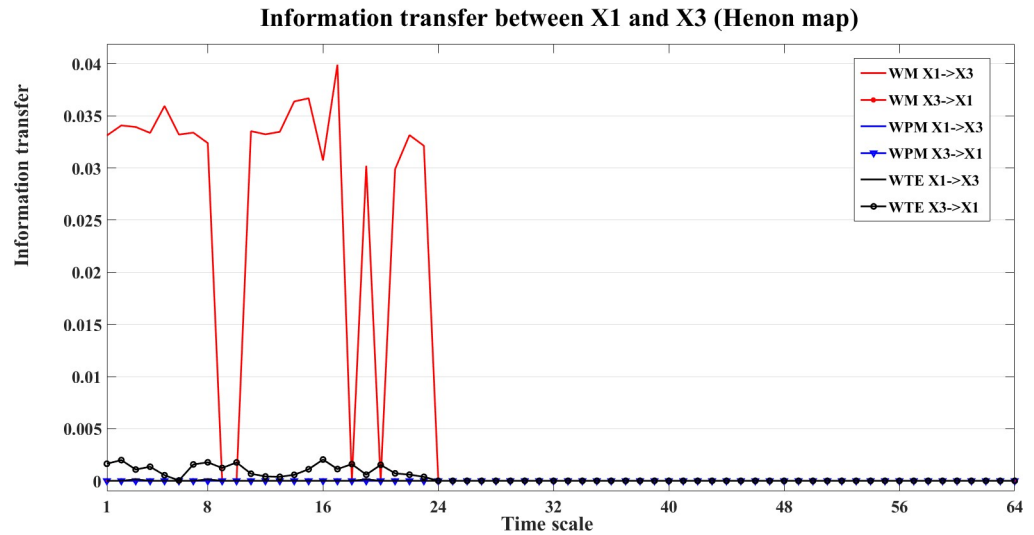


Fig 10. Information transfer between X1 and X3 for the four unidirectionally coupled Henon maps. In this figure, we present the bias corrected information transfer values between X1 and X3. The curves in different colors separately show the corrected WM (red), WPM (blue) and WTE (black) information transfer between X1 and X3, which are plotted against the different time scales.

<https://doi.org/10.1371/journal.pone.0208423.g010>

although the average information transfer values for $X3 \rightarrow X1$ and $X3 \rightarrow X2$ are positive, but they are too weak compared to the strength of $X1 \rightarrow X2$, and cannot be shown in this color map.

An example of the information transfer between X1 and X3 is shown in Figs 10 and 11. Fig 10 shows the bias reduced information transfer values for WM, WPM and WTE between X1 and X3. We can see from Fig 10 that WM presents positive information transfer for $X1 \rightarrow X3$ and zero information transfer for $X3 \rightarrow X1$. Since $X1 \rightarrow X3$ is an indirect direction of interaction, and WPM is a direct measure, therefore no positive information transfer is detected by

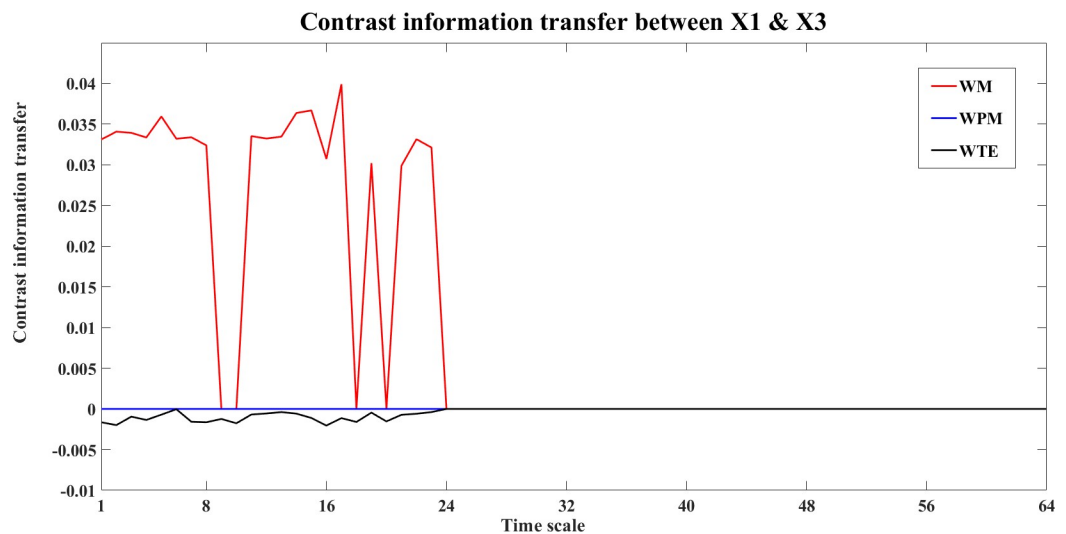


Fig 11. The contrast information transfer values for $X1 \rightarrow X3$ for the four unidirectionally coupled Henon maps. This figure shows the contrast WM (red), WPM (blue) and WTE (black) information transfer between X1 and X3 at different time scales s_i ($i = 1, 2, \dots, 64$). WM (red) indicate the indirect information transfer from $X1 \rightarrow X3$ (positive curves), while WPM (blue) presents strictly vanished $\Omega_{WPM, X1 \rightarrow X3}(s_i)$ for all scales, which indicates no direct information transfer between X1 and X3. WTE (black) presents subtle but negative $\Omega_{WTE, X1 \rightarrow X3}(s_i)$ which indicates false information transfer from $X3 \rightarrow X1$.

<https://doi.org/10.1371/journal.pone.0208423.g011>

WPM between X1 and X3. WTE presents false positiveness for X3->X1 but zeros information transfer for X1->X3.

The contrast information transfer values for X1->X3 ($\Omega_{X1 \rightarrow X3}(s_i), i = 1, \dots, 64$) are shown in Fig 11. In this figure, WM (red) indicates the indirect information transfer from X1->X3, while WPM (blue) indicates no direct information transfer between X1 and X3. WTE (black) presents negative $\Omega_{WTE, X2 \rightarrow X3}(s_i)$, which indicates false direction from X3->X1.

Also, we note that the information transfer values of WM and WPM decline to zero for large time scales. This may due to many reasons. One possible reason is that when time scale increases, the frequency and resolution decrease, details of the time series are smeared out, hence the causal interactions become too weak to be detected. The other reason is because of the characteristic correlation time of Henon maps [51–52]: if the time lag of Henon maps exceeds the characteristic cross-correlation time, the directed influence disappears [54]. The threshold of time scales may indicate the cross-correlation time of Henon maps [54]. Alternatively, it may due to the full correlation between the coupled time series that if the time series are fully correlated, the system becomes deterministic, whose values of information transfer between the wavelet coefficients become zero. The signal correlation at certain common frequencies [24, 39–43] is also be a possible reason for the vanishing causalities.

3.1.3 A system of three coupled variables. The system of three coupled variables is given by the equations [50]:

$$\begin{aligned} x_{1,t} &= 3.4x_{1,t-1}(1 - x_{1,t-1})^2 \exp(-x_{1,t-1}^2) + 0.4\epsilon_{1,t} \\ x_{2,t} &= 3.4x_{2,t-1}(1 - x_{2,t-1})^2 \exp(-x_{2,t-1}^2) + 0.5x_{1,t-1}x_{2,t-1} + 0.4\epsilon_{2,t} \\ x_{3,t} &= 3.4x_{3,t-1}(1 - x_{3,t-1})^2 \exp(-x_{3,t-1}^2) + 0.3x_{2,t-1} + 0.5x_{1,t-1}^2 + 0.4\epsilon_{3,t} \end{aligned} \tag{18}$$

where $X_2 \rightarrow X_3$ is the linear interaction of the system and $X_1 \rightarrow X_2$ and $X_1 \rightarrow X_3$ are nonlinear interactions, $\epsilon_{i,t}$ ($i = 1, \dots, 3$) are Gaussian random white noises. The data of this example can be found in S3 Dataset. The WM and WPM parameters are the same to the previous example.

The average information transfer values (over all time scales) are computed for the system which are shown in Fig 12. In this figure, all measures correctly identify the linear ($X_2 \rightarrow X_3$) and nonlinear ($X_1 \rightarrow X_2$ and $X_1 \rightarrow X_3$) interactions.

The bias corrected information transfer values are computed for each direction. The WM and WPM information transfer for the X1->X2 are similar, and all three measures present the correct interaction from X1->X2. We take the direction of X1->X3 as an example. The information transfer for both X1->X3 and X3->X1 are presented in the Fig 13. In this figure, we can see that all three measures detect the correct nonlinear interaction from X1->X3. This can also be seen from the plot of the contrast information transfer between X1 and X3 in Fig 14.

For the interaction between X2 and X3, the contrast information transfer between X2 and X3 is shown in Fig 15.

3.1.4 Three coupled Lorenz systems with nonlinear couplings. The three coupled Lorenz systems with nonlinear couplings $X_1 \rightarrow X_2$ and $X_2 \rightarrow X_3$ are given by the following equations [50]:

$$\begin{aligned} \dot{x}_1 &= 10(y_1 - x_1) \\ \dot{y}_1 &= 28x_1 - y_1 - x_1z_1 \\ \dot{z}_1 &= x_1y_1 - \frac{8}{3}z_1 \end{aligned} \tag{19}$$

Average information transfer for the system of three coupled variables

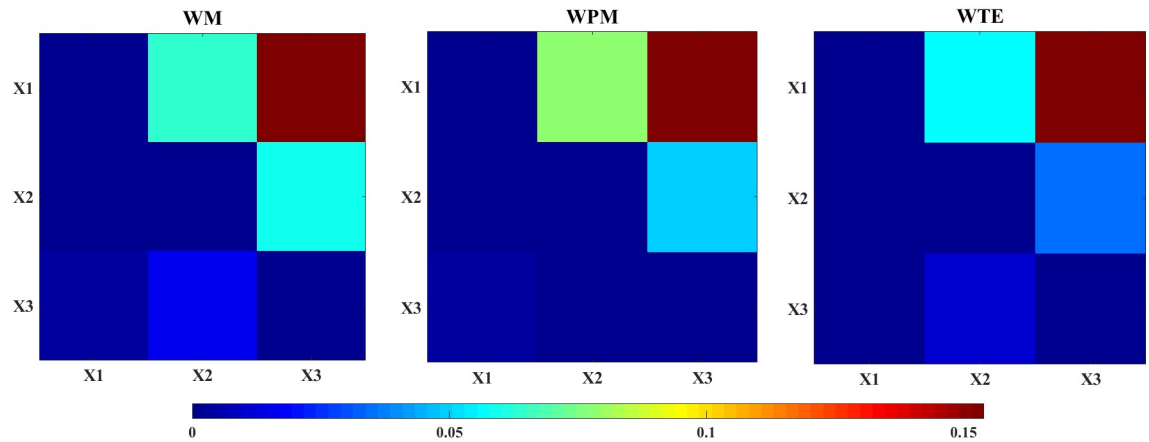


Fig 12. The color-map of the average information transfer values for the system of three coupled variables. The three color-maps show the average WM (left), WPM (middle) and WTE (right) values for the system of three coupled variables. The directed direction of each lattice is indicated from the row channel to the column channel. Comparing the information transfer values between opposite directions, all measures identify the correct linear (X2->X3) and nonlinear (X1->X2, X1->X3) interactions.

<https://doi.org/10.1371/journal.pone.0208423.g012>

$$\begin{aligned}
 \dot{x}_2 &= 10(y_2 - x_2) + k(x_1 - x_2) \\
 \dot{y}_2 &= 28x_2 - y_2 - x_2z_2 \\
 \dot{z}_1 &= x_2y_2 - \frac{8}{3}z_2
 \end{aligned}
 \tag{20}$$

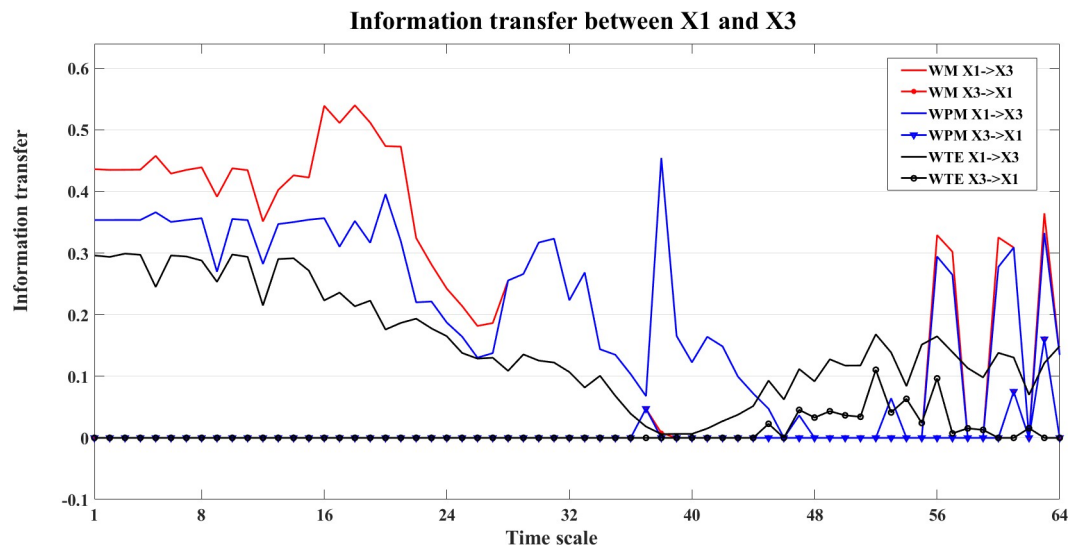


Fig 13. Information transfer between X1 and X3 the nonlinear interaction X1->X3. In this figure, we present the bias corrected WM (red), WPM (blue) and WTE (black) information transfer between X1 and X3 at different time scales s_i ($i = 1, 2, \dots, 64$). By comparing the strength between opposite directions, all three measures indicate the correct nonlinear interaction from X1->X3.

<https://doi.org/10.1371/journal.pone.0208423.g013>

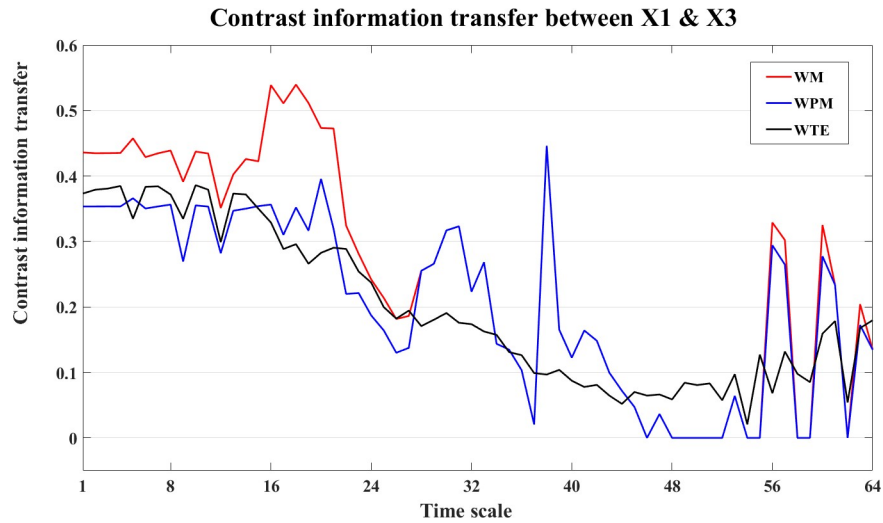


Fig 14. The contrast information transfer values for the nonlinear interaction X1->X3. The three curves show the contrast WM (red), WPM (blue) and WTE (black) information transfer between X1 and X3 at different time scales s_i ($i = 1, 2, \dots, 64$). All three measures indicate the correct nonlinear interaction from X1->X3.

<https://doi.org/10.1371/journal.pone.0208423.g014>

$$\begin{aligned}
 \dot{x}_3 &= 10(y_3 - x_3) + k(x_2 - x_3) \\
 \dot{y}_3 &= 28x_3 - y_3 - x_3z_3 \\
 \dot{z}_3 &= x_3y_3 - \frac{8}{3}z_3
 \end{aligned}
 \tag{21}$$

where k ($k = 0, 1, 3, 5$) is the coupling strength regulates the interaction from $X_1 \rightarrow X_2$ and $X_2 \rightarrow X_3$. All time series become completely synchronized when the coupling strengths $k \geq 8$.

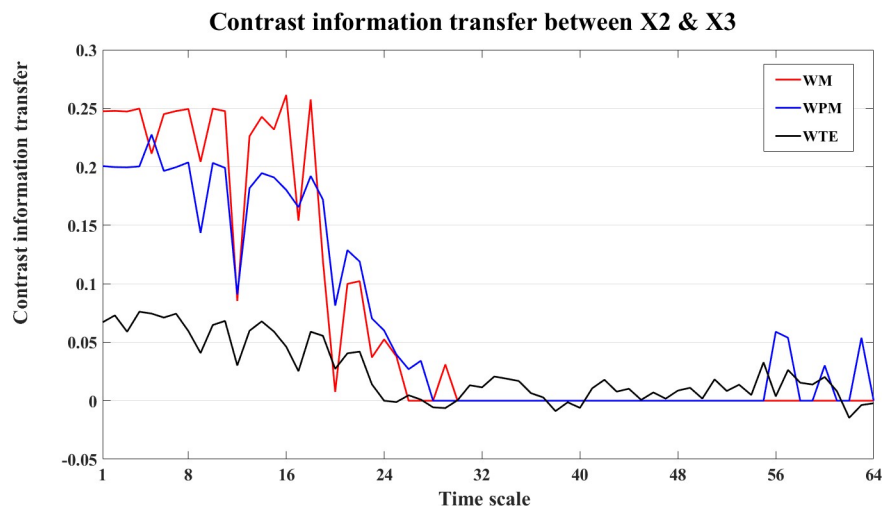


Fig 15. The contrast information transfer values for the linear interaction X2->X3. The three curves show the contrast WM (red), WPM (blue) and WTE (black) information transfer between X2 and X3 at different time scales s_i ($i = 1, 2, \dots, 64$). All measures indicate X2->X3 at small time scale s_i ($1 \leq i < 30$), and contrast values decline to zero at middle time scales. The contrast WM keeps zero at all larger scales, while the contrast WPM rises positive again at some large time scales. The contrast WTE fluctuates around zero level with more positive than negative values. All three measures are able to indicate the correct linear interaction from X2->X3.

<https://doi.org/10.1371/journal.pone.0208423.g015>

Average information transfer for the three coupled Lorenz systems (k=3)

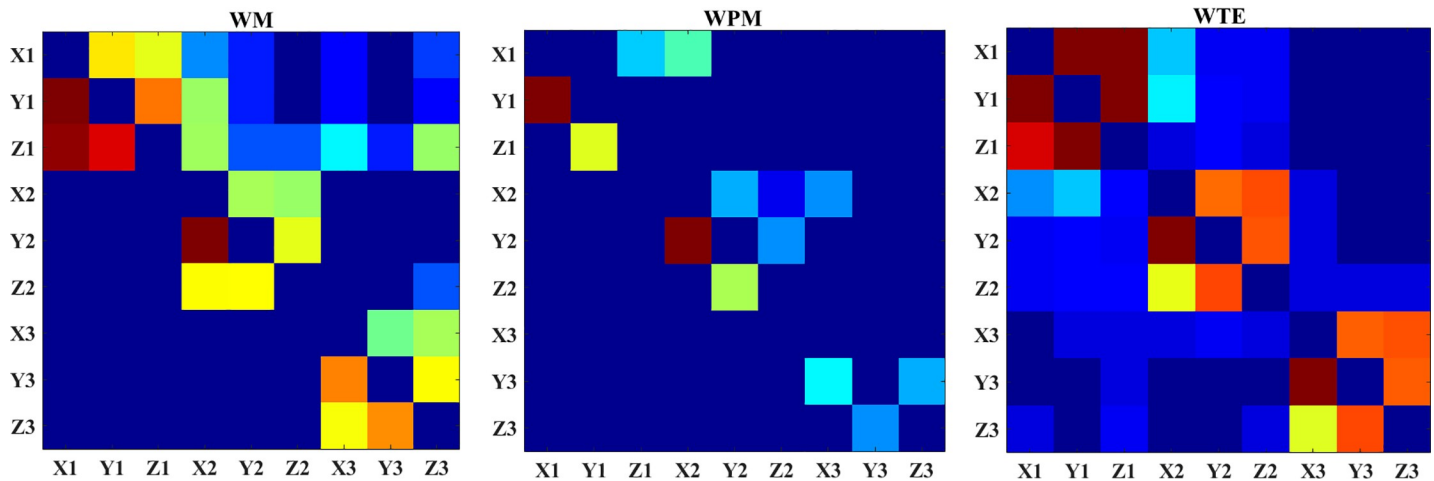


Fig 16. The color-map of the average information transfer for the three coupled Lorenz systems (k = 3). This figure shows the color-map for the average WM (left), WPM (middle) and WTE (right) information transfer (coupling strength $k = 3$) over time scales. The directed direction of each lattice is indicated from the row to the column channel. The bright diagonal blocks of WM and WPM indicate internal interaction within each system. In the WM and WPM graphs, the upper-right blocks are comparatively brighter than the lower-left blocks, which indicates the interactions between the systems. WPM indicates only direct interactions from the first to the second system and from the second to the third system, but no indirect interaction from the first to the third system. WTE presents similar strength of information transfer on both directions between coupled variables.

<https://doi.org/10.1371/journal.pone.0208423.g016>

The WM and WPM parameters are the same to the previous example. The data of this example can be found in [S4 Dataset](#).

The average values of information transfer (over all time scales) for the Lorenz systems ($c = 3$) are shown in [Fig 16](#). In this figure, both WM and WPM indicate clear interactions within each system and the cross interactions between different systems. WTE also indicates the internal and external interaction, but with less directional manner.

[Fig 17](#) shows the bias corrected information transfer between X1 and X2. In this figure, we can see that the WM indicates $X1 \rightarrow X2$ at middle time scales, while WPM indicates $X1 \rightarrow X2$ at lower time scales. WTE presents similar and subtle information transfer between X1 and X2. Similarly, the information transfer for the indirect interaction between X1 and X3 is shown in [Fig 18](#).

The contrast information transfers between the systems are shown in [Figs 19–24](#). [Figs 19–21](#) present the contrast information transfer between X1 and X2, in these figures, we see that both WM and WPM give clear positive information transfers for $X1 \rightarrow X2$, while WTE presents fluctuate information transfers between X1 and X2. Similarly, [Figs 22–24](#) present the contrast information transfer between X2 and X3, in these figures, both WM and WPM present the positive contrast which indicate the correct directional inference for $X2 \rightarrow X3$, the WTE again gives biased contrast that cannot indicate a clear direction.

3.2 Real world time series

In this section, we use two real-world data examples to demonstrate the analysis of WM and WPM. The examples include a set of EEG data measured from experiments and a financial data set observed from real market.

3.2.1 The reading experiment. The reading experiment is comprised of a reader and a listener whose EEG data are measured when the reader is reading a short story to the listener.

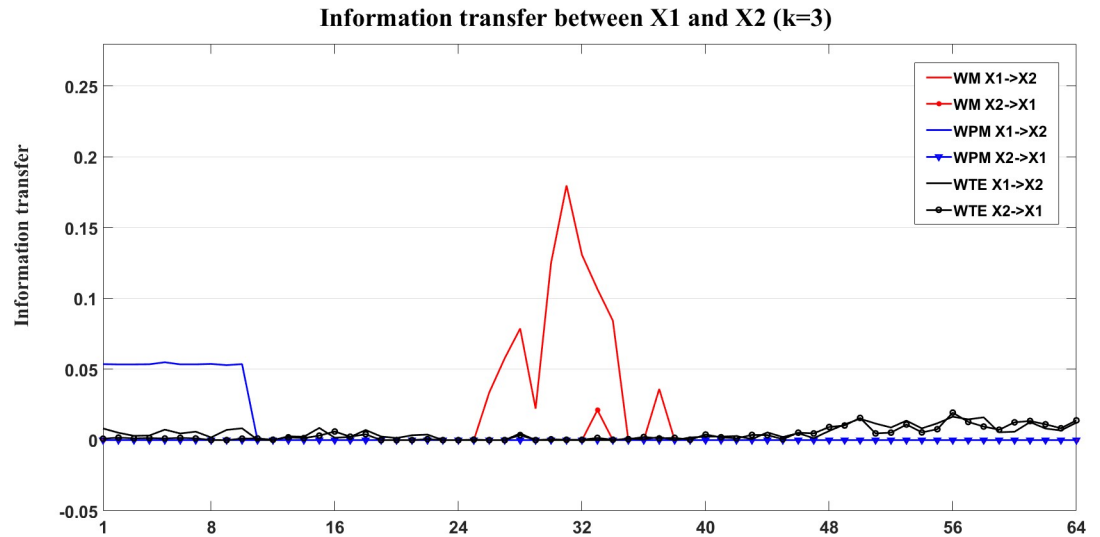


Fig 17. Information transfer between X1 and X2. In this figure, the bias corrected WM (red), WPM (blue) and WTE (black) information transfer between X1 and X2 are plotted at different time scales s_i ($i = 1, 2, \dots, 64$). By comparing the strength of information transfer between X1 and X2, WPM identifies $X1 \rightarrow X2$ at small time scales, while WM identifies $X1 \rightarrow X2$ at middle time scales. WTE gives similar and small strength on both directions that hard to identify a clear direction for the interaction.

<https://doi.org/10.1371/journal.pone.0208423.g017>

This experiment has been reported in [26, 38] for information flow test. The EEG data is made up of 10 channels for each participant, which are measured from 10 international standard electrodes [26, 38] at 100Hz frequency. The set of EEG time series are typical nonlinear and non-stationary [2–5]. The reader and the listener together form a “driver-responder” system. Here, we use WM and WPM to test the information transfer for the EEG data. The data of the reading experiment can be found in S5 Dataset.

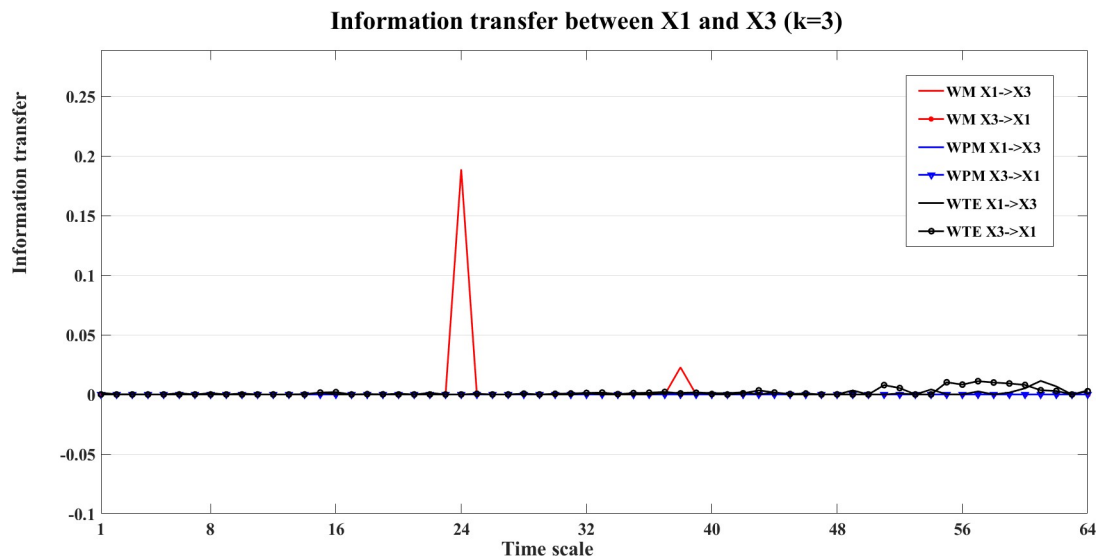


Fig 18. Information transfer between X1 and X3. In this figure, the bias corrected WM (red), WPM (blue) and WTE (black) information transfer between X1 and X3 are plotted at different time scales s_i ($i = 1, 2, \dots, 64$). By comparing the strength between X1 and X3, only WM identifies the indirect direction of interaction from $X1 \rightarrow X3$, WPM is a direct measure that presents zero information transfer between X1 and X3. WTE presents almost zero information transfer between X1 and X3 with slightly higher $X3 \rightarrow X1$ than $X1 \rightarrow X3$.

<https://doi.org/10.1371/journal.pone.0208423.g018>

Contrast WM between X1 and X2

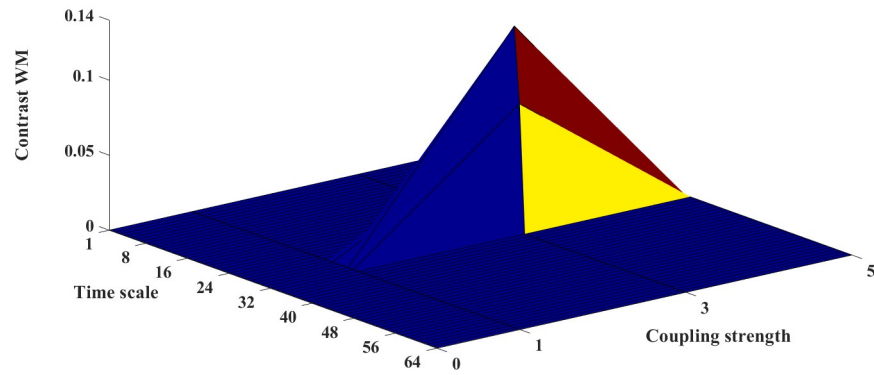


Fig 19. The contrast WM information transfer for X1→X2. The 3D surface presents the contrast WM information transfer ($\Omega_{WM,X1 \rightarrow X2}(s_i,k) = WM_{C,X1 \rightarrow X2}(s_i,k) - WM_{C,X2 \rightarrow X1}(s_i,k)$, $i = 1,2, \dots, 64, k = 0,1,3,5$) between X1 and X2 at different couplings and time scales. The non-negative surface indicates clear interactions from X1→X2.

<https://doi.org/10.1371/journal.pone.0208423.g019>

To view the instantaneous dynamics of the system, the entire data is split into equal-space time windows of 4 seconds [26, 38]. The entire data contains 30 such time windows, we use intermediate 20 consecutive time windows (from the 6th to the 25th windows) to demonstrate the analysis. WM and WPM are supposed to detect the directed interaction from the reader to the listener [26, 38].

Fig 25 shows the average values of information transfer (over windows and scales) for the 20 channel EEGs of reader and the listener. The color lattices indicate the magnitudes of the average information transfer values, whose direction is read from the row channel to the column channel. In this figure, the 20x20 color-matrix are divided into two 10x10 diagonal blocks presenting “intra-brain” interactions within the participants and two 10x10 off-diagonal blocks presenting the “cross-brain” interactions between different participants. Here, both WM and WPM indicate information transfer from the reader to the listener, because the

Contrast WPM between X1 and X2

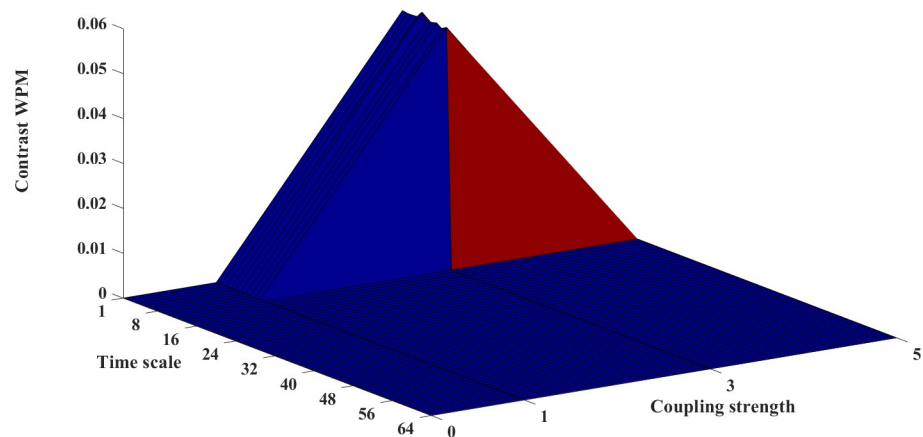


Fig 20. The contrast WPM information transfer for X1→X2. The 3D surface presents the contrast WPM information transfer ($\Omega_{WPM,X1 \rightarrow X2}(s_i,k) = WPM_{C,X1 \rightarrow X2}(s_i,k) - WPM_{C,X2 \rightarrow X1}(s_i,k)$, $i = 1,2, \dots, 64, k = 0,1,3,5$) between X1 and X2 at different couplings and time scales. The non-negative surface indicates clear interactions from X1→X2.

<https://doi.org/10.1371/journal.pone.0208423.g020>

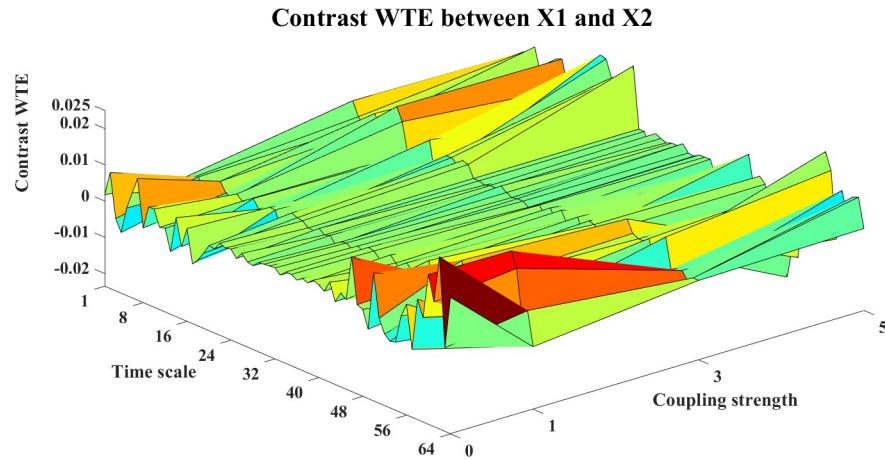


Fig 21. The contrast WTE information transfer for X1→X2. The 3D surface presents the contrast WTE information transfer ($\Omega_{WTE,X1 \rightarrow X2}(s_i, k) = WTE_{C,X1 \rightarrow X2}(s_i, k) - WTE_{C,X2 \rightarrow X1}(s_i, k)$, $i = 1, 2, \dots, 64, k = 0, 1, 3, 5$) between X1 and X2 at different couplings and time scales. The fluctuant surface around the zero plane fails to indicate a clear direction of the interaction.

<https://doi.org/10.1371/journal.pone.0208423.g021>

upper-right block (Reader->Listener) is bit brighter than the lower-right block (Listener->Reader). The color-matrix of WTE is nearly symmetric, which is hard to detect a clear direction.

To analyze the directed interaction between the reader and the listener, the contrast information transfers for Reader→Listener are plotted in Figs 26–28. We can see from Figs 26 and 27, both WM and WPM present positive ridges for the contrast information transfer, which indicate dominant information transfer from Reader→Listener. The WTE results as shown in Fig 28 present fluctuant surface around the zero plane, which is hard to indicate a clear direction.

3.2.2 Fixed incomes. The fixed incomes data are composed of 10 sovereign bond futures issued by different countries and with different maturities. The labels of these futures are listed as follows: CAN10 (Canadian, 10 years maturity), GER10 (German, 10 years maturity), GER5

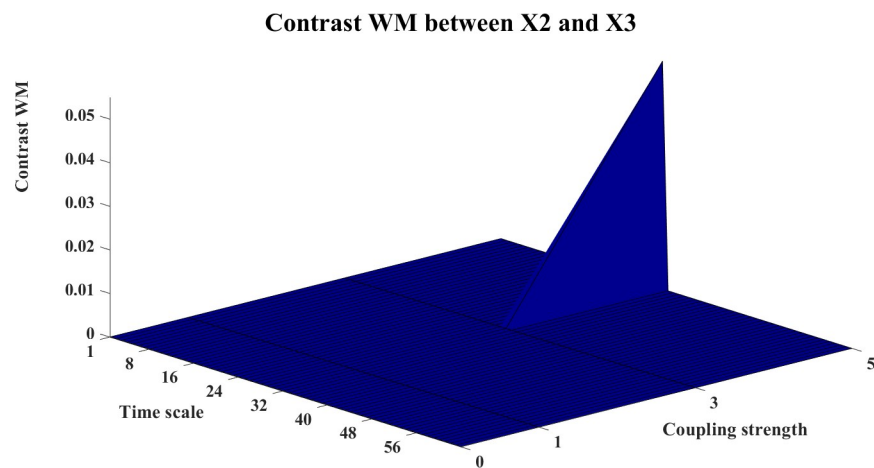


Fig 22. The contrast WM information transfer for X2→X3. The 3D surface presents the contrast WM ($\Omega_{WM,X2 \rightarrow X3}(s_i, k) = WM_{C,X2 \rightarrow X3}(s_i, k) - WM_{C,X3 \rightarrow X2}(s_i, k)$, $i = 1, 2, \dots, 64, k = 0, 1, 3, 5$) between X2 and X3 at different couplings and time scales. The positive ridge of the surface indicates the directed interaction from X2→X3.

<https://doi.org/10.1371/journal.pone.0208423.g022>

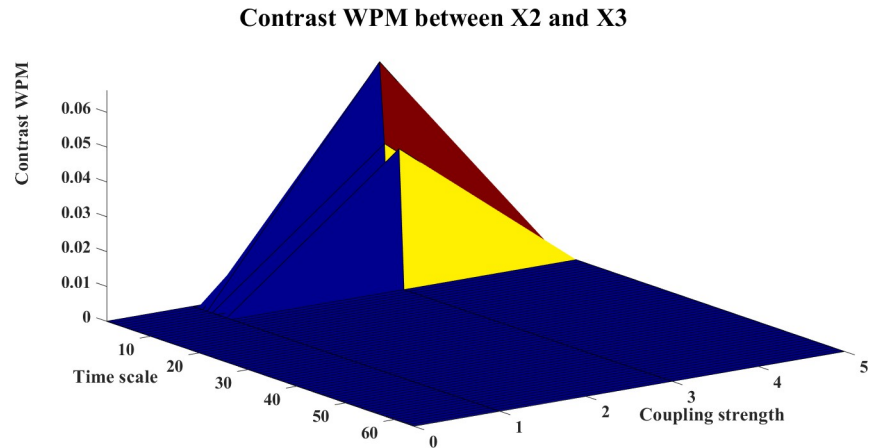


Fig 23. The contrast WPM information transfer for X2→X3. The 3D surface presents the contrast WPM ($\Omega_{WPM,X2 \rightarrow X3}(s_i,k) = WPM_{C,X2 \rightarrow X3}(s_i,k) - WPM_{C,X3 \rightarrow X2}(s_i,k), i = 1,2, \dots, 64, k = 0,1,3,5$) between X2 and X3 at different couplings and time scales. The positive surface indicates the directed information flows from X2→X3.

<https://doi.org/10.1371/journal.pone.0208423.g023>

(German, 5 years maturity), GER2 (German, 2 years maturity), US15 (US, 15 years maturity), US25 (US, 25 years maturity), US10 (US, 10 years maturity), US5 (US, 5 years maturity), US2 (US, 2 years maturity). The data is extracted from E-Signal on November 13th at 12:51 and goes back to October 18th at 13:00, on a minute-by-minute basis and in the unit of US dollars. We use a synchronized segment of the data to demonstrate the analysis. The data of the fixed incomes can be found in [S6 Dataset](#).

Financial time series are often nonlinear and non-stationary [6], we use WM and WPM to analyze the inter dynamics between different bond futures. Fig 29 shows the average information transfer (over time scales) for the Germany bond futures with different maturities. Both WM and WPM indicate the information transfer from GER5->GER10, GER5->GER2, and GER2->GER10. WTE also indicates GER5->GER10 and GER2->GER10, but with another direction from GER2->GER5.

Fig 30 shows the average information transfer values (over time scales) for the US bond futures. Both WM and WPM indicate the long-year US bond futures influence the short-year

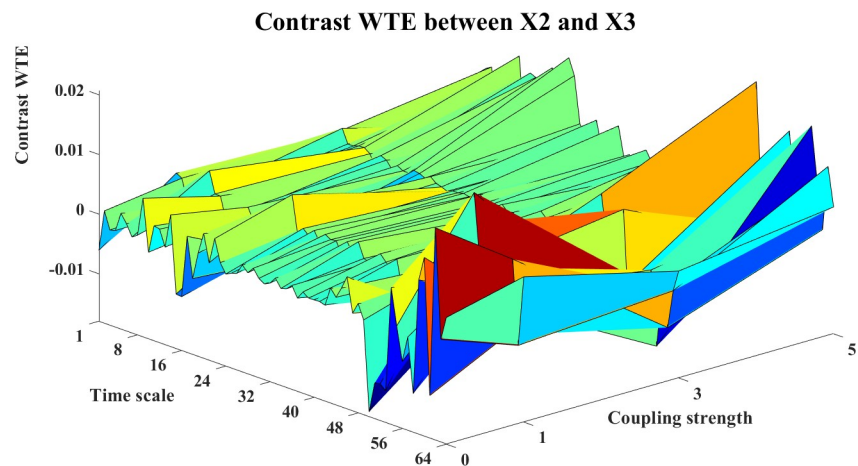


Fig 24. The contrast WTE information transfer for X2→X3. The 3D surface presents the contrast WTE ($\Omega_{WTE,X2 \rightarrow X3}(s_i,k) = WTE_{C,X2 \rightarrow X3}(s_i,k) - WTE_{C,X3 \rightarrow X2}(s_i,k), i = 1,2, \dots, 64, k = 0,1,3,5$) for X2→X3 at different couplings and time scales. The fluctuant sign of the surface indicates no clear direction of interactions.

<https://doi.org/10.1371/journal.pone.0208423.g024>

Average information transfer between the reader and the listener

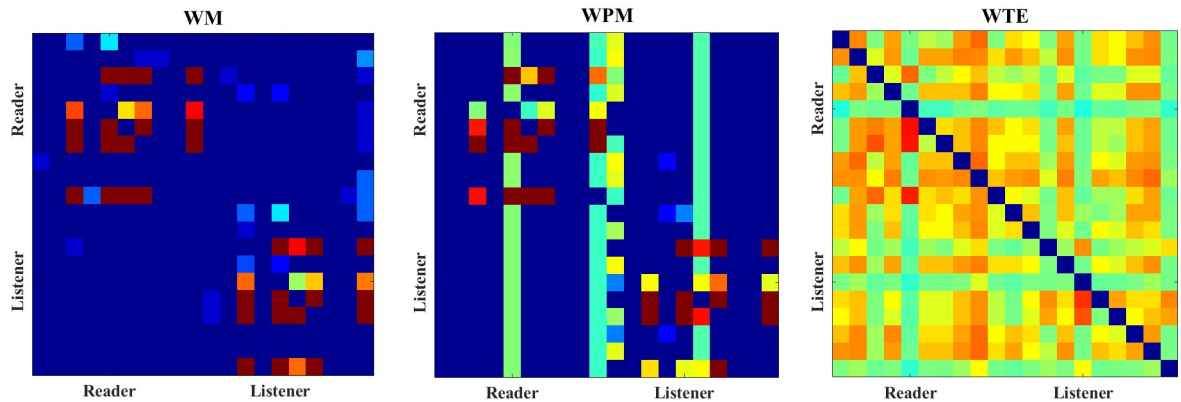


Fig 25. The color-map for the average information transfer between the reader and the listener. The color-graphs separately show the average WM (left), WPM (middle) and WTE (right) values for the 20 channel EEGs. For each lattice the directional inference is read from the row channel to the column channel. In the color-matrices, the diagonal blocks present the “intra-brain” information transfers within each participant, while the off-diagonal blocks present the “cross-brain” interactions across different participants.

<https://doi.org/10.1371/journal.pone.0208423.g025>

US bond futures such as US10 and US2. WTE identifies strong influence from US5 to US25 and US15, but it fails to inference the other directions.

Fig 31 plots the average information transfer values (over time scales) between all 10-year bond futures issued by different countries. All three measures indicate US10→CAN10, WM also indicates GER10->US10, while WPM and WTE indicate CAN10->GER10 and US10→GER10.

To study the dynamics at different time scales, we also compute the contrast information transfer for the three measures, an example of the Germany bond futures is shown in Figs 32–34.

In Fig 32, WM indicates clear interactions from GER5->GER2 (black) and GER5->GER10 (red), and fluctuant interaction between GER10 and GER2 (blue). WPM (Fig 33) also indicates clear interaction from GER5->GER2 (black) and GER5->GER10 (red), and also GER2->GER10 (blue). WTE (Fig 34) indicates GER5->GER10 (red), GER2->GER10 (blue) and GER2->GER5 (black).

Contrast WM between the reader and the listener

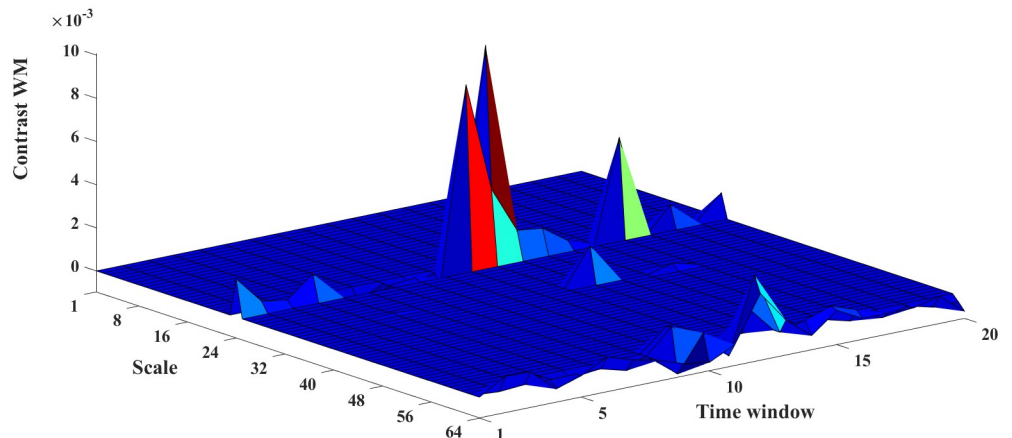


Fig 26. The contrast WM for Reader->Listener. The 3D surface presents the contrast WM for Reader→Listener at different scales and time windows ($\Omega_{WM,R \rightarrow L}(s_i, w) = WM_{C,R \rightarrow L}(s_i, w) - WM_{C,R \rightarrow L}(s_i, w)$, $i = 1, 2, \dots, 64, w = 1, 2, \dots, 20$).

<https://doi.org/10.1371/journal.pone.0208423.g026>

Contrast WPM between the reader and the listener

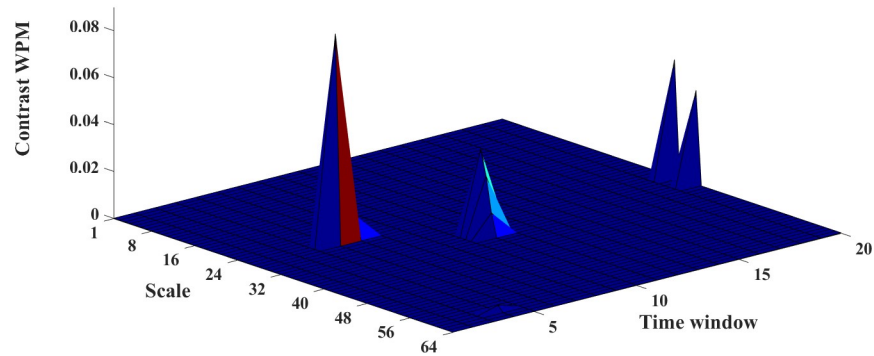


Fig 27. The contrast WPM for Reader->Listener. The 3D surface presents the contrast WPM for Reader→Listener at different scales and time windows ($\Omega_{WPM,R-L}(s_i,w) = WPM_{C,R-L}(s_i,w) - WPM_{C,R-L}(s_i)$, $i = 1,2, \dots, 64$, $w = 1,2, \dots, 20$).

<https://doi.org/10.1371/journal.pone.0208423.g027>

We also examine the information transfer for each independent direction. An example of the information transfer between Germany bond futures is shown in Fig 35. In this figure, the three graphs separately show the directed information transfer between the three Germany bond futures. In these graphs, we can see that both WM and WPM give higher information transfer values for GER5->GER10, WTE presents some of the positiveness for GER5->GER10, but the strength of WTE is quite subtle that almost vanish. For the interaction between GER10 and GER2, both WM and WPM indicate dominant information flows from GER2->GER10 at middle time scales, again WTE presents subtle information transfers that approximately vanish. In the third graph, both WM and WPM show dominant information flow from GER5->GER2, while WTE fails to identify the interactions.

The overall results of the directed interaction for the bond futures agree with the outcomes found by MIME [55]. The difference is that the WM and WPM can interpret the directed interaction at multi-scales. Since the scales are related to the frequencies, one can use these multi-scale measures to specify the interactions or time series correlation at specific scales or

Contrast WTE between the reader and the listener

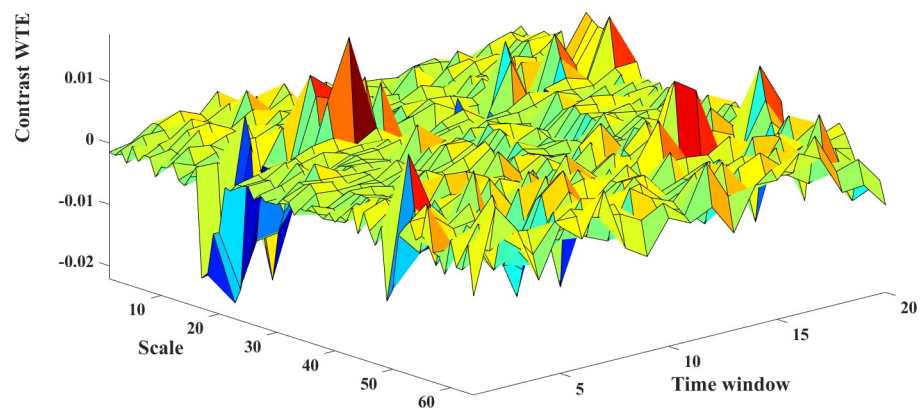


Fig 28. The contrast WTE for Reader->Listener. The 3D surface presents the contrast WTE for Reader→Listener at different scales and time windows ($\Omega_{WTE,R-L}(s_i,w) = WTE_{C,R-L}(s_i,w) - WTE_{C,R-L}(s_i)$, $i = 1,2, \dots, 64$, $w = 1,2, \dots, 20$).

<https://doi.org/10.1371/journal.pone.0208423.g028>

Average information transfer for the Germany bond futures

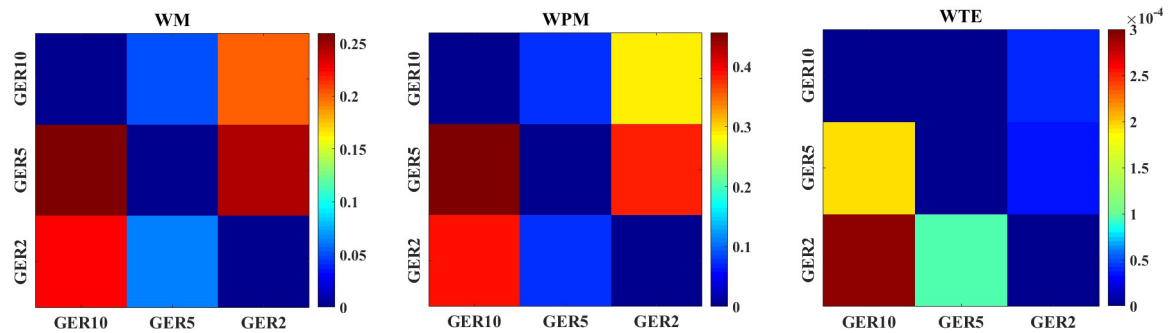


Fig 29. The color-map for the average information transfer between the Germany bond futures. The three color-graphs present the color-map of the average WM (left), WPM (middle) and WTE (right) information transfer between the Germany bond futures. The direction of information flow is read from the row channel to the column channel for each lattice.

<https://doi.org/10.1371/journal.pone.0208423.g029>

frequencies, the multi-scale interpretation of information transfer may help one to find more information from given datasets.

Discussion

In this paper, we propose two multi-scale information transfer measures, namely the WM and WPM, which are extended from conditional mutual information measures by using MORLET wavelet. By a series of simulation studies, the two measures are proved to be efficient and accurate in directional inference, which are computational efficient and outperform the wavelet extension of transfer entropy (WTE) in various situations. Particularly, the two measures are very useful in real-world data analysis.

The two proposed measures have many good advantages. By using wavelet, the proposed measures are able to inference the directed interactions at multi-scales. This not only helps to discover more information between coupled time series, but also solves the problem for non-stationary and discontinuous data analysis [24, 39–43]. One reason that we use MORLET wavelet to do the measure extensions is that MORLET wavelet is believed to be closely related to human perception and has vital applications to medicine [39–43], which may have wide-

Average information transfer for the US bond futures

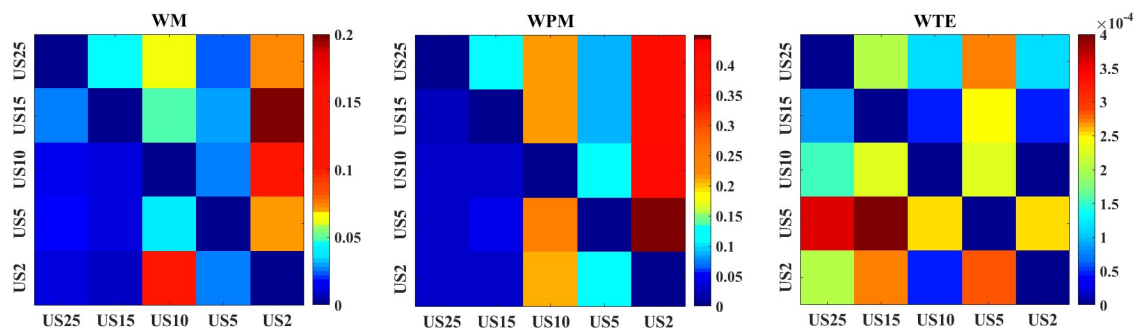


Fig 30. The color-map for the average information transfer values of the US bond futures. The three color-graphs present the color-map for the average WM (left), WPM (middle) and WTE (right) information transfer for the US bond futures. The direction of interaction is read from the row channel to the column channel for each lattice.

<https://doi.org/10.1371/journal.pone.0208423.g030>

Average information transfer for the 10 year bond futures

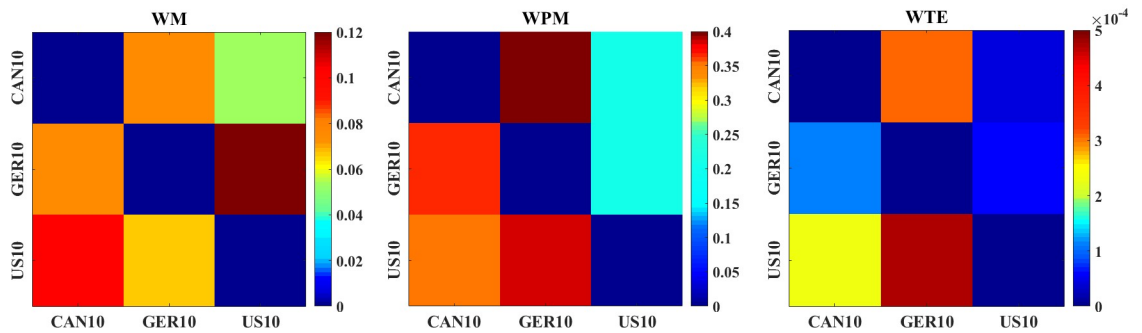


Fig 31. The color-map for the average information transfer values of the 10-year bond futures. The three color-graphs present the color-map for the average WM (left), WPM (middle) and WTE (right) values for the 10-year bond futures. The direction of interaction is read from the row channel to the column channel for each lattice.

<https://doi.org/10.1371/journal.pone.0208423.g031>

applications in medicine and biological data analysis e.g. EEG data analysis. The other reason is that M. Lungarella and A. Pitti have successfully introduced MORLET wavelet to Transfer Entropy, which is proved to be feasible in information transfer detection [24] of non-stationary and discontinuous data analysis. Since real-world time series are sometimes non-stationary and discontinuous, it is necessary to have this wavelet-extension that suits for practical data analysis.

Due to the basis of conditional mutual information measures, the proposed multi-scale measures can have wide-applications to various data models. Information transfer measures are a type of very useful measures. Although we note that information transfer is a different concept from that of the causal effect [31–34], but the efficiency of directional inference and the model-free advantage of the information transfer measures guarantee their wide-applications in various types of data analysis particularly the real-world data analysis [26].

Contrast WM between the Germany bond futures

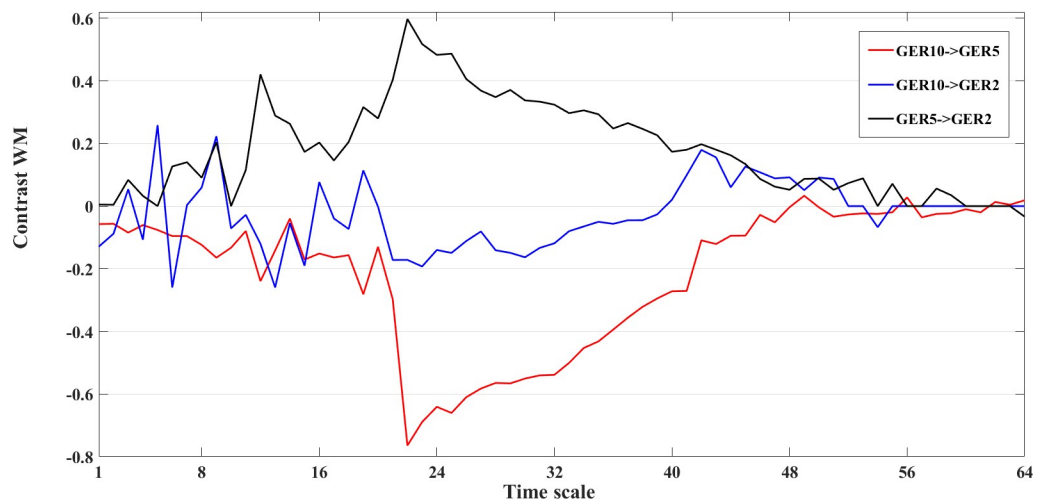


Fig 32. The contrast WM information transfer values for the Germany bond futures. The line graphs show the contrast WM values between the Germany bond futures at different time scales s_i ($i = 1, 2, \dots, 64$).

<https://doi.org/10.1371/journal.pone.0208423.g032>

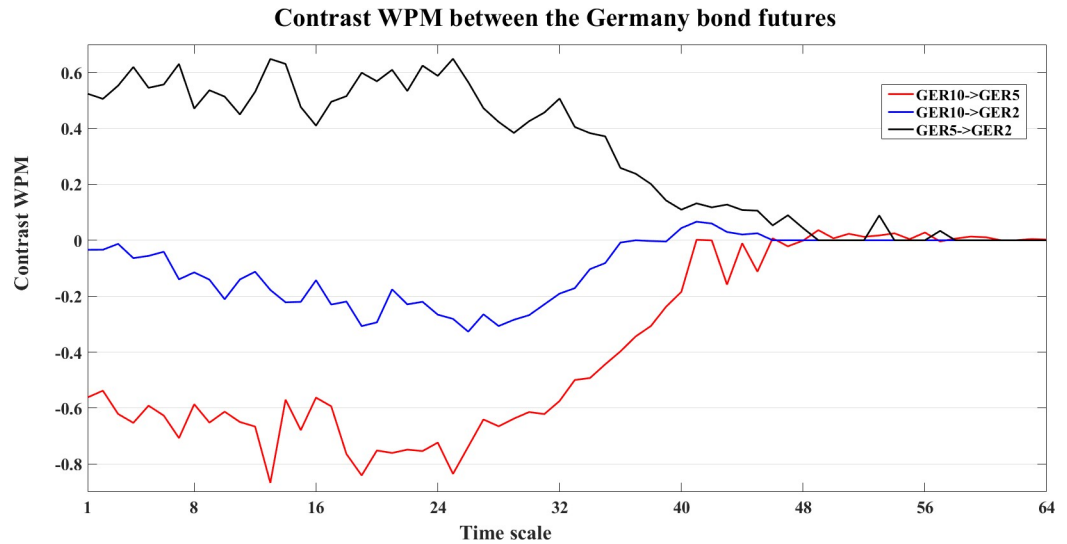


Fig 33. The contrast WPM information transfer values for the Germany bond futures. The line graphs show the contrast WPM values between the Germany bond futures at different time scales s_i ($i = 1, 2, \dots, 64$).

<https://doi.org/10.1371/journal.pone.0208423.g033>

Transfer entropy (TE) is a fundamental information transfer measure, which is a good prototype for other derivative information measures. The work by M. Lungarella and A. Pitti [24] shows the wavelet derivation of TE for non-stationary and discontinuous data analysis. However, WTE uses uniform state-space embedding vector [25–27] which is computational redundant and costs long time of computation. This drawback not only affects the accuracy in information flow detection but also limits WTE from large dataset applications.

To avoid the computational redundancy and improve the speed and accuracy, we do the extension of wavelet on two conditional mutual information measures of mixed embeddings [22, 29]. The two prototype measures are the MIME and PMIME, while the latter is the direct

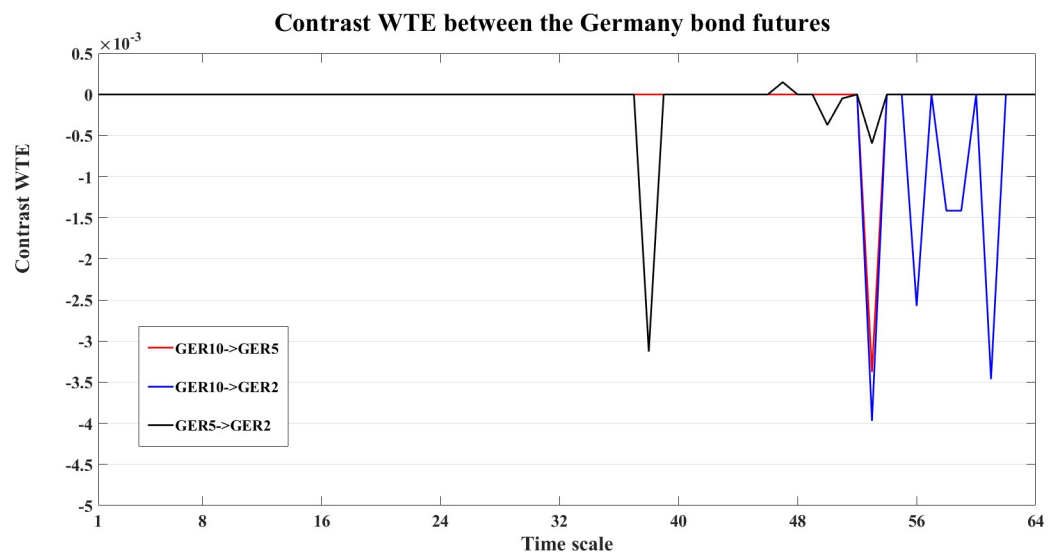


Fig 34. The contrast WTE information transfer values for the Germany bond futures. The line graphs show the contrast WTE values between the Germany bond futures at different time scale s_i ($i = 1, 2, \dots, 64$).

<https://doi.org/10.1371/journal.pone.0208423.g034>

Information transfer between Germany bond futures

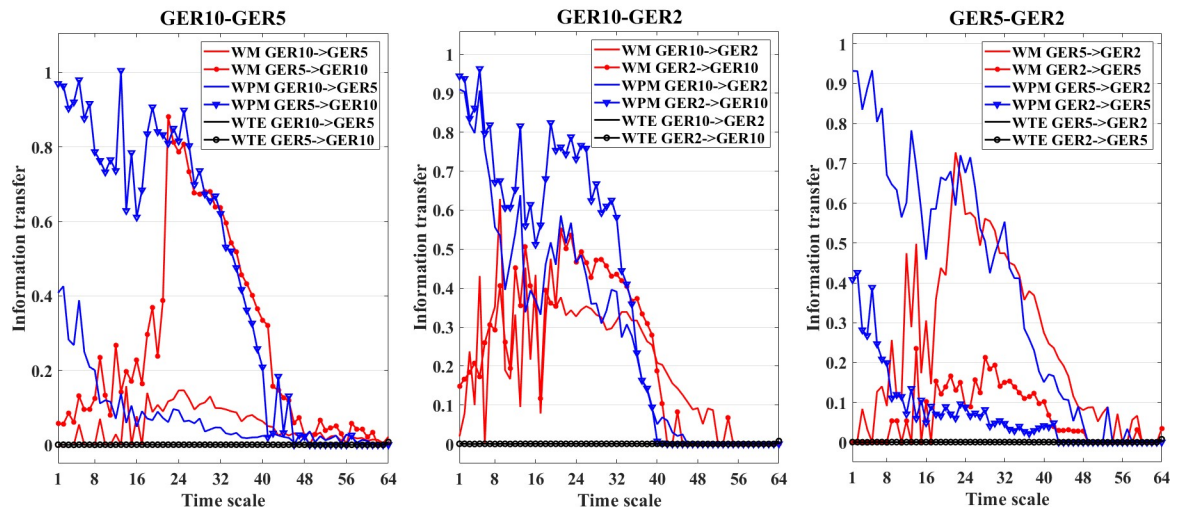


Fig 35. Information transfer between the Germany bond futures. In this figure, the three graphs show the information transfer between the three Germany bond futures (GER10, GER5, GER2). In each graph, the WM (red), WPM (blue) and WTE (black) information transfer values are plotted against different time scales $s_i, i = 1, 2, \dots, 64$.

<https://doi.org/10.1371/journal.pone.0208423.g035>

version of the former. These two measures are proved to be accurate and efficient in various data analysis [22, 26, 29, 35–38]. They use a progressive scheme and a stopping criterion to select only useful embedding components to be included and also to prevent false causalities [22, 29]. This stopping criterion is used as a threshold that balances the inclusion of useful lagged elements and the exclusion of useless lagged elements [22,26,29,38,46]. By using the selected non-uniform state-space embedding vectors, MIME and PMIME exclude the redundant components in information transfer evaluation, which not only guarantees the accuracy but also removes the computational redundancy.

In the simulation study, the parameters used are referenced from early studies [22, 29, 24, 38, 50]. Other parameters are plausible but depend on specific type of the applications. In this paper, we particularly studied the influence of the stopping criterion on the information transfer detection. The stopping criterion A is the proportionality of the conditional mutual information rate between the current and the past iterative cycles. We see that appropriate choices of this criterion help to ensure the true interactions to be identified, and false positiveness are prevented. Via simulation studies, we found that $A = 0.97$ and $A = 0.95$ are good choices for the stopping criterion. The stopping criterion A should not be too large or too small, because a too large A for instance $A > 0.97$ will cause false interactions to be detected, while a too small A for instance $A < 0.95$ will be too rigid for the criterion that it often fails to identify the interactions that truly exist. However, the choice of the stopping criterion may also depend on the datasets to be analyzed.

By definition of the WM and WPM methods, we should note that both WM and WPM evaluate the information transfer between wavelet coefficients of the same scales rather than different scales. This may be a limit of these two measures. However, the detection of information transfers across scales will be our next stage study.

In the synthetic data analysis, we note that the information transfer declines to zero for large time scales. Many reasons can explain for this phenomenon. One reason is that, when time scale grows, the frequency and resolution decrease, hence details of the time series are smeared out, the directed interaction becomes too weak to be detected. The other reason is the

characteristic correlation time of theoretic maps [51–52, 54]. When the time scale exceeds the characteristic correlation time, wavelet coefficients of the coupled time series become fully correlated and the system becomes deterministic, information transfer measures such as MIME and PMIME are vanished for deterministic systems [22, 29, 50]. Another reason is that, the time scale threshold may correspond to the frequency where two time series have large mutual information between each other, the wavelet coefficients are correlated at this frequency and the system becomes deterministic. This concept of correlation is one of the key features of wavelet [24, 39–43] that makes WM and WPM special in this case. We are interested to make the bold hypothesis that the WM and WPM may be able to inference the cross-correlation time between time series and may also be able to identify the common frequencies for signal correlations.

Conclusion

In this paper, we have proposed two multi-scale information transfer measures, namely the WM and WPM, which are the MORLET wavelet extension of conditional mutual information from mixed embedding measures. Both measures are model-free and accurate in information transfer detection of various datasets. By using non-uniform state-space embeddings, both WM and WPM are computational efficient which outperform WTE in both accuracy and speed. Due to the nature of wavelet, the proposed measures may have wide-applications including also non-stationary and discontinuous data analysis.

Supporting information

S1 Dataset. Data for the cosine map.

(MAT)

S2 Dataset. Data for the Henon maps.

(MAT)

S3 Dataset. Data for the system of three coupled variables.

(MAT)

S4 Dataset. Data for the Lorenz systems.

(MAT)

S5 Dataset. Data for the reading experiment.

(MAT)

S6 Dataset. Data for the fixed incomes.

(MAT)

Acknowledgments

The authors acknowledge Prof. D. Kugiumitzis (Aristotle University of Thessaloniki, Greece) for his precious advice on MIME and PMIME and their software packages. Also thanks for Prof. H. Jensen (Imperial College London, UK) for advice and support on this research.

Author Contributions

Conceptualization: Xiaogeng Wan.

Data curation: Xiaogeng Wan.

Formal analysis: Xiaogeng Wan.

Funding acquisition: Xiaogeng Wan.

Investigation: Xiaogeng Wan.

Methodology: Xiaogeng Wan.

Project administration: Xiaogeng Wan.

Resources: Xiaogeng Wan, Lanxi Xu.

Software: Xiaogeng Wan.

Validation: Lanxi Xu.

Writing – original draft: Xiaogeng Wan.

Writing – review & editing: Xiaogeng Wan.

References

1. Vejmelka M and Palus M. Inferring the directionality of coupling with conditional mutual information. *Physical Review E*. 2008; 77(2): 026214.
2. Breakspear M and Stam CJ. Dynamics of a neural system with a multiscale architecture. *Philosophical Transactions of the Royal Society B*. 2005; 360 (1457): 1051–1074.
3. Kugiumtzis D and Larsson PG. Linear and nonlinear analysis of EEG for the prediction of epileptic seizures. *Proceedings of the 1999 Workshop “Chaos in Brain?”*, World Scientific, Singapore. 2000; 329–333.
4. Palus M, Dvorak I, David I. Spatio-temporal dynamics of human EEG. *Physica A*. 1992; 185: 433–438.
5. Pereda E, Quiroga RQ and Bhattacharya J. Nonlinear multivariate analysis of neurophysiological signals. *Progress in Neurobiology*. 2005; 77(1): 1–37.
6. Marschinski R and Kantz H. Analysing the information flow between financial time series—an improved estimator for transfer entropy. *European Physical Journal B*. 2002; 30: 275–281.
7. Granger CWJ. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*. 1969; 37 (3): 424–438.
8. Seth A. Granger causality. *Scholarpedia*. 2007; 2 (7): 1667.
9. Ancona N, Marinazzo D and Stramaglia S. Radial basis function approach to nonlinear Granger causality of time series. *Physical Review E*. 2004; 70(2): 056221.
10. Chen Y, Bressler SL and Ding M. Frequency decomposition of conditional Granger causality and application to multivariate neural field potential data. *Journal of Neuroscience Methods*. 2006; 150(2): 228–237. <https://doi.org/10.1016/j.jneumeth.2005.06.011> PMID: 16099512
11. Chicharro D. On the spectral formulation of Granger causality. *Biological Cybernetics*. 2011; 105 (5–6): 331–347. <https://doi.org/10.1007/s00422-011-0469-z> PMID: 22249416
12. Faes L, Nollo G and Porta A. Information-based detection of nonlinear Granger causality in multivariate processes via a nonuniform embedding technique. *Physical Review E*. 2011; 83(1): 051112.
13. Baccala LA and Sameshima K. K. Partial directed coherence: a new concept in neural structure determination. *Biological Cybernetics*. 2001; 84(6): 463–474. <https://doi.org/10.1007/PL00007990> PMID: 11417058
14. Biazoli CE, Sturzbecher M, White T, Onias HH, Andrade KC, Araujo DB, et al. Application of the partial directed coherence to the analysis of resting-state EEG-fMRI data. *Brain Connectivity*. 2013; 3 (6), 563–568. <https://doi.org/10.1089/brain.2012.0135> PMID: 23724827
15. Sameshima K and Baccala L. A. LA. Using partial directed coherence to describe neuronal ensemble interactions. *Journal of Neuroscience Methods*. 1999; 94(1): 93–103. PMID: 10638818
16. Schelter B, Timmer J and Eichler M. Assessing the strength of directed influences among neural signals using renormalized partial directed coherence. *Journal of Neuroscience Methods*. 2009; 179: 121–130. <https://doi.org/10.1016/j.jneumeth.2009.01.006> PMID: 19428518
17. Schelter B, Winterhalder M, Eichler M, Peifer M, Hellwig B, Guschlbauer B, et al. Testing for directed influences among neural signals using partial directed coherence. *Journal of Neuroscience Methods*. 2005; 152: 210–219. <https://doi.org/10.1016/j.jneumeth.2005.09.001> PMID: 16269188
18. Takahashi DY, Baccala LA and Sameshima K. Frequency domain connectivity: an information theoretic perspective. *International Conference of the IEEE Engineering in Medicine and Biology*. 2010; 1726–1729.

19. Takahashi DY, Baccala LA and Sameshima K. Information theoretical interpretation of frequency domain connectivity measures. *Biological Cybernetics*. 2010; 103(6): 463–469. <https://doi.org/10.1007/s00422-010-0410-x> PMID: 21153835
20. Cover T and Thomas J. *Elements of Information Theory*. John Wiley & Sons. 2003.
21. Gelfand I and Yaglom A. Calculation of amount of information about a random function contained in another such function. *Uspehi Mat. Nauk*. 1959; 2(1): 3–52.
22. Kugiumtzis D. Direct coupling information measure from non-uniform embedding. *Physical Review E*. 2013; 87(6): 062918.
23. Kugiumtzis D. Transfer entropy on rank vectors. *Journal of Nonlinear Systems and Applications*. 2012; 3: 73–81.
24. Lungarella M and Pitti A. Information transfer at multiple scales. *Physical Review E*. 2007; 76(2): 056117.
25. Razak FA and Jensen HJ. Quantifying ‘causality’ in complex systems: understanding transfer entropy. *PLoS ONE*. 2014; 9(6): e99462. <https://doi.org/10.1371/journal.pone.0099462> PMID: 24955766
26. Razak FA, Wan X and Jensen HJ. Information theoretic measures of causality: Music performance as a case study. *Edward Elgar Handbook on Complexity Science Methods*. 2018.
27. Schreiber T. Measuring information transfer. *Physical Review Letters*. 200; 85 (2): 461–464. <https://doi.org/10.1103/PhysRevLett.85.461> PMID: 10991308
28. Staniek M and Lehnertz K. Symbolic transfer entropy. *Physical Review Letters*. 2008; 100(15): 158101. <https://doi.org/10.1103/PhysRevLett.100.158101> PMID: 18518155
29. Vlachos I and Kugiumtzis D. Nonuniform state-space reconstruction and coupling detection. *Physical Review E Statistical Nonlinear & Soft Matter Physics*. 2010; 82(1 Pt 2): 016207.
30. Zhou Y. *The basics of information theory*, 3rd Edition. Beijing University of Aeronautics and Astronautics Press. 2006.
31. Ay N and Polani D. Information flows in causal network. *Advances in Complex Systems*. 2008; 11 (01):17–41.
32. Pearl J. *Causality: Models, Reasoning and Inference*. Cambridge University Press. 2000.
33. Lizier JT and Prokopenko M. Differentiating information transfer and causal effect. *The European Physical Journal B*. 2010; 73(4):605–615.
34. Wibral M, Pampu N, Priesemann V, Siebenhühner F, Seiwert H, and Lindner M, et. al. Measuring Information-Transfer Delays. *PLoS ONE*. 2013; 8(2): e55809. <https://doi.org/10.1371/journal.pone.0055809> PMID: 23468850
35. Papan A, Kugiumtzis D, Larsson PD. Detection of direct causal effects and application to electroencephalogram analysis. *International Journal of Bifurcation and Chaos*. 2012; 22(9): 046203.
36. Koutlis C and Kugiumtzis D. Discrimination of coupling structures using causality networks from multivariate time series. *Chaos*. 2016; 26(9): 093–120.
37. Kugiumtzis D and Kimiskidis V. Direct Causal Networks for the Study of Transcranial Magnetic Stimulation Effects on Focal Epileptiform Discharges. *International Journal of Neural Systems*. 2015; 25 (5): 166–244.
38. Wan X, Cruts B and Jensen HJ. The causal inference of cortical neural networks during music improvisations. *PLoS ONE*. 2014; 9(12): e112776. <https://doi.org/10.1371/journal.pone.0112776> PMID: 25489852
39. Angelini L, Maestri R, Marinazzo D, Nitti L, Pellicoro M, Pinna GD, et al. Multiscale analysis of short-term heart beat interval, arterial blood pressure, and instantaneous lung volume time series. *Artificial Intelligence in Medicine*. 2007; 41(3): 237–250. <https://doi.org/10.1016/j.artmed.2007.07.012> PMID: 17950584
40. Ding J. A tutorial of the Morlet wavelet transform. Term paper, Institution of telecommunications. 2014; 1–21.
41. Liner C. An overview of wavelet transform concepts and applications. University of Houston. 2010; 1–17.
42. Najmi I and Sadowsky J. The continuous wavelet transform and variable resolution time-frequency analysis. *Johns Hopkins APL Technical Digest*. 1997; 18(1): 134–140.
43. Popinski W, Kosek W, Schuh H and Schmidt M. Comparison of two wavelet transform coherence and cross-covariance functions applied on polar motion and atmospheric excitation. *Studia Geophysica et Geodaetica*. 2002; 46(3): 455–468.
44. Lancaster G, Iatsenko D, Pidde A, Ticcinelli V and Stefanovska A. Surrogate data for hypothesis testing of physical systems. *Physics Reports*. 2018; 748: 1–60.

45. Prichard D and Theiler J. Generating surrogate data for time series with several simultaneously measured variables. *Physical Review Letters*. 1994; 73(7): 951. <https://doi.org/10.1103/PhysRevLett.73.951> PMID: 10057582
46. Schreiber T and Schmits A. Surrogate time series. *Physica D*. 2000; 142(3): 346–382.
47. Small M and Tse CK. Detecting determinism in Time series: the method of surrogate data. *IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications*. 2003; 50 (5): 663–672.
48. Theiler J, Eubank S, Longtin A, Galdrikian B and Farmer JD. Testing for nonlinearity in time series: the method of surrogate data. *Physica D*. 1992; 58(92): 77–94.
49. Yu SH, Wang DH, Li K and Zhao ZW. Estimation in autoregressive models with surrogate data and validation data. *Communications in Statistics-Theory and Methods*. 2017; 46(3): 1532–1545.
50. Papan A, Kyrtsov C, Kugiumtzis D and Diks C. Simulation study of direct causality measures in multivariate time series. *Entropy*. 2013; 15(7): 2635–2661.
51. Cvitanovic P, Gunaratne G and Procaccia I. Topological and metric properties of Henon-type strange attractors. *Physical Review A*. 1988; 38(3): 1503–1520.
52. Henon M. A two-dimensional mapping with a strange attractor. *Communications in Mathematical Physics*. 1976; 50(1): 69–77.
53. Lorenz EN. The Mechanics of Vacillation. *Journal of the atmospheric sciences*. 1963; 20(5): 448–465.
54. Schelter B, Winterhalder M and Timmer J. *Handbook of time series analysis: recent theoretical developments and applications*. WILEY-VCH. 2015.
55. Wan X. PhD Thesis: Time series causality analysis and EEG data analysis on music improvisation. Imperial College London. 2015.