




Causality Analysis with Information Geometry: A Comparison

Heng Jie Choong ^{1,*}, Eun-jin Kim ¹ and Fei He ²

¹ Centre for Fluid and Complex Systems, Coventry University, Coventry CV1 5FB, UK; ejk92122@gmail.com
² Centre for Computational Science and Mathematical Modelling, Coventry University, Coventry CV1 5FB, UK; fei.he@coventry.ac.uk
* Correspondence: choongh2@uni.coventry.ac.uk

Abstract: The quantification of causality is vital for understanding various important phenomena in nature and laboratories, such as brain networks, environmental dynamics, and pathologies. The two most widely used methods for measuring causality are Granger Causality (GC) and Transfer Entropy (TE), which rely on measuring the improvement in the prediction of one process based on the knowledge of another process at an earlier time. However, they have their own limitations, e.g., in applications to nonlinear, non-stationary data, or non-parametric models. In this study, we propose an alternative approach to quantify causality through information geometry that overcomes such limitations. Specifically, based on the information rate that measures the rate of change of the time-dependent distribution, we develop a model-free approach called information rate causality that captures the occurrence of the causality based on the change in the distribution of one process caused by another. This measurement is suitable for analyzing numerically generated non-stationary, nonlinear data. The latter are generated by simulating different types of discrete autoregressive models which contain linear and nonlinear interactions in unidirectional and bidirectional time-series signals. Our results show that information rate causality can capture the coupling of both linear and nonlinear data better than GC and TE in the several examples explored in the paper.

Keywords: causality; information geometry; Transfer Entropy; Granger Causality; information causal rate; signal processing; nonlinear models; non-stationary; probability distribution



Citation: Choong, H.J.; Kim, E.-j.; He, F. Causality Analysis with Information Geometry: A Comparison. *Entropy* **2023**, *25*, 806. <https://doi.org/10.3390/e25050806>

Academic Editor: Hector Zenil

Received: 6 March 2023

Revised: 12 May 2023

Accepted: 13 May 2023

Published: 16 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The study of causality aims to explore cause-and-effect relationships between processes. In general, causality can be investigated in two categories: causal discovery and causal analysis. Causal discovery shows the inherent causal relationship within the dataset by analyzing and creating causal models based on graph theory. The construction of these models can be achieved via algorithmic dynamics and evaluating the algorithms against the observed data [1–3]. Generally, the complexity of the models' networks is characterized and measured using graph-theoretic measures. A recent approach using classical information theory, such as Shannon entropy, has been considered as an alternative measurement, but its high dependence on the distribution can lead to spurious disparate values for the same complexity, as shown in [4].

Alternatively, causal analysis, which is the focus of this paper, studies the potential changes in a system caused by another. In practical terms, causal discovery requires more detailed background information about the provided dataset to accurately conduct the study. However, the causal analysis does not require detailed underlying knowledge of the connectivity embedded within the dataset [3,5]. Among the various analysis methods used to quantify causality, Granger causality (GC) and Transfer Entropy (TE) are extensively used in various fields, including economics, climate, education, meteorology, neuroscience, and more [5–16]. For instance, Rebecca M. Pullon et al. utilized GC to demonstrate the decrease in cortical information flow throughout the brain as the subject loses responsiveness [10]; Yunyuan Gao et al. used TE to investigate the coupling strength of the brain

between the motor and sensory areas [15]. Both of these methods are derived from the same notion of examining the reliability of a present process towards the past process(es). Despite sharing the same notion, their approaches to quantifying causality are completely different. GC evaluates the dependency based on signal models, which means modelling the signals incorrectly can lead to the wrong conclusion about the causality [11,17,18]. On the other hand, TE is model-free and employs the statistical dependency approach to access causality. However, it is subject to the estimation of the probability distribution, such as the number of bin size and the high dimensionality in the distribution due to the number of lags [19,20].

More specifically, inspired by Norbert Wiener, Clive Granger introduced GC [6,21], which suggested that the causality of two stochastic processes ($x_1(t)$ and $x_2(t)$) exists if and only if one of the processes is able to predict another. For instance, if including the past observation of $x_1(t)$ along with $x_2(t)$ improves the prediction of the current state of $x_2(t)$ compared to using only past information of $x_2(t)$, then this would suggest that $x_1(t)$ is Granger-causing $x_2(t)$. This predictivity is quantitatively studied by comparing the errors of the autoregressive model (linear regressive model) that is used in modelling the signals [6,18]. As natural time series such as electroencephalogram (EEG) often exhibit oscillatory aspects, it is interesting to study the spectrum of the GC. John Geweke has discussed the evaluation of GC in the frequency domain instead of the time domain [17,22,23]. Additionally, the time-varying frequency domain of the GC that is studied in [18,24,25] is able to show both the spectrum of the GC and the period of causation occurrence. The frequency domain of GC is the spectral of GC which shows the amplitude of GC correlated to a specific frequency of the entire signals, while the time-varying frequency of GC shows when the coupling of the signals occurs alongside the spectral of GC.

When utilizing the GC, it is necessary to model the provided signals using a linear autoregressive model [17,18]. However, this linear model cannot accommodate possible nonlinear and non-stationary effects within real-world signals, such as the brain EEG signals [26,27]. Therefore, instead of studying the causality through deterministic or parametric linear models, causality can be statistically quantified using information-theoretic measures like TE [19]. In the event of two stochastic processes, TE measures the additional information required from one process to achieve the current realization of another process [28]. Therefore, TE is not limited to any specific model or linearity assumption. Although TE uses a different approach in evaluating the causality, it can yield similar results to GC for a Gaussian process, as discussed analytically by Lionel Barnett et al. [29].

Alternatively, the causality can be investigated by examining the changes in the probability distribution that are caused by the inclusion of the other variables [30]. In a statistical space, the change in distribution is known as distance and can be quantified via Kullback–Leibler (KL) divergence. However, KL divergence cannot be well-defined as a distance, as it is not symmetrical and not path-dependent. Hence, information rate and information length are used with the intention to have the symmetrical and path-dependent measurement of the distance [30–33]. Information rate quantifies the rate of the temporal change in distributions, while information length measures the total statistical change in the process along the time [30,32]. Refs. [30,31] proposed an information-geometric causality measure, the so-called *information rate causality* based on the information rate, by quantifying causality through the effects of the information rate (see also [34]) and applied it to a solvable, linear Gaussian process (Kramer model) which has an exact time-dependent distribution.

In this paper, we will further develop this information rate causality for analyzing the numerically generated time-series signals that are in general time-varying and nonlinear and compare it with the GC and TE. Specifically, we will generate the signals by simulating the (discrete) autoregressive equations which contain unidirectional or bidirectional interactions. For the GC, both the frequency domain and time varying of GC will be presented. For the TE, the signals are evaluated as a whole at different lags. Next, the net TE is calculated by taking the difference in the TE from both directions. For instance, $T_{1,2}^{net} = T_{1 \rightarrow 2} - T_{2 \rightarrow 1}$

represents the net TE of signals 1 and 2. The net TE enables us to identify the appropriate lag(s) that yield significant results in TE. Using the knowledge of the appropriate lag(s) from the net TE, the TE is calculated through the window sliding approach, with each small window of the signals evaluating TE at the designated lag. Similarly, information rate causality will be evaluated through the window sliding approach with the information rate calculated within each window. In the scenario where the signals are coupled, it is expected that there will be spikes in the information rate causality, indicating changes in the distribution of one signal caused by another. Note that the methodology and concepts of the analyses used in this paper are mainly from the previous works [17–19,28,30,32].

The remainder of this paper is organized as follows. In Section 2, we briefly review the concepts for GC, TE, and information rate causality along with their implementation. The three different types of autoregressive models—unidirectional, interchange unidirectional, and bidirectional models—are introduced and analyzed in Sections 3 to 5, respectively. Section 6 contains our conclusions. Appendices A–C contain some background theory to make the paper self-contained. As a reference, the basic model of continuous coupling is studied and analyzed in Appendix D.

2. Methodologies

2.1. Granger Causality

The general idea behind GC is to evaluate the dependency of one process on another process [17,18]. This dependency is calculated by comparing the errors/noises of the modelled signals through the autoregressive model. Consider two stochastic processes, namely $x_1(t)$ and $x_2(t)$, which can be modelled using the information from their respective signals (as shown in Equations (1) and (2)) or including some information from each other (as shown in Equations (3) and (4)).

$$x_1(t) = \sum_{i=1}^n a_i x_1(t-i) + \epsilon_1(t), \tag{1}$$

$$x_2(t) = \sum_{i=1}^n c_i x_2(t-i) + \epsilon_2(t), \tag{2}$$

$$x_1(t) = \sum_{i=1}^n a_i^* x_1(t-i) + \sum_{i=1}^n b_i^* x_2(t-i) + \epsilon_1^*(t), \tag{3}$$

$$x_2(t) = \sum_{i=1}^n c_i^* x_1(t-i) + \sum_{i=1}^n d_i^* x_2(t-i) + \epsilon_2^*(t). \tag{4}$$

Here, $\{a_i, c_i, a_i^*, b_i^*, c_i^*, d_i^*\} \in \mathbb{R}$ are the constant parameters that represent the fractions of contribution from the past observations towards $x_1(t)$ and $x_2(t)$. $\epsilon_{i \in [1,2]}$ is the uncorrelated external additive noise needed for modelling the processes with variance $\Sigma_{ii, i \in [1,2]}$. $\epsilon_{i \in [1,2]}^*$ is the correlated noise with the variance $\Sigma_{ij, (i,j) \in [1,2]}^*$, which can be represented by

the covariance matrix $\Sigma = \begin{bmatrix} \Sigma_{11}^* & \Sigma_{12}^* \\ \Sigma_{21}^* & \Sigma_{22}^* \end{bmatrix}$.

The total interdependence or total causality index ($F_{x_1 \cdot x_2}$) between $x_1(t)$ and $x_2(t)$ is composed of two directional causalities ($F_{x_1 \rightarrow x_2}$ and $F_{x_2 \rightarrow x_1}$) and one instantaneous causality ($F_{x_1 \leftrightarrow x_2}$). These are defined in Equations (5) to (8).

$$F_{x_1 \cdot x_2} = \log \frac{\Sigma_{11} \Sigma_{22}}{|\Sigma|} = \log \frac{\Sigma_{11} \Sigma_{22}}{\Sigma_{11}^* \Sigma_{22}^* - (\Sigma_{12}^*)^2} = F_{x_1 \rightarrow x_2} + F_{x_2 \rightarrow x_1} + F_{x_1 \leftrightarrow x_2}, \tag{5}$$

$$F_{x_1 \rightarrow x_2} = \log \frac{\Sigma_{11}}{\Sigma_{11}^*}, \tag{6}$$

$$F_{x_2 \rightarrow x_1} = \log \frac{\Sigma_{22}}{\Sigma_{22}^*}, \tag{7}$$

$$F_{x_1 \leftrightarrow x_2} = \log \frac{\Sigma_{11}^* \Sigma_{22}^*}{|\Sigma|}. \tag{8}$$

In conjunction, the spectral of the GC can be evaluated in the frequency domain ($\omega = 2\pi f$, $f \equiv$ frequency) using the spectral matrix $S(\omega)$, transfer matrix $H(\omega)$, and the covariance matrix Σ . They are related according to Equation (9).

$$S(\omega) = \begin{bmatrix} S_{11}(\omega) & S_{12}(\omega) \\ S_{21}(\omega) & S_{22}(\omega) \end{bmatrix} = \langle x(\omega)x^\dagger(\omega) \rangle = H(\omega)\Sigma H^\dagger(\omega). \tag{9}$$

Here, $x(\omega) = \begin{bmatrix} x_1(\omega) \\ x_2(\omega) \end{bmatrix}$, $H(\omega) = \begin{bmatrix} H_{11}(\omega) & H_{12}(\omega) \\ H_{21}(\omega) & H_{22}(\omega) \end{bmatrix}$ = $\begin{bmatrix} 1 - \sum_{i=1}^n a_i^* e^{-j\omega i} & -\sum_{i=1}^n b_i^* e^{-j\omega i} \\ -\sum_{i=1}^n c_i^* e^{-j\omega i} & 1 - \sum_{i=1}^n d_i^* e^{-j\omega i} \end{bmatrix}$; $\langle \cdot \rangle$ is defined as an ensemble average and \dagger is denoted as the complex conjugate transpose of a matrix.

Similar to the analysis in the time domain, the total independence in the frequency domain between $x_1(\omega)$ and $x_2(\omega)$ ($I_{x_1 \cdot x_2}$) also consists of two directional causalities ($I_{x_1 \rightarrow x_2}$ and $I_{x_2 \rightarrow x_1}$) and one instantaneous causality ($I_{x_1 \leftrightarrow x_2}$). They are expressed as Equations (10) to (13).

$$I_{x_1 \cdot x_2} = \log \frac{S_{11}(\omega)S_{22}(\omega)}{|S(\omega)|} = I_{x_1 \rightarrow x_2} + I_{x_2 \rightarrow x_1} + I_{x_1 \leftrightarrow x_2}. \tag{10}$$

$$I_{x_1 \rightarrow x_2} = \log \frac{S_{22}(\omega)}{\hat{H}_{22}(\omega)\Sigma_{22}\hat{H}_{22}(\omega)}, \tag{11}$$

$$I_{x_2 \rightarrow x_1} = \log \frac{S_{11}(\omega)}{\hat{H}_{11}(\omega)\Sigma_{11}\hat{H}_{11}(\omega)}, \tag{12}$$

$$I_{x_1 \leftrightarrow x_2} = \log \frac{(\hat{H}_{11}(\omega)\Sigma_{11}\hat{H}_{11}(\omega))(\hat{H}_{22}(\omega)\Sigma_{22}\hat{H}_{22}(\omega))}{|S(\omega)|}. \tag{13}$$

Here, $\hat{H}_{22}(\omega) = H_{22}(\omega) + \frac{\Sigma_{12}^*}{\Sigma_{22}^*}H_{21}(\omega)$ and $\hat{H}_{11}(\omega) = H_{11}(\omega) + \frac{\Sigma_{12}^*}{\Sigma_{11}^*}H_{12}(\omega)$; $\overline{[\cdot]}$ is denoted as the complex conjugate of the element.

In this paper, the frequency domain and the time-varying frequency domain of GC will be calculated through the non-parametric method. The general idea of this method is to decompose the spectral matrix $S(\omega)$ into the transfer matrix $H(\omega)$ and the covariance of the noises Σ . This decomposition can be achieved using Wilson’s algorithm [18,24,25,35,36]. The spectral matrices for the frequency domain and the time-varying frequency domain are expressed as Equations (14) and (15), respectively.

$$S(\omega) = \frac{\langle x_i(\omega)\overline{x_j(\omega)} \rangle}{T}, \quad i, j \in \mathbb{R}, \tag{14}$$

$$S(t, \omega) = \frac{\langle s_i(t, \omega)\overline{s_j(t, \omega)} \rangle}{T}, \quad i, j \in \mathbb{R}. \tag{15}$$

Here, T denotes the total period of the signal. $x(\omega)_{n|n \in [i,j]}$ is the discrete Fourier transform of the signal and $s_{n|n \in [i,j]}(t, \omega)$ is the short-time Fourier transform of the signal $x(t)_{n|n \in [i,j]}$. Note that $\overline{[\cdot]}$ denotes the complex conjugate. The Hann function window is used when evaluating the short-time Fourier transform. The window will move through the whole series with half of the data points being overlapped. The general flow of the procedure is shown in Figure 1a.

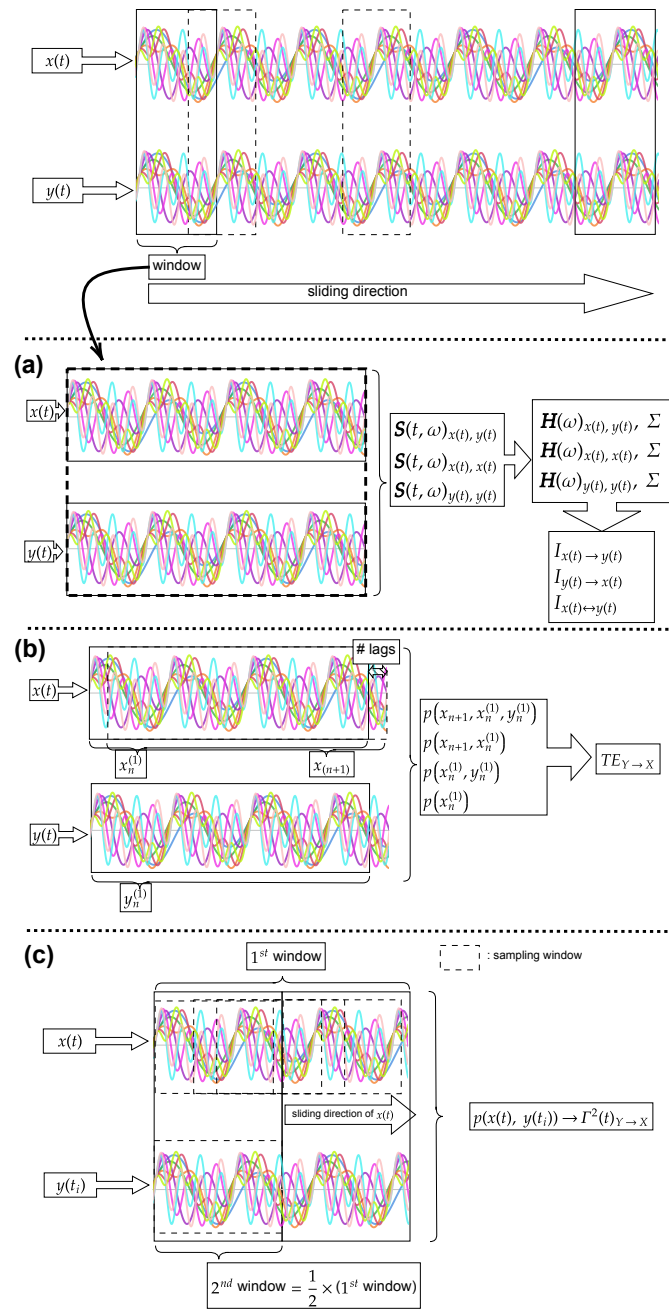


Figure 1. The procedure of implementing the causality analyses used in this paper, each colour of the lines in the image represents a different simulation. In this study, each window contains 0.5 s of data points and overlaps with the previous window by 0.25 s. The causality analyses are conducted within each window. (a) illustrates the components ($S(\dots)$, $H(\dots)$) calculated within the windows for the non-parametric GC analysis. Refer to Section 2.1 to know the corresponding components. (b) shows the estimation of distributions $p(\dots)$ for TE evaluation. The distributions $p(\dots)$ are estimated based on the samples $x_n^{(1)}$, $x_{n+1}^{(1)}$, and $y_n^{(1)}$. Refer to Section 2.2 for the definition of TE. (c) demonstrates the evaluation of information rate causality. Each window (labelled as 1st window) is divided into two windows with the first half labelled as 2nd window. The distribution $p(x(t), y(t_i))$ estimates the evolution of distribution $x(t)$ while $y(t_i)$ is fixed at the 2nd window. Refer to Section 2.3 for the definition of information rate causality.

2.2. Transfer Entropy

TE is a model-free approach used to calculate causality by evaluating the dependency between processes. The general expression of TE is given by Equation (17), which measures the additional information required for the realization of one state of the process $(x_1(t))$ depending on the past states of the processes (e.g., $x_1(t - 1)$ and $x_2(t - 1)$) [19,28,37].

$$TE_{x_2(t) \rightarrow x_1(t)} = \sum_{x \in x_1(t), y \in x_2(t)} p(x_{n+1}, x_n^{(k)}, y_n^{(l)}) \log_2 \frac{p(x_{n+1} | x_n^{(k)}, y_n^{(l)})}{p(x_{n+1} | x_n^{(k)})} \tag{16}$$

$$= \sum_{x \in x_1(t), y \in x_2(t)} p(x_{n+1}, x_n^{(k)}, y_n^{(l)}) \log_2 \frac{p(x_{n+1}, x_n^{(k)}, y_n^{(l)}) p(x_n^{(k)})}{p(x_{n+1}, x_n^{(k)}) p(x_n^{(k)}, y_n^{(l)})}. \tag{17}$$

Here, x_{n+1} represents the state of $x_1(t)$ at the $(n + 1)$ -th moment, $x_n^{(k)}$ represents the state at n -th moment of $x_1(t)$ consisting of k previous states of $x_1(t)$, $y_n^{(l)}$ represents the state at n -th moment of $x_2(t)$ consisting of l previous states. Note that the previous states of k and l are arbitrary depending on the interest of the study. For instance, in a collective observed stochastic process $X = (x_1, x_2, \dots, x_{10})$, $x_8^{(2)} = (x_n, x_m) | n, m \in [1, 8] \rightarrow (x_1, x_2)$ or (x_3, x_8) , etc. Thus, TE quantifies the additional information needed for the realization of $x_1(t)$ at $(n + 1)$ from $x_2(t)$ with the assumption that $x_1(t)$ is independent of $x_2(t)$. If $x_2(t)$ has no impact on the future evolution of $x_1(t)$, and hence $p(x_{n+1} | x_n^{(k)}, y_n^{(l)}) = p(x_{n+1} | x_n^{(k)})$, then $T_{Y \rightarrow X} = 0$.

For the TE calculation, the probability distribution $p(\dots)$ will be estimated via histogram with the bin size of 5, as this bin size is determined by the cubic root of Rice’s rule to accommodate the 3-dimensional of joint probability within TE ($p(x_{n+1}, x_n^{(1)}, y_n^{(1)})$). Using a larger number of bin sizes will not be able to properly depict the distribution and it will bring a spurious result in calculating the TE. To simplify the analysis, the number of past values of k and l is set to 1 ($k = l = 1$) in this paper. Two different evaluations will be conducted for TE. First, assuming that the processes are stationary, the net TE ($TE_{x_1(t), x_2(t)} = TE_{x_1(t) \rightarrow x_2(t)} - TE_{x_2(t) \rightarrow x_1(t)}$) is evaluated on the entire processes at different lags. This will enable us to choose the appropriate past value(s)/lag(s) in evaluating the TE based on the significance of the net TE. Second, using the knowledge of the appropriate lag from the net TE, the TE is next evaluated via a sliding window approach. This sliding window approach is used to calculate the TE at each instance of the interested window. In this sliding window approach, a small number of data points of the stochastic processes is sampled within the window and the TE is calculated. These calculations are sketched in Figure 1b.

2.3. Information Rate Causality

In a dimensionless statistical manifold, the distance between two probability distributions is defined by their statistical difference. One commonly used measure of this difference is the KL divergence/relative entropy, but this measure is not symmetrical and not path-dependent. For the time-dependent probability distribution, the changes in the distribution along with time can be measured through the information rate and information length. The information rate ($\Gamma(t)$) quantifies the rate of change of the distribution, while information length (\mathcal{L}) measures the total change of the distribution. These measures are defined and expressed as Equations (18) and (19) [30–32].

$$\Gamma^2(t) = \int dx p(x, t) [\partial_t \ln p(x, t)]^2 = 4 \int dx \left[\partial_t \sqrt{p(x, t)} \right]^2, \tag{18}$$

$$\mathcal{L} = \int dt \Gamma(t). \tag{19}$$

Here $p(x, t)$ is the probability distribution of a stochastic process x at the time t .

For studying causality, we consider two stochastic processes $x_1(t)$ and $x_2(t)$ and the information rate for a joint probability distribution is evaluated. The causality between these processes can be quantified by the changes in the information rate (statistical changes or the changes in distribution) while having another process remain static. For instance, the information rate of $x_2(t)$ causing $x_1(t)$ is defined as Equation (21) and called information rate causality in this paper.

$$\Gamma^2(t)_{x_2(t) \rightarrow x_1(t)} = \int dx_1 dx_2 p(x_1(t), x_2(t_i)) [\partial_t \ln p(x_1(t), x_2(t_i))]^2 \quad (20)$$

$$= 4 \int dx_1 dx_2 \left[\partial_t \sqrt{p(x_1(t), x_2(t_i))} \right]^2. \quad (21)$$

Referring to Figure 1c, here t_i is denoted as the reference time that remains static in the process. Therefore, $p(x_1(t), x_2(t_i))$ is a probability distribution that is sampled by having the process $x_2(t)$ remain static at time t_i while the process $x_1(t)$ evolves along the time t . Since information rate can only quantify the changes in the distribution, therefore the *information rate causality* is evaluated within each window of interest instead of the whole process of the time series to determine whether the coupling of the processes is still persists. To accommodate the calculation, a discretized version of Equation (21) is used and it is expressed as Equation (22).

$$\Gamma^2(t)_{x_2 \rightarrow x_1} = 4 \sum_{x_1 \in X_1, x_2 \in X_2} \frac{\Delta x_1 \Delta x_2}{[\Delta t]^2} \left[\sqrt{p(x_1(t + \Delta t), x_2(t_i))} - \sqrt{p(x_1(t), x_2(t_i))} \right]^2 \quad (22)$$

For calculating the information rate, the joint probability distribution for two processes is estimated via the histogram with the bin size (b) determined by the square root of Rice's rule, as expressed in Equation (23). Rice's rule is used because it is able to appropriately determines the bin size to sample the obtained data [38,39] for 1-dimensional distribution. Hence, the square root of Rice's rule is to accommodate the 2-dimensional distribution in this case.

$$b = \sqrt{[2\sqrt[3]{n}]}, \quad n = \text{number of sampled data.} \quad (23)$$

In this paper, we evaluate the information rate causality by employing Equation (22) to examine the impact of one process, which remains static in time, on the distribution of another process within a specific time window. To estimate the joint probability distribution utilized in Equation (22), a histogram is employed with a bin size determined by Equation (23). Figure 1c illustrates the overall procedure of this evaluation.

2.4. Summary of the Procedure: Data And Analysis

We further develop *information rate causality* (refer to Section 2.3) for the analysis of our numerically generated data and compare it with the non-parametric GC (refer to Section 2.1) and window sliding TE (refer to Section 2.2). Our numerical data are generated by simulating different discrete autoregressive models covering both linear and nonlinear models. The simulation is conducted for 10,000 trials with each trial running for 25 s (physical time) at 200 Hz (physical sampling frequency). Different coupling/causal relationships between the signals are considered in this paper, such as unidirectional (refer to Section 3), interchanging unidirectional (refer to Section 4), and bidirectional (refer to Section 5). Note that the noncoupling cases are also considered in this paper to check the robustness of the causality analyses; specifically, we refer to the uncoupling of $x_1(t)$ towards $x_2(t)$ after the physical time 10 s. Using simulated signals, the causality analyses (GC, TE, *information rate causality*) are conducted according to the sketch shown in Figure 1.

To be consistent with the data points used for each window, each sampling window will contain 0.5 s of data points with an overlap of 0.25 s (0.5×0.5 s), which is equivalent to 100 data points for one simulation. Since the models are simulated for 10,000 trials, each

window will consist of 1×10^6 sample points (100 points \times 10,000 trials). Note that this window is not applied to the frequency domain of GC and the net TE as both are calculated based on the whole series.

2.5. Summary of the Different Models and Key Results

Prior to the discussion of different models in Sections 3 to 5 with the implementation of different analysis methods, we provide the key results/findings in Table 1.

Table 1. Summary of the performance of GC, TE, and information rate causality for different models discussed in Sections 3 to 5. In general, GC performs well for linear and stationary signals; TE performs well only if the lags are chosen correctly; information rate causality performs well and can recover the underlying lags of the signals.

Method	Unidirectional (Section 3)		Interchange (Section 4)		Bidirectional (Section 5)	
	Linear	Nonlinear	Linear	Nonlinear	Linear	Nonlinear
GC	performs well	fails in nonstationary signals	performs well		fails to capture the bidirectional coupling of signals	
TE	performs well if correct lags are chosen or else produces spurious results					
Information rate causality	performs well and able to recover the underlying lags of coupling					

3. Unidirectional Model

In this part, unidirectional causality, where the first process ($x_1(t)$) influencing the second process ($x_2(t)$), is considered for both linear and nonlinear autoregressive models given by Equations (24) and (25), respectively. The causality between the processes occurs as the processes are coupled with each other through the Heaviside step function ($H(\dots)$).

Linear model:

$$\begin{aligned}
 x_1(t) &= 0.55x_1(t - 1) + \epsilon_1(t), \\
 x_2(t) &= H(\tau - 10)x_1(t - 2) + \epsilon_2(t).
 \end{aligned}
 \tag{24}$$

Nonlinear model:

$$\begin{aligned}
 x_1(t) &= 0.55x_1(t - 1)e^{1-|x_1(t-1)|} + \epsilon_1(t), \\
 x_2(t) &= H(\tau - 10)x_1(t - 2) + \epsilon_2(t).
 \end{aligned}
 \tag{25}$$

Here, t is the time steps and τ is the physical time. Both t and τ are related as $\tau = [\text{physical frequency}] \times t$. $H(\tau - 10)$ is a Heaviside step function that allows the coupling to occur starting at the physical time of 10 s. $\epsilon_{n|n \in [1,2]}$ is the Gaussian noise that perturbs the systems. In this study, the noises are set to have zero mean ($\langle \epsilon_n \rangle = 0 | n \in [1, 2]$) along with specific covariance/variance ($\Sigma_{nm} = \langle (\epsilon_n - \langle \epsilon_n \rangle)(\epsilon_m - \langle \epsilon_m \rangle) \rangle | n, m \in [1, 2]$), which will be expressed in covariance matrix $\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$. Due to the exponential term in the nonlinear model, the process will have non-stationary oscillation when the power of the exponential term becomes positive ($e^{\mathbb{R}} \rightarrow \text{diverge}$). Hence, two different values of noise covariance will be considered to study the stationary and non-stationary oscillation of the processes. They are labelled as large noise and small noise with the respective matrices expressed as $\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{bmatrix}$ and $\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} 0.005 & 0.0 \\ 0.0 & 0.005 \end{bmatrix}$. Note that the cross-covariance of the noises is set to zero ($\Sigma_{12} = \Sigma_{21} = 0.0$) to ensure the coupling is purely due to the intrinsic interaction in the model but not through the mutual noises. The linear model will have stationary oscillation for either large or small noises, and hence only one noise will be used to simulate the linear model and large noise is chosen. The general structure of the models is shown in Figure 2.

To explore all possible combinations of cases, we investigate the coupling and noncoupling cases for both linear and nonlinear models. To this end, we investigate the following six difference cases: [linear: couple], [linear: noncouple], [nonlinear: large noise, coupling], [nonlinear: large noise, noncoupling], [nonlinear: small noise, coupling], and [nonlinear: small noise, noncoupling].

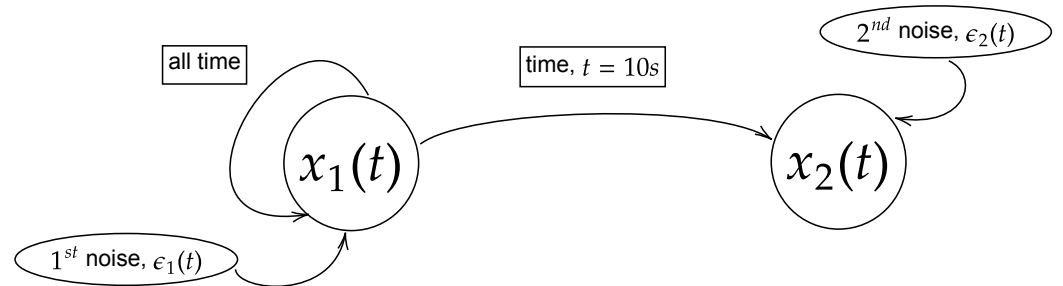


Figure 2. Model of the flow of information of the processes $x_1(t)$ and $x_2(t)$ for Equations (24) and (25). In this paper, the equations are simulated with the physical time of 25 s and sampling frequency of 200 Hz (5000 realizations) with either large noise or small noise. The coupling between the processes occurs at physical time of 10 s.

The result of the simulation for the Equations (24) and (25) is shown in Figure 3. Notice that the signals for the nonlinear model exhibit non-stationary oscillation when the perturbing noise is small, as shown in Figure 3e,f. This is due to the divergence of the positive power of the exponential term of Equation (25) ($e^{1-|x_1(t-1)|} = e^{\mathbb{R}} \rightarrow$ diverge $|\epsilon_1(t) \equiv$ small noise) and the influences of the previous state towards the current state ($x_1(t) \leftarrow 0.55x_1(t - 1)$). The rest of the simulated signals have stationary oscillation, as shown in the phase space where the observed states are localized around $(x_1, x_2) = (0, 0)$.

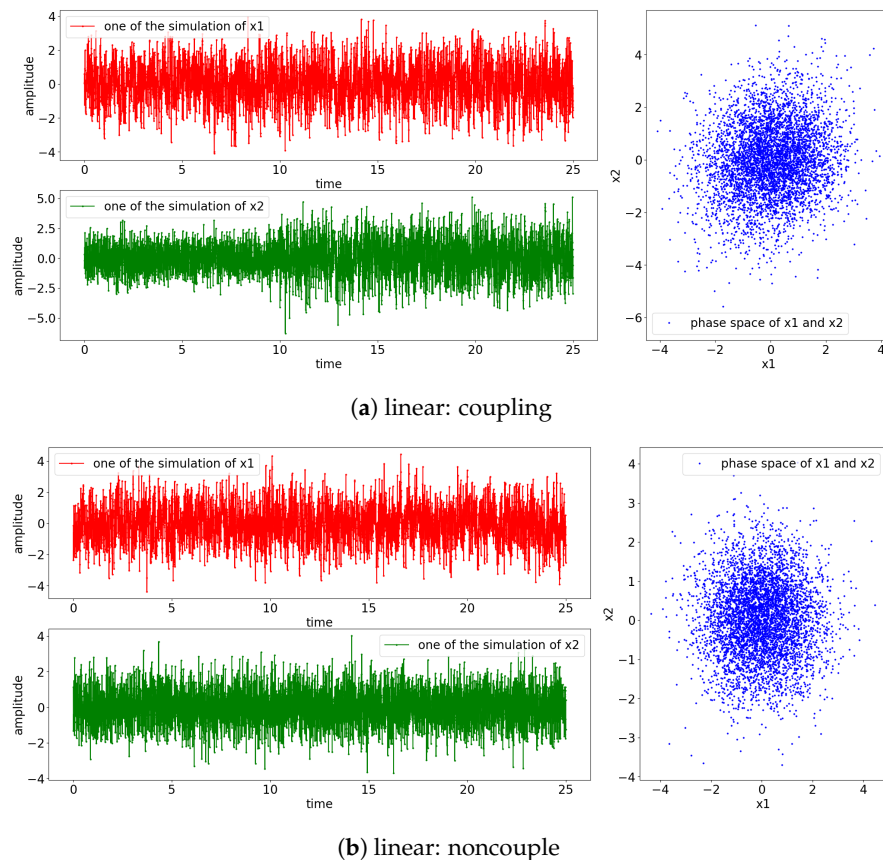
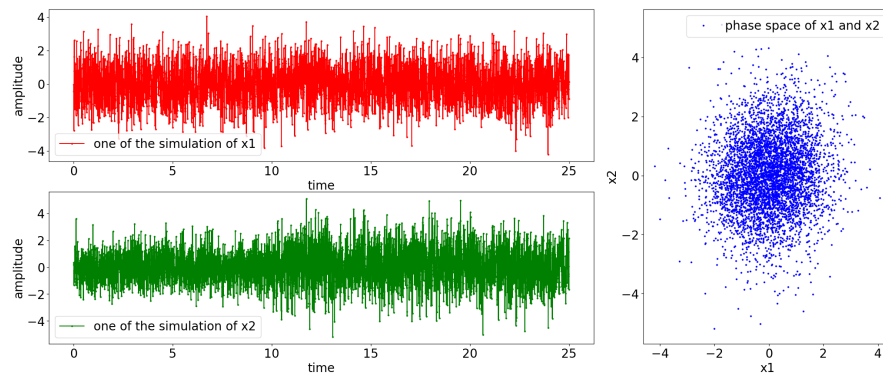
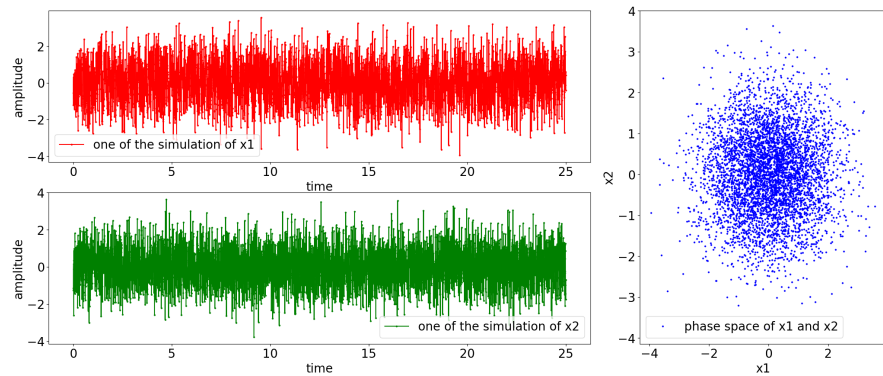


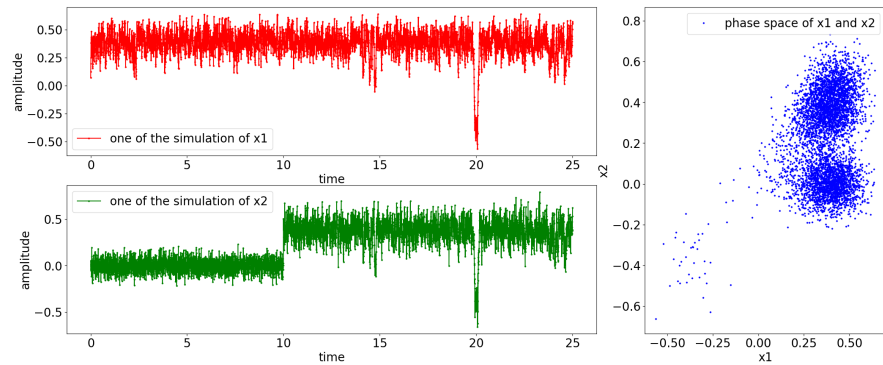
Figure 3. Cont.



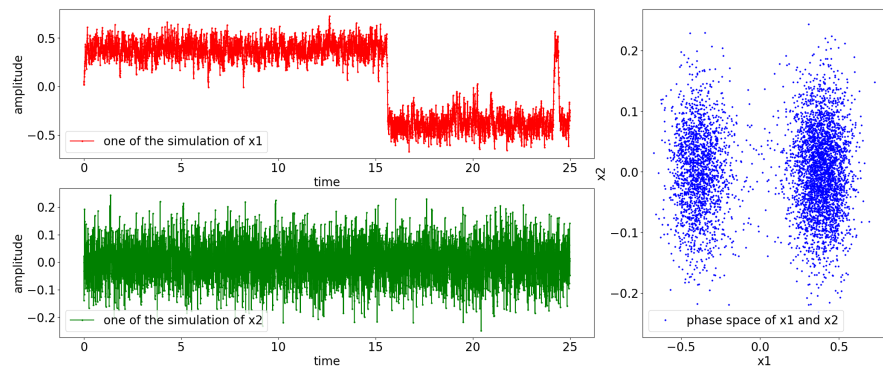
(c) nonlinear: large noise, couple



(d) nonlinear: large noise, noncouple



(e) nonlinear: small noise, couple



(f) nonlinear: small noise, noncouple

Figure 3. Result of the simulation based on Equations (24) and (25) (refer to Figure 2 for graphical explanation of the process). The [top left, red] is the result of x_1 and [bottom, green] is the result of x_2 ; [right blue] is the phase space of x_1 and x_2 . Note that the coupling of the processes occurs at 10 s.

3.1. Granger Causality

As shown in Figure 4, the coupling of x_1 to x_2 can be captured well through non-parametric GC analysis for stationary linear signals, as shown in Figure 4a. Concurrently, the detection of the causality works well for stationary nonlinear signals, as shown in Figure 4c. This is because the nonlinear exponential term is well-approximated to a constant value as the spectral matrix decomposed to the transfer matrix and covariance matrix of noise via Wilson’s algorithm. Note that the amplitude of the GC decreases for this nonlinear stationary signal. This is due to the contribution of the exponential term ($e^{-\mathbb{R}}$) that eventually affects the $S_{22}(\omega)$ element of the spectral matrix (refer to Equation (11)) and hence the amplitude of the causality ($I_{x_1 \rightarrow x_2}$). For non-stationary signals, the frequency domain of GC is still able to show the causality within the signals, but the time-varying frequency of GC cannot represent well the period of causation, as shown in Figure 4e. From the figure, we can see that the nonparametric GC gives a false result, as it suggests that the coupling between the processes occurs throughout the time. Similarly, the nonparametric GC analysis works fine in evaluating the non-causality cases for stationary cases but not so well for the non-stationary case (refer to Figure 4b,d,f. GC analysis fails to work on the signals oscillating non-stationary, as the signals cannot be modelled well via the linear autoregressive equations (refer to Equations (3) and (4)).

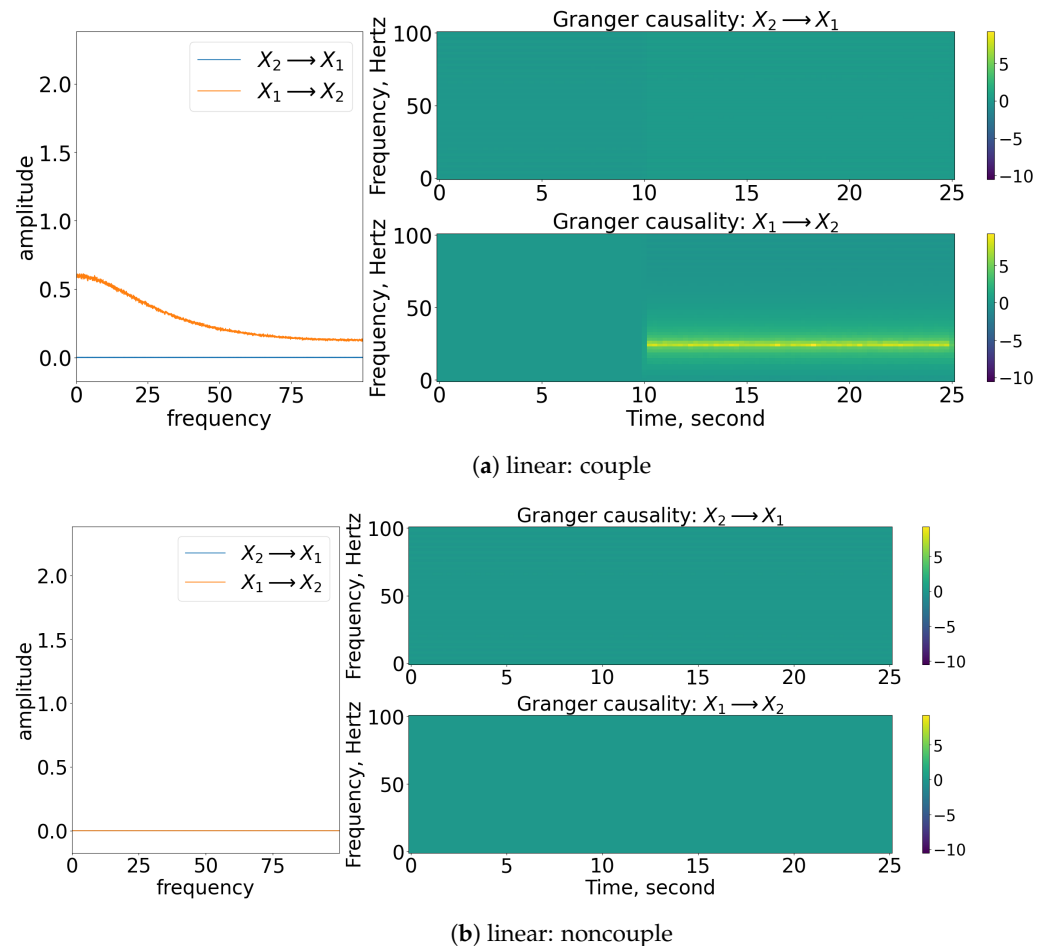
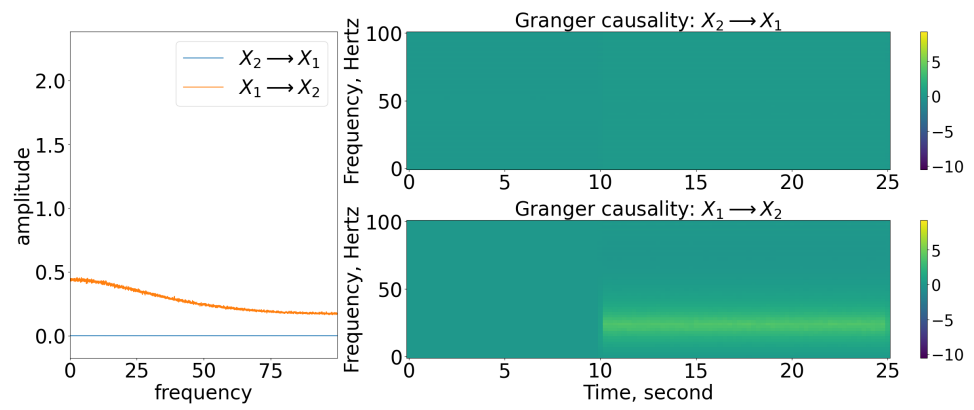
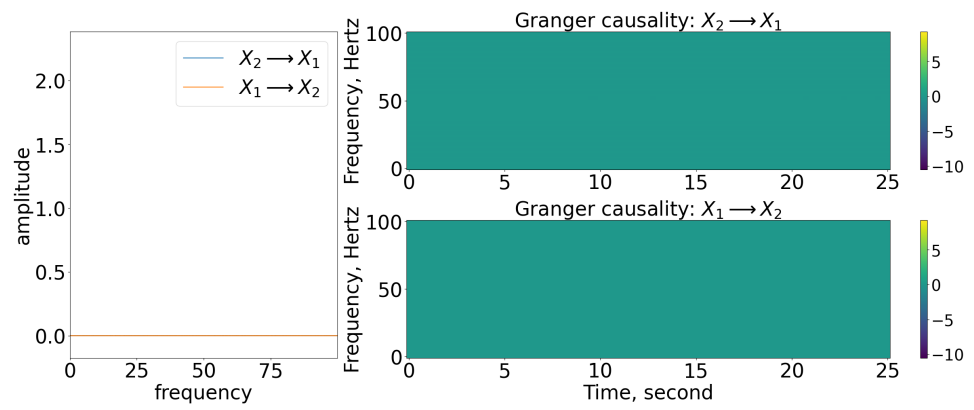


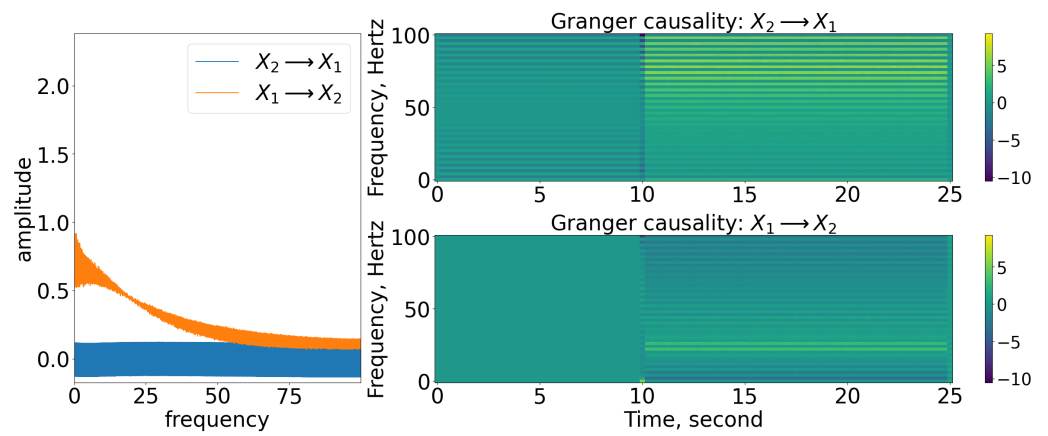
Figure 4. Cont.



(c) nonlinear: large noise, couple

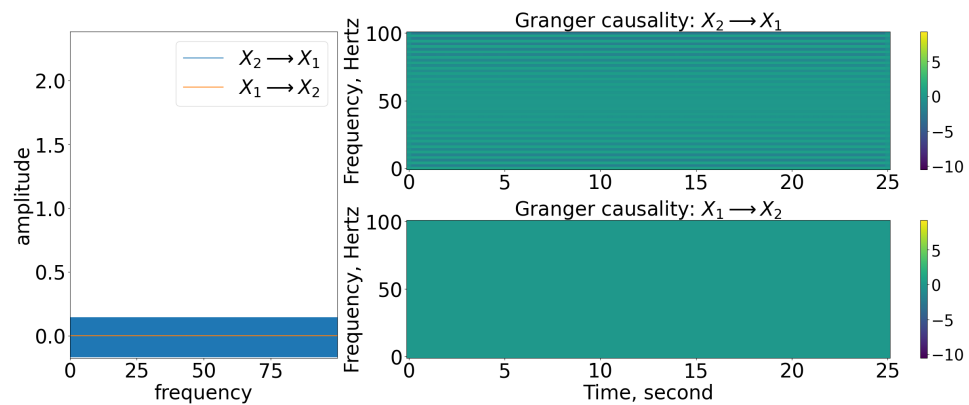


(d) nonlinear: large noise, noncouple



(e) nonlinear: small noise, couple

Figure 4. Cont.

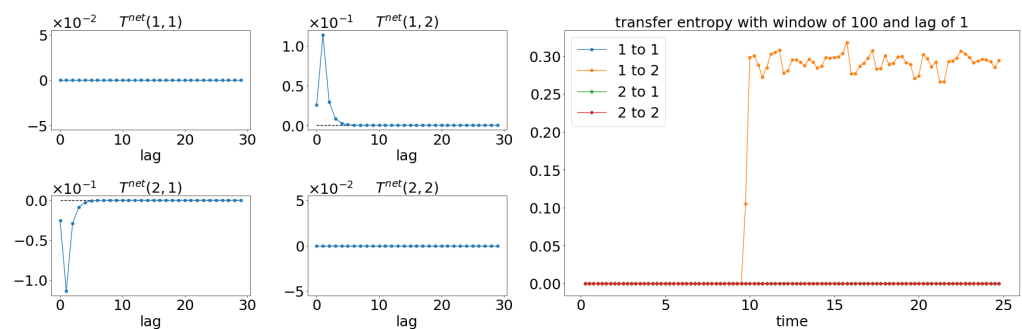


(f) nonlinear: small noise, noncouple

Figure 4. Result of the spectral and time-varying frequency of GC. Note that for each subfigure, the [left] shows the spectral (frequency domain) of GC, [blue] indicates x_2 causes x_1 , while [orange] indicates x_1 causes x_2 ; [top, right] shows the time-varying frequency of GC of x_2 to x_1 ; [bottom, right] shows the time-varying frequency of GC of x_1 to x_2 . The results shown are based on the processes in Equations (24) and (25). Refer to Figure 2 for the graphical explanation of the process.

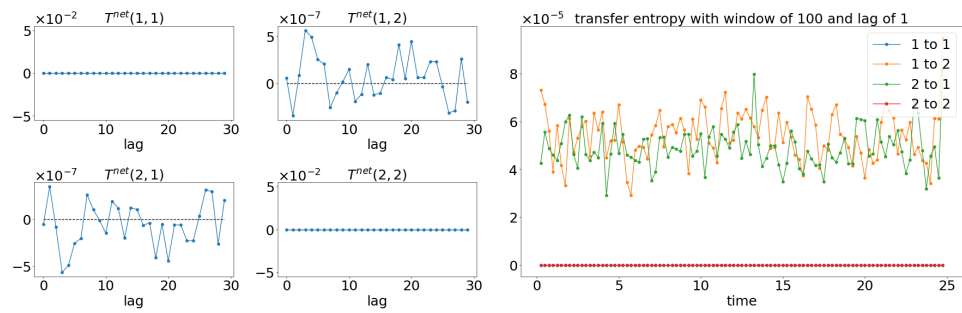
3.2. Transfer Entropy

The results from TE analysis are shown in Figure 5. By first assuming that the signals are stationary, the net TE ($TE_{X,Y} = TE_{X \rightarrow Y} - TE_{Y \rightarrow X}$) is first calculated to find the suitable lag for evaluating the sliding window TE later. Figure 5a,c,e show that the net TE between the signals is significant at the lag of 1. This is because the simulated models (Equations (24) and (25)) have the coupling occurring at one lag (compares $x_1(t) := x_1(t - 1)$ and $x_2(t) := x_1(t - 2)$). Next, the window sliding TE is conducted through the conditional probability (multidimensional probability) that contains the data from one previous lag (refer to Equation (17)) at each window of the evaluation of TE. As shown in Figure 5a,c,e, the window sliding TE is able to show the time (10 s physical time) when the causality occurs. Similarly, the TE is also able to capture the non-causality situation well for all the cases, as shown in Figure 5b,d,f. Even though TE is able to detect the causality between the signals for either linear stationary or nonlinear non-stationary cases, it requires extra inputs/assumptions (such as bin sizes, number of lags, and the dimension of the multidimensional probability) to work properly/accurately. The failure of TE in capturing the causality for the linear model (Equation (24)) is shown in Figure 6 when the lag is set to 9 instead of 1 when evaluating the TE.

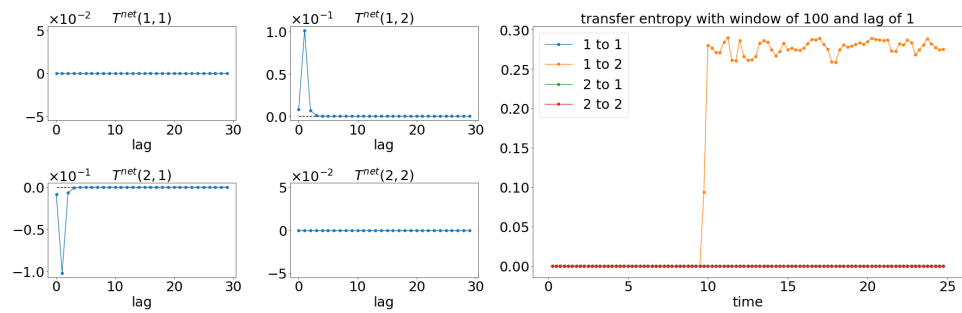


(a) linear: couple

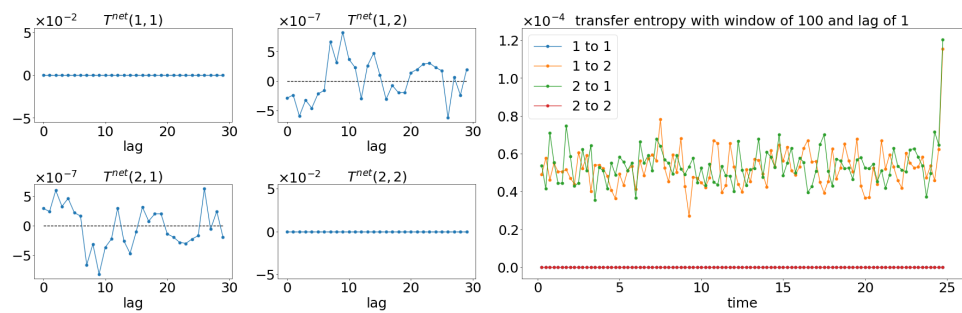
Figure 5. Cont.



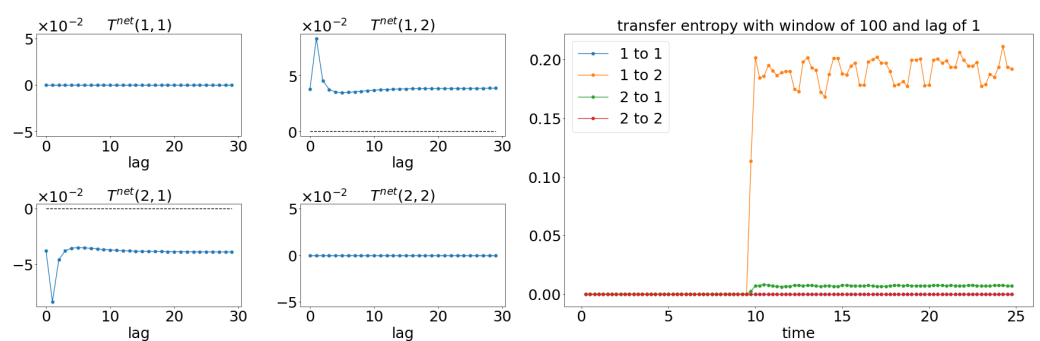
(b) linear: noncouple



(c) nonlinear: large noise, couple

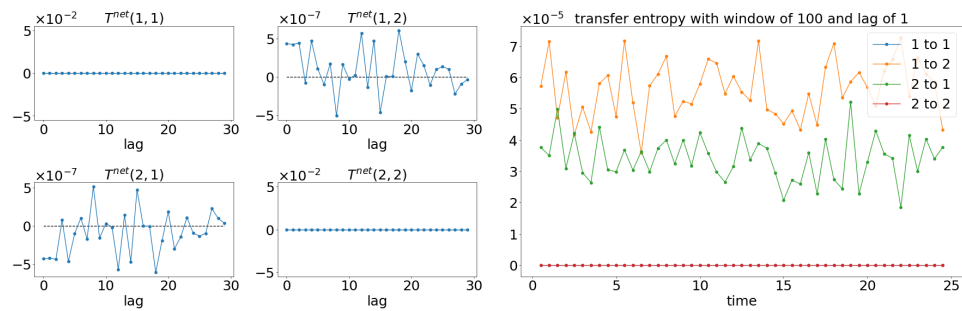


(d) nonlinear: large noise, noncouple



(e) nonlinear: small noise, couple

Figure 5. Cont.



(f) nonlinear: small noise, noncouple

Figure 5. Result of the net TE and window sliding TE. Note that for each subfigure, [left, 4 × 4] depicts the net TE for the whole signals; while the [right] shows the window sliding TE. The results shown are based on the processes in Equations (24) and (25). Refer to Figure 2 for the graphical explanation of the process.

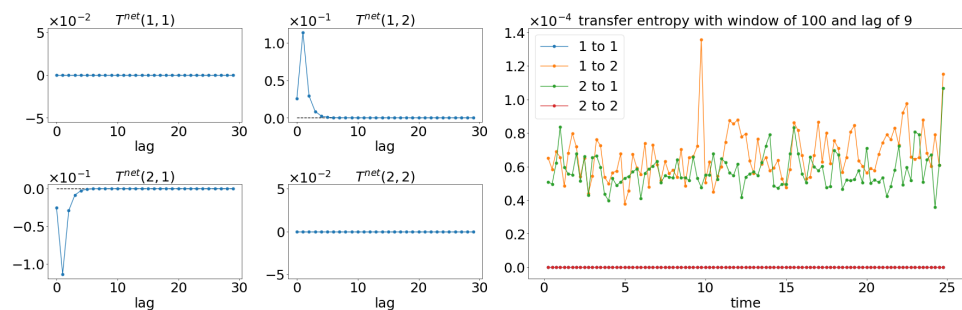
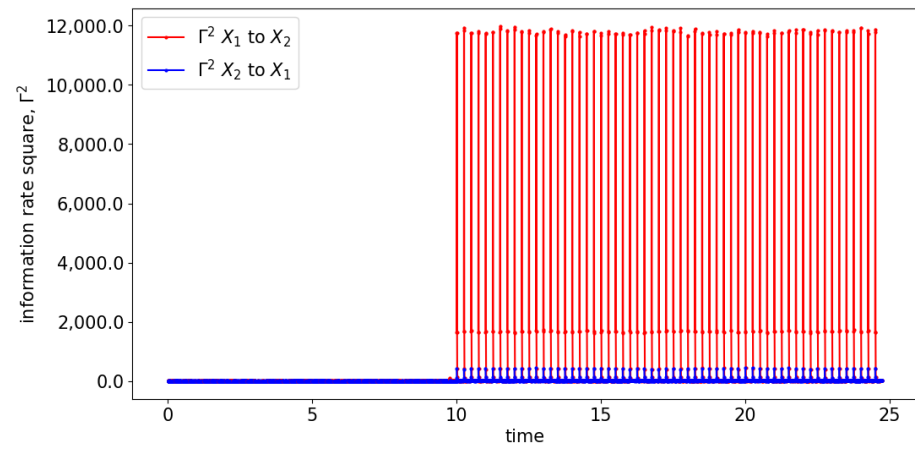


Figure 6. Failure of TE in capturing the causality between the signals x_1 and x_2 for [linear: couple] (refer to Equation (24)) when lag 9 is used for evaluation, as compared to Figure 5a, where the TE accurately shows that the coupling occurs after 10 s.

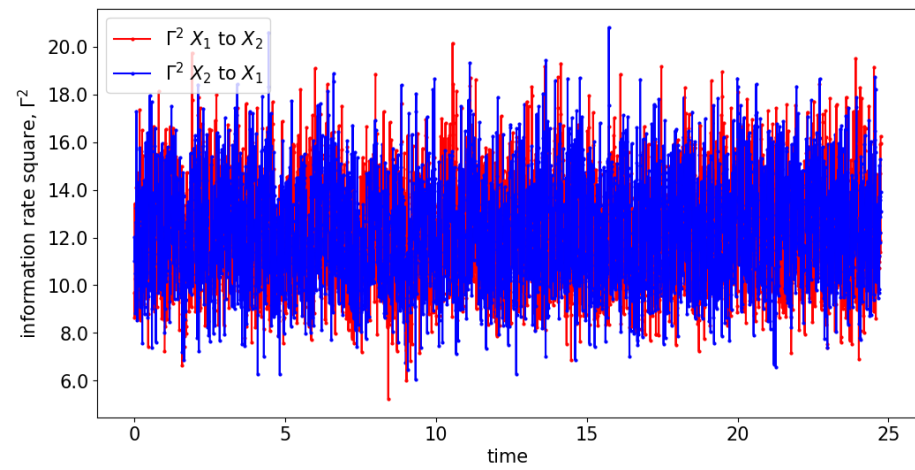
3.3. Information Rate Causality

Both GC and TE detect causality by examining the increase in the predictability of the stochastic processes (measured by the inverse of entropy) when the past information is included. In comparison, information rate causality evaluates causality by quantifying the rate of change of the probability distribution of one variable conditioned on another, as noted previously. The results are shown in Figures 7 and 8 where the changes of the distribution due to the causality between the processes are well-reflected via the changes in the information rate (Γ). For the cases of noncoupling, the information rate is not changing and remains constant, which would suggest that none of the signals cause the changes to the joint distribution (refer to Figure 7b,d,f). Hence, no causality occurs between the signals. Similarly, prior to the coupling of the signals, the causality is not seen, and hence the information rate again remains constant. Notice that the information rate does not stay at zero when no causality happens, it is due to the noises/spikes of the estimated distribution, as shown in Figure S1.

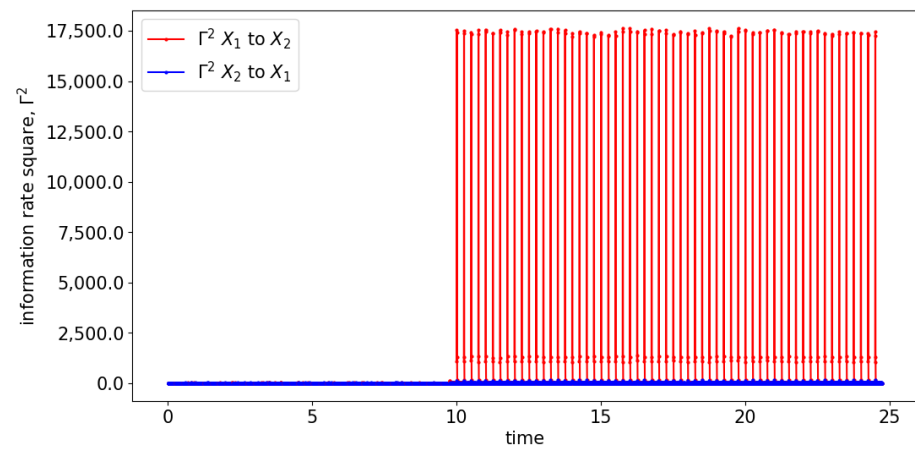
In the presence of the couplings, the changes in the distribution due to the causality among the signals can be observed via the changes in the information rate. For instance, the changes in the information rate due to the causality of $x_1(t)$ to $x_2(t)$ can be observed in Figure 7a,c,e (zoom-in: Figure 8a,c,e). Referring to the models, the signals are coupled with the lag difference of one ($x_1(t) = x_1(t - 1)$ and $x_2(t) = x_1(t - 2)$), and hence the peak of the information rate is observed after one lag. From Figure 8a,c,e, the changes in the information rate prior to the coupling (at physical time 10 s) are observed as the evaluation window of [9.75 s to 10.25 s] picks up the causality of the signals from [10 s to 10.25 s]. Referring to the starting time (near 0 s) of Figure 7e,f, there is the presence of a sharp difference in information rate. This is due to the non-stationary oscillation of $x_1(t)$ that causes the high disturbance of the distribution.



(a) linear: couple

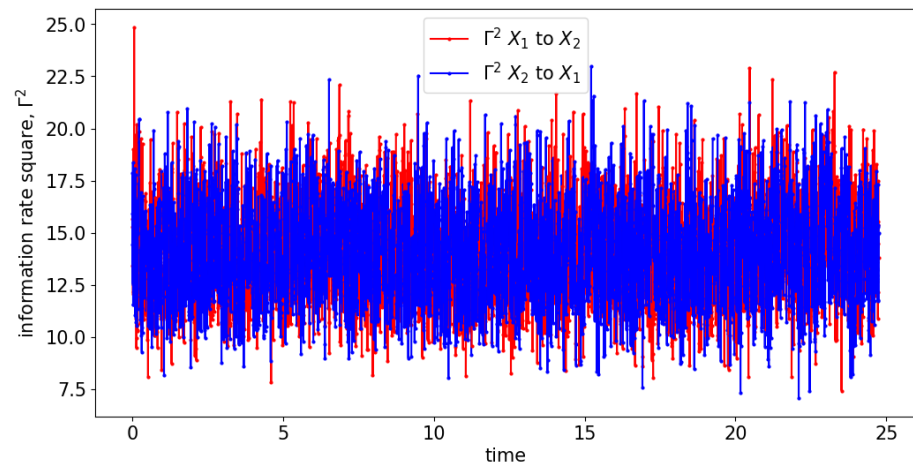


(b) linear: noncouple

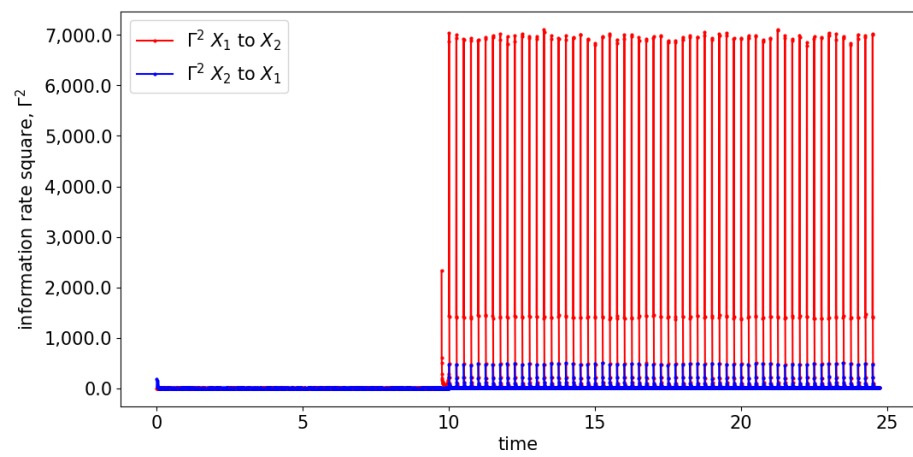


(c) nonlinear: large noise, couple

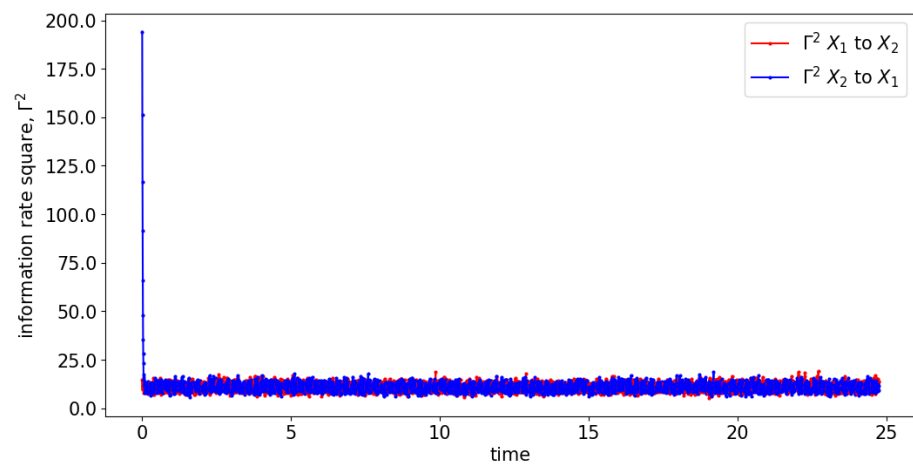
Figure 7. Cont.



(d) nonlinear: large noise, noncouple

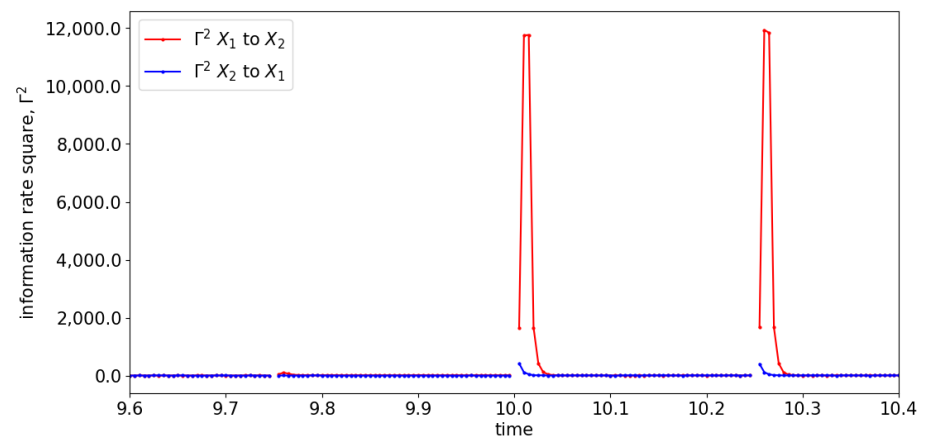


(e) nonlinear: small noise, couple

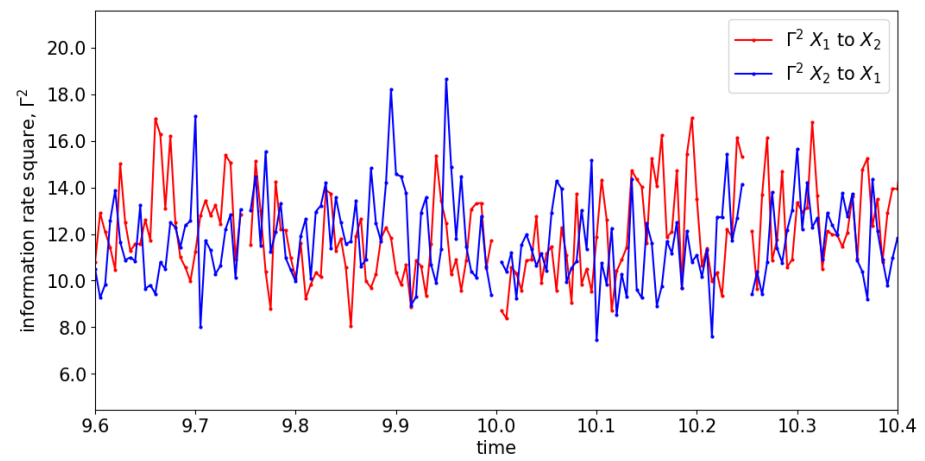


(f) nonlinear: small noise, noncouple

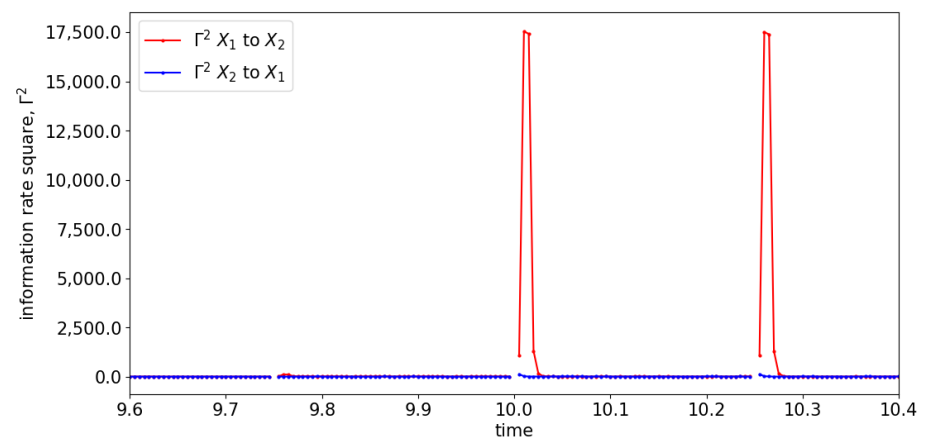
Figure 7. Result of information rate causality for Equations (24) and (25).



(a) linear: couple

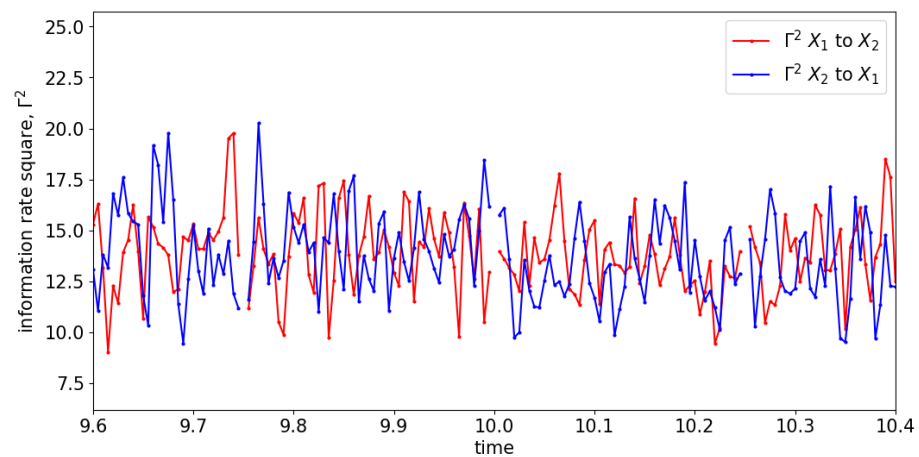


(b) linear: noncouple

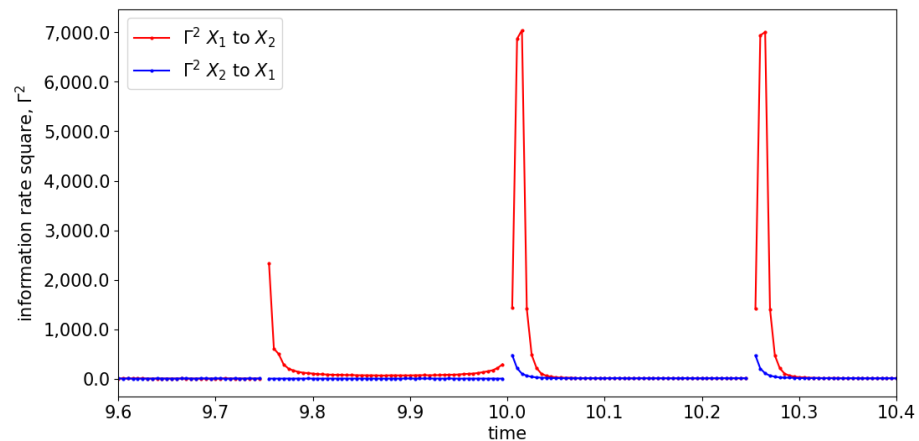


(c) nonlinear: large noise, couple

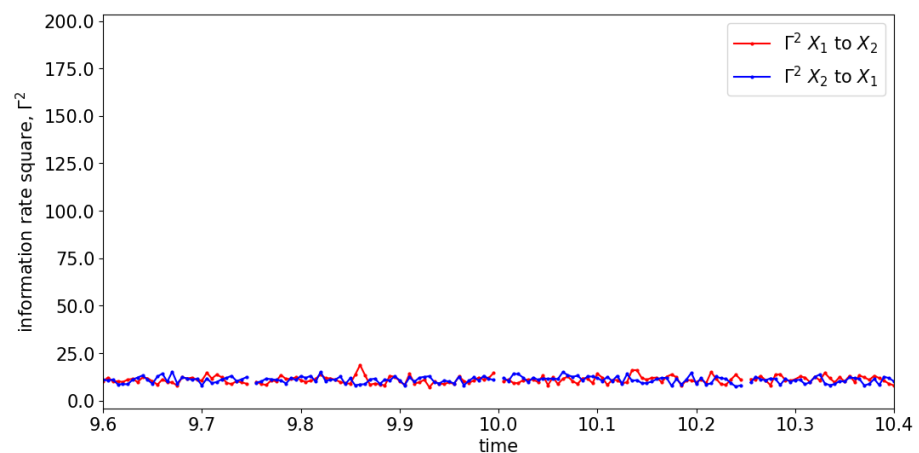
Figure 8. Cont.



(d) nonlinear: large noise, noncouple



(e) nonlinear: small noise, couple



(f) nonlinear: small noise, noncouple

Figure 8. Zoom-in of Figure 7 for the information rate causality between 9.6 s and 10.4 s. Note that the change in the causality occurs at 10 s.

To demonstrate the capability of information rate causality to detect lag differences within a signal, we simulated and analyzed Equation (26), which has the coupling starting at 10 s with a lag difference of 9 (compares $x_1 = x_1(t - 1)$ and $x_2 = x_1(t - 10)$). Figure 9 shows the section of the information rate from 9.6 s to 10.4 s. The figure reveals that the peak of the information rate appears at the 9th time step (0.045 s) at every window of evaluation, which accurately reflects the lag difference specified in Equation (26). This result suggests

that information rate causality can effectively capture the underlying causality of the signals at each evaluation window.

$$\begin{aligned} x_1(t) &= 0.55x_1(t - 1) + \epsilon_1(t), \\ x_2(t) &= H(\tau - 10)x_1(t - 10) + \epsilon_2(t). \end{aligned} \tag{26}$$

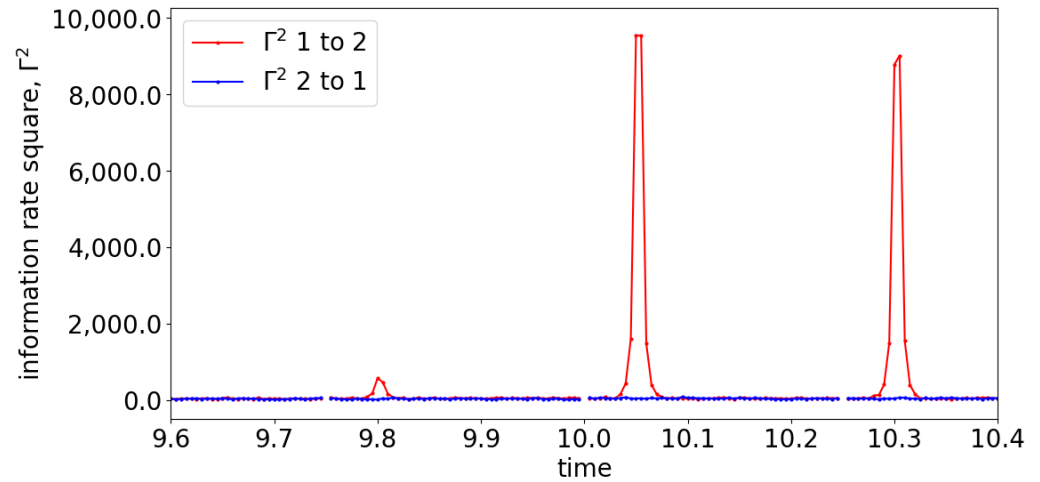


Figure 9. The information rate causality of Equation (26) between 9.6 s and 10.4 s where the causality occurs at 10 s.

4. Interchange of Unidirectional Model

In this part, similar models from Section 3 are used with slight modifications to study the interchange of the unidirectional causality. The models are modified to Equations (27) and (28) for linear and nonlinear models, respectively. The modification with the Heaviside step functions ($H(\dots)$) in the models portrays that the signal $x_2(t)$ causing $x_1(t)$ occurs prior to physical time (τ) 10 s, while the signal $x_1(t)$ causes $x_2(t)$ after 10 s. The general structure of Equations (27) and (28) is illustrated in Figure 10.

Linear model:

$$\begin{aligned} x_1(t) &= H(10 - \tau)0.55x_2(t - 1) + \epsilon_1(t), \\ x_2(t) &= H(\tau - 10)x_1(t - 2) + \epsilon_2(t). \end{aligned} \tag{27}$$

Nonlinear model:

$$\begin{aligned} x_1(t) &= H(10 - \tau)0.55x_2(t - 1)e^{1-|x_2(t-1)|} + \epsilon_1(t), \\ x_2(t) &= H(\tau - 10)x_1(t - 2) + \epsilon_2(t). \end{aligned} \tag{28}$$

Here t is the time-step and τ is the physical time. The Heaviside step functions are used with $H(10 - \tau)$ suggesting the coupling occurs from the beginning of the simulation till the physical time 10 s conversely $H(\tau - 10)$, suggesting the coupling occurs after 10 s. Similar to Section 3, 6 possible cases are simulated based on Equations (27) and (28). The simulated signals will be evaluated by GC, TE, and information rate causality.

The result of the simulations is shown in Figure 11, which suggests that all the signals oscillate stationary. Unlike Equation (25), in which the signal $x_1(t)$ is perturbed by its previous state by a factor of 0.55 ($0.55x_1(t - 1)$) and the small noise ($\epsilon_1(t)$), Equation (28) is perturbed by the past state of $x_2(t)$ ($x_2(t - 2)$) instead. Hence, the power of the exponential will always be negative; consequently, all the simulated signals have stationary oscillations. Note that the noncoupling is referring to the decoupling of $x_1(t)$ to $x_2(t)$ ($H(\tau - 10) = 0$).

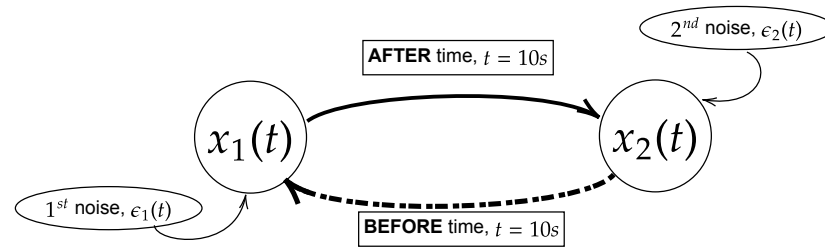
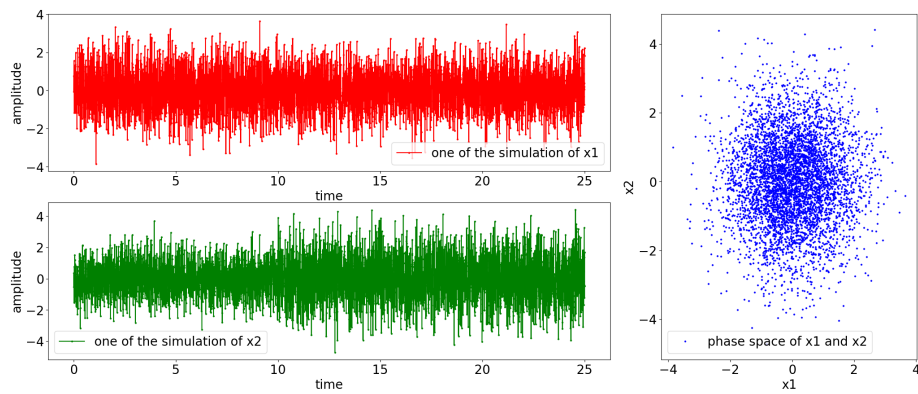
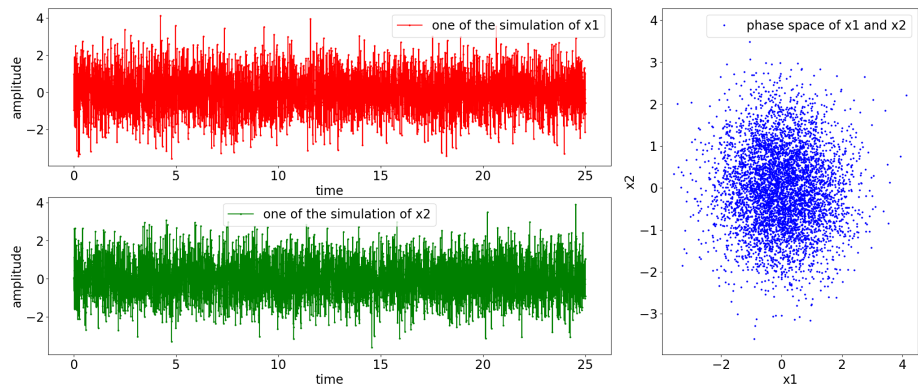


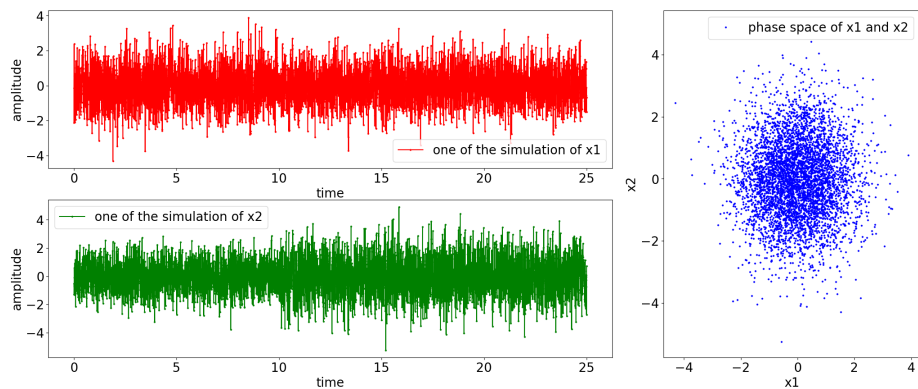
Figure 10. Model of the flow of information of the processes $x_1(t)$ and $x_2(t)$ for Equations (27) and (28). In this paper, the equations are simulated with the physical time of 25 s and sampling frequency of 200 Hertz (5000 realizations) with either large noise or small noise. Note that the process $x_2(t)$ coupling with $x_1(t)$ occurs before 10 s and the interchange of coupling occurs after 10 s. In the context of noncoupling, it is referring to $H(\tau - 10) = 0$.



(a) linear: couple

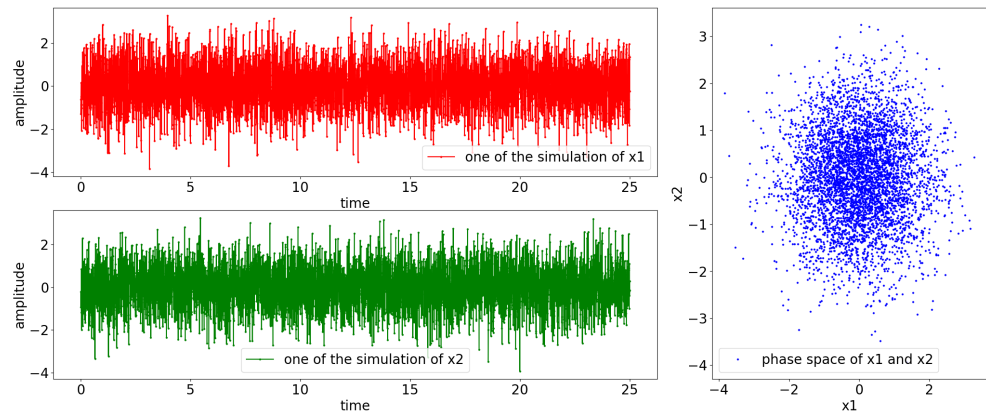


(b) linear: noncouple

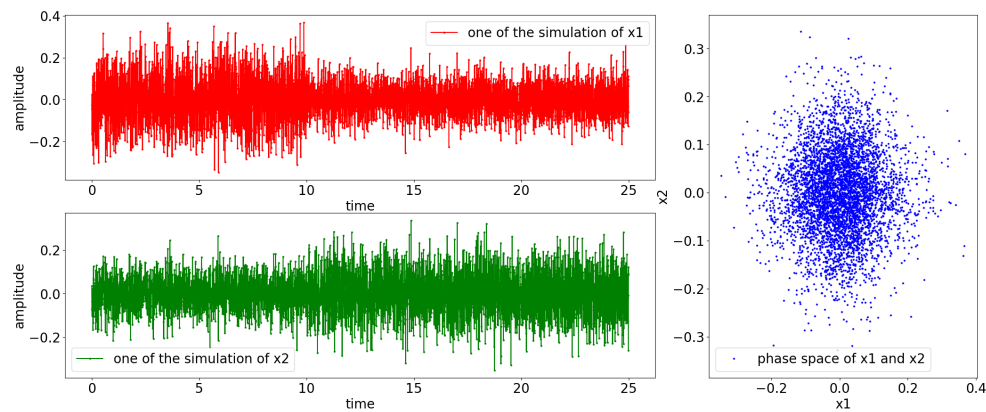


(c) nonlinear: large noise, couple

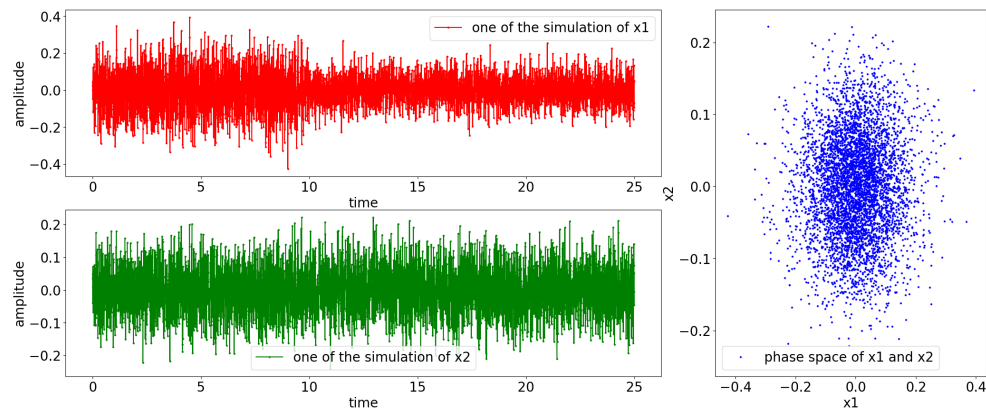
Figure 11. Cont.



(d) nonlinear: large noise, noncouple



(e) nonlinear: small noise, couple



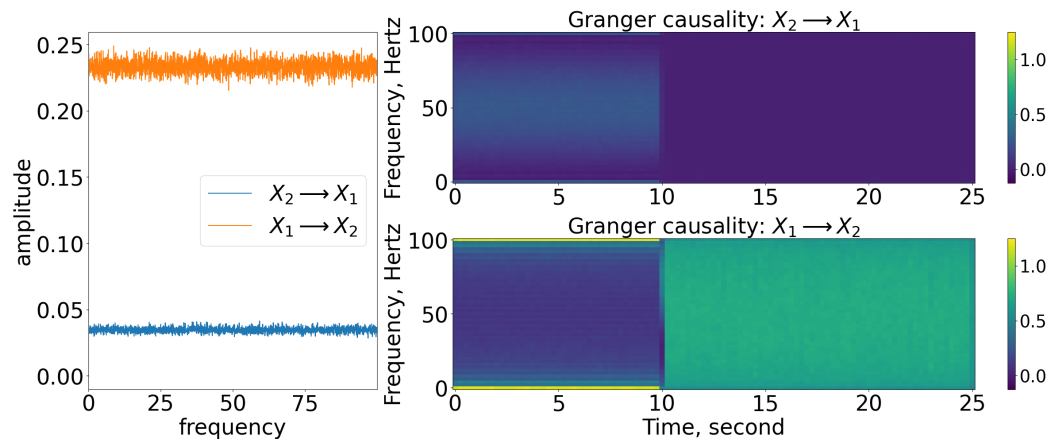
(f) nonlinear: small noise, noncouple

Figure 11. Result of simulation based on Equations (27) and (28) (refer to Figure 10 for graphical explanation of the process). The [top left, red] is the result of x_1 and [bottom, green] is the result of x_2 ; [right, blue] is the phase space of x_1 and x_2 . Note that in the context of noncoupling, it is referring to $H(\tau - 10) = 0$.

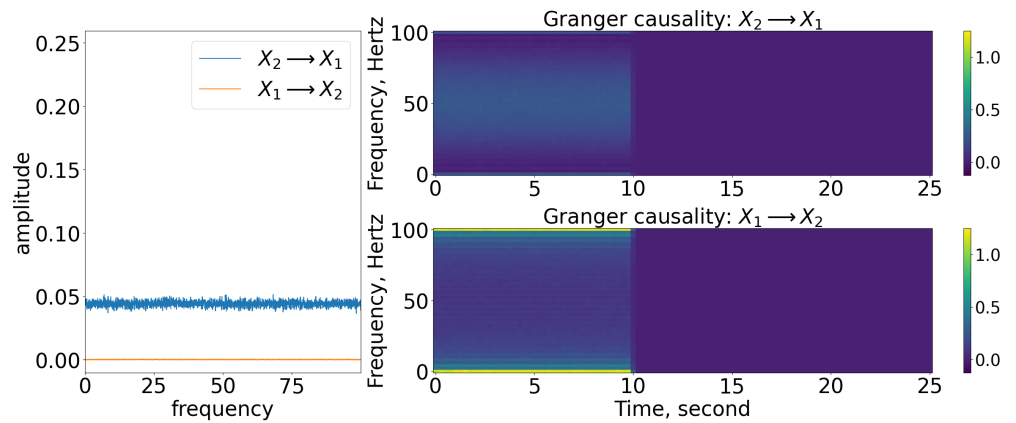
4.1. Granger Causality

Based on the models, the signal $x_2(t)$ will first cause $x_1(t)$ for the beginning of the process until the physical time of 10 s, and $x_1(t)$ causes $x_2(t)$ after the 10 s. This pattern is captured in the time-varying frequency of GC, as shown in Figure 12 for all the cases. Prior to 10 s, the $x_1(t)$ is caused by $x_2(t)$ by a factor of 0.55, and hence a similar factor of GC is shown in the time-varying frequency domain. Similarly, the coupling between the signals can also be observed in the frequency domain of GC. Notice that the [nonlinear:

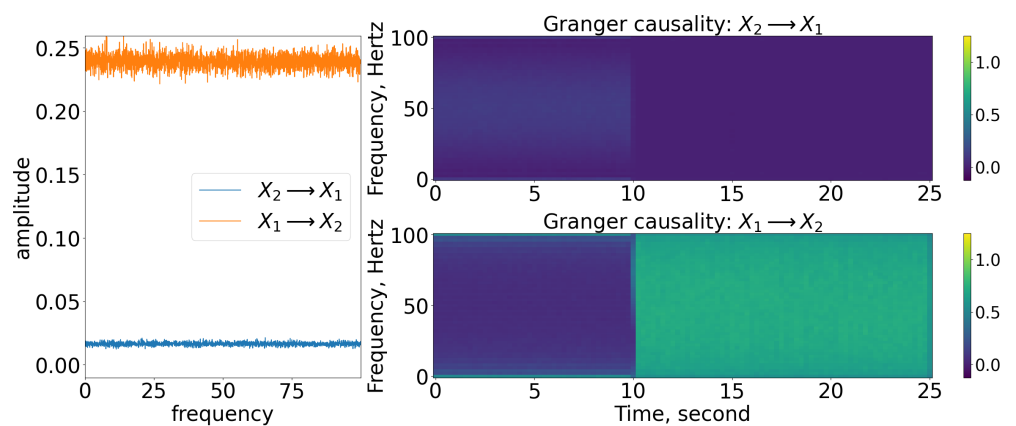
small noise] cases are able to show the significance of $x_2(t)$ causing $x_1(t)$ prior to 10 s because the small noise and the exponential term had $x_2(t)$ causing $x_1(t)$ increased to about 1.5 ($e^1 \times 0.55 \approx 1.5$).



(a) linear: couple



(b) linear: noncouple



(c) nonlinear: large noise, couple

Figure 12. Cont.

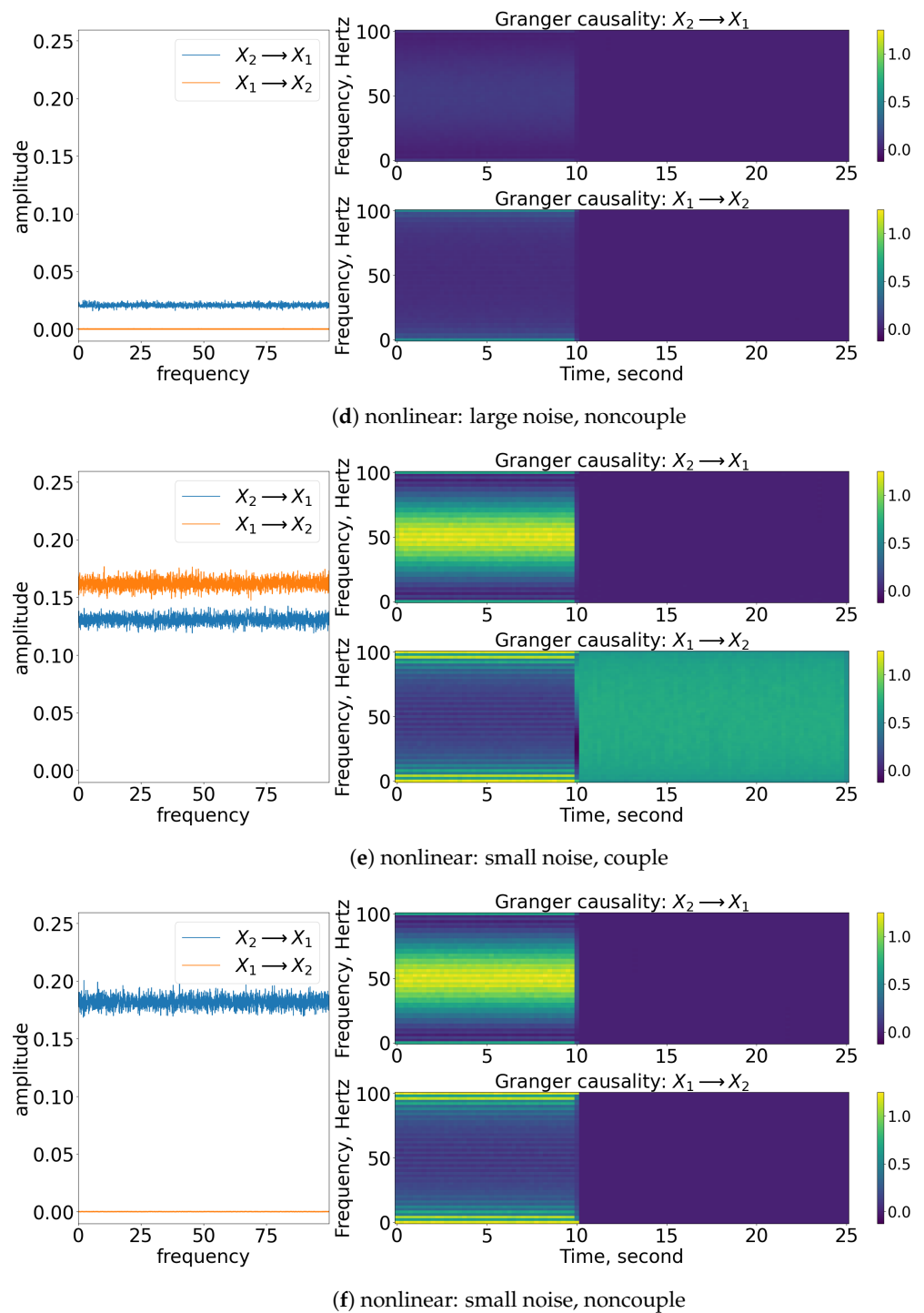
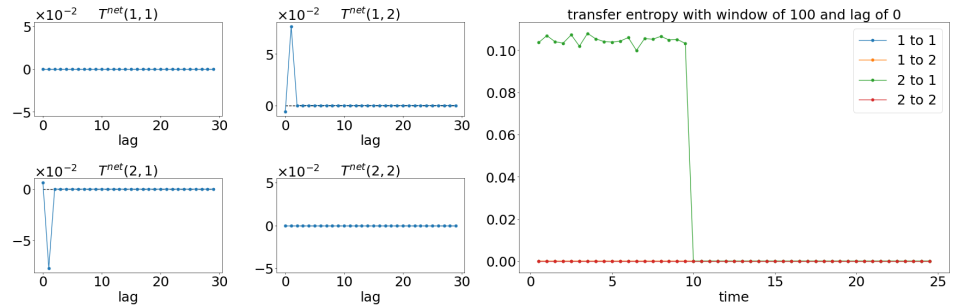


Figure 12. Result of the spectral and time-varying frequency of GC. Note that for each subfigure, the [left] shows the spectral (frequency domain) of GC, [blue] indicates x_2 causes x_1 , while [orange] indicates x_1 causes x_2 ; [top, right] shows the time-varying frequency of GC of x_2 to x_1 ; [bottom, right] shows the time-varying frequency of GC of x_1 to x_2 . The results shown are based on the processes in Equations (27) and (28). Refer to Figure 10 for the graphical explanation of the process. Note that the noncoupling is referring to $H(\tau - 10) = 0$ for all the cases.

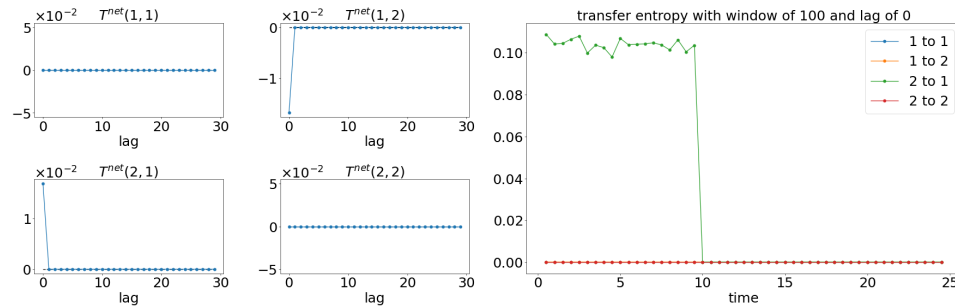
4.2. Transfer Entropy

TE is able to show the correction direction of the causality between the signals if the correct lag is chosen. Therefore, having models (like Equations (27) and (28)) with the coupling occurring at different time lags cannot properly show the direction of the causality via the window sliding TE. For instance, the window sliding TE shows that $x_2(t)$ is causing

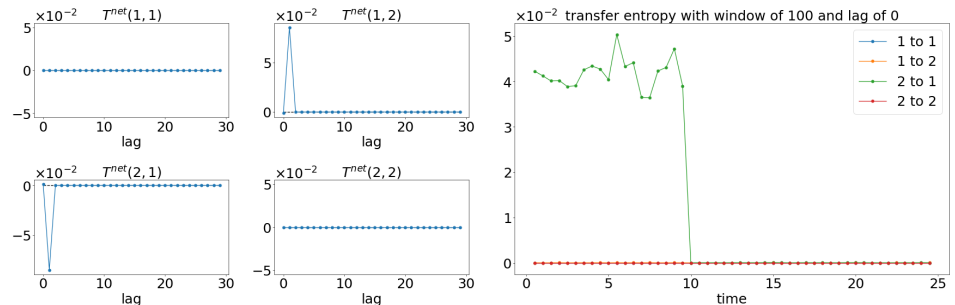
the $x_1(t)$ when lag 0 is chosen, while the reverse is shown when lag 1 is used, as shown in Figure 13 (for lag 0) and Figure 14 (for lag 1). This is depicted well in the underlining models (Equations (27) and (28)), in which $x_1(t)$ causes $x_2(t)$ one lag later. As a result, TE cannot show the whole causality of the signals by just one lag.



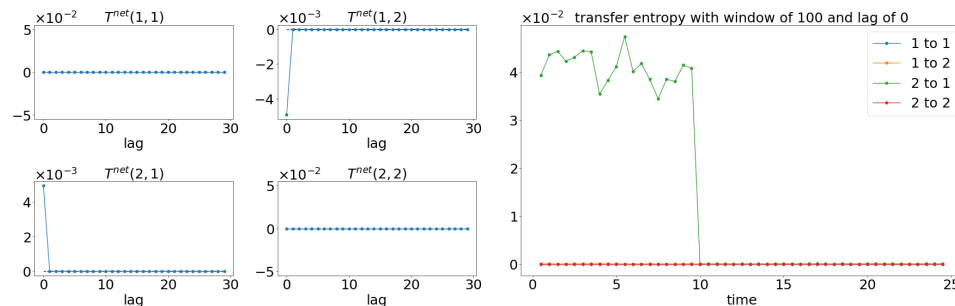
(a) linear: couple



(b) linear: noncouple

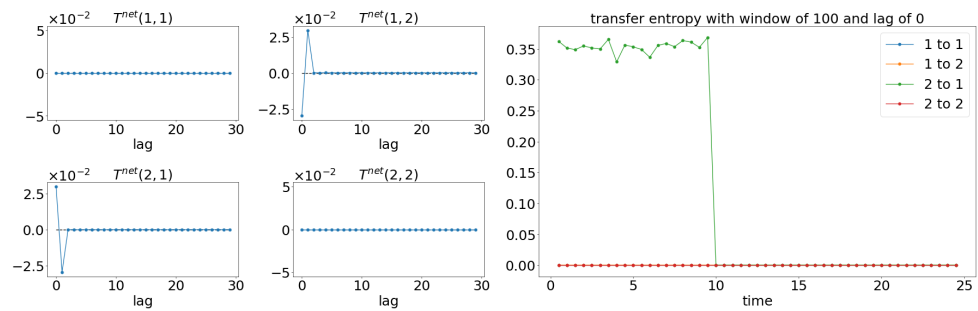


(c) nonlinear: large noise, couple

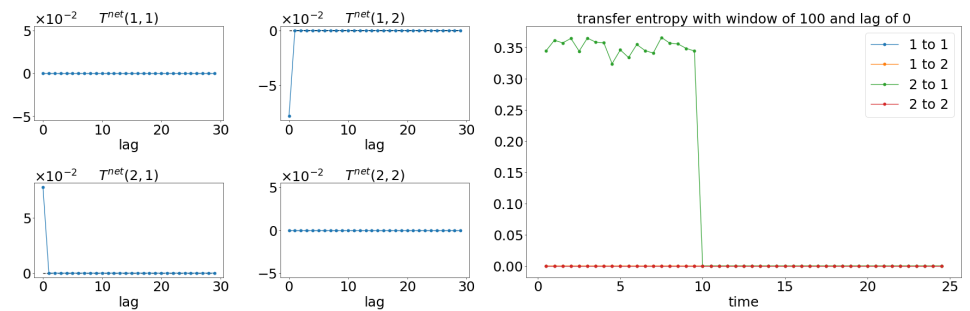


(d) nonlinear: large noise, noncouple

Figure 13. Cont.

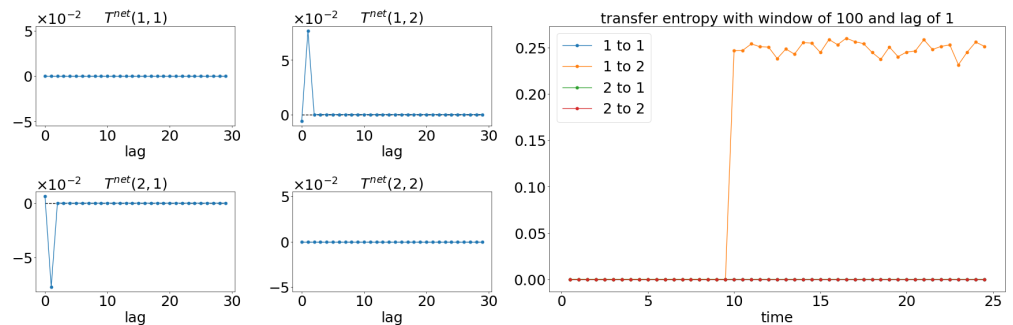


(e) nonlinear: small noise, couple

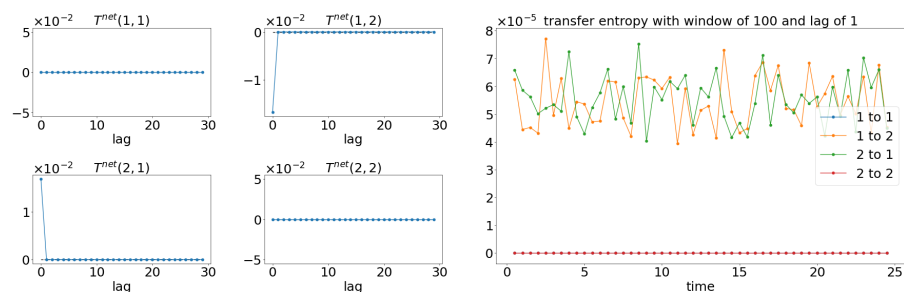


(f) nonlinear: small noise, noncouple

Figure 13. Result of the net TE and window sliding TE with the lag of 0. Note that for each subfigure, [left, 4×4] depicts the net TE for the whole signals, while the [right] shows the window sliding TE. The results shown are based on the processes in Equations (27) and (28). Refer to Figure 10 for the graphical explanation of the process. Note that the noncoupling is referring to $H(\tau - 10) = 0$ for all the cases.

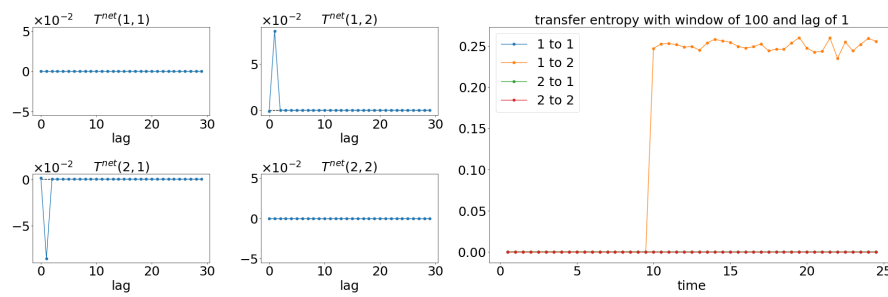


(a) linear: couple

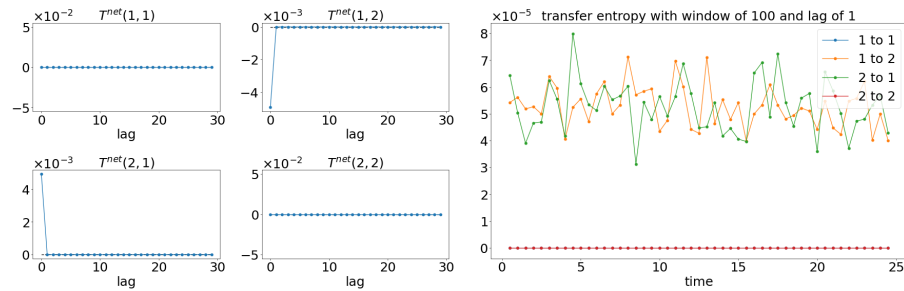


(b) linear: noncouple

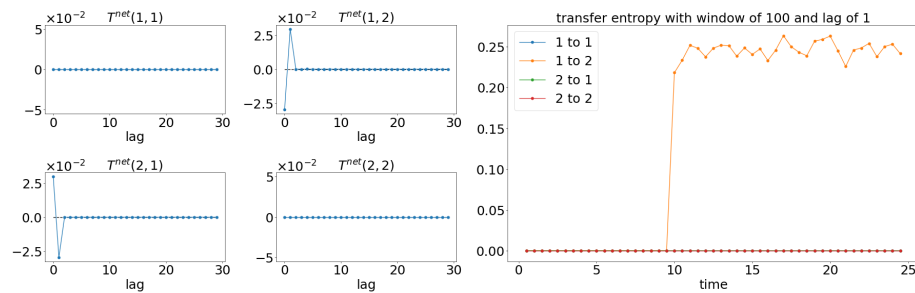
Figure 14. Cont.



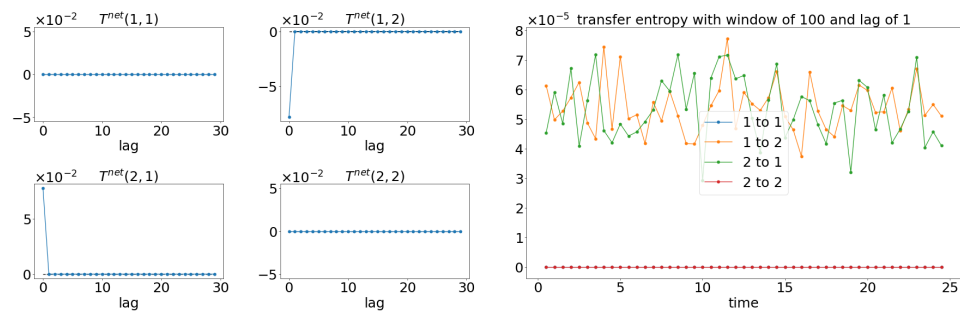
(c) nonlinear: large noise, couple



(d) nonlinear: large noise, noncouple



(e) nonlinear: small noise, couple



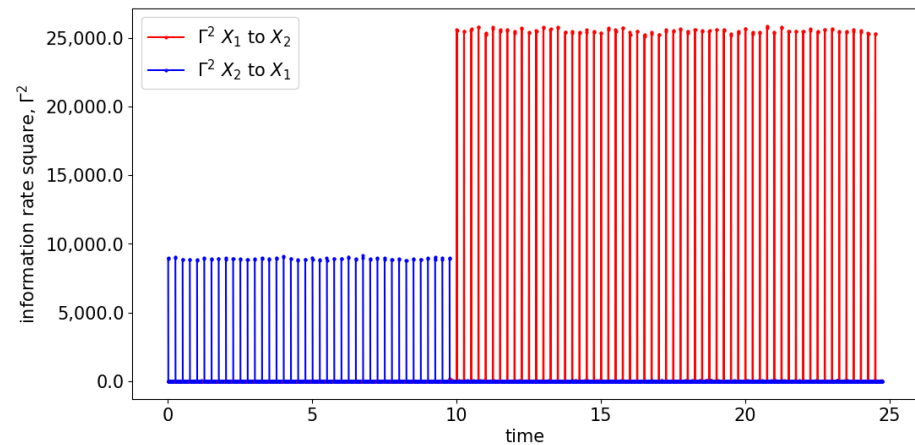
(f) nonlinear: small noise, noncouple

Figure 14. Result of the net TE and window sliding TE with the lag of 1. Note that for each subfigure, [left, 4 × 4] depicts the net TE for the whole signals, while the [right] shows the window sliding TE. The results shown are based on the processes in Equations (27) and (28). Refer to Figure 10 for the graphical explanation of the process. Note that the noncoupling is referring to $H(\tau - 10) = 0$ for all the cases.

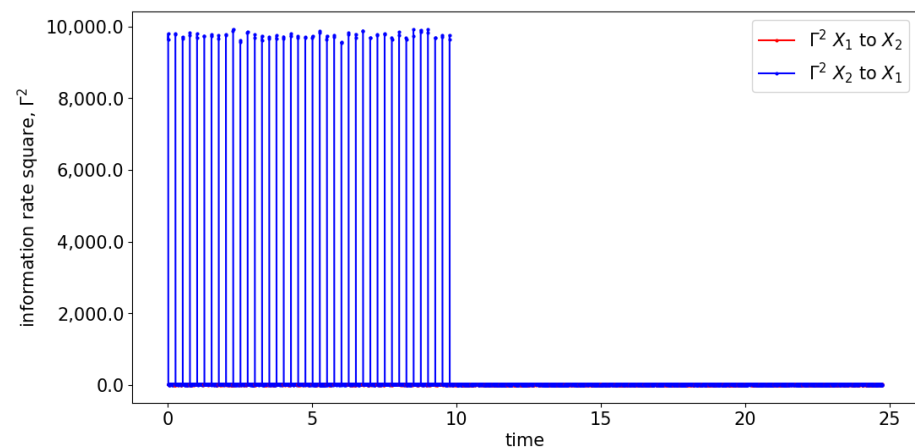
4.3. Information Rate Causality

Alternatively, the information rate causality is able to show the coupling of the signals along with the lag of coupling in each window of the evaluation. As discussed, the information rate causality evaluates the causality by observing the changes in the probability distribution of one signal after it is influenced by another signal. As shown in Figures 15 and 16, the information rate causality can captures the interchange of the coupling among the signals of the simulated models (Equations (27) and (28)). As discussed

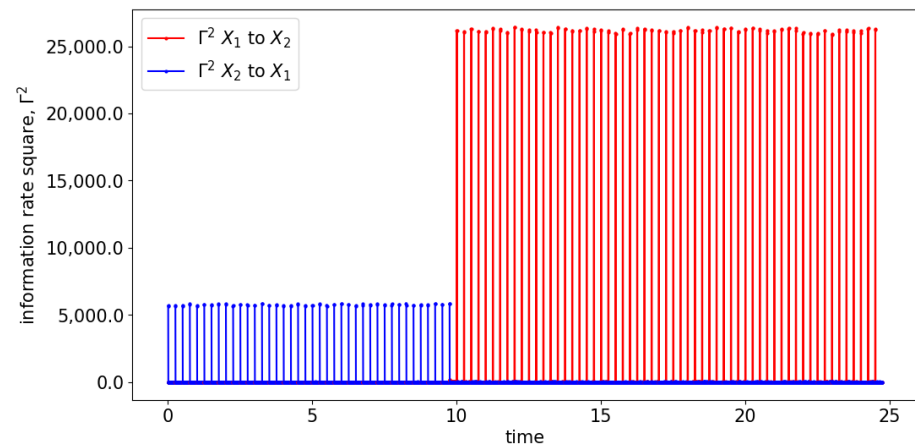
in TE, signal $x_2(t)$ is causing $x_1(t)$ at a lag of 0, and $x_1(t)$ causes $x_2(t)$ at a lag of 1. These lags can be observed at each window of the information rate evaluation. For instance, as shown in Figure 16, the information rate of 2 to 1 (Γ^2 2 to 1) has a spike at the beginning of the time, while the information rate of 1 to 2 (Γ^2 1 to 2) has a spike after one time step (one lag). Hence, the information rate causality is capable of showing the underlying causality of the signals.



(a) linear: couple

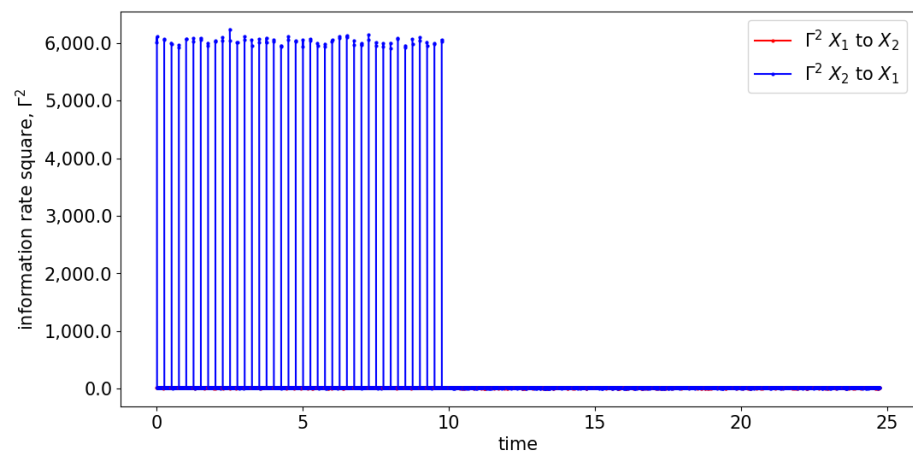


(b) linear: noncouple

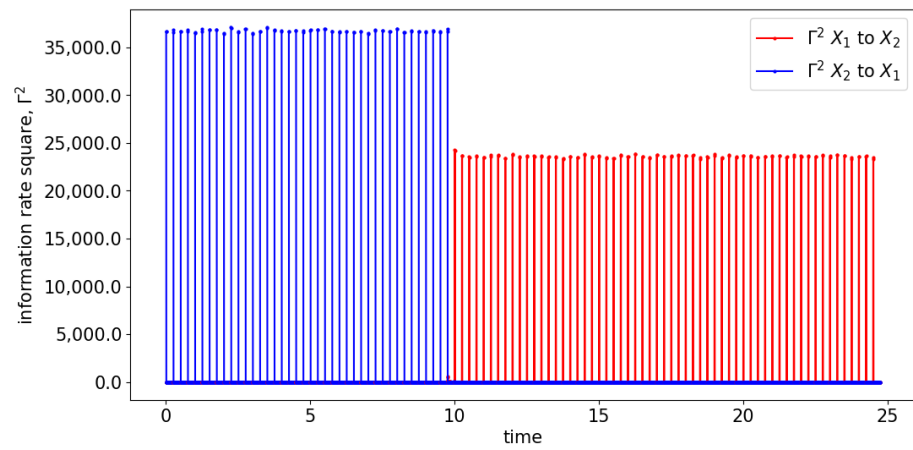


(c) nonlinear: large noise, couple

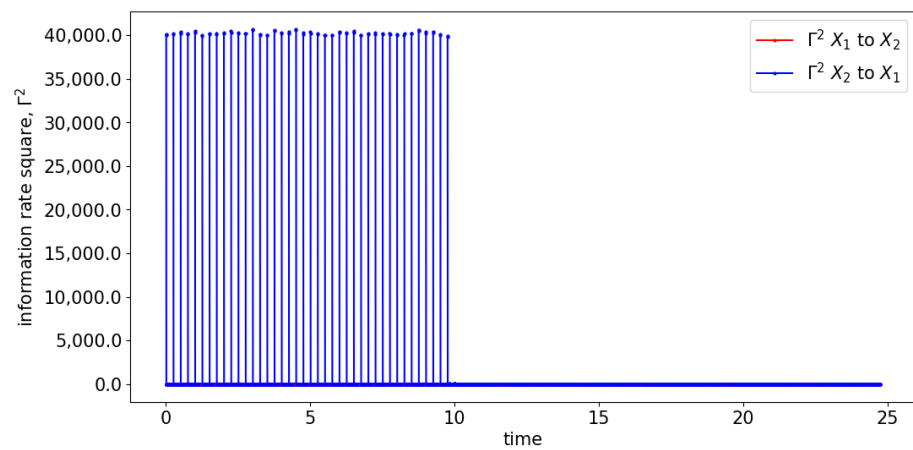
Figure 15. Cont.



(d) nonlinear: large noise, noncouple

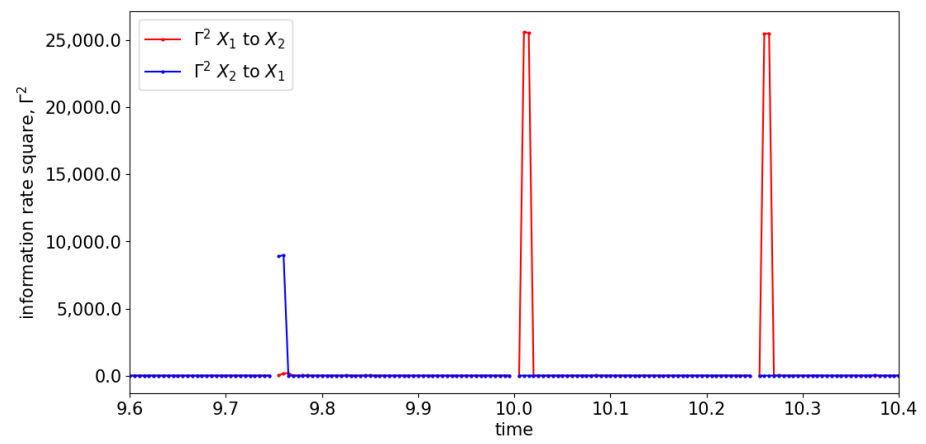


(e) nonlinear: small noise, couple

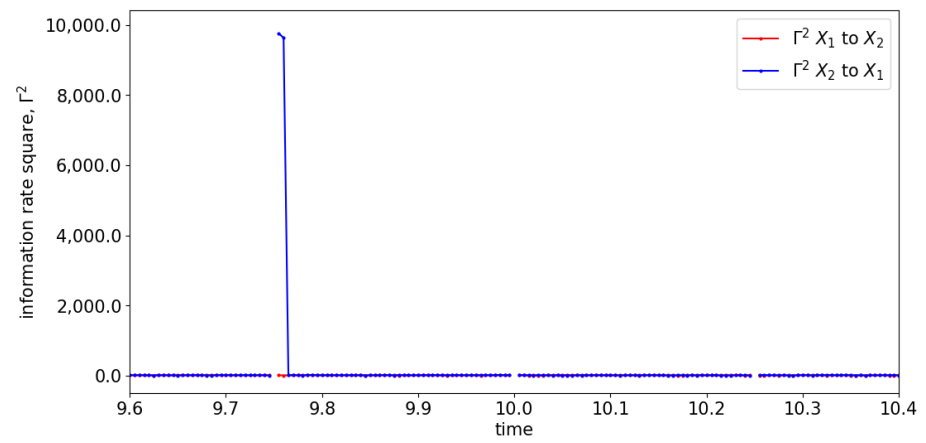


(f) nonlinear: small noise, noncouple

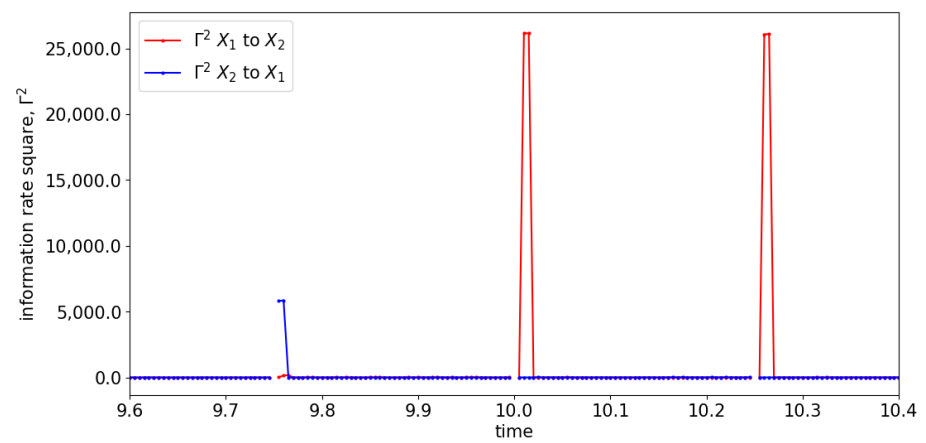
Figure 15. Result of information rate causality for Equations (27) and (28). Note that the interchange of the causality occurs at 10 s.



(a) linear: couple

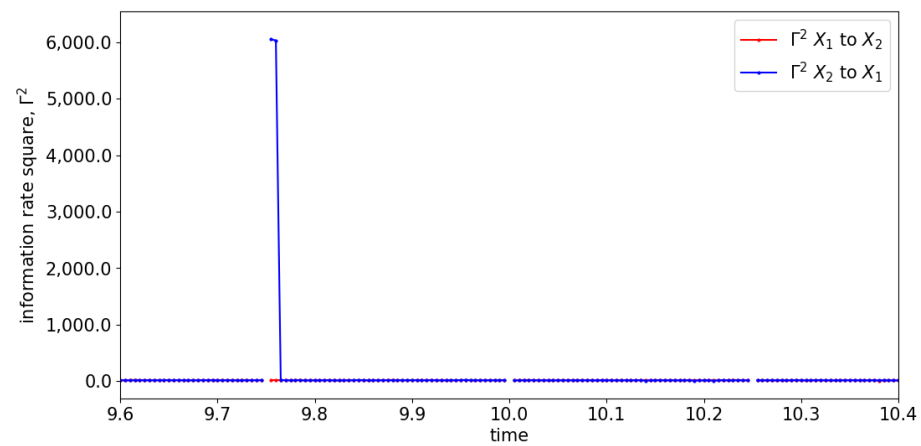


(b) linear: noncouple

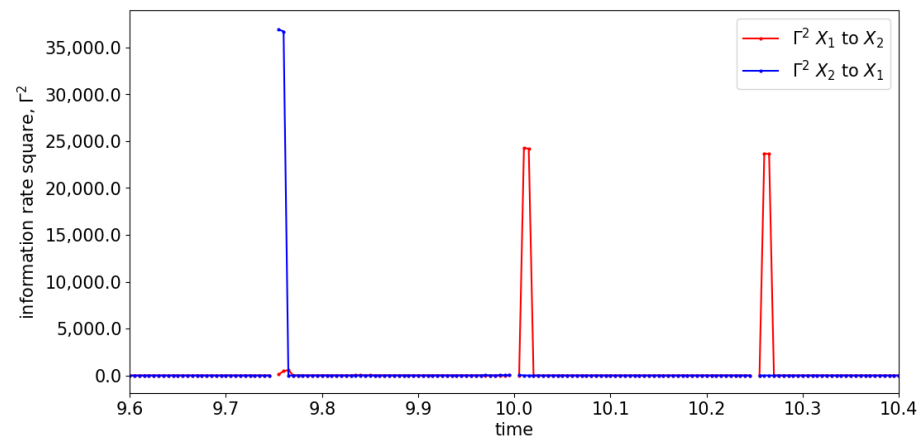


(c) nonlinear: large noise, couple

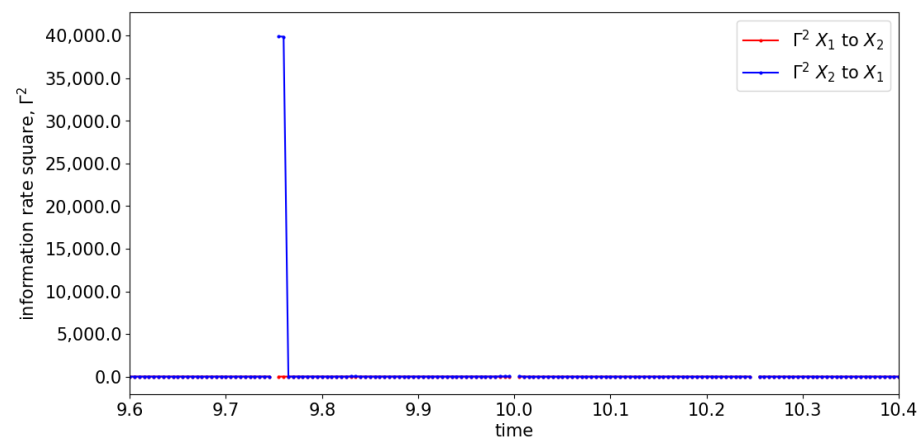
Figure 16. Cont.



(d) nonlinear: large noise, noncouple



(e) nonlinear: small noise, couple



(f) nonlinear: small noise, noncouple

Figure 16. Zoom-in of Figure 15 for the information rate causality between 9.6 s and 10.4 s. Note that the interchange of the causality occurs at 10 s.

5. Bidirectional Model

In this section, the bidirectional causality is studied by modifying the linear and nonlinear models from Section 3 as Equations (29) and (30), respectively. These modified models have the signals $x_2(t)$ continuously influencing $x_1(t)$ at all times $t \geq 1$. In addition, at time 10 s, the Heaviside step function ($H(\tau - 10)$) enables the coupling of $x_1(t)$ and

$x_2(t)$, which allows $x_1(t)$ to influence $x_2(t)$ while the previous coupling remains intact. The general structure of the Equations (29) and (30) is illustrated in Figure 17.

Linear model:

$$\begin{aligned} x_1(t) &= 0.55x_2(t-1) + \epsilon_1(t), \\ x_2(t) &= H(\tau - 10)x_1(t-2) + \epsilon_2(t). \end{aligned} \tag{29}$$

Nonlinear model:

$$\begin{aligned} x_1(t) &= 0.55x_2(t-1)e^{1-|x_2(t-1)|} + \epsilon_1(t), \\ x_2(t) &= H(\tau - 10)x_1(t-2) + \epsilon_2(t). \end{aligned} \tag{30}$$

Here, t is the time step and τ is the physical time; $H(\tau - 10)$ is the Heaviside step function that allows the coupling of the signal starting at a physical time of 10 s. Similar to the previous model, 6 cases are simulated and the signals are used to evaluate through GC, TE, and information rate causality. Recall that the noncoupling cases here are referring to the decoupling of $x_1(t)$ to $x_2(t)$ ($H(\tau - 10) = 0$).

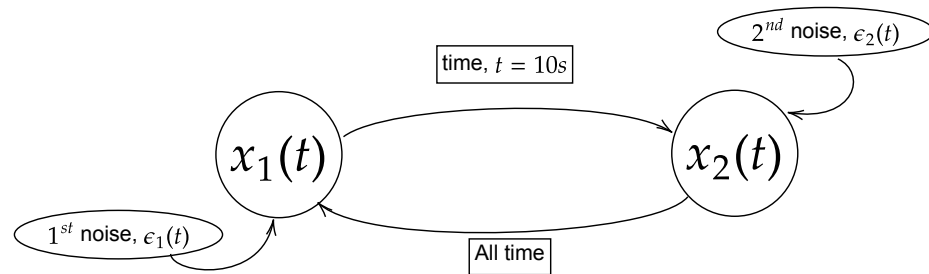
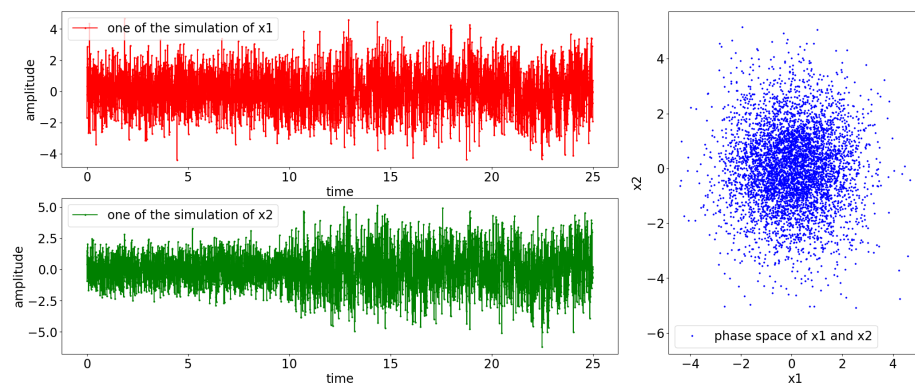


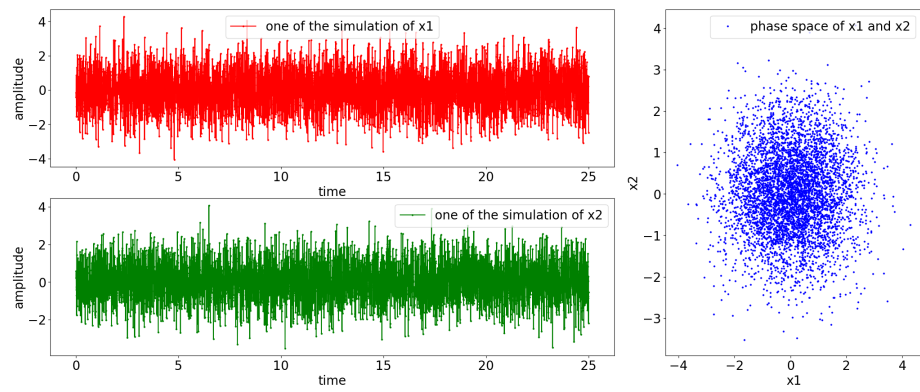
Figure 17. Model of the flow of information of the processes $x_1(t)$ and $x_2(t)$ for Equations (29) and (30). In this paper, the equations are simulated with physical time of 25 s and sampling frequency of 200 Hz (5000 realizations) with either large noise or small noise. Note that the $x_2(t)$ coupled with $x_1(t)$ throughout the process and the occurrence of bidirectional causality happens after 10 s. In the context of noncoupling, it is referring to $H(\tau - 10) = 0$.

The simulation result from all aforementioned cases is shown in Figure 18. From the observation of the individual signal and the phase space, the scenario of [nonlinear: small noise, couple] has non-stationary oscillation after the coupling, as shown in Figure 18e. This is due to the divergence caused by the nonlinear exponential term embedded within the simulated equation ($e^{1-|x_2(t-1)|} = e^{\mathbb{R}}$) along with the factor of 0.55 of the previous state of $x_2(t)$ ($0.55x_2(t-1) \dots$). On the other hand, [nonlinear: small noise, noncouple] has a stationary oscillation, as shown in Figure 18f; this is because $x_1(t)$ only receives the input from $x_2(t)$, which keeps the power of the exponential term negative and hence results in a stationary oscillation.

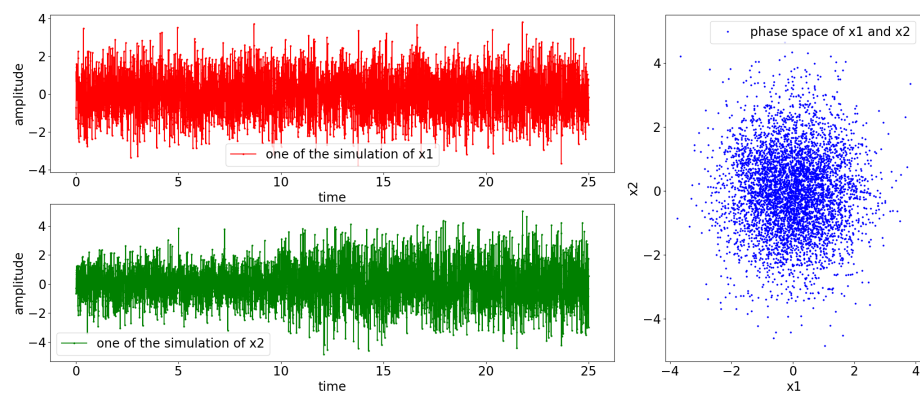


(a) linear: couple

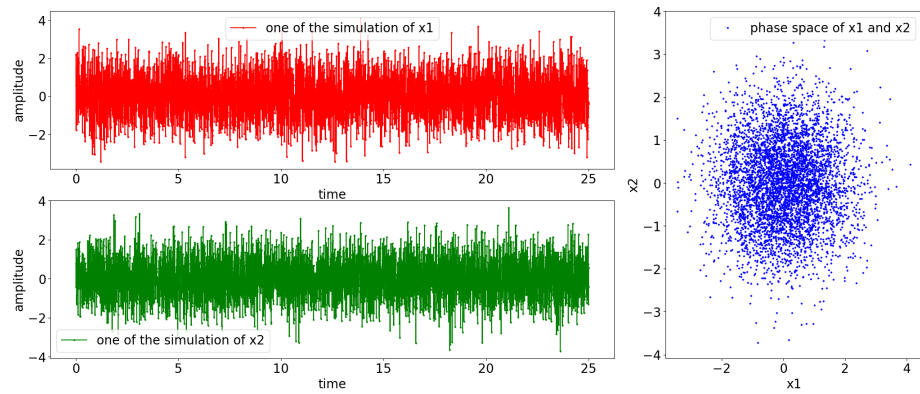
Figure 18. Cont.



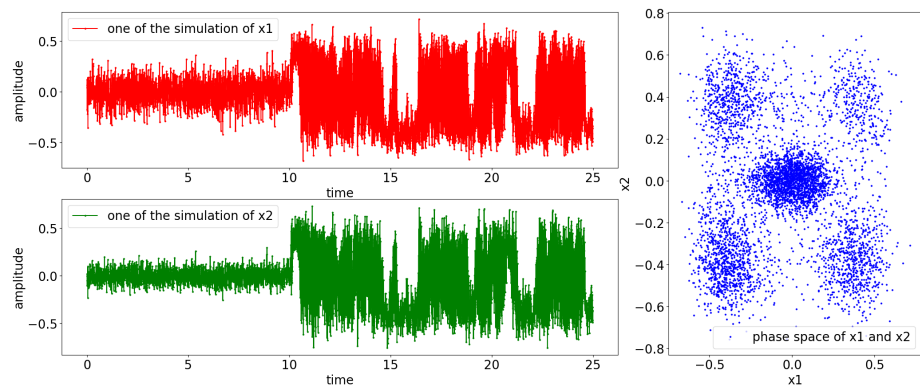
(b) linear: noncouple



(c) nonlinear: large noise, couple

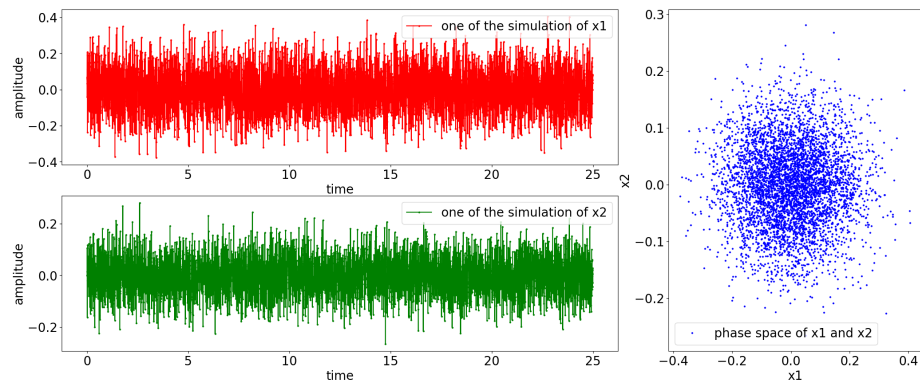


(d) nonlinear: large noise, noncouple



(e) nonlinear: small noise, couple

Figure 18. Cont.

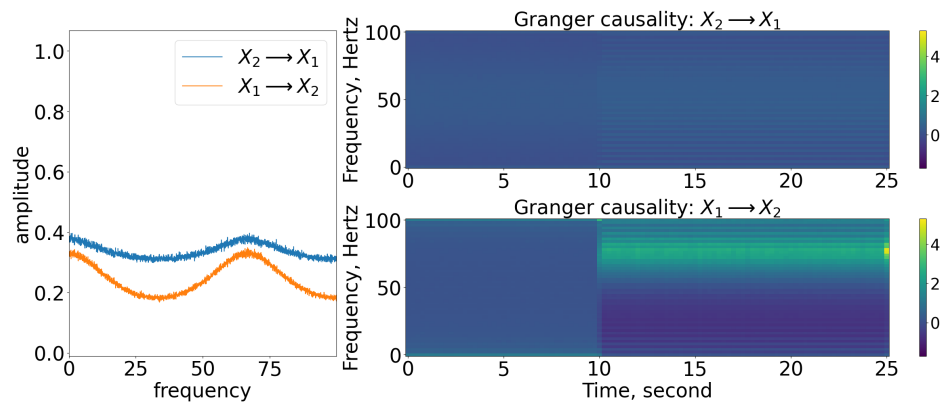


(f) nonlinear: small noise, noncouple

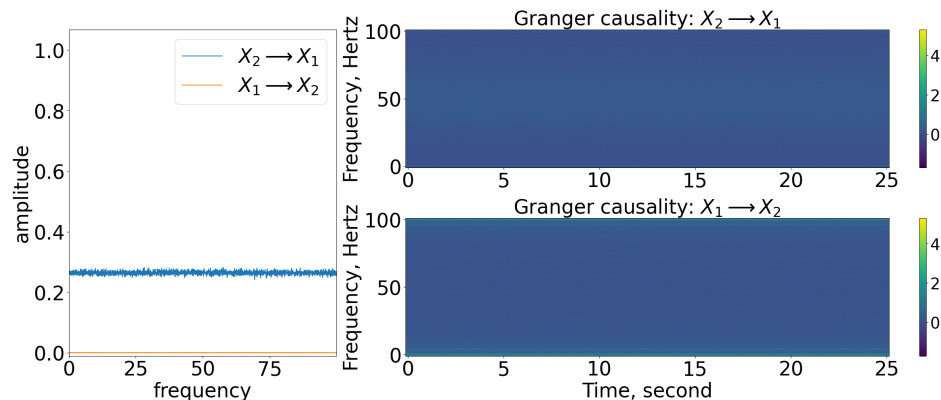
Figure 18. Result of simulation based on Equations (29) and (30) (refer to Figure 17 for graphical explanation of the process). The [top left, red] is the result of x_1 ; [bottom, green] shows the result of x_2 ; [right, blue] shows the phase space of x_1 and x_2 . Note that the noncoupling in the context is referring to $H(\tau - 10) = 0$.

5.1. Granger Causality

From the GC analysis, the coupling from $x_2(t)$ to $x_1(t)$ cannot be captured well in it except for the [nonlinear: small noise] cases. This result is due to a similar reason as discussed in Section 4, i.e., the coupling between the signals is affected by a factor of 1.5 ($e^1 \times 0.55 \approx 1.5$). Therefore, the rest of the cases cannot show the significant contrast of the causality from [$x_2(t)$ to $x_1(t)$] as shown in Figure 19. Furthermore, referring to the time-varying GC, the bidirectional coupling/causality between $x_1(t)$ and $x_2(t)$ that occurs after 10 s are not able to capture all the coupling cases.

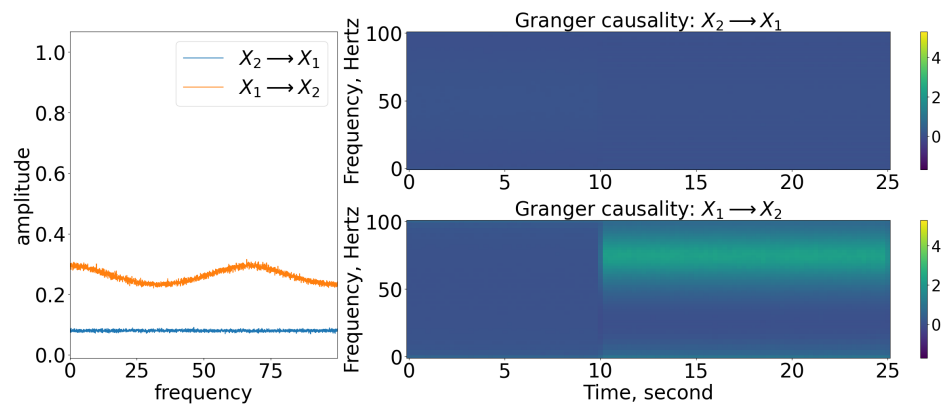


(a) linear: couple

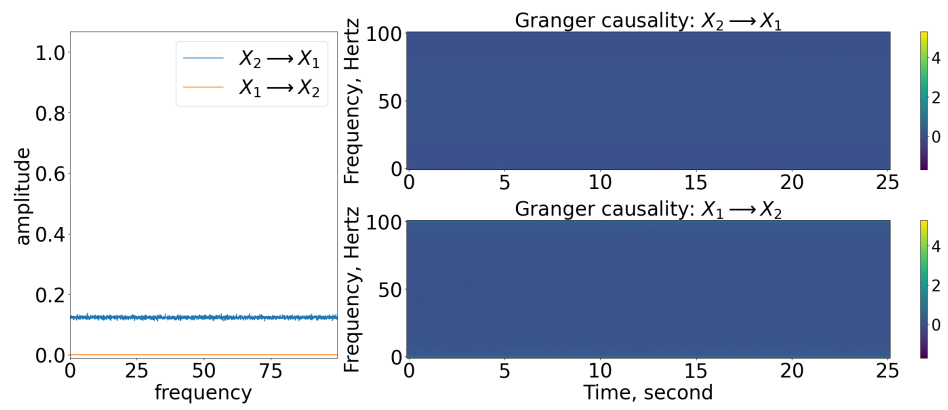


(b) linear: noncouple

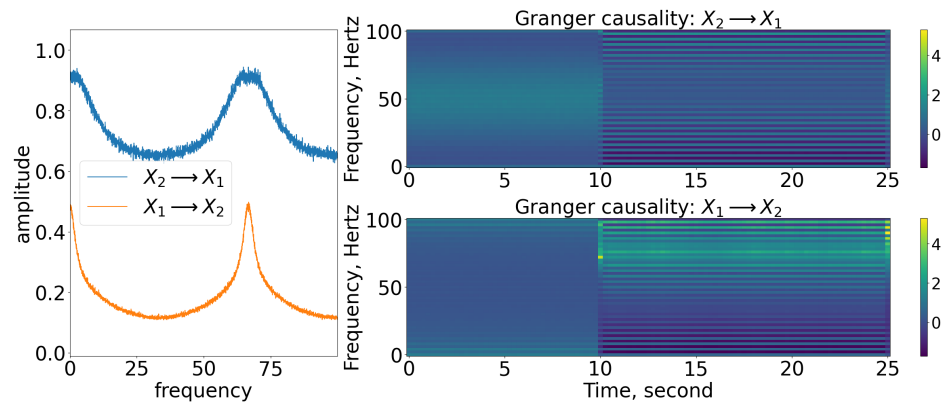
Figure 19. Cont.



(c) nonlinear: large noise, couple

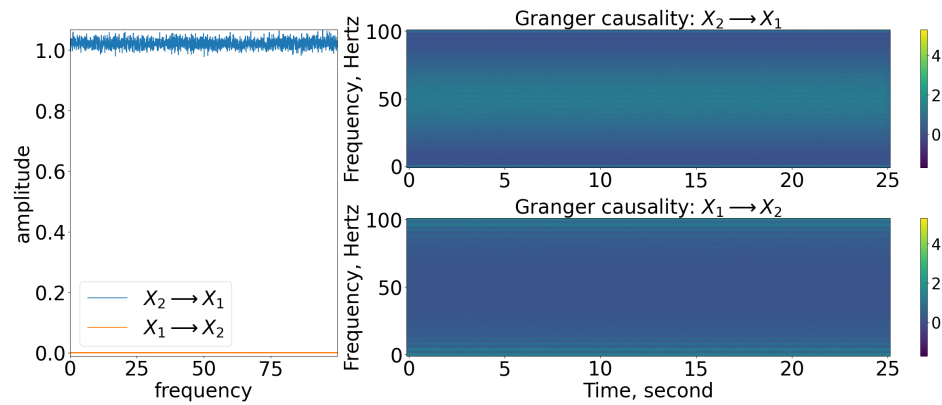


(d) nonlinear: large noise, noncouple



(e) nonlinear: small noise, couple

Figure 19. Cont.

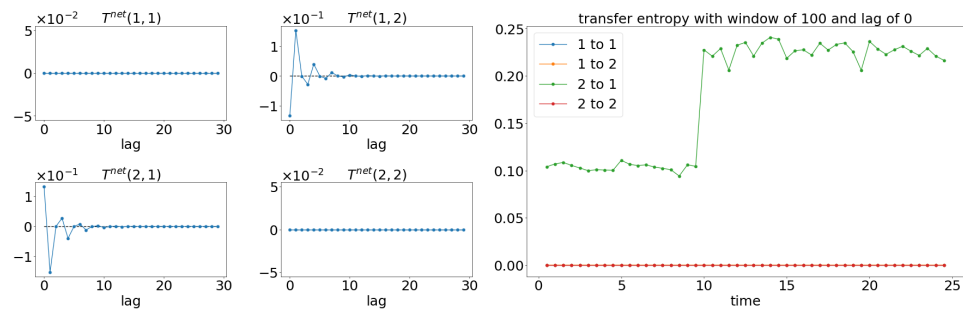


(f) nonlinear: small noise, noncouple

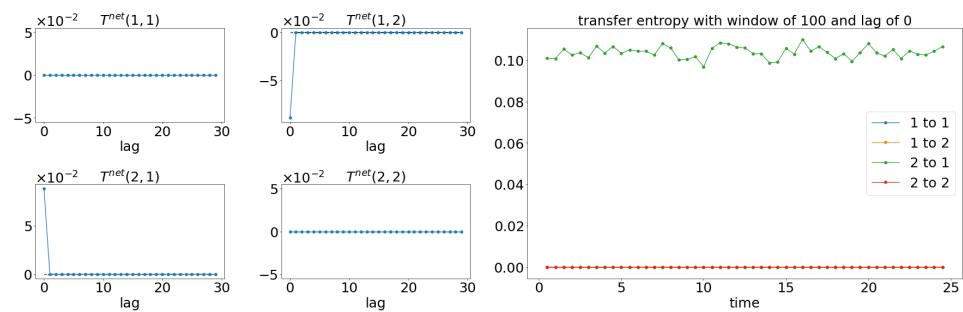
Figure 19. Result of the spectral and time-varying frequency of GC. Within each subfigure, the [left] shows the spectral (frequency domain) of GC, with [blue] showing that x_2 causes x_1 ($I_{x_2 \rightarrow x_1}$) and [orange] indicating that x_1 causes x_2 ($I_{x_1 \rightarrow x_2}$). The [top, right] figure shows that the time-varying frequency of GC of x_2 causes x_1 and [bottom, right] shows that x_1 causes x_2 . The results shown are based on the expression in Equations (29) and (30). Refer to Figure 17 for the graphical explanation of the process. Note that the noncoupling is referring to $H(\tau - 10) = 0$ for all the cases.

5.2. Transfer Entropy

The results of TE shown in Figures 20 and 21 once again are quite similar to those shown in Section 4. The window sliding TE, evaluated with lag 0, shows that signal $x_2(t)$ causes $x_1(t)$ throughout the time, as shown in Figure 20. For the coupling cases, the amplitude of the TE changes when $x_1(t)$ feedbacks to $x_2(t)$ (after time 10 s). Similar to Section 4, the window sliding TE only shows that signal $x_1(t)$ is causing $x_2(t)$ when evaluated at lag 1 for the coupling cases shown in Figure 21. Hence, TE cannot fully capture the causality between the signals with just one lag.

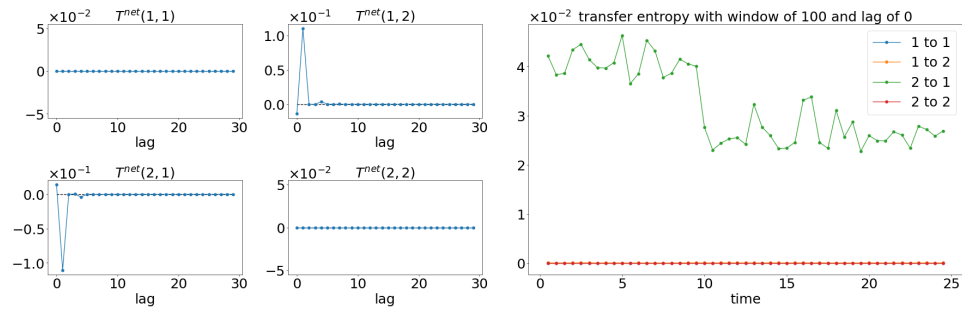


(a) linear: couple

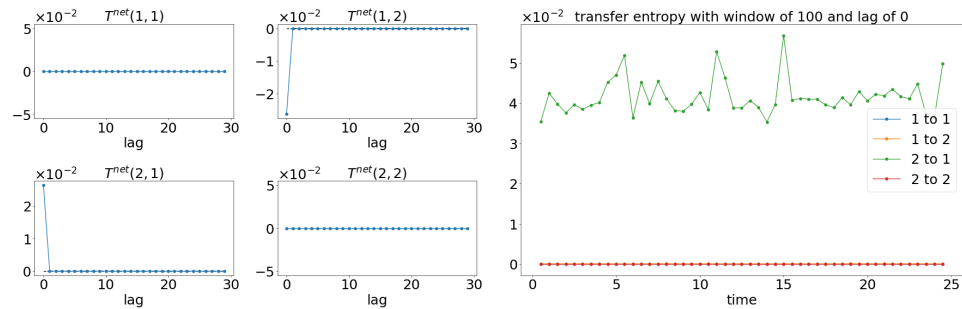


(b) linear: noncouple

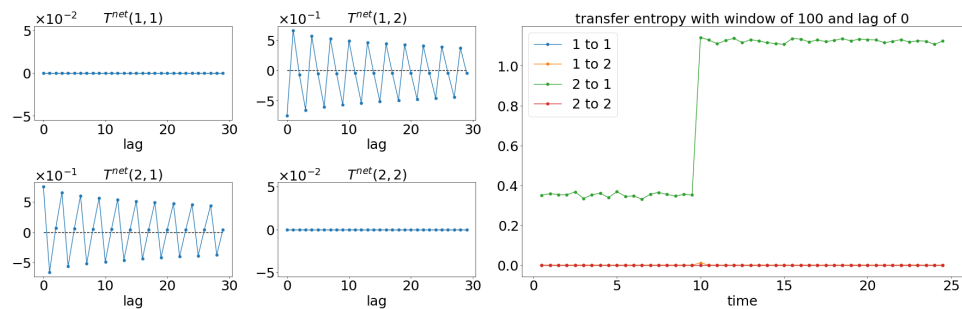
Figure 20. Cont.



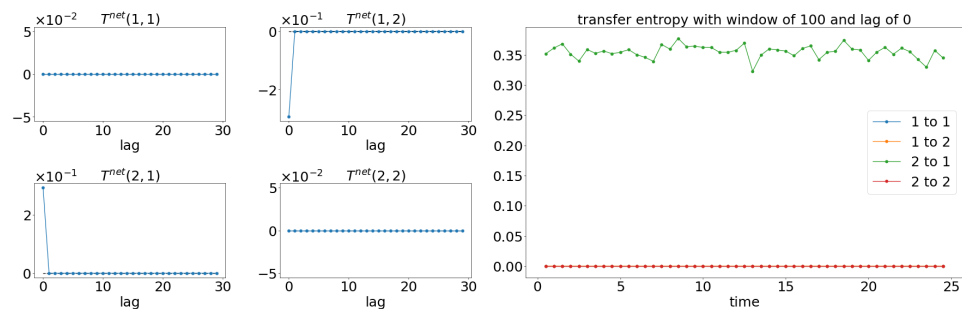
(c) nonlinear: large noise, couple



(d) nonlinear: large noise, noncouple

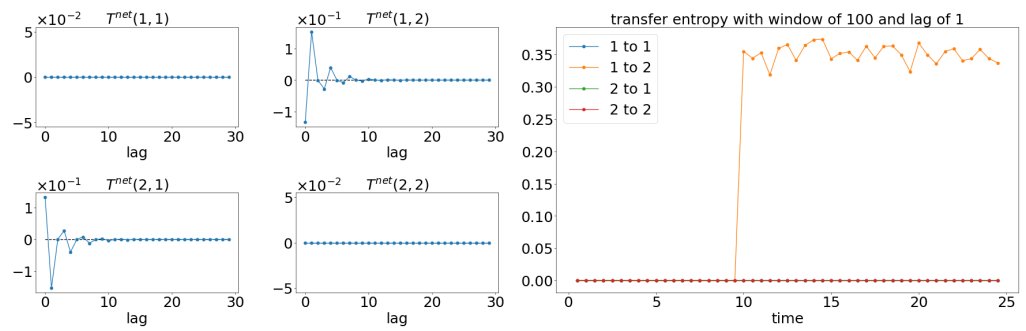


(e) nonlinear: small noise, couple

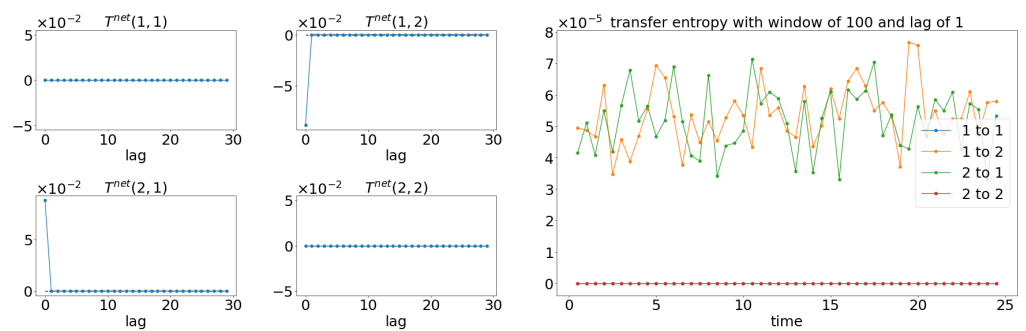


(f) nonlinear: small noise, noncouple

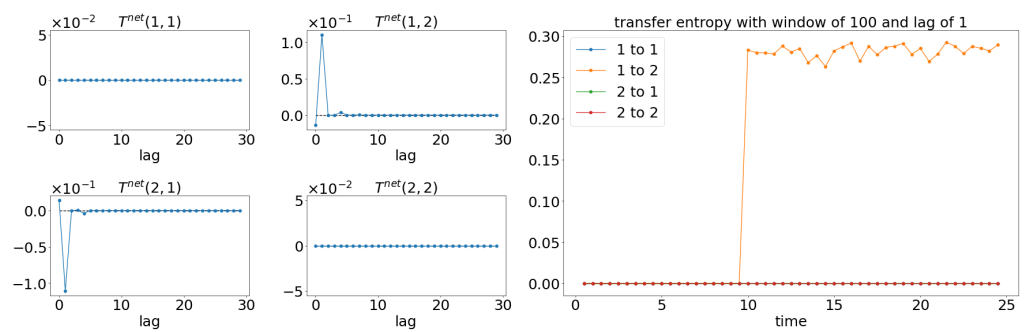
Figure 20. Result of the net TE and window sliding TE with the lag at 0. Within each subfigure, the [left, 4 × 4] figures show the net TE for the whole signals, while the [right] shows the window sliding TE. The results shown are based on the expression in Equations (29) and (30). Refer to Figure 17 for the graphical explanation of the process. Note that the noncoupling is referring to $H(\tau - 10) = 0$ for all the cases.



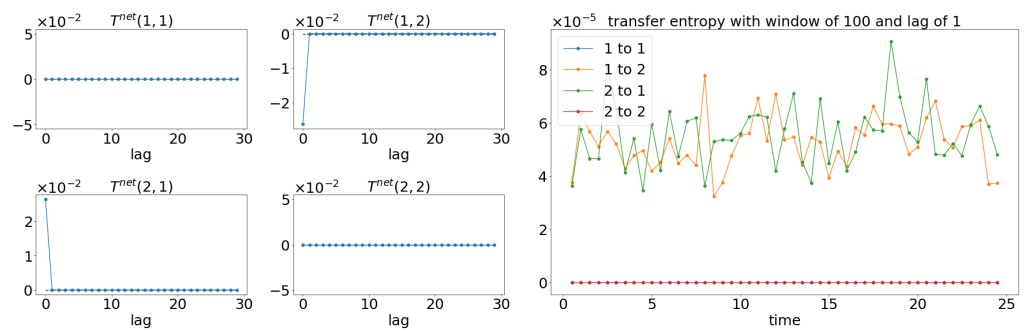
(a) linear: couple



(b) linear: noncouple

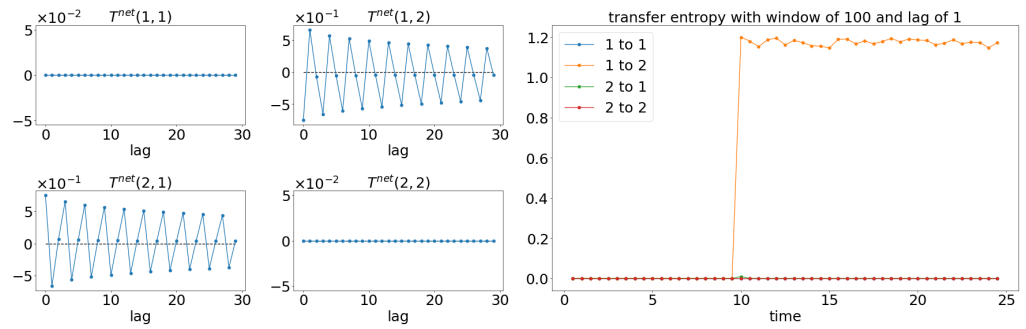


(c) nonlinear: large noise, couple

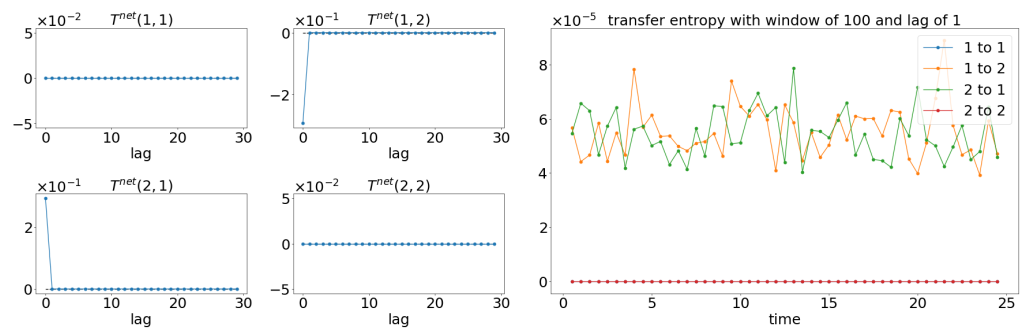


(d) nonlinear: large noise, noncouple

Figure 21. Cont.



(e) nonlinear: small noise, couple

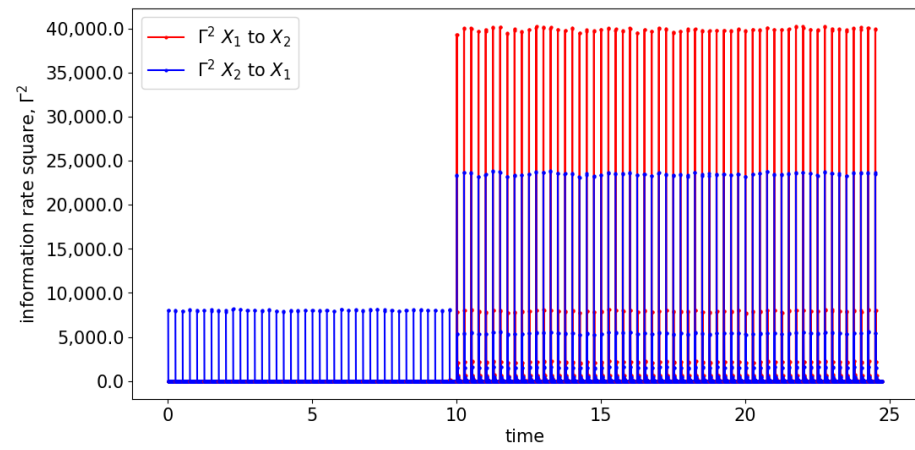


(f) nonlinear: small noise, noncouple

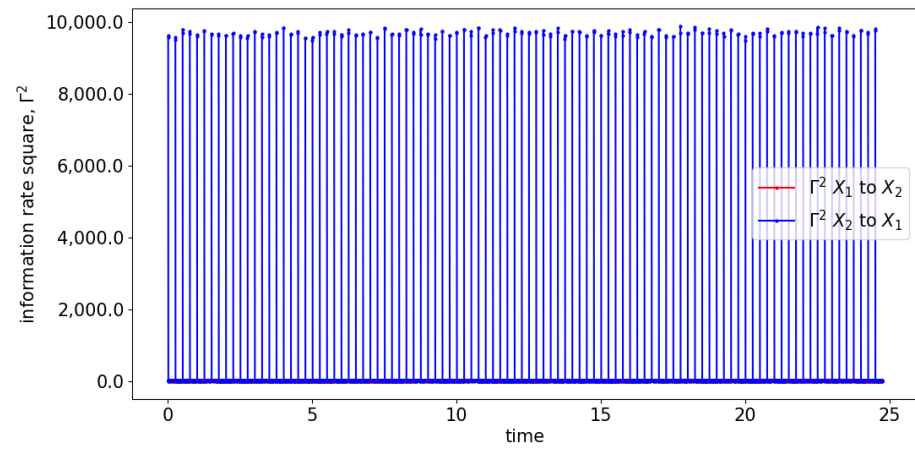
Figure 21. Result of the net TE and window sliding TE with the lag at 1. Within each subfigure, the [left, 4 × 4] figures show the net TE for the whole signals, while the [right] shows the window sliding TE. The results shown are based on the expression in Equations (29) and (30). Refer to Figure 17 for the graphical explanation of the process. Note that the noncoupling is referring to $H(\tau - 10) = 0$ for all the cases.

5.3. Information Rate Causality

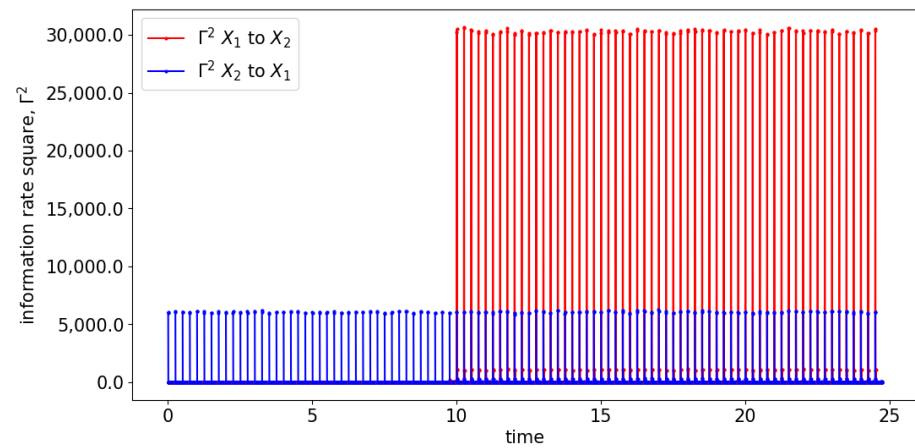
The information rate causality analysis evaluates the causality through the change in probability distribution of an original signal after including another signal assuming it causes the original signal. The information rate causality is able to show the coupling among the signals well, as shown in Figures 22 and 23. Similar to Section 4, prior to time 10 s, the causality from signal $x_2(t)$ to $x_1(t)$ can be depicted with the presence of the peak of Γ^2 2 to 1. The peak presented at the beginning of the evaluation window suggests that the coupling occurs at lag 0. After 10 s, the mutual feedback between the signals is shown within the evaluation window of the information rate causality. This mutual causality between the signals is observed through the alternating oscillations of the peak occurrence in the information rate within the window. Hence, this can reflect well the underlying model of the signals. Note that this is different than Section 4, which is an interchange of unidirectional causality that has the feedback solely from one signal to the other without retaining information from itself.



(a) linear: couple

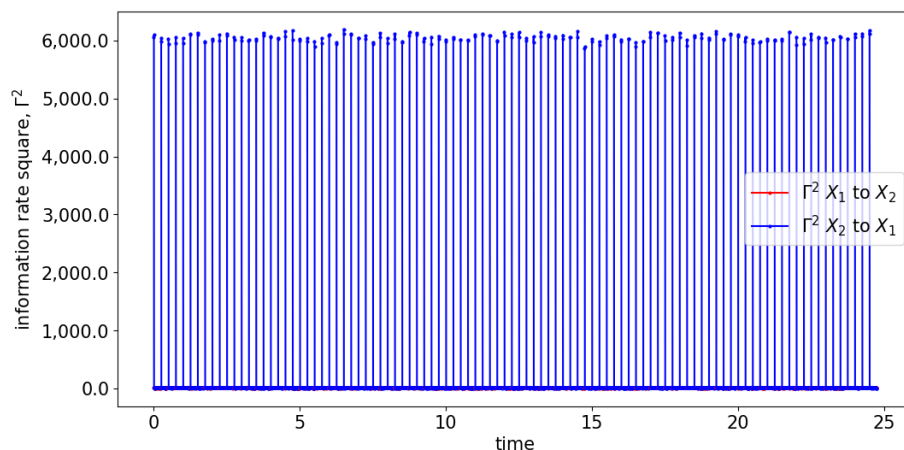


(b) linear: noncouple

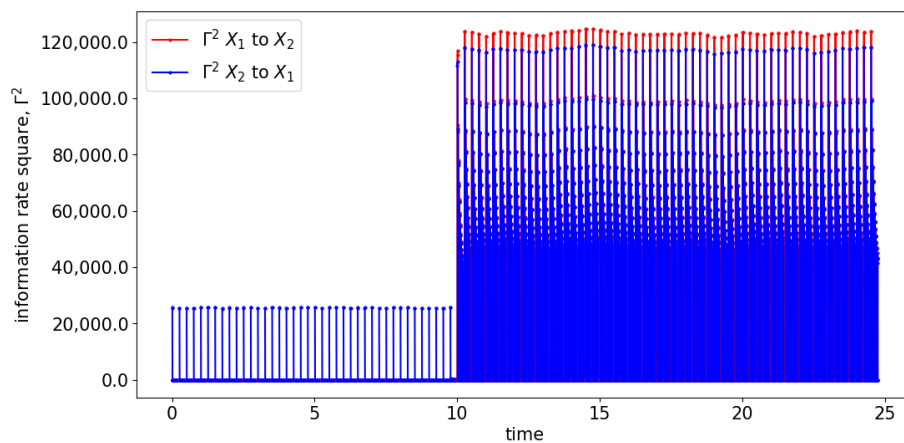


(c) nonlinear: large noise, couple

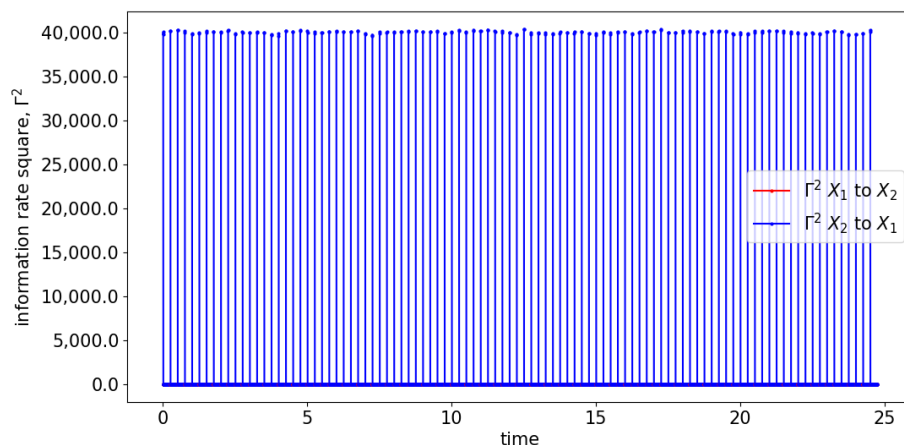
Figure 22. Cont.



(d) nonlinear: large noise, noncouple

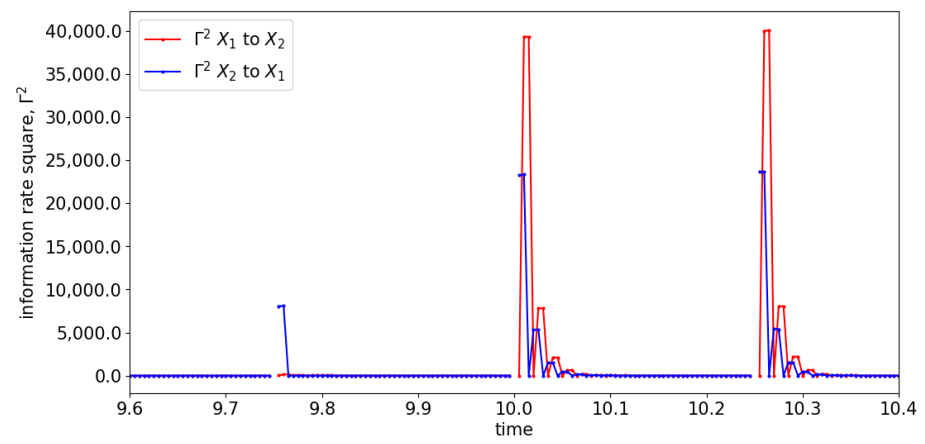


(e) nonlinear: small noise, couple

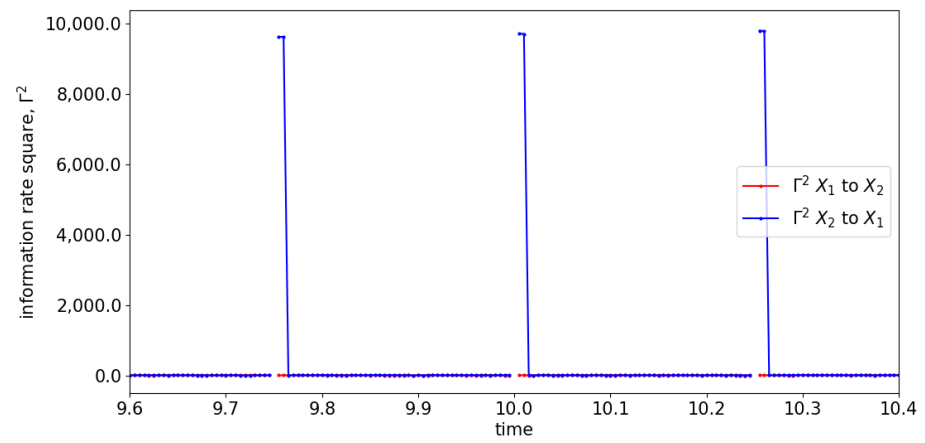


(f) nonlinear: small noise, noncouple

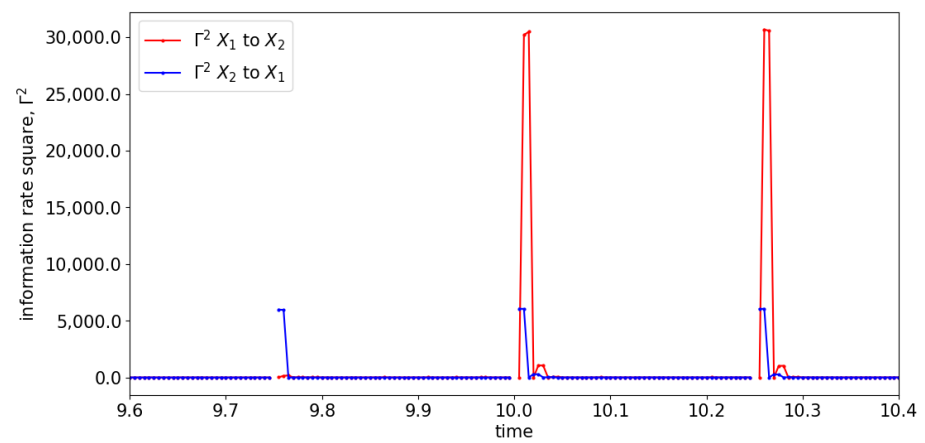
Figure 22. Result of information rate causality for Equations (29) and (30). Note that the bidirectional causation begins at 10 s. The results shown are based on the expression in Equations (29) and (30). Refer to Figure 17 for the graphical explanation of the process. Note that the noncoupling is referring to $H(\tau - 10) = 0$ for all the cases.



(a) linear: couple

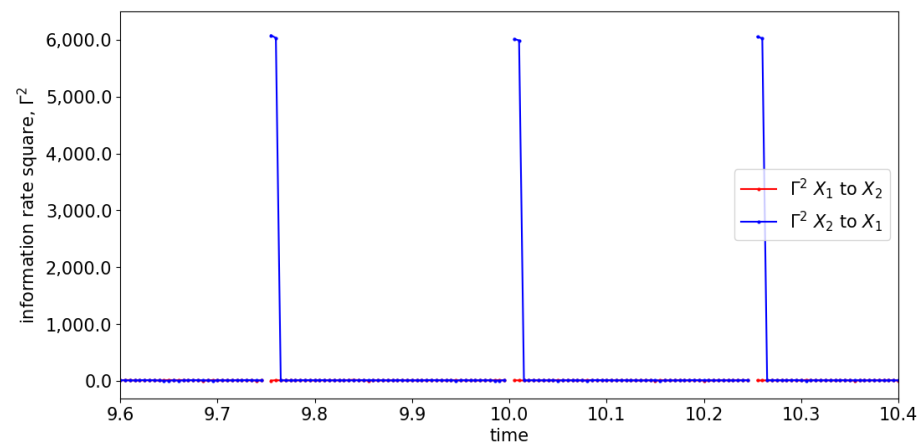


(b) linear: noncouple

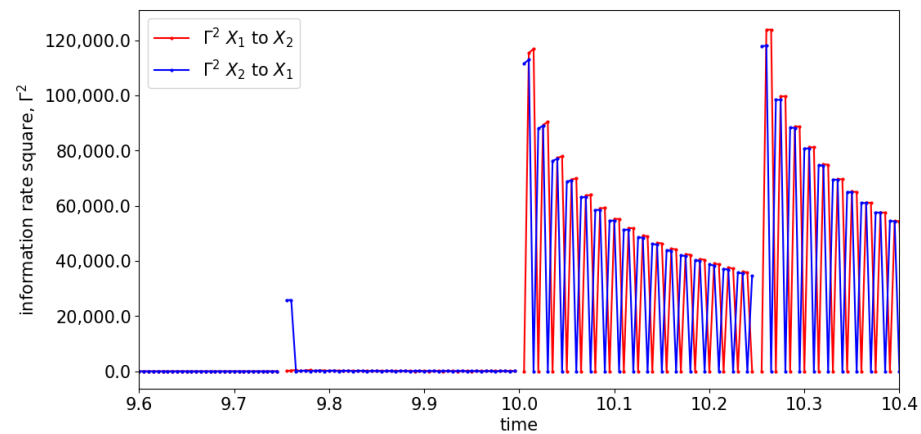


(c) nonlinear: large noise, couple

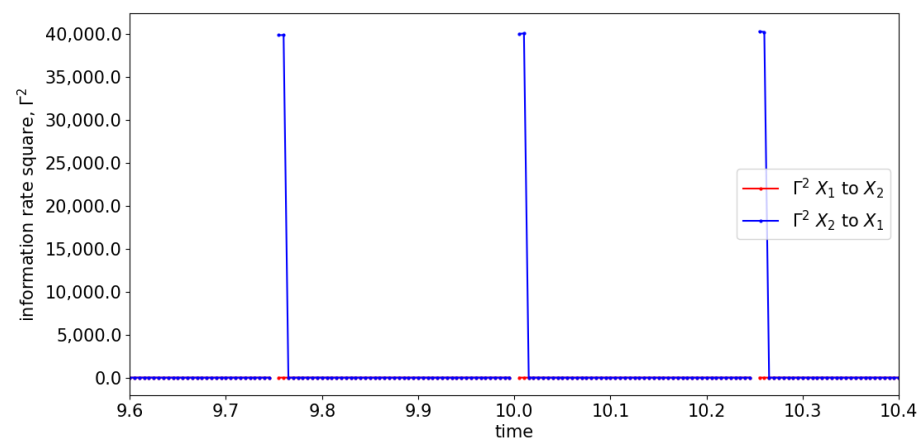
Figure 23. Cont.



(d) nonlinear: large noise, noncouple



(e) nonlinear: small noise, couple



(f) nonlinear: small noise, noncouple

Figure 23. Zoom-in of Figure 22 for the information rate causality between 9.6 s and 10.4 s. Note that the bidirectional causation begins at 10 s. The results shown are based on the expression in Equations (29) and (30). Refer to Figure 17 for the graphical explanation of the process. Note that the noncoupling is referring to $H(\tau - 10) = 0$ for all the cases.

6. Conclusions

In this paper, we proposed an alternative causal analysis method to the popular GC and TE for causality quantification. Specifically, based on information rate, which measures the rate of temporal change in the time-dependent distribution, we developed a model-free,

information-geometric measure of causality—information rate causality—that is suitable for analyzing numerically generated non-stationary, nonlinear data. To compare the GC, TE, and information rate causality, we applied the methods to numerical data generated by simulating different types of discrete autoregressive models which contain linear and nonlinear interactions. In Section 3, we showed that information rate causality performs equally well compared to GC in the standard linear stationary signals. Later, we showed that the GC did not perform well for non-stationary and/or nonlinear signals. Furthermore, it failed to capture mutual feedback between the signals, as shown specifically in Section 5 in all the coupling cases. TE performed slightly better than GC, since it is model-independent. However, the information on the lag is needed to properly evaluate the TE, as shown in Sections 4 and 5. In comparison with GC and TE, information rate causality has shown to be able to uncover the underlying mechanism of causality in the signals, such as the interchanging oscillatory feedback between the signals that is discussed in Section 5, which was not captured by either GC or TE.

While our results in this paper were obtained based on the interaction of two time series of different types of (non)linearity and (non)stationary, they have at least pointed out some limitations of GC and TE that can be resolved by employing information rate causality. It remains as future work to extend this work and to investigate a larger class of time series data, including real data such as EEG signals.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/e25050806/s1>, Figure S1: Estimation of the joint probability distribution via histogram.

Author Contributions: Conceptualization, E.-j.K.; Methodology, E.-j.K., F.H. and H.J.C.; software, H.J.C.; Writing—original draft, H.J.C.; Writing—review & editing, E.-j.K.; Supervision, E.-j.K. and F.H. All authors have read and agreed to the published version of the manuscript.

Funding: E.-j.K. and F.H. are partly supported by EPSRC Grant EP/W036770/1.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

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

Appendix A. Granger Causality

Appendix A.1. Autoregressive Process

Given a stochastic process $(x(t))$, it can be modelled through an autoregressive (AR) model with available/appropriate n number of observations from the process (e.g., AR(n) known as autoregressive of n th order). The general expression of the AR(n) model is expressed as Equation (A1).

$$x(t) = \sum_{i=1}^n \alpha_i x(t-i) + \epsilon(t), \quad (\text{A1})$$

such that t is the time step, and α_i is the constant coefficient which suggested the weight of contribution from the i th past observation towards the process $x(t)$. $\epsilon(t)$ is the additive noise that has the variance Σ , which indicated the proportion of the external influence towards the process. Hence, a process that is strongly related to past observations will have small noise (e.g., small Σ).

Appendix A.2. Granger Causality: Time Domain

Given two stochastic processes $(x_1(t))$ and $(x_2(t))$, they can be modelled independently, as shown in Equation (A1) and expressed as Equations (A2) and (A3).

$$x_1(t) = \sum_{i=1}^n a_i x_1(t-i) + \epsilon_1(t); \tag{A2}$$

$$x_2(t) = \sum_{i=1}^n c_i x_2(t-i) + \epsilon_2(t), \tag{A3}$$

such that $\epsilon_1(t)$ and $\epsilon_2(t)$ are the noises with variances of Σ_{11} and Σ_{22} , respectively; a_i and c_i are the constant parameters that suggested the fractions of the contribution from the past observations towards $x_1(t)$ and $x_2(t)$, respectively.

If there is a present causal relationship between two processes, they can be modelled with the past values from each other, as shown in Equations (A4) and (A5).

$$x_1(t) = \sum_{i=1}^n a_i x_1(t-i) + \sum_{i=1}^n b_i x_2(t-i) + \epsilon_1^*(t), \tag{A4}$$

$$x_2(t) = \sum_{i=1}^n c_i x_1(t-i) + \sum_{i=1}^n d_i x_2(t-i) + \epsilon_2^*(t). \tag{A5}$$

Here $a_i, b_i, c_i,$ and d_i are the constants parameters. The noises ($\epsilon_1^*(t)$ and $\epsilon_2^*(t)$) would have the variance presented as the covariance matrix Σ shown in Equation (A6):

$$\Sigma = \begin{bmatrix} \Sigma_{11}^* & \Sigma_{12}^* \\ \Sigma_{21}^* & \Sigma_{22}^* \end{bmatrix}. \tag{A6}$$

Similar to Equations (A2) and (A3), Σ_{11}^* and Σ_{22}^* are the variance of the $\epsilon_1^*(t)$ and $\epsilon_2^*(t)$, respectively, while $\Sigma_{12}^* = \Sigma_{21}^*$ is the covariance of the noises ($\epsilon_1^*(t)$ and $\epsilon_2^*(t)$).

The causal relationship between $x_1(t)$ and $x_2(t)$ can be determined by evaluating the noises from Equations (A2) to (A5). For instance, $\Sigma_{11}^* < \Sigma_{11}$ would indicate that the predictability of $x_1(t)$ increases as the past information of $x_2(t)$ is included. Hence, this would suggest that there is a causal relationship from $x_2(t)$ to $x_1(t)$ which is known as GC ($x_2(t)$ Granger-causes $x_1(t)$, $F_{x_2 \rightarrow x_1}$). Similarly, $\Sigma_{22}^* < \Sigma_{22}$ would suggest $x_1(t)$ Granger-causes $x_2(t)$, $F_{x_1 \rightarrow x_2}$. The covariance of the noises $\Sigma_{12}^* = \Sigma_{21}^*$ would suggest that the processes are correlated by one another via a common external force that drives the processes; consequently, it shows an instantaneous causality ($F_{x_1 \leftrightarrow x_2}$).

The idea behind GC is to evaluate the total interdependence (also known as total causality index) between $x_1(t)$ and $x_2(t)$, which consists of two directional causalities ($F_{x_1 \rightarrow x_2}$ and $F_{x_2 \rightarrow x_1}$) and one instantaneous causality ($F_{x_1 \leftrightarrow x_2}$). The total interdependence between $x_1(t)$ and $x_2(t)$ is defined as Equation (A7).

$$F_{x_1 \cdot x_2} = \log \frac{\Sigma_{11} \Sigma_{22}}{|\Sigma|} = \log \frac{\Sigma_{11} \Sigma_{22}}{\Sigma_{11}^* \Sigma_{22}^* - (\Sigma_{12}^*)^2} = F_{x_1 \rightarrow x_2} + F_{x_2 \rightarrow x_1} + F_{x_1 \leftrightarrow x_2}, \tag{A7}$$

such that $|\Sigma|$ is the determinant of the covariance matrix Σ . If either one of the signals Granger-causes or correlates to one another, the value of $F_{x_1 \cdot x_2} > 0$ because the $\Sigma_{11}^* < \Sigma_{11}$, $\Sigma_{22}^* < \Sigma_{22}$, or $\Sigma_{12}^* \neq 0$. On the contrary, $F_{x_1 \cdot x_2} = 0$ indicates there is no GC relationship and correlation between the signals.

The directional causalities and instantaneous causality are defined as Equations (A8) to (A10).

$$F_{x_1 \rightarrow x_2} = \log \frac{\Sigma_{11}}{\Sigma_{11}^*}, \tag{A8}$$

$$F_{x_2 \rightarrow x_1} = \log \frac{\Sigma_{22}}{\Sigma_{22}^*}, \tag{A9}$$

$$F_{x_1 \leftrightarrow x_2} = \log \frac{\Sigma_{11}^* \Sigma_{22}^*}{|\Sigma|}. \tag{A10}$$

Similar to Equation (A7), the individual causalities from Equations (A8) to (A10) have the range of $[0, \infty)$, with $F(\dots) = 0$ indicating the absence of causality relationship.

Appendix A.3. Granger Causality: Frequency Domain

The spectral of the GC is studied at the frequency domain of the AR model. Unlike the time domain, the frequency domain of GC required both the constant parameters and the variance of the noises for the evaluation. To understand the formulation, the lag operator (L^k) is used and defined as Equation (A11).

$$L^k x(t) = x(t - k). \tag{A11}$$

Hence, by applying the lag operator expressed in Equation (A11) to Equations (A4) and (A5), they can be rewritten as

$$\begin{cases} x_1(t) = (\sum_{i=1}^n a_i L^i) x_1(t) + (\sum_{i=1}^n b_i L^i) x_2(t) + \epsilon_1^*(t), \\ x_2(t) = (\sum_{i=1}^n c_i L^i) x_1(t) + (\sum_{i=1}^n d_i L^i) x_2(t) + \epsilon_2^*(t), \end{cases} \tag{A12}$$

$$\begin{cases} (1 - \sum_{i=1}^n a_i L^i) x_1(t) + (-\sum_{i=1}^n b_i L^i) x_2(t) = \epsilon_1^*(t), \\ (-\sum_{i=1}^n c_i L^i) x_1(t) + (1 - \sum_{i=1}^n d_i L^i) x_2(t) = \epsilon_2^*(t), \end{cases} \tag{A13}$$

$$\begin{cases} \begin{bmatrix} a(L) & b(L) \\ c(L) & d(L) \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} \epsilon_1^*(t) \\ \epsilon_2^*(t) \end{bmatrix}, \end{cases} \tag{A14}$$

Fourier transform $\rightarrow \begin{cases} \begin{bmatrix} a(\omega) & b(\omega) \\ c(\omega) & d(\omega) \end{bmatrix} \begin{bmatrix} x_1(\omega) \\ x_2(\omega) \end{bmatrix} = \begin{bmatrix} \epsilon_1^*(\omega) \\ \epsilon_2^*(\omega) \end{bmatrix} \equiv \mathbf{A}(\omega)\mathbf{x}(\omega) = \boldsymbol{\epsilon}^*(\omega), \end{cases} \tag{A15}$

$$\begin{cases} \begin{bmatrix} x_1(\omega) \\ x_2(\omega) \end{bmatrix} = \begin{bmatrix} a(\omega) & b(\omega) \\ c(\omega) & d(\omega) \end{bmatrix}^{-1} \begin{bmatrix} \epsilon_1^*(\omega) \\ \epsilon_2^*(\omega) \end{bmatrix} = \begin{bmatrix} H_{11}(\omega) & H_{12}(\omega) \\ H_{21}(\omega) & H_{22}(\omega) \end{bmatrix} \begin{bmatrix} \epsilon_1^*(\omega) \\ \epsilon_2^*(\omega) \end{bmatrix}, \end{cases} \tag{A16}$$

$$\begin{cases} \mathbf{x}(\omega) = \mathbf{H}(\omega)\boldsymbol{\epsilon}^*(\omega), \end{cases} \tag{A17}$$

such that $\mathbf{A}(\omega)$ is known as the coefficient matrix. The elements in the coefficient matrix are expressed as

$$a(\omega) = 1 - \sum_{i=1}^n a_i e^{-j\omega i}, \tag{A18}$$

$$b(\omega) = - \sum_{i=1}^n b_i e^{-j\omega i}, \tag{A19}$$

$$c(\omega) = - \sum_{i=1}^n c_i e^{-j\omega i}, \tag{A20}$$

$$d(\omega) = 1 - \sum_{i=1}^n d_i e^{-j\omega i}, \tag{A21}$$

which is obtained via the Fourier transform from the time domain to the frequency domain. Note that j is the imaginary unit while ω is the angular frequency that is defined as $\omega = 2\pi f$, f = frequency. The matrix $\mathbf{H}(\omega)$ is known as the transfer matrix. The spectral matrix ($\mathbf{S}(\omega)$) can be obtained by proper ensemble averaging of Equation (A17), which is expressed as Equation (A22):

$$\mathbf{S}(\omega) = \begin{bmatrix} S_{11}(\omega) & S_{12}(\omega) \\ S_{21}(\omega) & S_{22}(\omega) \end{bmatrix} = \langle \mathbf{x}(\omega)\mathbf{x}^\dagger(\omega) \rangle = \mathbf{H}(\omega)\boldsymbol{\Sigma}\mathbf{H}^\dagger(\omega), \tag{A22}$$

such that $\langle \cdot \rangle$ is defined as an ensemble average, and \dagger is denoted as the complex conjugate transpose of a matrix. The element $S_{nm|n=m}$ is known as auto spectra and $S_{nm|n \neq m}$ is known as cross spectra. Similar to Equation (A7), the total interdependence between $x_1(\omega)$ and $x_2(\omega)$ is defined as Equation (A23), which consists of two directional causalities ($I_{x_1 \rightarrow x_2}$ and $I_{x_2 \rightarrow x_1}$) and one instantaneous causality ($I_{x_1 \leftrightarrow x_2}$).

$$I_{x_1 \cdot x_2} = \log \frac{S_{11}(\omega)S_{22}(\omega)}{|S(\omega)|} = I_{x_1 \rightarrow x_2} + I_{x_2 \rightarrow x_1} + I_{x_1 \leftrightarrow x_2}, \tag{A23}$$

such that $|S(\omega)|$ is the determinant of spectral matrix $S(\omega)$. The individual causalities are defined as

$$I_{x_1 \rightarrow x_2} = \log \frac{S_{22}(\omega)}{\hat{H}_{22}(\omega)\Sigma_{22}\hat{H}_{22}^*(\omega)}, \tag{A24}$$

$$I_{x_2 \rightarrow x_1} = \log \frac{S_{11}(\omega)}{\tilde{H}_{11}(\omega)\Sigma_{11}\tilde{H}_{11}^*(\omega)}, \tag{A25}$$

$$I_{x_1 \leftrightarrow x_2} = \log \frac{(\tilde{H}_{11}(\omega)\Sigma_{11}\tilde{H}_{11}^*(\omega))(\hat{H}_{22}(\omega)\Sigma_{22}\hat{H}_{22}^*(\omega))}{|S(\omega)|}. \tag{A26}$$

Here $\hat{H}_{22}(\omega) = H_{22}(\omega) + \frac{\Sigma_{12}}{\Sigma_{22}}H_{21}(\omega)$ and $\tilde{H}_{11}(\omega) = H_{11}(\omega) + \frac{\Sigma_{12}}{\Sigma_{11}}H_{12}(\omega)$; $*$ is denoted as the complex conjugate of the element [17].

As shown in Equations (A23) to (A26), the elements from the transfer matrix and covariance matrix of the noises are needed to evaluate the GC in the frequency domain. The elements can be obtained via the constant parameters of Equations (A4) and (A5), which can be estimated via Yule–Walker equations [18,40]. This approach of evaluating GC through the knowledge of constant parameters is known as the parametric method; it has the disadvantage that the information of lags (number of past values) is needed when estimating the constant parameters. Alternatively, the non-parametric method conducts the GC analysis by directly decomposing the spectral matrix ($S(\omega)$) into the transfer matrix ($H(\omega)$) and the covariance matrix of the noises (Σ). The decomposition is conducted via a recursive algorithm known as the Wilson algorithm [18,35]. The derivation of the Wilson algorithm will be discussed in the next section.

Appendix A.4. Wilson Algorithm

The spectral matrix $S(\omega)$ that is defined in the range $[-\pi, \pi]$ is known as the spectral density function. It can be expressed as Equation (A27).

$$S(\omega) = \sum_{i=-\infty}^{\infty} C_i e^{j\omega i}, \tag{A27}$$

$$C_i = \frac{1}{2\pi} \int_{-\pi}^{\pi} S(\omega) e^{-j\omega i} d\omega, \tag{A28}$$

such that the C_k is known as the correlation coefficient obtained via inverse Fourier transform Equation (A28).

Following Wilson’s factorization theorem (theorem 1.1 in [35]), the spectral density function is the product of generating function $\psi(e^{j\omega})$ with its complex conjugate transpose that is defined on a unit circle $|z| = 1$:

$$S(\omega) = \psi(e^{j\omega})\psi(e^{j\omega})^\dagger, \tag{A29}$$

$$\psi(e^{j\omega}) = \sum_{i=0}^{\infty} A_i e^{j\omega i}, \tag{A30}$$

$$A_i = \frac{1}{2\pi} \int_{-\pi}^{\pi} \psi(e^{j\omega}) e^{-j\omega i} d\omega. \tag{A31}$$

Here A_k is the moving average coefficient, which can be obtained by inverse Fourier transform as expressed in Equation (A31). With Equation (A30) being holomorphic extended to the inner of the unit circle ($|z| < 1$), it can be rewritten as Equation (A32).

$$\psi(e^{j\omega}) = \psi(z) = \sum_{i=0}^{\infty} A_i z^i = A_0 + \sum_{i=1}^{\infty} A_i z^i. \tag{A32}$$

$$\therefore S(z) = \psi(z)\psi(z)^\dagger = \left(\sum_{i=0}^{\infty} A_i z^i \right) \left(\sum_{i'=0}^{\infty} A_{i'}^\top z^{-i'} \right) = \sum_{i=0}^{\infty} \sum_{i'=0}^{\infty} A_i A_{i'}^\top z^{i-i'}. \tag{A33}$$

Comparing Equations (A22) and (A33) when $z = e^{j\omega} = 0$:

$$S(0) = \psi(0)\psi(0)^\dagger = \mathbf{H}(0)\mathbf{\Sigma}\mathbf{H}^\dagger(0) = \mathbf{I}\mathbf{\Sigma}\mathbf{I}^\top = \mathbf{\Sigma} = \mathbf{A}_0\mathbf{A}_0^\top, \tag{A34}$$

such that \mathbf{I} is the identity matrix, and $[\cdot]^\top$ indicates the transpose of the matrix. Since $\mathbf{A}_0^{-1}\mathbf{A}_0\mathbf{A}_0^\top\mathbf{A}_0^{-\top} = \mathbf{I}$, the transfer matrix $\mathbf{H}(\omega)$ can be obtained via

$$S(z) = \psi(z)\mathbf{I}\psi(z)^\dagger = \psi(z)\mathbf{A}_0^{-1}\mathbf{A}_0\mathbf{A}_0^\top\mathbf{A}_0^{-\top}\psi(z)^\dagger = \psi(z)\mathbf{A}_0^{-1}\mathbf{\Sigma}\mathbf{A}_0^{-\top}\psi(z)^\dagger, \tag{A35}$$

$$\therefore \mathbf{H}(\omega) = \psi(e^{j\omega})\mathbf{A}_0^{-1}, \mathbf{H}(\omega)^\dagger = \mathbf{A}_0^{-\top}\psi(e^{j\omega})^\dagger. \tag{A36}$$

As shown, the transfer matrix $\mathbf{H}(\omega)$ and covariance matrix of the noises $\mathbf{\Sigma}$ can be obtained once the generating function $\psi(e^{j\omega})$ is known. G.T. Wilson introduces the solution of $\psi(e^{j\omega})$ by iteration along with the plus operator ($[\cdot]^+$) of a given spectral density function $f(\omega)$, as expressed in Equations (A37) to (A39).

$$f(\omega) = \sum_{i=-\infty}^{\infty} \beta_i e^{j\omega i}, \beta_i = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\omega) e^{-j\omega i} d\omega, \tag{A37}$$

$$[f(\omega)]^+ = 0.5\beta_0 + \sum_{i=1}^{\infty} \beta_i e^{j\omega i}, \tag{A38}$$

$$\psi_{i+1} = \psi_i \underbrace{\left[\psi_i^{-1} \mathbf{S} \left(\psi_i^{-1} \right)^\dagger + \mathbf{I} \right]^+}_{[f(\omega)]^+}. \tag{A39}$$

From the previous work [35], it has been shown that the iteration of Equation (A39) will be uniquely converged if the initial generating function (ψ_0) is an upper triangle of the matrix which can be obtained via Cholesky factorization.

Appendix B. Transfer Entropy

Given a stochastic process (X), the average information/uncertainty needed to describe the process can be defined by Shannon entropy (I_X) and expressed as Equation (A40).

$$I_X = - \sum_{x \in X} p(x) \log_2 p(x), \tag{A40}$$

such that $p(x)$ is the discrete probability distribution at state x . The unit of the Shannon entropy varies with the base of the logarithm (i.e., the unit is bits when the logarithm is based 2) [28,41,42].

For the situation of two stochastic processes (X and Y), the statistical dependency of these two processes can be evaluated via Mutual Information (MI) and defined as Equation (A41).

$$MI = \sum_{x \in X, y \in Y} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)}. \tag{A41}$$

The definition of *MI* is derived from the Kullback–Leibler (KL) divergence. KL divergence measures the differences between two probability distributions ($p(x)$ and $q(x)$) non-symmetrically; it is defined as Equation (A42) [37].

$$D_{KL}(P||Q) = \sum_{x \in X} p(x) \log_2 \frac{p(x)}{q(x)}. \tag{A42}$$

Hence, *MI* measures the difference between the marginal probability distribution ($p(x)p(y)$) and joint probability distribution ($p(x, y)$) that quantifies the common information shared between the processes. Therefore, the *MI* will be zero when both processes are statistically independent and no common information is presented between the processes as the joint probability distribution will be equal to the product of marginal probability distribution ($p(x, y) = p(x)p(y) \rightarrow MI = \sum_{x \in X, y \in Y} p(x, y) \log_2 1 = 0$) [19,37].

As discussed above, *MI* only measured the commonly shared information, which subsequently does not indicate the directional flow of information. On the contrary, Transfer Entropy (*TE*) is able to reveal the directional flow of information by considering the conditional probability $p(\cdot|\cdot)$ and is defined as Equation (A43).

$$TE_{Y \rightarrow X} = \sum_{x \in X, y \in Y} p(x_{n+1}, x_n^{(k)}, y_n^{(l)}) \log_2 \frac{p(x_{n+1}|x_n^{(k)}, y_n^{(l)})}{p(x_{n+1}|x_n^{(k)})}, \tag{A43}$$

such that x_{n+1} is the state of X at the $(n + 1)$ -th moment; $x_n^{(k)}$ is the state at n -th moment of X consists of k previous states of X ; $y_n^{(l)}$ is the state at n -th moment of Y consisting of l previous states. Note that the previous states of k and l are arbitrary and depend on the interest of the study. For instance, given a collective observed stochastic process $X = (x_1, x_2, \dots, x_{10}), x_8^{(2)} = (x_n, x_m) | n, m \in [1, 8] \rightarrow (x_1, x_2)$ or (x_3, x_8) , etc. Hence, *TE* measures the additional information needed for the realization of X at $(n + 1)$ from Y with the assumption that X is independent of Y . If Y does not affect the future evolution of X , then $p(x_{n+1}|x_n^{(k)}, y_n^{(l)}) = p(x_{n+1}|x_n^{(k)})$ and subsequently $T_{Y \rightarrow X} = 0$. Using the general definition of conditional probability distribution ($p(x|y) = \frac{p(x,y)}{p(y)}$), Equation (A43) can be rewritten as Equation (A44) [19].

$$TE_{Y \rightarrow X} = \sum_{x \in X, y \in Y} p(x_{n+1}, x_n^{(k)}, y_n^{(l)}) \log_2 \frac{p(x_{n+1}, x_n^{(k)}, y_n^{(l)})p(x_n^{(k)})}{p(x_{n+1}, x_n^{(k)})p(x_n^{(k)}, y_n^{(l)})}. \tag{A44}$$

Appendix C. Information Rate, Information Length, and Information Rate Causality

In a dimensionless statistical manifold, the distance of two probability distributions is defined by their differences. One of the distances is discussed in Appendix B; this is KL divergence/relative entropy with the expression as Equation (A42). As mentioned, the KL divergence measured the differences non-symmetrically, and hence it is hardly considered as a metric distance because it does not satisfy the triangle inequality [32]. The symmetric version of KL divergence is introduced as Jensen divergence and expressed as Equation (A45). The square root of Jensen divergence has shown to be a metric distance that satisfies triangle inequality [32,43,44].

$$D_{JS} = \frac{1}{2}[D_{KL}(P||Q) + D_{KL}(Q||P)]. \tag{A45}$$

As depicted in Equations (A42) and (A45), the distance is measured at the ending points (e.g., initial and final probabilities), and hence it is not path-dependent. The path-dependent distance can be evaluated with the *information length* (\mathcal{L}), which is computed through the knowledge of *information rate* (Γ). The detailed discussion and derivation are shown in [32]. An essential brief discussion and derivation of Γ and \mathcal{L} are shown in the next part.

Consider a time-dependent probability distribution $(p(x, t))$ that evolves over time. For instance, $p(x, t)$ changes to $p(x, t + dt)$ with the small time step of $dt \rightarrow 0$. This evolution of difference can be expressed as Equation (A46) with the dt as the leading order.

$$\lim_{dt \rightarrow 0} \frac{1}{(dt)^2} D_{KL}(p(x, t + dt) || p(x, t)) = \lim_{dt \rightarrow 0} \frac{1}{(dt)^2} D_{KL}(p(x, t) || p(x, t + dt)) = \lim_{dt \rightarrow 0} \frac{1}{(dt)^2} D_{JS}, \tag{A46}$$

$$\lim_{dt \rightarrow 0} \frac{1}{(dt)^2} \int dx p(x, t + dt) \ln \frac{p(x, t + dt)}{p(x, t)}, \tag{A47}$$

$$\begin{aligned} \lim_{dt \rightarrow 0} \frac{1}{(dt)^2} \int dx \left[p(x, t) + (dt)\partial_t p + \frac{1}{2}(dt)^2(\partial_{tt} p) + \mathcal{O}((dt)^3) \right] \\ \ln \left[1 + \frac{(dt)\partial_t p}{p(x, t)} + \frac{1}{2} \frac{(dt)^2(\partial_{tt} p)}{p(x, t)} + \mathcal{O}((dt)^3) \right], \end{aligned} \tag{A48}$$

$$\lim_{dt \rightarrow 0} \frac{1}{(dt)^2} \int dx \left[p(x, t) + (dt)\partial_t p \right] \left[\frac{(dt)\partial_t p}{p(x, t)} - \frac{1}{2} \left[\frac{(dt)\partial_t p}{p(x, t)} \right]^2 \right], \tag{A49}$$

$$\lim_{dt \rightarrow 0} \frac{1}{(dt)^2} \int dx \left[p(x, t) \frac{(dt)\partial_t p}{p(x, t)} - \frac{1}{2} \left[\frac{(dt)\partial_t p}{p(x, t)} \right]^2 \right] p(x, t) + (dt)\partial_t p \frac{(dt)\partial_t p}{p(x, t)} - \frac{1}{2} \left[\frac{(dt)\partial_t p}{p(x, t)} \right]^2 (dt)\partial_t p, \tag{A50}$$

$$\lim_{dt \rightarrow 0} \frac{1}{(dt)^2} \int dx \left[(dt)\partial_t p - \frac{1}{2} \frac{(dt)^2(\partial_{tt} p)^2}{p(x, t)} + \frac{(dt)^2(\partial_{tt} p)^2}{p(x, t)} - \frac{1}{2} \frac{(dt)^3(\partial_{tt} p)^3}{(p(x, t))^2} \right], \tag{A51}$$

$$\lim_{dt \rightarrow 0} \frac{1}{(dt)^2} \int dx \left[(dt)\partial_t p + \frac{1}{2} \frac{(dt)^2(\partial_{tt} p)^2}{p(x, t)} - \frac{1}{2} \frac{(dt)^3(\partial_{tt} p)^3}{(p(x, t))^2} \right], \tag{A52}$$

$$\lim_{dt \rightarrow 0} 0 \int dx \frac{\partial_t p}{dt} + \int dx \frac{1}{2} \frac{(\partial_{tt} p)^2}{p(x, t)} - 0 \int dx \frac{1}{2} \frac{(dt)(\partial_{tt} p)^3}{(p(x, t))^2}, \tag{A53}$$

$$\therefore \lim_{dt \rightarrow 0} \frac{1}{(dt)^2} D_{JS} = \frac{1}{2} \int dx p(x, t) (\partial_t \ln p(x, t))^2 \equiv \frac{1}{2} \Gamma^2(t), \tag{A54}$$

$$\therefore \int dx \partial_t p(x, t) = \int dx \partial_{tt} p(x, t) = 0 \therefore \int dx p(x, t) = 1, \tag{A55}$$

$$\therefore \Gamma^2(t) = \int dx p(x, t) [\partial_t \ln p(x, t)]^2 = 4 \int dx \left[\partial_t \sqrt{p(x, t)} \right]^2 \tag{A56}$$

Note that Taylor series is implemented in Equation (A48) and Mercator series is used in Equation (A49). The final result $\Gamma(t)$ is defined as the information rate which measures the rate of the statistical change (e.g., due to the new input of information) from one state to another. Hence, the total statistical change in $p(x, t)$ or the total path-dependent distance of $p(x, t)$ that evolves through the time is expressed as Equation (A57) and defined as information length (\mathcal{L}).

$$\mathcal{L} = \int dt \Gamma(t). \tag{A57}$$

To accommodate the computation, the continuous differentiation ($\frac{df(t)}{dt}, \frac{\partial f(x,t)}{\partial t}$) and integration ($\int dt f(t), \int \partial t f(x,t)$) are discretized, as shown in Equations (A58) to (A61).

$$\frac{df(t)}{dt} = \frac{f(t + \Delta t) - f(t)}{\Delta t}, \tag{A58}$$

$$\frac{\partial f(x,t)}{\partial t} = \frac{f(x, t + \Delta t) - f(x, t)}{\Delta t}, \tag{A59}$$

$$\int dt f(t) = \sum_{t_i=0}^N \Delta t f(t_i), \Delta t = t_{i+1} - t_i, N \in \mathbb{R} \tag{A60}$$

$$\int \partial t f(x,t) = \sum_{t_i=0}^N \Delta t f(x, t_i), \Delta t = t_{i+1} - t_i, N \in \mathbb{R}. \tag{A61}$$

Hence, the discretized version of information rate (Γ), information length (\mathcal{L}), and information rate causality ($\Gamma_{x_n \rightarrow x_m} | n \neq m \& (n,m) \in [1,2]$), shown in Equations (A62) to (A64), are used when computing the results that are included in the paper.

$$\Gamma^2 = 4 \sum_{x \in X} \frac{\Delta x}{[\Delta t]^2} \left[\sqrt{p(x, t + \Delta t)} - \sqrt{p(x, t)} \right]^2, \tag{A62}$$

$$\mathcal{L} = \sum_{t \in T} \Delta t \Gamma(t), \tag{A63}$$

$$\Gamma_{x_2 \rightarrow x_1}^2 = 4 \sum_{x_1 \in X_1, x_2 \in X_2} \frac{\Delta x_1 \Delta x_2}{[\Delta t]^2} \left[\sqrt{p(x_1(t + \Delta t), x_2(t))} - \sqrt{p(x_1(t), x_2(t))} \right]^2. \tag{A64}$$

Appendix D. Basic Model—Stationary and Continuous Coupling

In this part, we investigate the coupling between two processes, namely $x_1(t)$ and $x_2(t)$, where the first process ($x_1(t)$) influences the second process ($x_2(t)$) throughout the entire process. We consider both linear and nonlinear cases, which are represented by Equations (A65) and (A66), respectively.

Linear model:

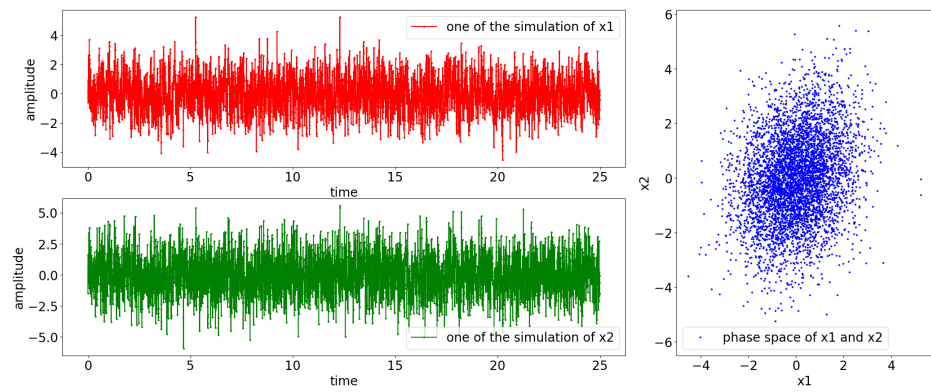
$$\begin{aligned} x_1(t) &= 0.55x_1(t - 1) + \epsilon_1(t), \\ x_2(t) &= x_1(t - 2) + \epsilon_2(t). \end{aligned} \tag{A65}$$

Nonlinear model:

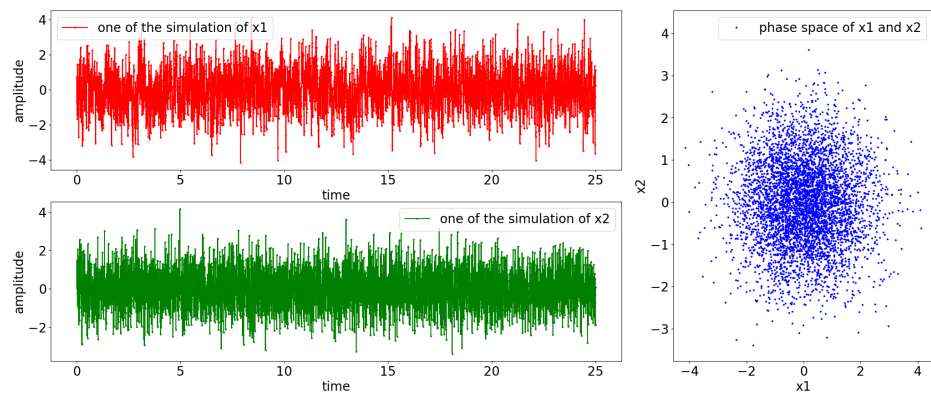
$$\begin{aligned} x_1(t) &= 0.55x_1(t - 1)e^{1-|x_1(t-1)|} + \epsilon_1(t), \\ x_2(t) &= x_1(t - 2) + \epsilon_2(t). \end{aligned} \tag{A66}$$

Here, t is the time steps and $\epsilon_n | n \in [1,2]$ is the Gaussian white noise. The white noise has zero mean ($\langle \epsilon_n \rangle = 0 | n \in [1,2]$) with unit variance ($\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{bmatrix}, \Sigma_{nm} = \langle (\epsilon_n - \langle \epsilon_n \rangle)(\epsilon_m - \langle \epsilon_m \rangle) \rangle | n, m \in [1,2]$). Additionally, we also consider the noncoupling scenario where the process $x_2(t)$ solely depends on the noise $\epsilon_2(t)$ ($x_2(t) = \epsilon_2(t) \because x_1(t - 2) = 0$).

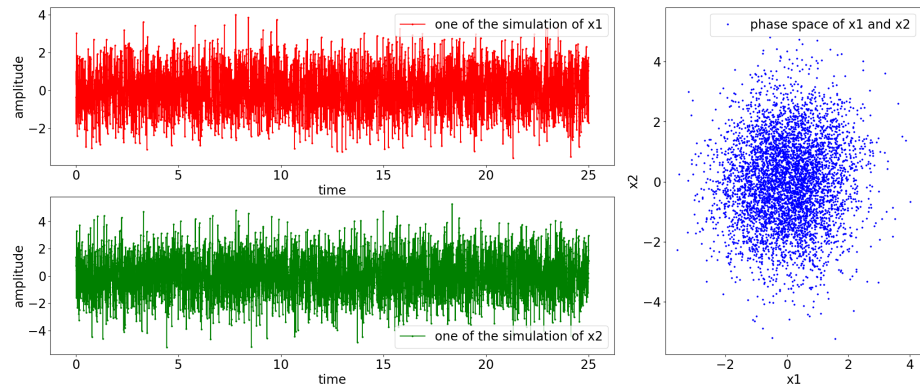
The simulation results for Equations (A65) and (A66) are presented in Figure A1. From the figures, it can be observed that both the linear and nonlinear processes exhibit stationary oscillation, regardless of whether they are coupled or noncoupled.



(a) linear: coupling

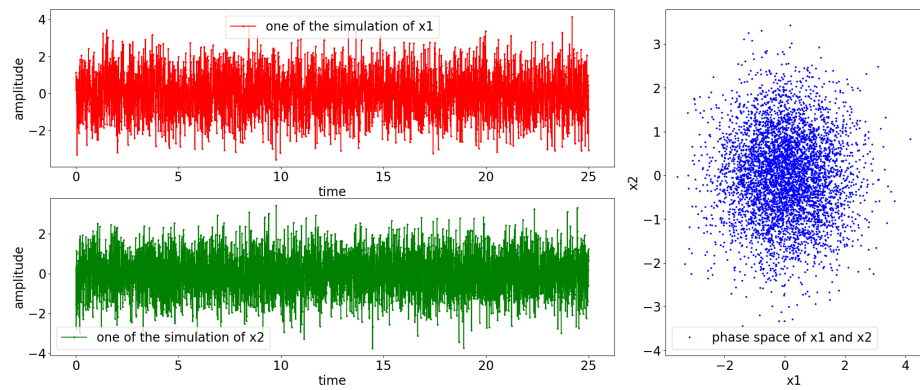


(b) linear: noncoupling



(c) nonlinear: coupling

Figure A1. Cont.

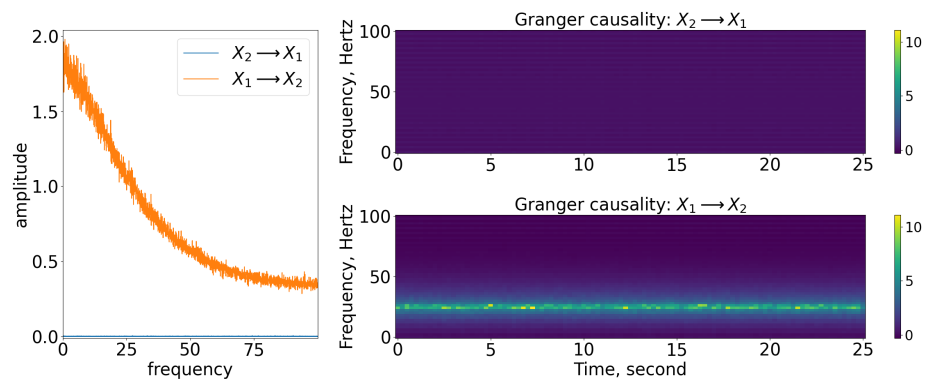


(d) nonlinear: noncoupling

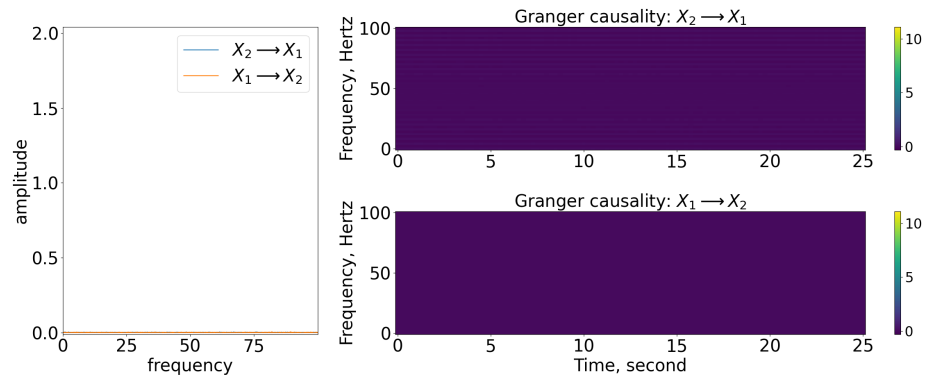
Figure A1. Result of the simulation based on Equations (A65) and (A66). The [top left, red] is the result of x_1 , while the [bottom left, green] shows the result of x_2 . The phase space of x_1 and x_2 is shown in [right blue] figure. Note that the noncoupling here refers to the processes oscillating by themselves without sharing any information with each other.

Appendix D.1. Granger Causality

The results of the GC are shown in Figure A2 for all the cases. Based on the results, both the coupling and noncoupling cases can be depicted well via the nonparametric GC approach (refer to Section 2.1). The continuous causality from x_1 to x_2 can be depicted well in Figure A2a,c. Similarly, the noncoupling between the signals is shown well in Figure A2b,d.

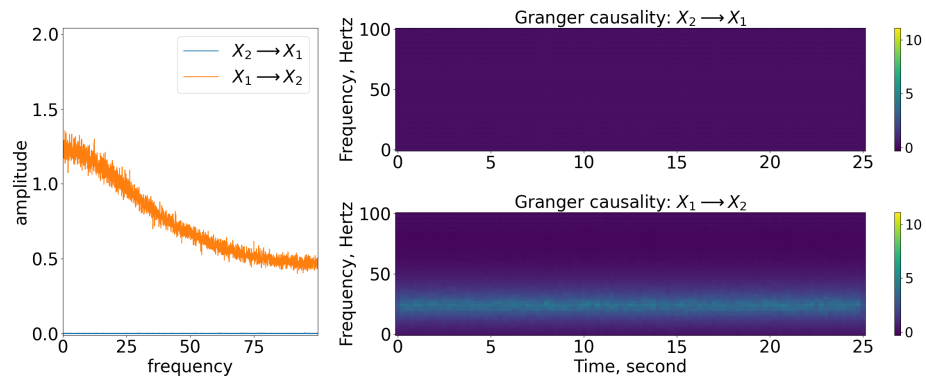


(a) linear: coupling

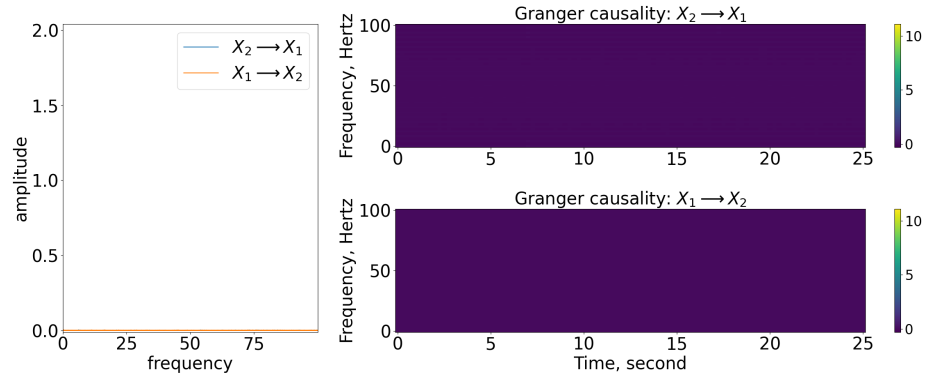


(b) linear: noncoupling

Figure A2. Cont.



(c) nonlinear: coupling

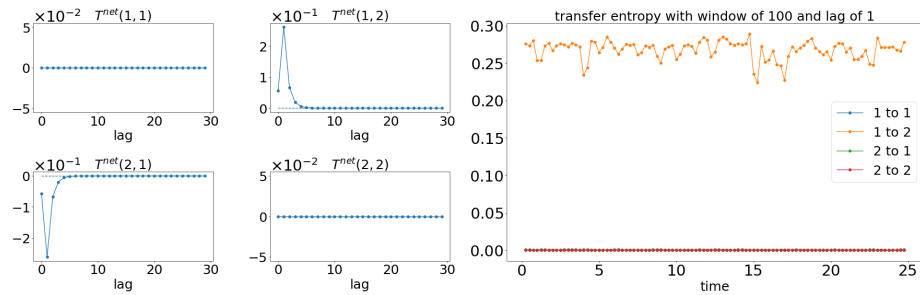


(d) nonlinear: noncoupling

Figure A2. Result of the spectral and time-varying frequency of GC. In each subfigure, the [left panel] shows the spectral (frequency domain) of GC with [blue line] indicating that x_2 causes x_1 , while [orange line] indicates that x_1 causes x_2 . The [top, right] subfigure illustrates the time-varying frequency of GC of x_2 to x_1 , while [bottom, right] shows x_1 to x_2 . The results are based on Equations (A65) and (A66).

Appendix D.2. Transfer Entropy

The results of TE shown in Figure A3 also show similar findings as GC for both coupling and noncoupling of the signals. As depicted in Figure A3a,c, TE can effectively capture the coupling from x_1 to x_2 when appropriate lag is selected. The suitable lag can be determined by observing the peak of Net TE ($TE_{x_1, x_2} = TE_{x_1 \rightarrow x_2} - TE_{x_2 \rightarrow x_1}$), as it indicates the lag at which a significant result can be expected. Similarly, the noncoupling cases are depicted well in Figure A3b,d, where the TE between the two signals is extremely low, in the order of magnitude of $\times 10^{-4}$.



(a) linear: coupling

Figure A3. Cont.

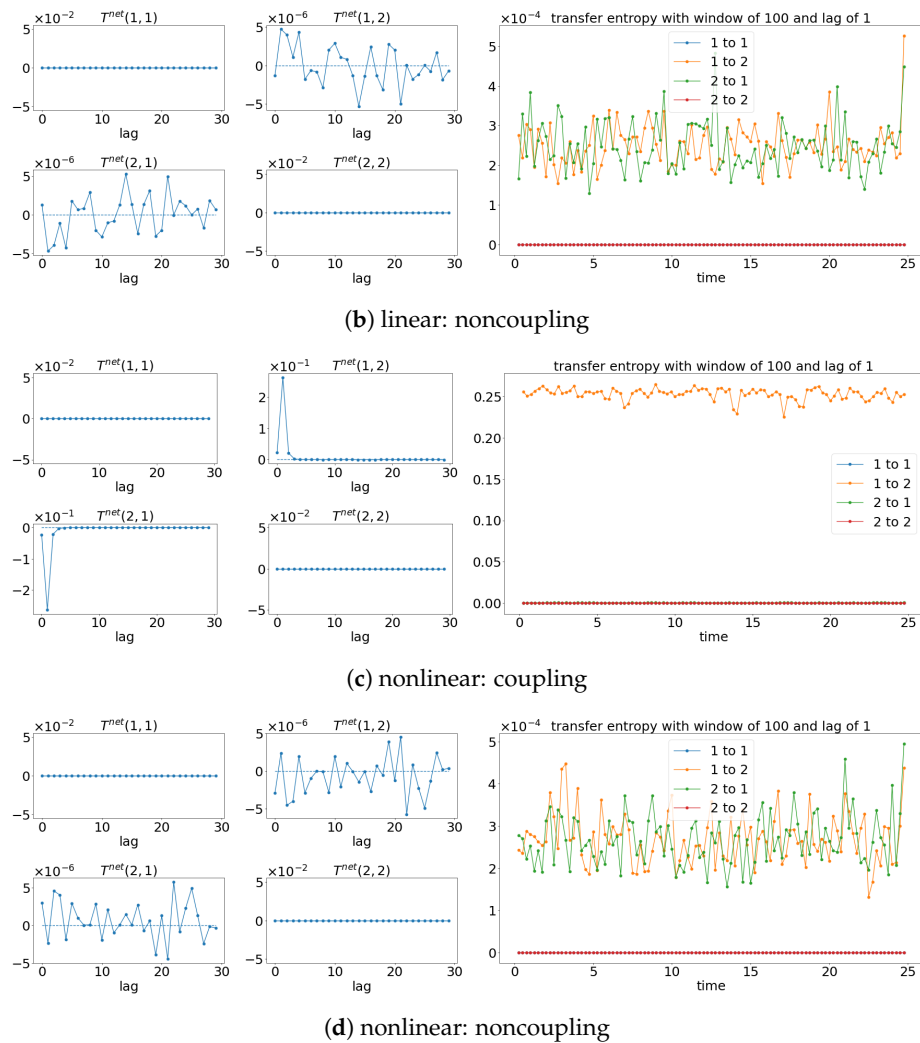
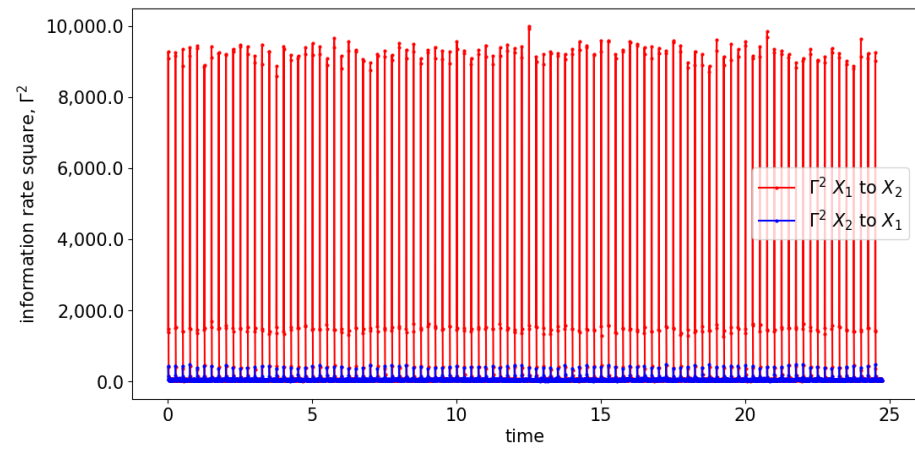


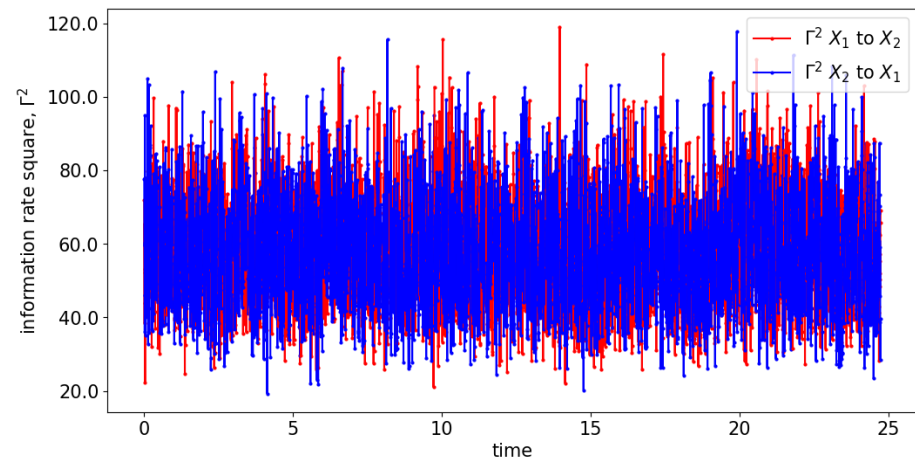
Figure A3. Result of the net TE and window sliding TE. Note that for each subfigure, [left, 4×4] depicts the net TE for the whole signals. The [right] subfigure shows the window sliding TE. The results shown are based on Equations (A65) and (A66).

Appendix D.3. Information Rate Causality

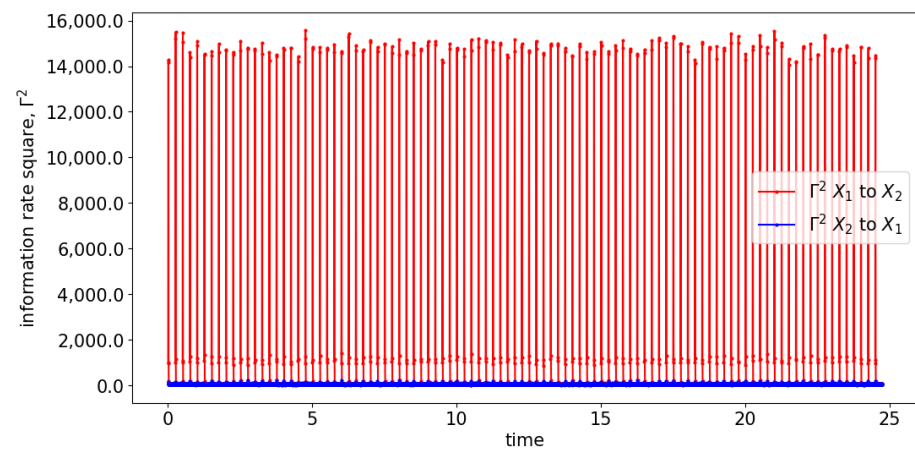
Information rate causality is a quantitative measure of the changes in probability distribution caused by one signal influencing another. In the cases of coupling, a high peak is expected to indicate a significant change in the distribution due to the causal relationship between the signals, as depicted in Figure A4a,c. The magnification of these figures is shown in Figure A5a,c. In this scenario, the causation from $x_1(t)$ to $x_2(t)$ is represented well by the peaks at an order of magnitude of $\times 10^4$. In noncoupling cases, as shown in Figure A4b,d (with magnification shown in Figure A5b,d), information rate causality does not reduce to zero due to the noise present in the estimation of the probability distribution (refer to Figure S1). However, this noise does not affect the evaluation, as it does not result in a high order of peaks observed in coupling cases.



(a) linear: coupling

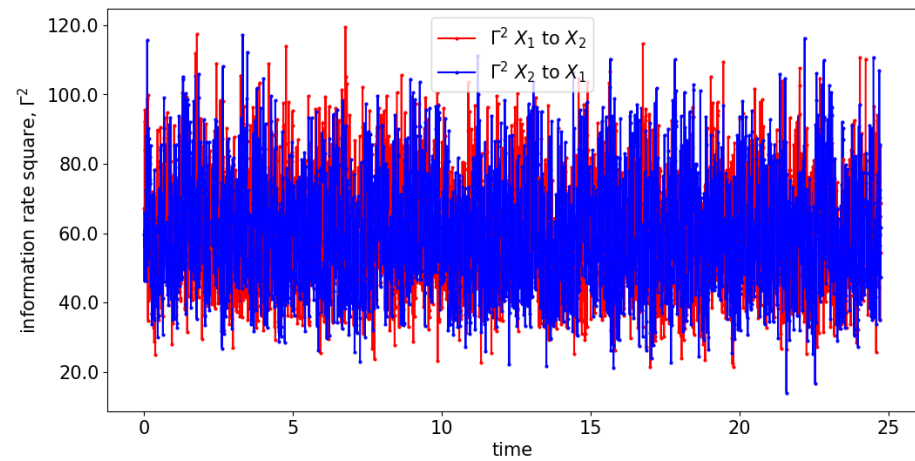


(b) linear: noncoupling



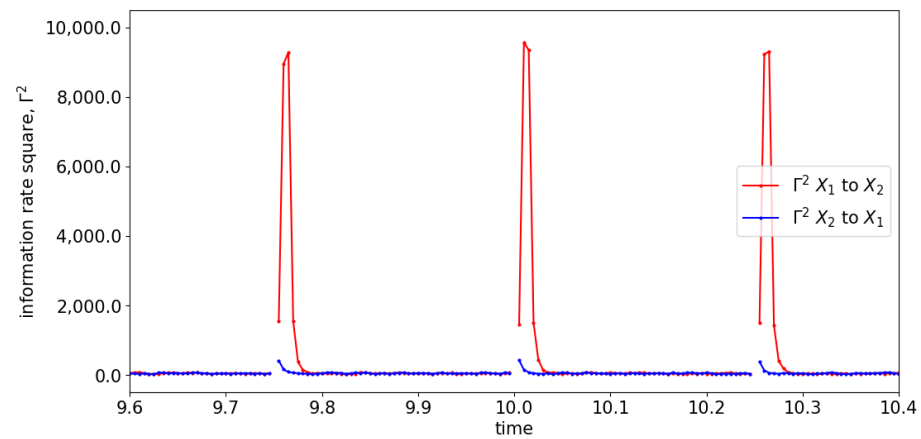
(c) nonlinear: coupling

Figure A4. Cont.

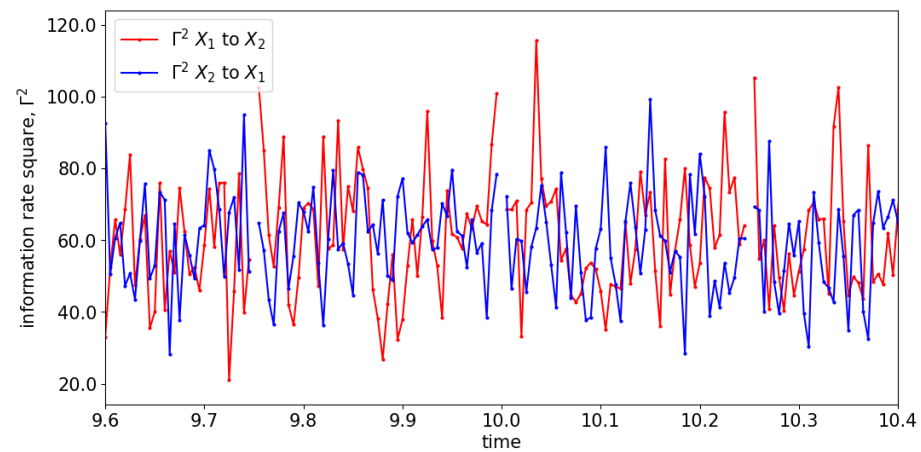


(d) nonlinear: noncoupling

Figure A4. Result of information rate causality for Equations (A65) and (A66).

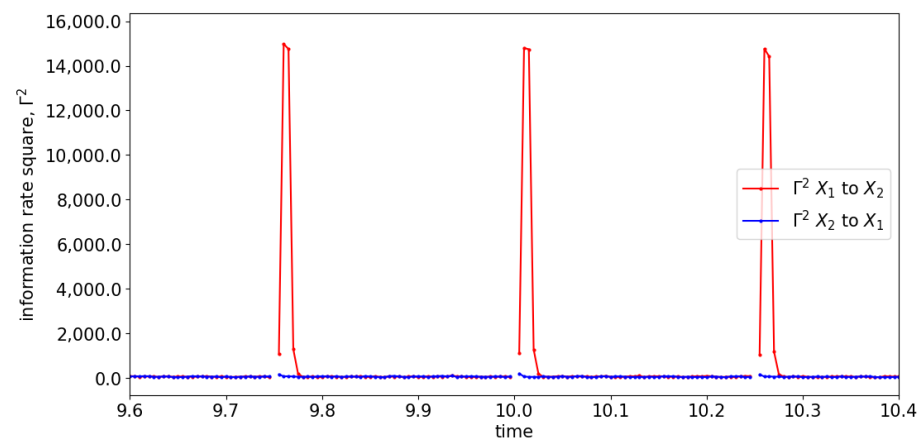


(a) linear: coupling

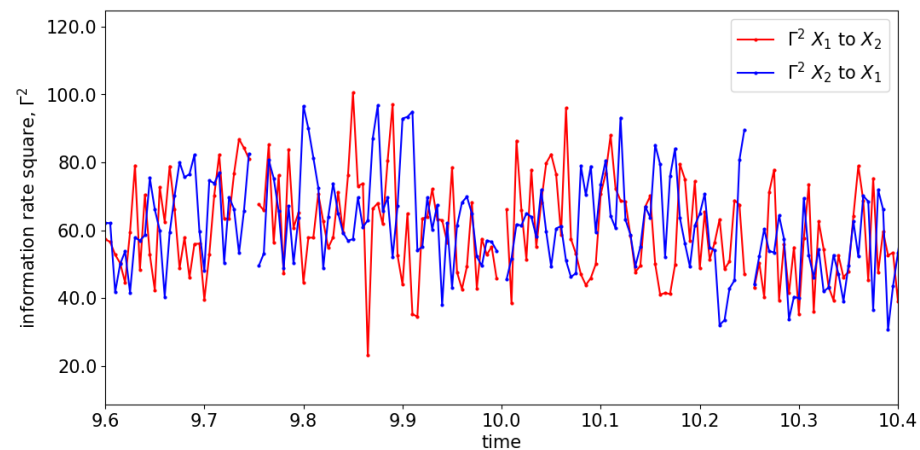


(b) linear: noncoupling

Figure A5. Cont.



(c) nonlinear: coupling



(d) nonlinear: noncoupling

Figure A5. Zoom-in of Figure A4 for the information rate causality between 9.6 s and 10.4 s.

Reference

- Zenil, H.; Kiani, N.A.; Marabita, F.; Deng, Y.; Elias, S.; Schmidt, A.; Ball, G.; Tegnér, J. An Algorithmic Information Calculus for Causal Discovery and Reprogramming Systems. *iScience* **2019**, *19*, 1160–1172. [[CrossRef](#)] [[PubMed](#)]
- Zenil, H.; Kiani, N.A.; Zea, A.A.; Tegnér, J. Causal deconvolution by algorithmic generative models. *Nat. Mach. Intell.* **2019**, *1*, 58–66. [[CrossRef](#)]
- Nogueira, A.R.; Pugnana, A.; Ruggieri, S.; Pedreschi, D.; Gama, J. Methods and tools for causal discovery and causal inference. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2022**, *12*, e1449. [[CrossRef](#)]
- Zenil, H.; Kiani, N.A.; Tegnér, J. Low-algorithmic-complexity entropy-deceiving graphs. *Phys. Rev. E* **2017**, *96*, 012308. [[CrossRef](#)] [[PubMed](#)]
- Danks, D.; Davis, I. Causal inference in cognitive neuroscience. *WIREs Cogn. Sci.* **2023**, e1650. [[CrossRef](#)]
- Granger, C.W.J. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econom. J. Econom. Soc.* **1969**, *37*, 424–438. [[CrossRef](#)]
- Guo, R.; Cheng, L.; Li, J.; Hahn, P.R.; Liu, H. A Survey of Learning Causality with Data: Problems and Methods. *ACM Comput. Surv. (CSUR)* **2020**, *53*, 1–37. [[CrossRef](#)]
- Kaufmann, R.K.; Stern, D.I. Evidence for human influence on climate from hemispheric temperature relations. *Nature* **1997**, *388*, 39–44. [[CrossRef](#)]
- Imbens, G.W. Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review*. *Rev. Econ. Stat.* **2004**, *86*, 4–29. [[CrossRef](#)]
- Pullon, R.M.; Yan, L.; Sleigh, J.W.; Warnaby, C.E. Granger causality of the electroencephalogram reveals abrupt global loss of cortical information flow during propofol-induced loss of responsiveness. *Anesthesiology* **2020**, *133*, 774–786. [[CrossRef](#)]
- Zhao, Y.; Billings, S.A.; Wei, H.; He, F.; Sarrigiannis, P.G. A new NARX-based Granger linear and nonlinear casual influence detection method with applications to EEG data. *J. Neurosci. Methods* **2013**, *212*, 79–86. [[CrossRef](#)] [[PubMed](#)]
- Ebert-Uphoff, I.; Deng, Y. Causal discovery for climate research using graphical models. *J. Clim.* **2012**, *25*, 5648–5665. [[CrossRef](#)]

13. Dehejia, R.H.; Wahba, S. Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs. *J. Am. Stat. Assoc.* **1999**, *94*, 1053–1062. [[CrossRef](#)]
14. Shovon, M.H.I.; Nandagopal, N.; Vijayalakshmi, R.; Du, J.T.; Cocks, B. Directed Connectivity Analysis of Functional Brain Networks during Cognitive Activity Using Transfer Entropy. *Neural Process. Lett.* **2017**, *45*, 807–824. [[CrossRef](#)]
15. Gao, Y.; Su, H.; Li, R.; Zhang, Y. Synchronous analysis of brain regions based on multi-scale permutation transfer entropy. *Comput. Biol. Med.* **2019**, *109*, 272–279. [[CrossRef](#)]
16. Seth, A.K.; Barrett, A.B.; Barnett, L. Granger causality analysis in neuroscience and neuroimaging. *J. Neurosci.* **2015**, *35*, 3293–3297. [[CrossRef](#)]
17. Ding, M.; Chen, Y.; Bressler, S.L. Granger Causality: Basic Theory and Application to Neuroscience; In *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2006.
18. Lima, V.; Dellajustina, F.J.; Shimoura, R.O.; Girardi-Schappo, M.; Kamiji, N.L.; Pena, R.F.; Roque, A.C. Granger causality in the frequency domain: Derivation and applications. *Rev. Bras. Ensino Fis.* **2020**, *42*. [[CrossRef](#)]
19. van Milligen, B.P.; Birkenmeier, G.; Ramisch, M.; Estrada, T.; Hidalgo, C.; Alonso, A. Causality detection and turbulence in fusion plasmas. *Nucl. Fusion* **2014**, *54*, 023011. [[CrossRef](#)]
20. Sigttermans, D. Is Information Theory Inherently a Theory of Causation? *arXiv* **2020**, arXiv:2010.01932.
21. Bressler, S.L.; Seth, A.K. Wiener-Granger Causality: A well established methodology. *Neuroimage* **2011**, *58*, 323–329. [[CrossRef](#)]
22. Geweke, J. Measurement of Linear Dependence and Feedback Between Multiple Time Series Measurement of Linear Dependence and Feedback Between Multiple Time Series. *J. Am. Stat. Assoc.* **1982**, *77*, 304–313. [[CrossRef](#)]
23. Geweke, J.F. Measures of Conditional Linear Dependence and Feedback Between Time Series. *J. Am. Stat. Assoc.* **1984**, *79*, 907–915. [[CrossRef](#)]
24. Dhamala, M.; Rangarajan, G.; Ding, M. Estimating granger causality from fourier and wavelet transforms of time series data. *Phys. Rev. Lett.* **2008**, *100*, 018701. [[CrossRef](#)]
25. Dhamala, M.; Rangarajan, G.; Ding, M. Analyzing information flow in brain networks with nonparametric Granger causality. *NeuroImage* **2008**, *41*, 354–362. [[CrossRef](#)] [[PubMed](#)]
26. Klonowski, W. Everything you wanted to ask about EEG but were afraid to get the right answer. *Nonlinear Biomed. Phys.* **2009**, *3*, 2. [[CrossRef](#)] [[PubMed](#)]
27. He, F.; Yang, Y. Nonlinear System Identification of Neural Systems from Neurophysiological Signals. *Neuroscience* **2021**, *458*, 213–228. [[CrossRef](#)]
28. Schreiber, T. Measuring Information Transfer. *Phys. Rev. Lett.* **2000**, *85*, 461–464. [[CrossRef](#)]
29. Barnett, L.; Barrett, A.B.; Seth, A.K. Granger causality and transfer entropy Are equivalent for gaussian variables. *Phys. Rev. Lett.* **2009**, *103*, 238701. [[CrossRef](#)] [[PubMed](#)]
30. Kim, E.J.; Guel-Cortez, A.J. Causal information rate. *Entropy* **2021**, *23*, 1087. [[CrossRef](#)]
31. Kim, E.J. Information geometry and non-equilibrium thermodynamic relations in the over-damped stochastic processes. *J. Stat. Mech. Theory Exp.* **2021**, *2021*, 93406. [[CrossRef](#)]
32. Kim, E.J. Information geometry, fluctuations, non-equilibrium thermodynamics, and geodesics in complex systems. *Entropy* **2021**, *23*, 1393. [[CrossRef](#)]
33. Ito, S.; Dechant, A. Stochastic Time Evolution, Information Geometry, and the Cramér-Rao Bound. *Phys. Rev. X* **2020**, *10*, 021056. [[CrossRef](#)]
34. O'donnell, R.T.; Korb, K.B.; Allison, L. Causal KL: Evaluating Causal Discovery. *arXiv* **2021**, arXiv:2111.06029.
35. Wilson, G.T. The Factorization of Matricial Spectral Densities. *SIAM J. Appl. Math.* **1972**, *23*, 420–426. [[CrossRef](#)]
36. Detto, M.; Molini, A.; Katul, G.; Stoy, P.; Palmroth, S.; Baldocchi, D. Causality and persistence in ecological systems: A nonparametric spectral Granger causality approach. *Am. Nat.* **2012**, *179*, 524–535. [[CrossRef](#)] [[PubMed](#)]
37. Gencaga, D.; Knuth, K.H.; Rossow, W.B. A recipe for the estimation of information flow in a dynamical system. *Entropy* **2015**, *17*, 438–470. [[CrossRef](#)]
38. Sahann, R.; Muller, T.; Schmidt, J. Histogram binning revisited with a focus on human perception. In Proceedings of the 2021 IEEE Visualization Conference (VIS), New Orleans, LA, USA, 24–29 October 2021. [[CrossRef](#)]
39. Terrell, G.R.; Scott, D.W. Oversmoothed nonparametric density estimates. *J. Am. Stat. Assoc.* **1985**, *80*, 209–214. [[CrossRef](#)]
40. Schelker, B.; Winterhalder, M.; Timmer, J. *Handbook of Time Series Analysis*; Wiley-VCH Verlag GmbH & Co KGaA: Weinheim, Germany, 2006; pp. 335–491.
41. Shannon, C.E. A Mathematical Theory of Communication PART III: MATHEMATICAL PRELIMINARIES. *Bell Syst. Tech. J.* **1948**, *27*, 623–656. [[CrossRef](#)]
42. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [[CrossRef](#)]
43. Virostek, D. The metric property of the quantum Jensen-Shannon divergence. *Adv. Math.* **2021**, *380*, 107595. [[CrossRef](#)]
44. Briët, J.; Harremoës, P. Properties of classical and quantum Jensen-Shannon divergence. *Phys. Rev. A—At. Mol. Opt. Phys.* **2009**, *79*, 052311. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.