

## Article

# Estimation of Information Flow-Based Causality with Coarsely Sampled Time Series

X. San Liang <sup>1,2</sup> 
<sup>1</sup> Department of Atmospheric and Oceanic Sciences, Fudan University, Shanghai 200438, China; xsliang@fudan.edu.cn

<sup>2</sup> The Artificial Intelligence Department, Division of Frontier Research, Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519082, China

## Abstract

The past decade has seen growing applications of the information flow-based causality analysis, particularly with the concise formula of its maximum likelihood estimator. At present, the algorithm for its estimation is based on differential dynamical systems, which, however, may raise an issue for coarsely sampled time series. Here, we show that, for linear systems, this is suitable at least qualitatively, but, for highly nonlinear systems, the bias increases significantly as the sampling frequency is reduced. This study provides a partial solution to this problem, showing how causality analysis can be made faithful with coarsely sampled series, provided that the statistics are sufficient. The key point here is that, instead of working with a Lie algebra, we turn to work with its corresponding Lie group. An explicit and concise formula is obtained, with only sample covariances involved. It is successfully applied to a system comprising a pair of coupled Rössler oscillators. Particularly remarkable is the success when the two oscillators are nearly synchronized. As more often than not observations may be scarce, this solution, albeit partial, is very timely.

**Keywords:** quantitative causality; information flow; coarsely sampled time series; synchronization; Rössler system; Lie group; Frobenius-Perron operator

## 1. Introduction

Causality analysis is a fundamental task in scientific research. Though traditionally formulated as a statistical problem (see, for example, the classics by Granger [1], Pearl [2], and Rubin [3,4]) in data science and computer science, among other disciplines, formalisms within the framework of dynamical systems have also been established; refer to a focus issue of *Chaos* [5] for relevant references. Particularly, in terms of information flow/transfer, it has been argued that causality is “a real notion in physics that can be derived *ab initio*” [6]. A comprehensive study with generic systems has been conducted, and explicit formulas have been attained in closed form [6,7] with the aid of the Frobenius-Perron operator (e.g., [8]), a technique also exploited in similar studies (e.g., [9]). These formulas have been validated with benchmark systems such as the baker transformation, Hénon map, etc., and have been applied successfully to real-world problems in diverse disciplines, such as global climate change, dynamic meteorology, land-atmosphere interaction, data-driven prediction, near-wall turbulence, neuroscience, financial analysis, and quantum information; see [10] for a brief review of the applications and [11–16] for some updates.



Academic Editor: Boris Ryabko

Received: 21 October 2025

Revised: 13 December 2025

Accepted: 24 December 2025

Published: 26 December 2025

**Copyright:** © 2025 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

For the purpose of this study, we first give a brief introduction to the theory within the framework of a differential dynamical system (also available for discrete-time mappings; see [6]). Let

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}, t) + \mathbf{B}(\mathbf{x}, t)\dot{\mathbf{w}}, \quad (1)$$

be a  $d$ -dimensional continuous-time stochastic system for  $\mathbf{x} = (x_1, \dots, x_d)$  (we do not distinguish notations for random and deterministic variables), where  $\mathbf{F} = (F_1, \dots, F_d)$  may be arbitrary nonlinear differentiable functions of  $\mathbf{x}$  and  $t$ ,  $\dot{\mathbf{w}}$  is a vector of white noise, and  $\mathbf{B} = (b_{ij})$  is the matrix of perturbation amplitudes, which may also be any differentiable functions of  $\mathbf{x}$  and  $t$ . This setting, particularly the setting in the absence of stochasticity, avails us of the arsenal from physics in approaching the problem. For example, the concept of symmetry has been found to play an important role [17]. Using the Frobenius–Perron operator, Liang (2016) [6] proved that the rate of information flowing from  $x_j$  to  $x_i$  (in nats per unit time) is

$$\begin{aligned} T_{j \rightarrow i} &= -E \left[ \frac{1}{\rho_i} \int_{\mathbb{R}^{d-2}} \frac{\partial(F_i \rho_{\tilde{\mathbf{x}}})}{\partial x_i} d\mathbf{x}_{\tilde{\mathbf{x}}} \right] + \frac{1}{2} E \left[ \frac{1}{\rho_i} \int_{\mathbb{R}^{d-2}} \frac{\partial^2(g_{ii} \rho_{\tilde{\mathbf{x}}})}{\partial x_i^2} d\mathbf{x}_{\tilde{\mathbf{x}}} \right], \\ &= - \int_{\mathbb{R}^d} \rho_{j|i}(x_j|x_i) \frac{\partial(F_i \rho_{\tilde{\mathbf{x}}})}{\partial x_i} d\mathbf{x} + \frac{1}{2} \int_{\mathbb{R}^d} \rho_{j|i}(x_j|x_i) \frac{\partial^2(g_{ii} \rho_{\tilde{\mathbf{x}}})}{\partial x_i^2} d\mathbf{x}, \end{aligned} \quad (2)$$

where  $d\mathbf{x}_{\tilde{\mathbf{x}}}$  signifies  $dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_{j-1} dx_{j+1} \dots dx_n$ ,  $E$  stands for mathematical expectation,  $g_{ii} = \sum_{k=1}^n b_{ik} b_{ik}$ ,  $\rho_i = \rho_i(x_i)$  is the marginal probability density function (pdf) of  $x_i$ ,  $\rho_{j|i}$  is the pdf of  $x_j$  conditioned on  $x_i$ , and  $\rho_{\tilde{\mathbf{x}}} = \int_{\mathbb{R}} \rho(\mathbf{x}) dx_j$ . The algorithm for the information flow-based causal inference is as follows: if  $T_{j \rightarrow i} = 0$ , then  $x_j$  is not causal to  $x_i$ ; otherwise, it is causal, and the absolute value measures the magnitude of the causality from  $x_j$  to  $x_i$ . This is guaranteed by a property called the “principle of nil causality”. Another property, as proven by Liang (2018) (see [10]), regards the invariance upon coordinate transformation, indicating that the obtained information flow (IF) is an intrinsic property in nature. Also established is that [6], for a linear model, i.e., for  $\mathbf{F}(\mathbf{x}, t) = \mathbf{A}\mathbf{x}$ , with  $\mathbf{A} = (a_{ij})$  and  $\mathbf{B} = (b_{ij})$  being constant matrices in (1), the information flow rate from  $x_j$  to  $x_i$  is

$$T_{j \rightarrow i} = a_{ij} \frac{\sigma_{ij}}{\sigma_{ii}},$$

where  $\sigma_{ij}$  is the population covariance of  $x_i$  and  $x_j$ . By this, it follows that, in the linear sense, causation implies correlation, but not vice versa. In an explicit expression, this corollary expresses mathematically the debate on causation vs. correlation ever since George Berkeley (1710) [18].

In the case with only  $d$  time series  $x_1, x_2, \dots, x_d$ , the quantitative causality, i.e., the IF, between them can be estimated using maximum likelihood estimation (see [19,20]). Under the assumption of a linear system with additive noise, the maximum likelihood estimator (mle) of (2) for  $T_{j \rightarrow i}$  is [20]

$$\hat{T}_{j \rightarrow i} = \frac{1}{\det \mathbf{C}} \cdot \sum_{v=1}^d \Delta_{jv} C_{v,di} \cdot \frac{C_{ij}}{C_{ii}}, \quad (3)$$

where  $C_{ij}$  is the sample covariance between  $x_i$  and  $x_j$ ,  $\Delta_{ij}$  are the cofactors of the matrix  $\mathbf{C} = (C_{ij})$ , and  $C_{i,dj}$  is the sample covariance between  $x_i$  and a series derived from  $x_j$  using the Euler forward differencing scheme:  $\dot{x}_{j,n} = (x_{j,n+k} - x_{j,n}) / (k\Delta t)$ , with being  $k \geq 1$  some integer. Equation (3) is rather concise in form, involving only the common statistics, i.e., sample covariances. The transparent formula makes causality analysis, which

otherwise would be complicated, very easy and computationally efficient. Note, however, that Equation (3) cannot replace Equation (2); it is merely the maximum likelihood estimator (mle) of the latter. Statistical significance tests can be performed for the estimators. This is achieved with the aid of a Fisher information matrix. See [19,20] for details.

Originally, the formalism was established within the framework of a differential system; in other words, it is with infinitesimal time increments. (The formalism with discrete mappings was also established by Liang (2016) [6], but no estimation has ever been made so far.) A question naturally arises about its applicability in the case of coarsely sampled time series. Indeed, it is not unusual that the given series may be coarsely sampled due to limited observations. As will be seen in the following section, this may pose a problem for nonlinear systems if the sample interval is large. This paper henceforth attempts to address this issue in the original linear framework. In the following, we first check the applicability of (3) for series from a linear system and a highly nonlinear system (Section 2), with a variety of sampling intervals. A new approach is presented in Section 3, which is then utilized to reconduct the causal inferences in Section 2. Some remaining issues are discussed in Section 5.

## 2. The Issue with Coarsely Sampled Series

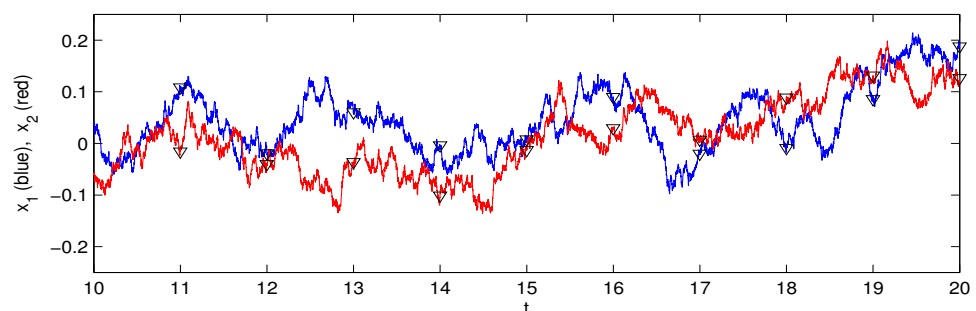
### 2.1. Time Series from Linear Systems

We first test the applicability of (3), as the sampling interval increases, with a well-studied linear system whose IF rates have been found half-analytically. This is the validation example in [19]:

$$\frac{dx_1}{dt} = -x_1 + 0.5x_2 + 0.1\dot{w}_1 \quad (4a)$$

$$\frac{dx_2}{dt} = -x_2 + 0.1\dot{w}_2 \quad (4b)$$

where  $\dot{w}_i$ ,  $i = 1, 2$  are components of independent white noise. It has been shown that the rates of information flow per unit time are  $T_{2 \rightarrow 1} \rightarrow 0.11$  as  $t \rightarrow \infty$ , and  $T_{1 \rightarrow 2} = 0$  for all  $t$ , reflecting accurately the one-way causality from  $x_2$  to  $x_1$ . Now, using the same sample path as that in Liang (2014) [19], we resample the series with low frequencies to obtain new series. (Note that, because of the pseudorandom number generator, the generated sample path using the normal differencing scheme may not be satisfactory. To see whether the obtained sample path is correct, one may check the resulting covariances, which can be rather accurately obtained by solving a deterministic ODE. Here, the data for the generated sample path can be downloaded from <http://www.ncoads.org/article/show/68.aspx> under the item PRE\_2014.dat (accessed on 1 January 2020).) Shown in Figure 1 is part of the sample path, with triangles marking the sampling points.



**Figure 1.** A segment of the sample path generated by Equation (4), with a time step  $\Delta t = 0.001$  (blue:  $x_1$ ; red:  $x_2$ ). The time spans from 0 to 100. The triangles mark the sampling points. As we are only interested in the limit IF as  $t \rightarrow \infty$ , the first 5000 steps, i.e., those for  $t \leq 5$ , are discarded.

**Table 1.** The estimated information flow rates within the linear system (4) for different sampling intervals (SIs). Also shown are the confidence intervals at a level of 90%.

| S.I. (# of Points)          | 1                  | 10                 | 50                 | 100                | 300                | 500                |
|-----------------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| $\hat{T}_{2 \rightarrow 1}$ | $0.110 \pm 0.051$  | $0.113 \pm 0.051$  | $0.099 \pm 0.050$  | $0.090 \pm 0.048$  | $0.055 \pm 0.036$  | $0.055 \pm 0.034$  |
| $\hat{T}_{1 \rightarrow 2}$ | $-0.002 \pm 0.056$ | $-0.001 \pm 0.055$ | $-0.015 \pm 0.054$ | $-0.011 \pm 0.053$ | $-0.008 \pm 0.038$ | $-0.015 \pm 0.034$ |

The computed IFs for different sampling intervals are listed in Table 1. We also compute the confidence intervals at a level of 90% (at a significance level of 0.1). First, the estimators  $\hat{T}_{2 \rightarrow 1}$  for all SIs here are significantly distinct from zero, while those in the opposite case,  $\hat{T}_{1 \rightarrow 2}$ , are not significant at a level of 90%. Thus, the causality in a qualitative sense can be faithfully recovered even with very low sampling frequencies (large SIs). (In fact, even with SI = 1000, the result is still correct; we do not consider cases beyond SI = 500 since the sample size is too small for SI > 500, resulting in insufficient statistics.)

Since this example actually has a half-analytical solution ( $T_{2 \rightarrow 1} \approx 0.11$ ,  $T_{1 \rightarrow 2} = 0$ ), further conclusions can be made about the computed results. Generally, the result of  $\hat{T}_{1 \rightarrow 2}$  appears satisfactory. For  $\hat{T}_{2 \rightarrow 1}$ , it is rather accurate for SI  $\leq 100$ . Beyond 100, it is no longer accurate.

## 2.2. Time Series from Synchronized Chaotic Oscillators

The following example is from the synchronization problem as examined by Paluš et al. [21]. The system is composed of two Rössler oscillators,  $\mathbf{x} = (x_1, x_2, x_3)$  and  $\mathbf{y} = (y_1, y_2, y_3)$ , where

$$\frac{dx_1}{dt} = -\omega_1 x_2 - x_3, \quad (5a)$$

$$\frac{dx_2}{dt} = \omega_1 x_1 + 0.15 x_2, \quad (5b)$$

$$\frac{dx_3}{dt} = 0.2 + x_3(x_1 - 10), \quad (5c)$$

is the master system, and

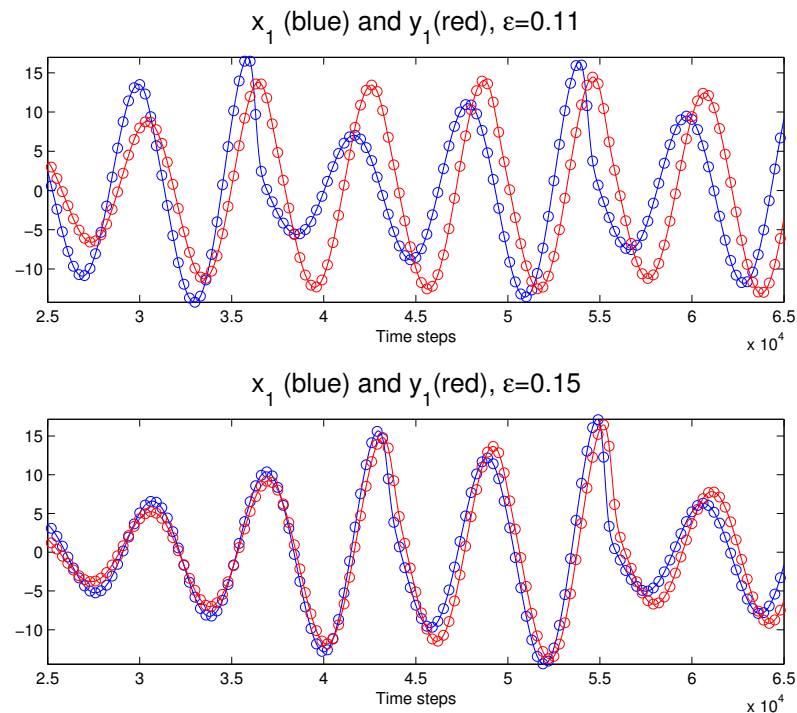
$$\frac{dy_1}{dt} = -\omega_2 y_2 - y_3 + \varepsilon(x_1 - y_1), \quad (6a)$$

$$\frac{dy_2}{dt} = \omega_2 y_1 + 0.15 y_2, \quad (6b)$$

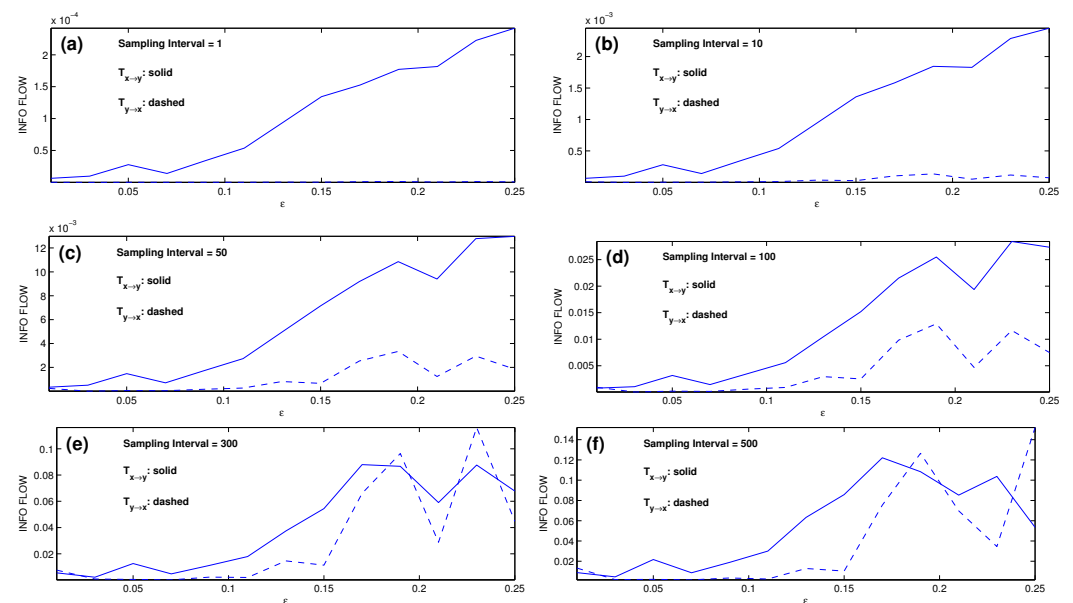
$$\frac{dy_3}{dt} = 0.2 + y_3(y_1 - 10), \quad (6c)$$

is the driven one. Following [21], we choose  $\omega_1 = 1.015$  and  $\omega_2 = 0.985$ . Using the Runge–Kutta scheme and choosing a time step  $\Delta t = 0.001$ , the coupled six-dimensional system can be solved rather accurately with different  $\varepsilon$ . Figure 2 plots the solutions of  $x_1$  and  $y_1$  when the coupling strength  $\varepsilon = 0.11$  (upper panel) and  $\varepsilon = 0.15$  (lower panel). As shown in the latter case, the two subsystems become synchronized if  $\varepsilon \geq 0.15$ . Again, we choose to study the problem for  $t \in [0, 100]$  ( $10^5$  time steps in total).

For each  $\varepsilon$ , we generate six time series of  $10^5$  steps and evaluate the IFs according to Equation (3) ( $k = 1$  is chosen). The IFs as functions of  $\varepsilon$  are then obtained, and they are plotted in Figure 3a, which accurately shows that the master is  $\mathbf{x}$ , and  $\mathbf{y}$  is the slave. An interesting observation is that this causality inference works even when the two oscillators are nearly synchronized as  $\varepsilon > 0.15$ , demonstrating the power of this rigorously formulated causality analysis.



**Figure 2.** Part of the time series of  $x_1$  and  $y_1$  of the coupled Rössler systems for  $\varepsilon = 0.11$  (**top**) and  $\varepsilon = 0.15$  (**bottom**). The circles indicate the sampling points (every 300 steps here, corresponding to 18 points in each period). The two oscillators become synchronized as  $\varepsilon \geq 0.15$ .



**Figure 3.** Absolute information flow rates  $|\hat{T}_{x \rightarrow y}|$  and  $|\hat{T}_{y \rightarrow x}|$  (dashed) between the two Rössler oscillators as functions of the coupling coefficient  $\varepsilon$  with different sampling intervals ((a)–(f) correspond to SI = 1, 10, 50, 100, 300, 500, respectively). Units are nats per unit time. With the preset causality, the dashed line should coincide with the abscissa.

We subsample the series at every SI step, SI = 10, 50, 100, 300, 500, and recompute the computation using the same scheme. The resulting IF rates are shown in Figure 3b–f, respectively. With the preset causality, the dashed line should be the zero line. Clearly, the causal inference works well for  $SI \leq 10$ . The computed IF becomes biased for  $SI \geq 50$ , and the bias grows significantly as the SI increases. If we focus on  $\varepsilon \leq 0.15$ , i.e., when the systems are not synchronized (see [21]), the causal inference still functions well for  $SI \leq 50$ .

If the synchronized cases are taken into account ( $\varepsilon > 0.15$ ), then the inferences in the cases for  $50 \leq \text{SI} \leq 100$  are significantly biased, and those for SIs exceeding 300, corresponding to an approximate sampling frequency of 20 per period, are no longer correct.

### 3. Approaching a Partial Solution

As shown above, if the sampling frequency of the time series is low, the resulting linear IF for nonlinear series may be biased. Indeed, in the case with high nonlinearity, the linear assumption is always easy to blame. While theoretically it is not a problem (causality is guaranteed, as it is proven in a theorem), we agree that, before a fully nonlinear algorithm is developed, this will be a continuing issue. What we want to show here is, how much room there is for improvement. At present, the algorithm documented in [19], and later in [20], is based on the Bernstein–Euler differencing scheme, which is, of course, very rudimentary due to the first-order differencing. If a time series is coarsely sampled, the error could be large.

A theorem established by Liang (2008) [7] reads that, if the noise is additive in Equation (1), i.e., if  $\mathbf{B}$  is a constant matrix, then the noise itself does not appear in the formula of  $T_{j \rightarrow i}$ . Thus, under the additive noise assumption, we can estimate the IF within the framework of a deterministic system. In this case, note that the linear equation set can be solved for an interval  $[t, t + \Delta t]$ , regardless of the size of  $\Delta t$ . This gives insight regarding a solution to the low sampling frequency problem.

Consider

$$\frac{d\mathbf{x}}{dt} = \mathbf{f} + \mathbf{A}\mathbf{x}, \quad (7)$$

where  $\mathbf{A} = (a_{ij})$  is a  $d \times d$  matrix. Let us assume that  $\mathbf{f} = \mathbf{0}$ , since the time series can always be pretreated by removing the linear trend, and it has been proven that this removal does not alter the IF rates. In this case, on the interval  $[t, t + \Delta t]$ , we actually have a mapping  $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}^d$  that takes the state  $\mathbf{x}(t)$  to the state  $\mathbf{x}(t + \Delta t)$  at  $t + \Delta t$ , with the propagating operator

$$\Phi = e^{\mathbf{A}\Delta t} = e^{\begin{bmatrix} a_{11} & \cdots & a_{1d} \\ \vdots & \ddots & \vdots \\ a_{d1} & \cdots & a_{dd} \end{bmatrix} \Delta t} \equiv \begin{bmatrix} \alpha_{11} & \cdots & \alpha_{1d} \\ \vdots & \ddots & \vdots \\ \alpha_{d1} & \cdots & \alpha_{dd} \end{bmatrix}. \quad (8)$$

It is not easy to estimate  $a_{ij}$ , but it is easy to estimate  $\alpha_{ij}$  instead, by observing the relation

$$\begin{bmatrix} \alpha_{11} & \cdots & \alpha_{1d} \\ \vdots & \ddots & \vdots \\ \alpha_{d1} & \cdots & \alpha_{dd} \end{bmatrix} \mathbf{x}(n) = \mathbf{x}(n+1), \quad n = 0, 1, 2, \dots, N. \quad (9)$$

This is written in matrix form as

$$\begin{bmatrix} x_1(0) & \cdots & x_d(0) \\ \vdots & \ddots & \vdots \\ x_1(N-1) & \cdots & x_d(N-1) \end{bmatrix} \begin{bmatrix} \alpha_{i1} \\ \vdots \\ \alpha_{id} \end{bmatrix} = \begin{bmatrix} x_i(1) \\ \vdots \\ x_i(N) \end{bmatrix}$$

for  $i = 1, \dots, d$ . Averaging all rows of the algebraic equation set, and subtracting the mean from each row, we get

$$\begin{bmatrix} x_1(0) - \bar{x}_1 & \dots & x_d(0) - \bar{x}_d \\ \vdots & \ddots & \vdots \\ x_1(N-1) - \bar{x}_1 & \dots & x_d(N-1) - \bar{x}_d \end{bmatrix} \begin{bmatrix} \alpha_{i1} \\ \vdots \\ \alpha_{id} \end{bmatrix} = \begin{bmatrix} x_i(1) - \bar{x}_{i+} \\ \vdots \\ x_i(N) - \bar{x}_{i+} \end{bmatrix},$$

where  $\bar{x}_i = \frac{1}{N} \sum_{n=0}^{N-1} x_i(n)$ ,  $\bar{x}_{i+} = \frac{1}{N} \sum_{n=1}^N x_i(n)$ , i.e., the series  $\{x_{i+}(n)\}$  is the series  $\{x_i(n)\}$  advanced by one step. Let  $i$  run through  $\{1, 2, \dots, d\}$ . We have the following  $d$  overdetermined equation sets:

$$\begin{bmatrix} x_1(0) - \bar{x}_1 & \dots & x_d(0) - \bar{x}_d \\ \vdots & \ddots & \vdots \\ x_1(N-1) - \bar{x}_1 & \dots & x_d(N-1) - \bar{x}_d \end{bmatrix} \begin{bmatrix} \alpha_{11} & \dots & \alpha_{d1} \\ \vdots & \ddots & \vdots \\ \alpha_{1d} & \vdots & \alpha_{dd} \end{bmatrix} = \begin{bmatrix} x_1(1) - \bar{x}_{1+} & \vdots & x_d(1) - \bar{x}_{d+} \\ \vdots & \ddots & \vdots \\ x_1(N) - \bar{x}_{1+} & \vdots & x_d(N) - \bar{x}_{d+} \end{bmatrix}. \quad (10)$$

Denote by  $\mathbf{\Lambda}$  the matrix  $(\alpha_{ij})$ ; then, the matrix of unknowns in the above equation sets is  $\mathbf{\Lambda}^T$ . Left multiplication by

$$\begin{bmatrix} x_1(0) - \bar{x}_1 & \dots & x_d(0) - \bar{x}_d \\ \vdots & \ddots & \vdots \\ x_1(N-1) - \bar{x}_1 & \dots & x_d(N-1) - \bar{x}_d \end{bmatrix}^T$$

on both sides yields  $d \times d$  equation sets:

$$\mathbf{C}\mathbf{\Lambda}^T = \tilde{\mathbf{C}}, \quad (11)$$

where  $\mathbf{C} = (C_{ij})$  is the sample covariance matrix of  $\mathbf{x}$ ,  $\tilde{\mathbf{C}} = (C_{i,j+})$ , and  $C_{i,j+}$  is the sample covariance between  $x_i$  and  $x_{j+}$ , i.e.,  $x_j$  advanced by one time step. The least-square solutions of the overdetermined sets (10) are the solutions of (11),

$$\mathbf{\Lambda}^T = \mathbf{C}^{-1}\tilde{\mathbf{C}},$$

and hence

$$\mathbf{\Lambda} = (\mathbf{C}^{-1}\tilde{\mathbf{C}})^T = \tilde{\mathbf{C}}^T\mathbf{C}^{-1}.$$

The estimator of  $\mathbf{\Lambda}$  is, therefore,

$$\hat{\mathbf{\Lambda}} = \frac{1}{\Delta t} \log(\tilde{\mathbf{C}}^T\mathbf{C}^{-1}). \quad (12)$$

(Caution should be used in cases of singularity. The irrelevant imaginary part also should be discarded.)

Once we have obtained  $\mathbf{\Lambda}$ , and hence the coefficients  $(a_{ij})$ , we substitute  $a_{ij}$  for the whole part

$$\frac{1}{\det \mathbf{C}} \sum_{k=1}^d \Delta_{jk} C_{k,di}$$

in Equation (3), i.e., we multiply  $a_{ij}$  by  $C_{ij}/C_{ii}$  to arrive at the desideratum,  $\hat{T}_{j \rightarrow i}$ . If we denote by  $[\mathbf{A}]_{ij}$  the extraction of the  $(i, j)^{th}$  entry of the matrix  $\mathbf{A}$ , this is

$$\hat{T}_{j \rightarrow i} = \frac{1}{\Delta t} \left[ \log(\tilde{\mathbf{C}}^T \mathbf{C}^{-1}) \right]_{ij} \cdot \frac{C_{ij}}{C_{ii}}. \quad (13)$$

(Note here that  $\log$  is the matrix logarithm. In MATLAB, the function is `logm`.)

#### 4. The Coarsely Sampled Series Problem Revisited

As demonstrated above, for series generated from linear systems, the estimation of the IF is satisfactory qualitatively. Nevertheless, we want to examine how the new scheme may improve the results. Shown in Table 2 is a recalculation of the estimates. Since this case has a ground truth (half analytical) ( $\approx 0.11$  nats per unit time), we can see that the result is accurate enough for all SIs.

**Table 2.** Estimation of the information flow rates within the system (4) with the new scheme.

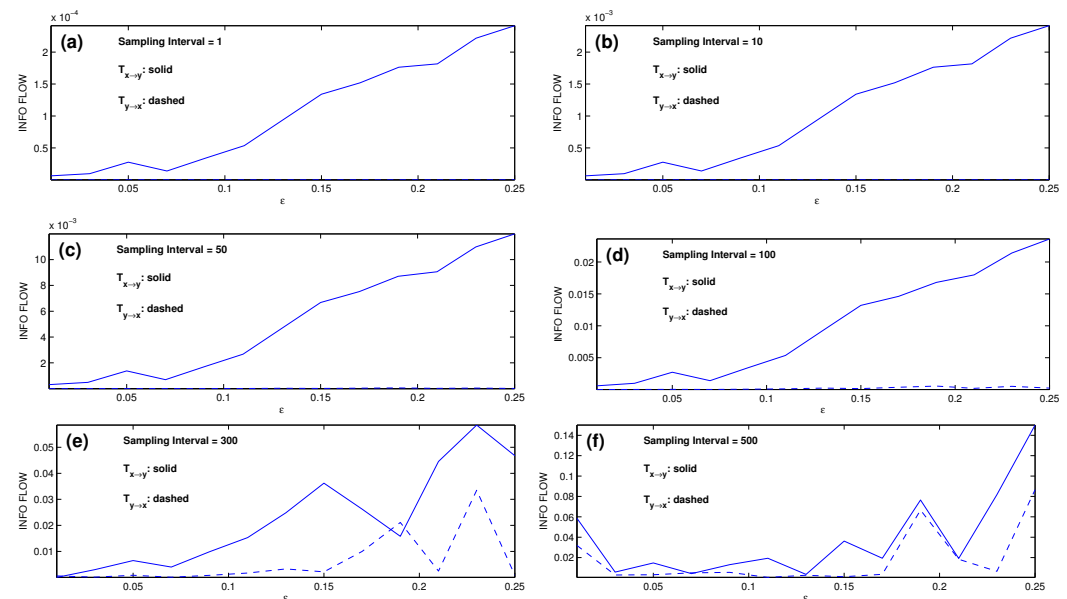
| Sampling Interval           | 1     | 10    | 50     | 100    | 300    | 500    |
|-----------------------------|-------|-------|--------|--------|--------|--------|
| $\hat{T}_{2 \rightarrow 1}$ | 0.114 | 0.118 | 0.109  | 0.106  | 0.082  | 0.098  |
| $\hat{T}_{1 \rightarrow 2}$ | 0.007 | 0.008 | −0.007 | −0.002 | −0.002 | −0.015 |

The new scheme for the estimation is particularly satisfactory for the nonlinear case. For the pair of Rössler oscillators, the computed results are plotted in Figure 4. Compared to Figure 3, it can be seen that the performance is significantly improved. Consider the cases with  $SI \leq 100$  first. To see the improvement more clearly, we introduce a ratio  $r = |T_{y \rightarrow x}|/|T_{x \rightarrow y}|$  to measure the performance of the one-way causal inference; the smaller the value, the more accurate the result, with  $r < 0.1$  taken as insignificant. By this standard, the cases with  $SI \leq 10$  (Figure 3a,b) appear satisfactory, but the causal relations as shown in Figure 3c,d are inaccurate or even incorrect. Specifically, for  $\varepsilon > 0.15$ , i.e., when the system is synchronized,  $r_{\max}$  reaches  $\frac{1}{3}$  and  $\frac{1}{2}$  for the cases of  $SI = 50$  and  $SI = 100$ , respectively. For  $\varepsilon \leq 0.15$ , the inference appears satisfactory, albeit inaccurate. However, at  $\varepsilon = 0.01$ ,  $r \approx 1$ , the inference is incorrect. In contrast, the inferred causalities in Figure 4a–d are rather accurate for the coupling strengths  $\varepsilon$  considered (both synchronized and nonsynchronized), with all the  $r$ 's being insignificant.

For the case  $SI = 300$ , which corresponds to a sampling frequency of 20 points per period, the one-way causality is accurately recovered for the nonsynchronized cases ( $\varepsilon \leq 0.15$ ). However, beyond this, i.e.,  $\varepsilon > 0.15$ , the inference fails. In particular, when  $SI = 500$  (Figure 4f), the result is even worse than its counterpart obtained with the traditional scheme as plotted in Figure 3f. This, of course, may be due to the resulting small ensemble size, which causes singularity to the matrix logarithm. Consider the extreme limit when the series are completely synchronized. In this case, it is not possible to determine whether there exists a causal relation or not, as they are identical, impossible to be differentiated. Correspondingly, this means a singular covariance matrix. Now, in the above problem, the series are nearly synchronized. Given the length, the size of the ensemble thus formed reduces as the sampling interval increases. This tends to increase the condition number of the resulting covariance matrix. Whenever a matrix is ill-conditioned, its logarithm is very sensitive, leading to large errors in the subsequent computation.

Nonetheless, the success in applying the linear estimator to such a highly nonlinear system is remarkable. The reason for this is probably the same as that for linearization; that is to say, on a small interval, a linearized system can provide a good approximation to an otherwise nonlinear system. While its applicability is yet to be proven, or investigated with more nonlinear problems, by our experience, it indeed works well, provided that

the statistics are sufficient (stochastic systems or chaotic deterministic systems) and the sampling interval is not too large.



**Figure 4.** As in Figure 3, but the information flow rates are computed with the new scheme. Subfigures (a)–(f) correspond to the cases with SI = 1, 10, 50, 100, 300, 500, respectively.

## 5. Discussion

The maximum likelihood estimator of the information flow (IF), i.e., Equation (3), provides an easy way toward causal inference. Theoretically, it is based on a linear assumption, but, practically, it has shown considerable success with series generated from highly nonlinear systems; anyhow, linearization piecewise in time proves to be an efficient asymptote to an otherwise nonlinear system. In reality, series may be coarsely sampled; the time resolution may be low. An issue thus arises, as this formalism is theoretically on the basis of infinitesimal time increments. In this case, as we have shown, it still works for linear systems in a qualitative sense; however, for a highly nonlinear system composed of two Rössler oscillators, the bias becomes increasingly significant as the sampling frequency is reduced.

A new scheme has been proposed to address this problem and provide a partial solution. Due to a property of the IF, as proven in [7], that additive noise does not alter the IF in form, it is reasonable to directly estimate the IF without paying attention to the stochasticity. Instead of estimating through the vector field using Euler–Bernstein differencing, we choose to consider the propagator on the finite time interval, i.e., to estimate the Lie group members. In doing so, the original Formula (3), which is rewritten here for ease of reference,

$$\hat{T}_{j \rightarrow i} = \frac{1}{\det \mathbf{C}} \cdot \sum_{v=1}^d \Delta_{jv} C_{v,di} \cdot \frac{C_{ij}}{C_{ii}},$$

is replaced with (13),

$$\hat{T}_{j \rightarrow i} = \frac{1}{\Delta t} \left[ \log(\tilde{\mathbf{C}}^T \mathbf{C}^{-1}) \right]_{ij} \cdot \frac{C_{ij}}{C_{ii}},$$

where  $\tilde{\mathbf{C}} = (C_{i,j+})$ , and  $C_{i,j+}$  is the sample covariance between  $x_i$  and  $x_{j+}$ , i.e.,  $x_j$  advanced by one time step. Note that here log is the matrix logarithm; in MATLAB (Version 23.2 or earlier), the function is logm. (The spurious imaginary part, if it arises, should be discarded.)

In this way, the preset causality within the coupled system of chaotic oscillators can be rather accurately reproduced even when the sampling interval is large (sampling frequency is low). For convenience, this is summarized as an algorithm (Algorithm 1).

---

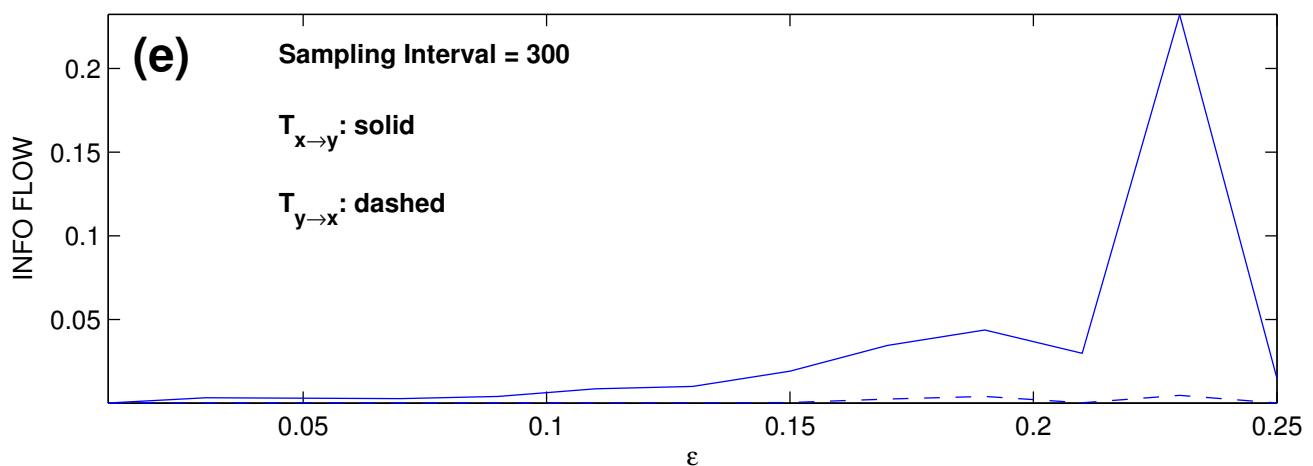
**Algorithm 1:** Causal inference through information flow estimation

---

**input:**  $d$  time series and time step size  $\Delta t$   
**output:** a causal graph  $\mathcal{G} = (V, E)$  and IFs along edges  
 evaluate the covariance matrices  $\mathbf{C}$  and  $\tilde{\mathbf{C}}$ ;  
 compute  $\mathbf{P} = \log(\tilde{\mathbf{C}}^T \mathbf{C}^{-1})$ ;  
**for each**  $(i, j) \in V \times V$  **do**  
   compute  $\hat{T}_{i \rightarrow j}$  by (13), i.e.,  
      $\hat{T}_{j \rightarrow i} = \frac{1}{\Delta t} P_{iC_{ii}}^{C_{ij}}$ ;  
**end**  
 return  $\mathcal{G}$ , together with the IFs  $\hat{T}_{j \rightarrow i}$

---

There is still much room for improvement in the above approach. For example, the estimation of the covariances in the quotient  $\sigma_{ij}/\sigma_{ii}$  is achieved by replacing the population covariances with sample covariances, while the sample is formed from the time series. While this is satisfactory for stochastic systems under the ergodic assumption, this may not be suitable for deterministic chaos, such as in the case of the Rössler oscillators studied here. The reason is obvious: the time mean of the series in Figure 2 is zero, but one can imagine that the ensemble mean of all possible paths is by no means zero; rather, it should be a function of time (like the series itself), which may be close to the asymptotic linear system solution. Thus, it is more reasonable to treat the linear system solution as the mean. As such, we have attempted to improve the estimation by replacing the covariances of  $\mathbf{x}$  with those of  $\mathbf{x} - \bar{\mathbf{x}}$ , where  $\bar{\mathbf{x}}$  stands for the resulting linear system solution. With this, we obtain another causal inference result for SI = 300; the resulting IFs are plotted in Figure 5. As one can see, the result appears rather accurate, as expected, in contrast to Figure 4e.



**Figure 5.** As in Figure 4e, but the covariances are estimated using the residuals of  $x_i$  relative to the model result, instead of  $x_i$  themselves. Here, SI = 300 implies a sampling frequency of 18 points each period. The label “(e)” in the figure corresponds to the same label in Figure 4e.

We, however, do not claim that we have solved the problem. What we want to show here is, how much room there is for improvement within a linear framework in estimating Equation (2), the original fully nonlinear formula  $T_{j \rightarrow i}$ . Indeed, in the case of high nonlinearity, the linear assumption is always easy to blame. While theoretically it is not a problem (causality is guaranteed as proven in a theorem; see [6] and other references),

it is believed that, before a fully nonlinear algorithm is developed (when this paper was prepared, ref. [22] was not published), this will be a continuing issue.

It should be pointed out that the above algorithm works only when the sampling interval does not exceed the scale in question. For processes with multiple scales involved, different samplings may result in different physically meaningful information flow rates. In this case, it is not just a computational issue any more; it is a physical problem that deserves for an in-depth investigation. We will leave this problem, among others, for future research.

**Funding:** This research was funded by National Science Foundation of China (Grant No. 42230105), Shanghai Institute for Mathematics and Interdisciplinary Sciences (Grant No. SIMIS-ID-2025-CA), Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (Grant Nos. SML2023SP203, 221425004, 221425008), and Fudan University (No. IDH2318027Y).

**Data Availability Statement:** The codes are available, and will be updated, at [www.ncoads.org/article/show/67.aspx](http://www.ncoads.org/article/show/67.aspx) (accessed on 1 January 2020).

**Acknowledgments:** The author is grateful to Milan Paluš for his inspiring question, which motivated this research.

**Conflicts of Interest:** The author declares no conflicts of interest.

## References

1. Granger, C.W.J. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **1969**, *37*, 424–438. [\[CrossRef\]](#)
2. Pearl, J. *Causality: Models, Reasoning, and Inference*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2000; 400p.
3. Rubin, D.B. Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Edu. Psych.* **1974**, *66*, 688–701. [\[CrossRef\]](#)
4. Imbens, G.W.; Rubin, D.B. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*; Cambridge University Press: Cambridge, UK, 2015. [\[CrossRef\]](#)
5. Bollt, E.M.; Sun, J.; Runge, J. Introduction to Focus Issue: Causation inference and information flow in dynamical systems: Theory and applications. *Chaos* **2018**, *28*, 075201. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Liang, X.S. Information flow and causality as rigorous notions ab initio. *Phys. Rev. E* **2016**, *94*, 052201. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Liang, X.S. Information flow within stochastic dynamical systems. *Phys. Rev. E* **2008**, *78*, 031113. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Lasota, A.; Mackey, M.C. *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*, 2nd ed.; Springer: New York, NY, USA, 1994; 472p.
9. Bollt, E.M. Open or closed? Information flow decided by transfer operators and forecastability quality metric. *Chaos* **2018**, *28*, 075309. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Liang, X.S.; Chen, D.; Zhang, R. Quantitative causality, causality-aided discovery, and causal machine learning. *Ocean-Land-Atmos. Res.* **2023**, *2*, 26. [\[CrossRef\]](#)
11. Doquier, D.; Massonnet, F.; Richaud, B.; Fichet, T.; Goosse, H.; Messina, B.; Topál, D.; Vannitsem, S. Drivers of summer Antarctic sea-ice extent at interannual time scale in CMIP6 large ensembles based on information flow. *Clim. Dyn.* **2025**, *63*, 10. [\[CrossRef\]](#)
12. Vannitsem, S.; Dalaiden, Q.; Goosse, H. Testing for dynamical dependence—Application to the surface mass balance over Antarctica. *Geophys. Res. Lett.* **2019**, *46*, 12125–12135. [\[CrossRef\]](#)
13. Hristopoulos, D.T.; Babul, A.; Babul, S.A.; Brucar, L.R.; Virji-Babul, N. Disrupted information flow in resting-state in adolescents with sports related concussion. *Front. Hum. Neurosci.* **2019**, *13*, 419. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Cong, J.; Zheng, W.; Liu, Y.; Yin, S.; Jia, H.; Yi, C.; Chen, K.; Xue, K.; Li, F.; Yao, D.; et al. Altered default mode network causal connectivity patterns in autism spectrum disorder revealed by Liang information flow analysis. *Hum. Brain Mapp.* **2023**, *44*, 2279–2293. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Lu, X.F.; Sun, J.; Guo, W.; Chang, C.-T. Causal interactions and financial contagion among the BRICS stock markets under rare events: A Liang causality analysis. *Int. J. Emerging Markets* **2025**, *20*, 2014–2041. [\[CrossRef\]](#)
16. Ghosh, R.; Yi, B.; Bose, S. Quantum Liang information flow probe of causality across critical points. *Phys. Rev. Lett.* **2025**, *134*, 150202. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Roy, S.; Jantzen, B. Detecting causality using symmetry transformations. *Chaos* **2018**, *28*, 075305. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Berkeley, G. *A Treatise Concerning the Principles of Human Knowledge*; Aaron Rhames: Dublin, Ireland, 1710.
19. Liang, X.S. Unraveling the cause-effect relation between time series. *Phys. Rev. E* **2014**, *90*, 052150. [\[CrossRef\]](#) [\[PubMed\]](#)

20. Liang, X.S. Normalized multivariate time series causality analysis and causal graph reconstruction. *Entropy* **2021**, *23*, 679. [[CrossRef](#)] [[PubMed](#)]
21. Paluš, M.; Krakovská, A.; Jakubík, J.; Chvosteková, M. Causality, dynamical systems and the arrow of time. *Chaos* **2018**, *28*, 075307. [[CrossRef](#)] [[PubMed](#)]
22. Pires, C.A.; Docquier, D.; Vannitsem, S. A general theory to estimate information transfer in nonlinear systems. *Phys. D* **2023**, *458*, 133988. [[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.