XXXVII SIMPÓSIO BRASILEIRO DE TELECOMUNICAÇÕES E PROCESSAMENTO DE SINAIS - SBrT2019, 29/09/2019–02/10/2019, PETRÓPOLIS, RJ

Estimation of Transfer Entropy for Capturing Connectivity and Causality in Industrial Processes

Milena Marinho Arruda, Francisco Marcos de Assis and Luciana Ribeiro Veloso

***Abstract*— This paper discusses the use of two information theory measures: transfer entropy and direct transfer entropy, as an approach for detection of causality and connectivity between continuous variables in industrial processes (continuous systems). These measures are asymmetric and identify and quantify linear or non-linear directional relationships between two variables. To estimate these measures, we used estimators based on distances between neighbors. The results obtained from simulations demonstrate the applicability of the measurements and their estimations in order to identify the connectivity map of two systems: autoregressive model (which can be compared with the analytical results) and four water tanks (an industrial systems).**

***Keywords*— Transfer entropy, causality, continuous processes, industrial processes.**

I. Introduction

Identifying the causal relationships between variables of a system from observations of time series is challenging when knowledge about the dynamics is partial. Although these relationships can be detected through mathematical modeling, for complex large-scale processes it is difficult to establish practical and precise mathematical models.

Mechanisms to automate the detection and diagnosis of plant wide abnormalities and disorders are challenges in the process industry. In a complex industrial process, the elements are not only interconnected, they are also mutually dependent [1], [2]. Therefore, it is possible to use the concepts of information theory to describe cause-effect relationships between variables through directional graphs, establishing a network called a causal map [3].

Paying attention to the detection of abnormalities in industrial processes is one way to avoid interfering with the overall process performance. So, identifying the spread of failures can prevent accidents that often yield both financial and human damages.

The concepts of information theory can also be used in this application, among them, directional information and transfer entropy. They have one important advantage: quantifies the directional causal influence between two processes. This analysis can support the diagnosis of abnormalities and disorders throughout the plant [1], [2], [3], [4], [5] and it has been recently used in different scientific fields, such as neuroscience [6], [7] and biomedical data [8], [9].

Milena Marinho Arruda is with the Department of Electrical Engineering, UFCG, Brazil, e-mail: milena.arruda@ee.ufcg.edu.br; Francisco Marcos de Assis is with the Department of Electrical Engineering, UFCG, Brazil, e-mail: fmarcos@dee.ufcg.edu.br; Luciana Ribeiro Veloso is with the Department of Electrical Engineering, UFCG, Brazil, e-mail: luciana.veloso@dee.ufcg.edu.br. This work was partially supported by CNPq.

Although the directional causal influence between two process can be detected using, for example, transfer entropy, the challenge is to distinguish whether the causal influence is direct or indirect. An indirect influence occurs when some intermediate variables transfer information between the two processes. Therefore, in a direct causality there is a direct information without any intermediate variables.

In addition, the processes can be related to a common source (spurious causality). In this case, if the process one causes process two and process one causes process three, the processes two e three can be related due to the influence of process one in both cases. Thus, to detect whether causality is true or spurious and/or direct or indirect is necessary for capturing the true process connectivity.

Therefore, an extension of the transfer entropy was proposed by [5] to detect the direct causality between two processes: direct transfer entropy. In general, transfer entropy represents the total causal influence and direct transfer entropy determine whether this influence is along direct or indirect pathways.

Although the concepts in information theory are relatively simple and mathematical formulations are objective, in practice, their estimation can be a complex process. For continuous variables, techniques that use nearest neighbors distances have been used for transfer entropy estimates [8], [10].

In such context, this paper attempts to investigate the use of estimator based in nearest neighbors distances for transfer entropy and direct transfer entropy and then construct directional graphs for information flow. First, an example base with a three dimensional autoregressive model will analyze the perform of estimator comparing with theoretical values. Next, an application with four water tank will be evaluated.

This paper is structured as follows. Section II organizes notations. Section III reviews definitions of transfer entropy and direct transfer entropy. Section IV introduces briefly estimators for these measures based in nearest neighbors distances. Section V presents the performed simulations. Finally, section VI concludes the paper.

II. Notation and Terminology

In this paper, we denote random variables by uppercase letters, their realizations by lowercase letters, stochastic processes by uppercase bold letters and realizations of a d-dimensional random variables by lowercase bold letters. The n-th output the process is indicated by subscripts, e. g., X_n. The finite length sequence of a random variable is defined by subscript and superscript, e. g., $X_{n-k}^{n} = \{X_{n-k}, \cdots, X_n\}$. Probability density function is denoted by $f(\cdot)$ and the set where $f(\cdot) > 0$ is called the support set.

III. Transfer Entropy and Direct Transfer Entropy

Transfer entropy is a nonlinear measure of information theory introduced by Thomas Schreiber [11]. It quantifies the exchange of information, depending on directional coupling, between two random processes when they are not independent and can be approximated by a stationary Markov process. In other words transfer entropy is an asymmetric measure which can determine linear or nonlinear coupling of two variables by quantifying the information transferred between them.

Initially, transfer entropy was defined for discrete processes [11] and extended for continuous processes [12]. Throughout this article, this measure will be used to detect causal relationships and information flow for continuous processes.

An interpretation about transfer entropy is: measure of the amount of information that future samples of one process contains about past samples of another process given the past samples of the first.

Before introducing its mathematical interpretation it is necessary to define the concept of entropy. For discrete random variables, entropy is a measure of the average uncertainty and the number of bits on average required to describe them. When the random variable is continuous we have differential entropy. It is related to the shortest description length of these variables. Considering a continuous random variable X with density $f(x)$ differential entropy is defined as [13],

$$h(X) = -\int_S f(x) \log f(x) dx, \tag{1}$$

where S is the support set of the random variable. Unlike entropy for discrete variables, differential entropy can be negative.

Now let's consider two continuous processes, the transfer entropy between them is defined as [12],

$$T_{\mathbf{X}\to\mathbf{Y}} = \int f\left(Y_{n+1}, Y_{n-l+1}^{n}, X_{n-k+1}^{n}\right) \log \frac{f\left(Y_{n+1}|Y_{n-l+1}^{n}, X_{n-k+1}^{n}\right)}{f\left(Y_{n+1}|Y_{n-l+1}^{n}\right)} d\mathbf{V}, \tag{2}$$

where $\mathbf{V}$ denotes the random vector $\left[Y_{n+1}, Y_{n-l+1}^{n}, X_{n-k+1}^{n}\right]$ and k and l are the embedding dimension of $\mathbf{X}$ and $\mathbf{Y}$ respectively. In this paper $k = l = 1$.

However, the transfer entropy is not sufficient to detect the connectivity. In order to detect the direct causality between two process with some possible intermediate variables, the direct transfer entropy should be used and it is defined as [5],

$$D_{\mathbf{X}\to\mathbf{Y}} = \int f\left(Y_{n+1}, Y_{n-l+1}^{n}, Z_{n-j_1+1}^{n}, \cdots, Z_{n-j_q+1}^{n}, X_{n-k+1}^{n}\right) \log \frac{f\left(Y_{n+1}|Y_{n-l+1}^{n}, Z_{n-j_1+1}^{n}, \cdots, Z_{n-j_q+1}^{n}, X_{n-k+1}^{n}\right)}{f\left(Y_{n+1}|Y_{n-l+1}^{n}, Z_{n-j_1+1}^{n}, \cdots, Z_{n-j_q+1}^{n}\right)} d\mathbf{W} \tag{3}$$

where $\mathbf{W}$ denotes the random vector $\left[Y_{n+1}, Y_{n-l+1}^{n}, Z_{n-j_1+1}^{n}, \cdots, Z_{n-j_q+1}^{n}, X_{n-k+1}^{n}\right]$, q is the number of intermediate processes and k, l and $j_1, \cdots, j_q$ are the embedding dimension of $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z_1}, \cdots, \mathbf{Z_q}$ respectively. In this paper $k = l = j_1 = \cdots = j_q = 1$.

Transfer entropy and direct transfer entropy can be rewritten as a sum of entropies or a conditioned mutual information. This will be useful during the study of estimators. To the first we have,

$$\begin{aligned} T_{\mathbf{X}\to\mathbf{Y}} &= h\left(X_{n-k+1}^{n}, Y_{n-l+1}^{n}\right) + h\left(Y_{n+1}^{n}, Y_{n-l+1}^{n}\right) \\ &\quad - h\left(Y_{n+1}, X_{n-k+1}^{n}, Y_{n-l+1}^{n}\right) - h\left(Y_{n-l+1}^{n}\right) \\ &= I(Y_{n+1}; X_{n-k+1}^{n}|Y_{n-l+1}^{n}), \end{aligned} \tag{4}$$

and, similarly, can be extended to direct transfer entropy,

$$\begin{aligned} D_{\mathbf{X}\to\mathbf{Y}} &= h\left(X_{n-k+1}^{n}, Y_{n-l+1}^{n}, Z_{n-j+1}^{n}\right) \\ &\quad + h\left(Y_{n+1}, Y_{n-l+1}^{n}, Z_{n-j+1}^{n}\right) \\ &\quad - h\left(Y_{n+1}, X_{n-k+1}^{n}, Y_{n-l+1}^{n}, Z_{n-j+1}^{n}\right) \\ &\quad - h\left(Y_{n-l+1}^{n}, Z_{n-j+1}^{n}\right) \\ &= I(Y_{n+1}; X_{n-k+1}^{n}|Y_{n-l+1}^{n}, Z_{n-j+1}^{n}). \end{aligned} \tag{5}$$

Suppose three continuous process where $T_{\mathbf{X}\to\mathbf{Y}}$, $T_{\mathbf{X}\to\mathbf{Z}}$, and $T_{\mathbf{Z}\to\mathbf{Y}}$ are all larger than zero then, $\mathbf{X}$ causes $\mathbf{Y}$, $\mathbf{X}$ causes $\mathbf{Z}$, and $\mathbf{Z}$ causes $\mathbf{Y}$. However, this result does not indicate whether the information flow from $\mathbf{X}$ to $\mathbf{Y}$ is along a direct or indirect pathway, in other words, if $\mathbf{Z}$ is an intermediate variable. For check that, the direct transfer entropy can be used. If $D_{\mathbf{X}\to\mathbf{Y}}$ is greater than zero, then there is a direct pathway between the process, otherwise, the causal influence is via the intermediate variable.

In addition, if there is direct causality from $\mathbf{X}$ to $\mathbf{Y}$ it is possible that $\mathbf{Z}$ is not a cause of $\mathbf{Y}$ and this spurious causality is generated by a common source of both, $\mathbf{X}$. Therefore, $D_{\mathbf{Z}\to\mathbf{Y}}$ needs to be calculated. After this procedures, the connectivity of the system can be reproduced in a directional causal map.

The direct transfer entropy represents the information about a future observation of a process $\mathbf{Y}$ obtained from the simultaneous observation of past values of another process $\mathbf{X}$ and possible intermediate processes $\mathbf{Z}$, after discarding the information about the future of $\mathbf{Y}$ obtained from the past of $\mathbf{Z}$ alone [5].

IV. Estimation of Transfer Entropy

The process of estimating Information Theory measures for continuous variables requires attention, simple solutions such as discretization of the data can compromise the estimates.

Among four estimators discussed in [14], the differential entropy estimator proposed by Kozachenko and Leonenko (1987) [15] was the one that presented the best performance for entropy estimation. This estimator considers that the probability distribution of the distances between x_i and its k-th nearest neighbor is trinomial and estimates the Shannon entropy by,

$$\widehat{h}(X) = -\psi(k) + \psi(N) + \log c_d + \frac{d}{N}\sum_{i=1}^{N} \log \epsilon(i), \tag{6}$$

where $\psi(\cdot)$ is digamma function, N is the sample size, c_d is the volume of the d-dimensional unit ball (for the maximum norm $c_d = 2^n$ and for Euclidean norm $c_d = \pi^{d/2}/\Gamma(1+d/2)$) and $\epsilon(i)$ is twice the distance from x_i to its k-th neighbor.

Two classes of improved estimators for mutual information based on Kozachenko and Leonenko estimator was presented by [16]. We have concentrated our applications using only the estimator given by,

$$\widehat{I}(X;Y) = \psi(k) - \langle \psi(n_x+1) + \psi(n_y+1) \rangle + \psi(N), \quad (7)$$

where $\langle \cdot \rangle$ denotes the mean and n_x and n_y are the number of samples within the region delimited by k-th nearest neighbor.

Using the same idea, it was possible develop an estimator for transfer entropy as [17],

$$\begin{aligned} \widehat{T}_{\mathbf{X}\to\mathbf{Y}} &= \widehat{I}(X;Y|Z) \\ &= \psi(k) + \langle \psi(n_z+1) \\ &\quad - \psi(n_{xz}+1) - \psi(n_{yz}+1) \rangle, \end{aligned} \quad (8)$$

where n_z, n_{xz} and n_{yz} are the number of samples within the region delimited by k-th nearest neighbor. By analogy with equation (4), X, Y and Z represents X^n_{n-k+1}, Y_{n+1} and Y^n_{n-l+1} respectively. Likewise, direct transfer entropy estimator can be defined as,

$$\begin{aligned} \widehat{D}_{\mathbf{X}\to\mathbf{Y}} &= \widehat{I}(X;Y|Z,W) \\ &= \psi(k) + \langle \psi(n_{zw}+1) \\ &\quad - \psi(n_{xzw}+1) - \psi(n_{yzw}+1) \rangle, \end{aligned} \quad (9)$$

where n_{zw}, n_{xzw} and n_{yzw} are the number of samples within the region delimited by k-th nearest neighbor. Again, by analogy with equation (5), X, Y, Z and W represents X^n_{n-k+1}, Y_{n+1}, Z^n_{n-j+1} and Y^n_{n-l+1} respectively.

V. Transfer Entropy for Detection of Information Flow Pathways

In this section, two applications for continuous processes will be described. First, an example base with a three dimensional autoregressive model will analyze the perform of estimator comparing with theoretical values. Next, an application with four water tank system will be evaluated. For both we will use the estimators already presented in order to describe, through directional graphs, the direct information flow of the systems.

For computational reasons, the embedding dimension of random variables sequences in equations (4) and (5) were considered equal to 1. Furthermore, for equations (8) and (9), we considered $k = 4$. The choice is based on [16] which suggests that in practical applications, one should choose $k > 1$ to reduce statistical errors, avoid high values of k and typically use $k = 2$, 3, or 4. When the estimates were negative or positive on the order of 10^{-3} we discard any causal relationship between the processes (in this case, we consider that the coupling between variables is very small).

A. Three-Dimensional Autoregressive Model

First we discuss an application of estimator to correlated Gaussians because we can compare with analytic results. So, we choose a system of equations which consists of three correlated Gaussians processes satisfying,

$$\left\{ \begin{array}{lllllll} X_{n+1} & = & \alpha X_n & + & \eta^X_n & & \\ Y_{n+1} & = & \beta X_n & + & \gamma Y_n & + & \eta^Y_n \\ Z_{n+1} & = & \delta Y_n & + & \epsilon Z_n & + & \eta^Z_n \end{array} \right. , \quad (10)$$

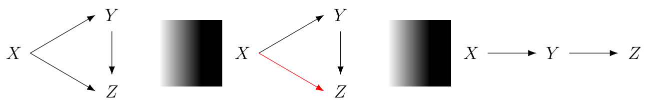

Fig. 1. Directional information flow graph for the three-dimensional autoregressive model with $\alpha = 0.8$, $\beta = 0.6$, $\gamma = 0.4$, $\delta = 0.5$ e $\epsilon = 0$. The first two graphs are the result of calculating the transfer entropy that represents the total causality of the system. The last graph is obtained after the analysis with the direct transfer entropy and corresponds to the direct and true causality, interpreted as the connectivity of the system.

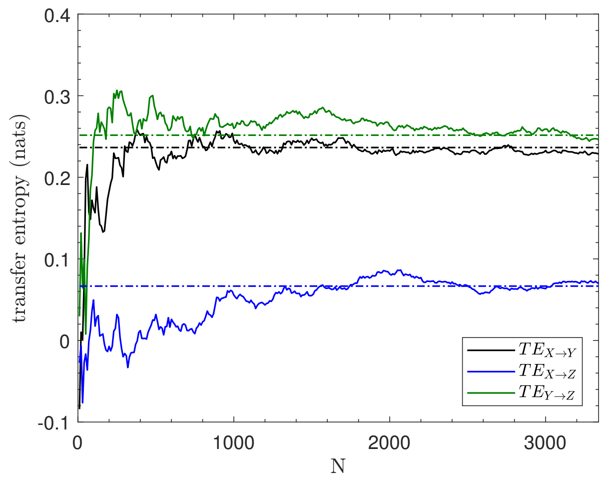

Fig. 2. Convergence of transfer entropy estimator to analytical values of $T_{\mathbf{X}\to\mathbf{Y}} = 0.2365$, $T_{\mathbf{Y}\to\mathbf{Z}} = 0.2516$ and $T_{\mathbf{X}\to\mathbf{Z}} = 0.0666$.

where $\eta^X_n, \eta^Y_n, \eta^Z_n \sim \mathcal{N}(0, 0.1)$. For $0 < \alpha, \beta, \gamma, \delta, \epsilon < 1$ there is direct information transfer from $\mathbf{X}$ to $\mathbf{Y}$ and from $\mathbf{Y}$ to $\mathbf{Z}$ and the information transfer from $\mathbf{X}$ to $\mathbf{Z}$ is indirect through an intermediate variable $\mathbf{Y}$.

The theoretical analysis of causality and connectivity of this system begins with the arbitrary definition of the set of parameters that describe it. Although, for this system, from any set of parameters can be made the analyzes, here we let's consider that $\alpha = 0.8$, $\beta = 0.6$, $\gamma = 0.4$, $\delta = 0.5$ and $\epsilon = 0$.

The analytical values for transfer entropy in this system are: $T_{\mathbf{X}\to\mathbf{Y}} = 0.2365$, $T_{\mathbf{Y}\to\mathbf{Z}} = 0.2516$, $T_{\mathbf{X}\to\mathbf{Z}} = 0.0666$ and all others relationships are equal to zero. Under these conditions, it is concluded that $\mathbf{X}$ causes $\mathbf{Y}$, $\mathbf{Y}$ causes $\mathbf{Z}$ and $\mathbf{X}$ causes $\mathbf{Z}$ and three path way for information flow is assumed, but it is not the true direct connectivity.

We need to determine whether there is direct causality from $\mathbf{X}$ to $\mathbf{Z}$ or if $\mathbf{Y}$ is just an intermediate variable. According to direct transfer entropy definition $D_{\mathbf{X}\to\mathbf{Z}} = 0$ so, there is no direct causality from $\mathbf{X}$ to $\mathbf{Z}$. Finally, the theoretical analysis is completed with directional graph of system connectivity, shown in Figure 1.

To analyze the convergence of the estimators, 8000 samples of each process were simulated, among which the first 3000 were discarded to guarantee the stationarity of the autoregressive processes [5]. The causal relationships were identified by the estimator of equation (8). The Figure 2 shows the convergence of estimated values to analytical values, when the estimates are greater than zero, as a function of the increasing number of samples analyzed (after discarding the 3000 samples).

As well as for theoretical values, we need to determine whether there is direct causality from $\mathbf{X}$ to $\mathbf{Z}$. According to equation (9) the direct transfer entropy $\widehat{D}_{\mathbf{X}\to\mathbf{Z}} = -0.0066$. So, in this case, there is no direct causality from $\mathbf{X}$ to $\mathbf{Z}$. The directional graph of connectivity of this system, obtained through simulation and disregarding any statistical knowledge of the processes, is exactly the same as that obtained analytically and shown in Figure 1.

B. Simulation Case Study: Four Water Tank System

The system consists of four connected water tanks as shown in Figure 3. The pump pumps water to tanks 4 and 3, and the outflow of these tanks is conducted to the inlets of tanks 1 and 2, respectively.

For each tank water enters from the top at a rate proportional to the voltage applied to the pump and leaves through an opening in the tank base at a rate proportional to the square root of the water height in the tank. The differential equation that describe each tank is,

$$\frac{d}{dt}Vol = A\frac{dh}{dt} = bx - a\sqrt{h} \tag{11}$$

where Vol is the volume of water (m^3), h is the height of water (m), x is the flow rate of the water out of the pump (m^3/s), A is the cross-sectional area (m^2), b is a constant related to the flow rate into the tank and a is a constant related to the flow rate out of the tank. The presence of the square root in the water flow rate results in a nonlinear plant.

The Table I describes the operation parameters used during the simulations of system. The simulations were performed using MATLAB software. The pump have normal random behavior, the flow rate of the water out of the pump varies with mean value 4 m^3/s and variance 0.1 m^3/s.

After simulating the processes of the system, causality and connectivity analysis are performed disregarding any knowledge about it. Therefore, the first step is to estimate the transfer entropy between each pair of processes $\mathbf{x}$, $\mathbf{h_1}$, $\mathbf{h_2}$, $\mathbf{h_3}$ and $\mathbf{h_4}$. So, the directional graph of the information flow for this system corresponds to the first graph of Figure 4 results from Table II.

However, the goal is to identify the true connectivity of the system, i.e., to determine if these relations are direct or indirect. The steps of identifying these information flow paths are shown in Figure 4 and in Table III.

It was first checked whether the causality between $\mathbf{h_1}$ and $\mathbf{h_2}$ is true and straightforward considering $\mathbf{h_3}$ as the possible intermediate variable and $\mathbf{h_4}$ and $\mathbf{x}$ as possible common source variables. Since the direct transfer entropy between $\mathbf{h_1}$ and $\mathbf{h_2}$ is very small, we conclude that there is no direct causality between them.

Then, we investigated four other direct causal relations of $\mathbf{h_2} \to \mathbf{h_1}$, $\mathbf{h_4} \to \mathbf{h_2}$, $\mathbf{h_3} \to \mathbf{h_1}$ and $\mathbf{h_1} \to \mathbf{h_3}$, however, all these estimates are either negative or small, thus excluding the connectivity of these variables (see Table III). In addition, it is not necessary to detect whether there is causality between

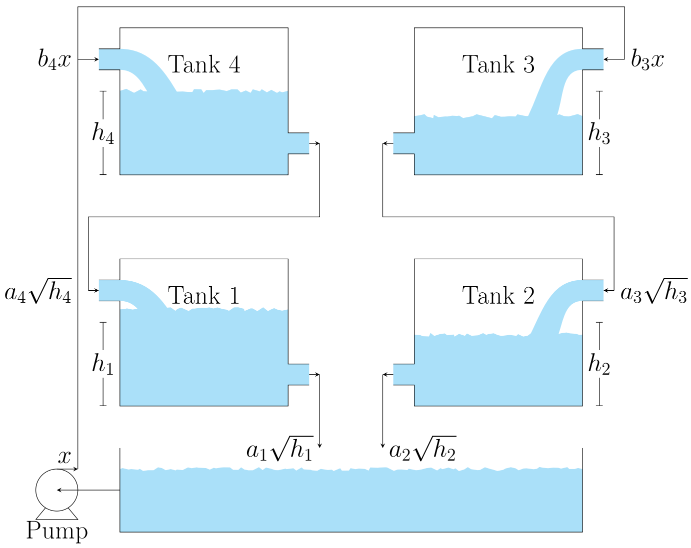

Fig. 3. Schematic of the system with four tanks (reservoirs) with water and a pump whose water flow rate varies with mean value 4m^3/s and variance 0.1m^3/s.

TABLE I
PARAMETER VALUES FOR FOUR WATER TANK SYSTEM IN WHICH $i = 1, 2, 3, 4$.

Parameter	Value
A_i (m^2)	1.44; 1.44; 2.25; 2.25
a_i	1; 1; 2; 2
b_i	0.85; 0.6; 0.4; 0.9

TABLE II
TRANSFER ENTROPY ESTIMATIONS USING EQUATION (8) BETWEEN EACH TWO VARIABLES FOR FOUR WATER TANK SYSTEM.

	$\widehat{T}_{lin\to col}$				
	$\mathbf{h_1}$	$\mathbf{h_2}$	$\mathbf{h_3}$	$\mathbf{h_4}$	$\mathbf{x}$
$\mathbf{h_1}$	-	0.1017	0.0171	0.0086	-0.0024
$\mathbf{h_2}$	0.3618	-	0.0013	-0.0051	-0.0003
$\mathbf{h_3}$	0.3136	0.4269	-	-0.0037	0,0004
$\mathbf{h_4}$	0.2001	0.2858	-0.0045	-	-0.0018
$\mathbf{x}$	-0.0487	-0.0462	0.9229	1.0876	-

TABLE III
DIRECT TRANSFER ENTROPY ESTIMATIONS USING EQUATION (9) FOR DIRECTIONAL RELATIONS THAT MAY HAVE ONE OR MORE INTERMEDIATE VARIABLES.

	Intermediate Processes	$\widehat{D}_{\mathbf{X}\to\mathbf{Y}}$
$\mathbf{h_1} \to \mathbf{h_2}$	$\mathbf{h_3}$, $\mathbf{h_4}$, $\mathbf{x}$	0.00005
$\mathbf{h_2} \to \mathbf{h_1}$	$\mathbf{h_4}$, $\mathbf{x}$	-0.0002
$\mathbf{h_4} \to \mathbf{h_2}$	$\mathbf{h_1}$, $\mathbf{h_3}$, $\mathbf{x}$	0.0002
$\mathbf{h_3} \to \mathbf{h_1}$	$\mathbf{h_2}$, $\mathbf{x}$	0.0071
$\mathbf{h_1} \to \mathbf{h_3}$	$\mathbf{x}$	0.0001
$\mathbf{h_4} \to \mathbf{h_1}$	$\mathbf{h_2}$	0.0432
$\mathbf{h_3} \to \mathbf{h_2}$	$\mathbf{h_1}$	0.0689

$\mathbf{x} \to \mathbf{h_3}$ and $\mathbf{x} \to \mathbf{h_4}$ since there is no intermediate variable or a variable that is common source in its pathways.

Finally, on the basis of all estimates, it is concluded that, except for the causality of $\mathbf{h_4} \to \mathbf{h_1}$ and $\mathbf{h_3} \to \mathbf{h_2}$, the other causalities detected are indirect or spurious. The last graph of Figure 4 corresponds to the true connectivity of the system.

VI. Conclusions

The identification of the causal relationships and the connectivity between continuous signals of a complex system is challenging, since, in general, knowledge about the underlying dynamics is partial. However, information theory can provides a variety of approaches to measure such relationships between multivariate time series, for example, transfer entropy.

Identify true connectivity of a system is important to distinguish whether causal influence occurs along a direct path without any intermediate variables, or indirectly through intermediate variables and/or variables that are common.

In this paper, the concepts of transfer entropy and direct transfer entropy were used to identify causality and connectivity relations, respectively, between continuous variables. According to the presented approach, non-zero transfer entropy values indicate a causal relationship between the analyzed signals and non-zero direct transfer entropy values indicate connectivity between the signals.

The analytical study of causality and connectivity of the three-dimensional autoregressive system, combined with simulations that disregard any statistical knowledge of the processes, is essential to demonstrate, in these cases, the reliability of the estimators based on distances of neighbors, (8) and (9).

In the study case, four water tank system and pump, it was possible, using information theory approaches and those estimators for connectivity analyzes, to faithfully recover the connectivity of the system. Considering this base application it is possible to continue the discussion about the applicability of this method in another industrial process.

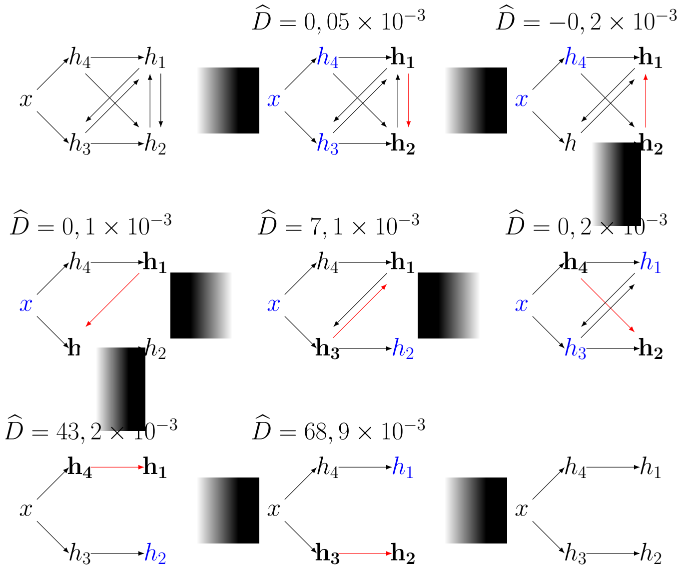

Fig. 4. Directional graphs of the information flow for the system with four tanks with water. The first graph is the result of calculating the transfer entropy that represents the total causality of the system. The following graphs are obtained sequentially from the connectivity analysis of the variables that are interconnected by means of a red arrow. The last graph corresponds to direct and true causality, interpreted as the connectivity of the system. In graph, we denote the processes analyzed by bold letters and possible intermediate processes by blue letters.

References

[1] F. Yang, P. Duan, S. L. Shah, and T. Chen, *Capturing Connectivity and Causality in Complex Industrial Processes*. SpringerBriefs in Applied Sciences and Technology, Springer International Publishing, 1 ed., 2014.

[2] P. Duan, F. Yang, S. L. Shah, and T. Chen, "Transfer zero-entropy and its application for capturing cause and effect relationship between variables," *IEEE Transactions on Control Systems Technology*, vol. 23, pp. 855–867, may 2015.

[3] M. Bauer, J. W. Cox, M. H. Caveness, J. J. Downs, and N. F. Thornhill, "Finding the direction of disturbance propagation in a chemical process using transfer entropy," *IEEE Transactions on Control Systems Technology*, vol. 15, pp. 12–21, Jan 2007.

[4] F. Yang and D. Xiao, "Progress in root cause and fault propagation analysis of large-scale industrial processes," *Journal of Control Science and Engineering*, vol. 2012, pp. 1–10, 2012.

[5] P. Duan, F. Yang, T. Chen, and S. L. Shah, "Direct causality detection via the transfer entropy approach," *IEEE Transactions on Control Systems Technology*, vol. 21, pp. 2052–2066, nov 2013.

[6] J. M. de Assis and F. M. de Assis, "An application of directed information to infer synaptic connectivity," in *Anais do XXXIV Simpósio Brasileiro de Telecomunicações e Processamento de Sinais, Santarém, Brazil*, p. 528–532, 2016.

[7] R. Vicente, M. Wibral, M. Lindner, and G. Pipa, "Transfer entropy—a model-free measure of effective connectivity for the neurosciences," *Journal of Computational Neuroscience*, vol. 30, pp. 45–67, Feb 2011.

[8] M. M. Arruda, L. R. Veloso, and F. M. de Assis, "Transfer entropy characterization of tbi cases," in *Anais do XXXV Simpósio Brasileiro de Telecomunicações e Processamento de Sinais, São Pedro, Brazil*, pp. 625–628, 2017.

[9] F. Marzbanrad, Y. Kimura, M. Endo, M. Palaniswami, and A. H. Khandoker, "Transfer entropy analysis of maternal and fetal heart rate coupling," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 7865–7868, Aug 2015.

[10] J. M. de Assis, M. O. Santos, and F. M. de Assis, "Auditory stimuli coding by postsynaptic potential and local field potential features," *PLOS ONE*, vol. 11, p. e0160089, aug 2016.

[11] T. Schreiber, "Measuring information transfer," *Phys. Rev. Lett.*, vol. 85, pp. 461–464, Jul 2000.

[12] A. Kaiser and T. Schreiber, "Information transfer in continuous processes," *Physica D: Nonlinear Phenomena*, vol. 166, no. 1-2, pp. 43–62, 2002.

[13] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, july, october 1948.

[14] M. M. Arruda, L. R. Veloso, and F. M. de Assis, "Differential entropy estimation via one-class svm," in *Anais do XXXVI Simpósio Brasileiro de Telecomunicações e Processamento de Sinais, Campina Grande, Brazil*, pp. 697–701, 2018.

[15] L. Kozachenko and N. N. Leonenko, "Sample estimate of the entropy of a random vector," *Problemy Peredachi Informatsii*, vol. 23, no. 2, pp. 9–16, 1987.

[16] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Phys. Rev. E*, vol. 69, p. 066138, Jun 2004.

[17] S. Frenzel and B. Pompe, "Partial mutual information for coupling analysis of multivariate time series," *Physical Review Letters*, vol. 99, nov 2007.