Direct Causality Detection via the Transfer Entropy Approach

Ping Duan, Fan Yang, *Member, IEEE*, Tongwen Chen, *Fellow, IEEE*, and Sirish L. Shah, *Member, IEEE*

***Abstract*— The detection of direct causality, as opposed to indirect causality, is an important and challenging problem in root cause and hazard propagation analysis. Several methods provide effective solutions to this problem when linear relationships between variables are involved. For nonlinear relationships, currently only overall causality analysis can be conducted, but direct causality cannot be identified for such processes. In this paper, we describe a direct causality detection approach suitable for both linear and nonlinear connections. Based on an extension of the transfer entropy approach, a direct transfer entropy (DTE) concept is proposed to detect whether there is a direct information flow pathway from one variable to another. Especially, a differential direct transfer entropy concept is defined for continuous random variables, and a normalization method for the differential direct transfer entropy is presented to determine the connectivity strength of direct causality. The effectiveness of the proposed method is illustrated by several examples, including one experimental case study and one industrial case study.**

***Index Terms*— Differential transfer entropy, direct causality, direct transfer entropy (DTE), information flow pathway, kernel estimation, normalization.**

NOMENCLATURE

x, y, z	Continuous random variables.
$\tilde{x}, \tilde{y}, \tilde{z}$	Quantized x, quantized y, and quantized z with quantization bin sizes Δ_x, Δ_y, and Δ_z, respectively.
$T_{x\to y}$	Differential transfer entropy(TE_{diff}) from x to y.
$D_{x\to y}$	Differential direct transfer entropy (DTE_{diff}) from x to y.
$t_{\tilde{x}\to\tilde{y}}$	Discrete transfer entropy (TE_{disc}) from $\tilde{x}$ to $\tilde{y}$.
$d_{\tilde{x}\to\tilde{y}}$	Discrete direct transfer entropy (DTE_{disc}) from $\tilde{x}$ to $\tilde{y}$.
$\text{NTE}_{\tilde{x}\to\tilde{y}}$	Normalized discrete transfer entropy (NTE_{disc}) from $\tilde{x}$ to $\tilde{y}$.
$\text{NTE}^c_{x\to y}$	Normalized differential transfer entropy (NTE_{diff}) from x to y.
$\text{NDTE}^c_{x\to y}$	Normalized differential direct transfer Entropy ($\text{NDTE}_{\text{diff}}$) from x to y.
Eq. (1):	Definition of the TE_{diff} from x to y.
Eq. (2):	Definition of the TE_{diff} from x to z.
Eq. (3):	Definition of the TE_{diff} from z to y.
Eq. (4):	Definition of the DTE_{diff} from x to y.
Eq. (5):	Definition of the DTE_{diff} from z to y.
Eq. (6):	Definition of the TE_{disc} from $\tilde{x}$ to $\tilde{y}$.
Eq. (10):	Relationship between the TE_{diff} from x to y and the TE_{disc} from $\tilde{x}$ to $\tilde{y}$.
Eq. (11):	Definition of the DTE_{disc} from $\tilde{x}$ to $\tilde{y}$.
Eq. (18):	Linear normalization of the TE_{diff}.
Eq. (19):	Nonlinear normalization of the TE_{diff}.
Eq. (20):	Normalization of the DTE_{diff}.
Eq. (21):	Definition of the DTE_{diff} from x to y with multiple intermediate variables.

Manuscript received May 2, 2012; revised October 11, 2012; accepted December 2, 2012. Manuscript received in final form December 7, 2012. Date of publication January 9, 2013; date of current version October 15, 2013. This work was supported by an NSERC Strategic Project, an NSFC Project under Grant 60904044 and the Tsinghua National Laboratory for Information Science and Technology Cross-Discipline Foundation. Recommended by Associate Editor J. Yu.

P. Duan and T. Chen are with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2V4, Canada (e-mail: pduan@ualberta.ca; tchen@ualberta.ca).

F. Yang is with the Department of Automation, Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084, China (e-mail: yangfan@tsinghua.edu.cn).

S. L. Shah is with the Department of Chemical and Materials Engineering, University of Alberta, Edmonton, AB T6G 2G6, Canada (e-mail: sirish.shah@ualberta.ca).

Digital Object Identifier 10.1109/TCST.2012.2233476

I. INTRODUCTION

WITH the increase in scale and complexity of process operations, the detection and diagnosis of plantwide abnormalities and disturbances are major problems in the process industry. Compared with the traditional fault detection, fault detection and diagnosis in a large-scale complex system are particularly challenging because of the high degree of interconnections among different parts in the system. A simple fault may easily propagate along information and material flow pathways and affect other parts of the system. To determine the root cause(s) of certain abnormality, it is important to capture the process connectivity and find the connecting pathways.

A qualitative process model in the form of a digraph has been widely used in root cause and hazard propagation analysis [1]. Digraph-based models usually express the causal relationships between faults and symptoms and define the fault propagation pathways by incorporating expert knowledge of the process [2]. A drawback is that extracting expert knowledge is very time consuming and that knowledge is not always easily available [3]. The modeling of digraphs can also be based on mathematical equations [4], [5], yet for large-scale complex processes it is difficult to establish practical and precise mathematical models.

Data-driven methods provide another way to find the causal relationships between process variables. A few data-based

methods are capable of detecting the causal relationships for linear processes [6]. In the frequency domain, directed transfer functions (DTFs) [7] and partial directed coherence (PDC) [8] are widely used in brain connectivity analysis. Other methods such as Granger causality [9], path analysis [10], and cross-correlation analysis with lag-adjusted variables [11] are commonly used.

The predictability improvement based on the nearest neighbors is proposed as an asymmetrical measure of interdependence in bivariate time series and applied to quantify the directional influences among physiological signals [12] and also industrial processes variables [13], [14]. Information theory provides a wide variety of approaches for measuring causal influence among multivariate time series [15]. Based on transition probabilities containing all information on causality between two variables, the transfer entropy (TE) was proposed to distinguish between driving and responding elements [16] and is suitable for both linear and nonlinear relationships; it has been successfully used in chemical processes [17] and neurosciences [18]. TE has two forms, discrete TE ($\mathrm{TE_{disc}}$) for discrete random variables [16], and differential TE ($\mathrm{TE_{diff}}$) for continuous random variables [19]. It has been shown in [20] that, for Gaussian distributed variables with linear relationships, Granger causality and TE are equivalent. The equivalence of the two causality measures has been extended under certain conditions on probability density distributions of the data [21]. In [22], comparisons are given for several causality detection methods; these methods include TE, extended and nonlinear Granger causality, and predictability improvement. That paper also includes a discussion on the usefulness of the methods for detecting asymmetric couplings and information flow directions in the deterministic chaotic systems. The authors conclude that, given a complex system with *a priori* unknown dynamics, the first method of choice might be TE. If a large number of samples are available, the alternative methods might be nonlinear Granger causality and predictability improvement.

Since information flow specifically means how variation propagates from one variable to another, it is valuable to detect whether the causal influence between a pair of variables is along a direct pathway without any intermediate variables or indirect pathways through some intermediate variables. In the frequency domain, a DTF/PDC-based method for quantification of direct and indirect energy flow in a multivariate process was recently proposed [23]. This method was based on vector auto-regressive or vector moving-average model representations, which are suitable for linear multivariate processes. In the time domain, a path analysis method was used to calculate the direct effect coefficients [24]. The calculation was based on a regression model of the variables, which captures only linear relationships.

For both linear and nonlinear relationships, based on a multivariate version of TE, the partial TE was proposed to quantify the total amount of indirect coupling mediated by the environment and was successfully used in neurosciences [25]. In [25]. The partial TE is defined such that all the environmental variables are considered as intermediate variables, which is not necessary in most cases; and in any case, this will increase the computational burden significantly. On the other hand, the utility of the partial TE is to detect unidirectional causalities, which is suitable for neurosciences; however, in industrial processes, feedback and bidirectional causalities, due to recycle streams, are common. Thus, the partial TE method cannot be directly used for direct/indirect causality detection in the process industry.

The main contribution of this paper is a TE-based methodology to detect and discriminate between direct and indirect causality relationships between process variables of both linear and nonlinear multivariate systems. Specifically, this method is able to uncover explicit direct and indirect, as if through intermediate variables, connectivity pathways between variables.

The rest of this paper is organized as follows. In Section II, we apply the $\mathrm{TE_{diff}}$ for continuous random variables to detect total causality and define a differential direct transfer entropy ($\mathrm{DTE_{diff}}$)[1] to detect direct causality. Calculation methods and the normalization methods are proposed for both the $\mathrm{TE_{diff}}$ and the $\mathrm{DTE_{diff}}$ in the same section. Section III describes three examples to show the effectiveness of the proposed direct causality detection method. An experimental case study and an industrial case study are introduced in Section IV to show the usefulness of the proposed method for detecting direct/indirect connecting pathways, followed by concluding remarks in Section V.

II. Detection of Direct Causality

In this section, an extension of the TE—direct transfer entropy (DTE)—is proposed to detect the direct causality between two variables. In addition to this, calculation methods and the normalization methods are also presented for the TE and the DTE, respectively.

A. DTE

In order to determine the information and material flow pathways to construct a precise topology of a process, it is important to determine whether the influence between a pair of process variables is along direct or indirect pathways. The direct pathway means direct influence without any intermediate or confounding variables.

The TE measures the amount of information transferred from one variable x to another variable y. This extracted transfer information represents the total causal influence from x to y. It is difficult to distinguish whether this influence is along a direct pathway or indirect pathways through some intermediate variables. For example, given three variables x, y, and z, if the calculated transfer entropies from x to y, from x to z, and from z to y are all larger than zero, then we can conclude that x causes y, x causes z, and z causes y. We can also conclude that there is an indirect pathway from x to y via the intermediate variable z which transfers information from x to y. However, we cannot distinguish whether there is a direct pathway from x to y (see Fig. 1), because it is possible that there exist both a direct pathway from x to y and an indirect pathway via the intermediate variable z.

[1]We caution the reader to be aware of the term $\mathrm{DTE_{diff}}$ for "differential direct transfer entropy" and that it is different from the term "discrete direct transfer entropy" ($\mathrm{DTE_{disc}}$) as it applies to discrete random variables.

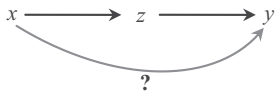


Fig. 1. Detection of direct causality from x to y.

In order to detect the direct and indirect pathways of the information transfer, the definition of a DTE is introduced as follows.

Since the data analyzed here is uniformly sampled data, as obtained from processes that are continuous, we only consider continuous random variables in this paper. Given three continuous random variables x, y, and z, let them be sampled at time instants i and denoted by $x_i \in [x_{\min}, x_{\max}]$, $y_i \in [y_{\min}, y_{\max}]$, and $z_i \in [z_{\min}, z_{\max}]$ with $i = 1, 2, \ldots, N$, where N is the number of samples. The causal relationships between each pair of these variables can be estimated by calculating the TEs [16].

Let y_{i+h_1} denote the value of y at time instant $i + h_1$, that is, h_1 steps in the future from i, and h_1 is referred to as the prediction horizon; $\mathbf{y}_i^{(k_1)} = [y_i, y_{i-\tau_1}, \ldots, y_{i-(k_1-1)\tau_1}]$ and $\mathbf{x}_i^{(l_1)} = [x_i, x_{i-\tau_1}, \ldots, x_{i-(l_1-1)\tau_1}]$ denote embedding vectors with elements from the past values of y and x, respectively (k_1 is the embedding dimension of y and l_1 is the embedding dimension of x); τ_1 is the time interval that allows the scaling in time of the embedded vector, which can be set to be $h_1 = \tau_1 \leq 4$ as a rule of thumb [17]; $f(y_{i+h_1}, \mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)})$ denotes the joint probability density function (pdf), and $f(\cdot|\cdot)$ denotes the conditional pdf, and thus $f(y_{i+h_1}|\mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)})$ denotes the conditional pdf of y_{i+h_1} given $\mathbf{y}_i^{(k_1)}$ and $\mathbf{x}_i^{(l_1)}$ and $f(y_{i+h_1}|\mathbf{y}_i^{(k_1)})$ denotes the conditional pdf of y_{i+h_1} given $\mathbf{y}_i^{(k_1)}$. The differential TE (TE_{diff}) from x to y, for continuous variables, is then calculated as follows:

$$T_{x\to y} = \int f(y_{i+h_1}, \mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)}) \cdot \log \frac{f(y_{i+h_1}|\mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)})}{f(y_{i+h_1}|\mathbf{y}_i^{(k_1)})} d\mathbf{w} \quad (1)$$

where the base of the logarithm is 2 and $\mathbf{w}$ denotes the random vector $[y_{i+h_1}, \mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)}]$. By assuming that the elements of $\mathbf{w}$ are $w_1, w_2, \ldots, w_s$, $\int(\cdot)d\mathbf{w}$ denotes $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}(\cdot)dw_1 \cdots dw_s$ for simplicity, and the following notations have the same meaning as this one.

The TE from x to y can be understood as the improvement when using the past information of both x and y to predict the future of y compared to only using the past information of y. In other words, the TE represents the information about a future observation of variable y obtained from the simultaneous observations of past values of both x and y, after discarding the information about the future of y obtained from the past values of y alone.

Similarly, the TE_{diff} from x to z is calculated as follows:

$$T_{x\to z} = \int f(z_{i+h_2}, \mathbf{z}_i^{(m_1)}, \mathbf{x}_i^{(l_2)}) \cdot \log \frac{f(z_{i+h_2}|\mathbf{z}_i^{(m_1)}, \mathbf{x}_i^{(l_2)})}{f(z_{i+h_2}|\mathbf{z}_i^{(m_1)})} d\eta \quad (2)$$

where h_2 is the prediction horizon, $\mathbf{z}_i^{(m_1)} = [z_i, z_{i-\tau_2}, \ldots, z_{i-(m_1-1)\tau_2}]$ and $\mathbf{x}_i^{(l_2)} = [x_i, x_{i-\tau_2}, \ldots, x_{i-(l_2-1)\tau_2}]$ are embedding vectors with time interval τ_2, and η denotes the random vector $[z_{i+h_2}, \mathbf{z}_i^{(m_1)}, \mathbf{x}_i^{(l_2)}]$.

The TE_{diff} from z to y is calculated as follows:

$$T_{z\to y} = \int f(y_{i+h_3}, \mathbf{y}_i^{(k_2)}, \mathbf{z}_i^{(m_2)}) \cdot \log \frac{f(y_{i+h_3}|\mathbf{y}_i^{(k_2)}, \mathbf{z}_i^{(m_2)})}{f(y_{i+h_3}|\mathbf{y}_i^{(k_2)})} d\zeta \quad (3)$$

where h_3 is the prediction horizon, $\mathbf{y}_i^{(k_2)} = [y_i, y_{i-\tau_3}, \ldots, y_{i-(k_2-1)\tau_3}]$ and $\mathbf{z}_i^{(m_2)} = [z_i, z_{i-\tau_3}, \ldots, z_{i-(m_2-1)\tau_3}]$ are embedding vectors with time interval τ_3, and ζ denotes the random vector $[y_{i+h_3}, \mathbf{y}_i^{(k_2)}, \mathbf{z}_i^{(m_2)}]$.

If $T_{x\to y}$, $T_{x\to z}$, and $T_{z\to y}$ are all larger than zero, then we conclude that x causes y, x causes z, and z causes y. In this case, we need to distinguish whether the causal influence from x to y is only via the indirect pathway through the intermediate variable z, or, in addition to this, there is another direct pathway from x to y, as shown in Fig. 1. We define a direct causality from x to y as x directly causing y, which means there is a direct information and/or material flow pathway from x to y without any intermediate variables.

In order to detect whether there is a direct causality from x to y, we define a differential DTE (DTE_{diff}) from x to y as follows:

$$D_{x\to y} = \int f(y_{i+h}, \mathbf{y}_i^{(k)}, \mathbf{z}_{i+h-h_3}^{(m_2)}, \mathbf{x}_{i+h-h_1}^{(l_1)}) \cdot \log \frac{f(y_{i+h}|\mathbf{y}_i^{(k)}, \mathbf{z}_{i+h-h_3}^{(m_2)}, \mathbf{x}_{i+h-h_1}^{(l_1)})}{f(y_{i+h}|\mathbf{y}_i^{(k)}, \mathbf{z}_{i+h-h_3}^{(m_2)})} d\mathbf{v} \quad (4)$$

where $\mathbf{v}$ denotes the random vector $[y_{i+h}, \mathbf{y}_i^{(k)}, \mathbf{z}_{i+h-h_3}^{(m_2)}, \mathbf{x}_{i+h-h_1}^{(l_1)}]$; the prediction horizon h is set to be $h = \max(h_1, h_3)$; if $h = h_1$, then $\mathbf{y}_i^{(k)} = \mathbf{y}_i^{(k_1)}$, if $h = h_3$, then $\mathbf{y}_i^{(k)} = \mathbf{y}_i^{(k_2)}$; the embedding vector $\mathbf{z}_{i+h-h_3}^{(m_2)} = [z_{i+h-h_3}, z_{i+h-h_3-\tau_3}, \ldots, z_{i+h-h_3-(m_2-1)\tau_3}]$ denotes the past values of z which can provide useful information for predicting the future y at time instant $i + h$, where the embedding dimension m_2 and the time interval τ_3 are determined by (3); the embedding vector $\mathbf{x}_{i+h-h_1}^{(l_1)} = [x_{i+h-h_1}, x_{i+h-h_1-\tau_1}, \ldots, x_{i+h-h_1-(l_1-1)\tau_1}]$ denotes the past values of x which can provide useful information to predict the future y at time instant $i+h$, where the embedding dimension l_1 and the time interval τ_1 are determined by (1). Note that the parameters in DTE_{diff} are all determined by the calculation of the TEs for consistency.

The DTE_{diff} represents the information about a future observation of y obtained from the simultaneous observation of past values of both x and z, after discarding the information about the future y obtained from the past z alone. This can mean that if the pathway from z to y is cut off, will the history of x still provide some helpful information to predict the future y? Obviously, if this information is nonzero (greater than zero), then there is a direct pathway from x to y. Otherwise, there is no direct pathway from x to y, and the causal influence from x to y is all along the indirect pathway via the intermediate variable z.

Note that the direct causality here is a relative concept; since the measured process variables are limited, the direct causality analysis is only based on these variables. In other words, even if there are intermediate variables in the connecting pathway between two measured variables, as long as none of these intermediate variables is measured, we still state that the causality is direct between the pair of measured variables.

After the calculation of $D_{x\to y}$, if there is direct causality from x to y, we need to further judge whether the causality from z to y is true or spurious, because it is possible that z is not a cause of y and the spurious causality from z to y is generated by x, i.e., x is the common source of both z and y. As shown in Fig. 2, there are still two cases of the information flow pathways between x, y, and z, and the difference is whether there is true and direct causality from z to y.

Thus, $\mathrm{DTE}_{\mathrm{diff}}$ from z to y needs to be calculated

$$D_{z\to y} = \int p(y_{i+h}, \mathbf{y}_i^{(k)}, \mathbf{x}_{i+h-h_1}^{(l_1)}, \mathbf{z}_{i+h-h_3}^{(m_2)}) \cdot \log \frac{p(y_{i+h}|\mathbf{y}_i^{(k)}, \mathbf{x}_{i+h-h_1}^{(l_1)}, \mathbf{z}_{i+h-h_3}^{(m_2)})}{p(y_{i+h}|\mathbf{y}_i^{(k)}, \mathbf{x}_{i+h-h_1}^{(l_1)})} d\mathbf{v} \quad (5)$$

where the parameters are the same as in (4). If $d_{z\to y} > 0$, then there is true and direct causality from z to y, as shown in Fig. 2(a). Otherwise, the causality from z to y is spurious, which is generated by the common source x, as shown in Fig. 2(b).

The need for detection of direct and indirect causality based on measured process variables is discussed. The traditional TE method only determines whether there is causality from x to y, but we cannot tell whether the causal influence is along a direct pathway or indirect pathways through some intermediate variables (see Fig. 1). The purpose of process causality analysis is to investigate propagation of faults, alarms events, and signals through material and information flow pathways (for example via feedback control) and in this respect it is important to know whether the connection between variables of interest is direct or indirect. As shown in Fig. 1, if direct causality from x to y is detected, then there should be a direct information and/or material flow pathway from x to y. Otherwise, there is no direct information and/or material flow pathway from x to y and the direct link should be eliminated. This is clearly illustrated in the experimental three-tank case study presented in Section IV. Such cases are common in industrial processes; the traditional TE approach will reveal a myriad of connections, as it is not able to discriminate between direct and indirect causality, whereas once one is able to detect direct paths, the number of connecting pathways reduces significantly.

Another important case is to detect the true or spurious causality as shown in Fig. 2. In fact, this can tell whether there is a direct information and/or material flow pathway from z to y or there is no information flow pathway from z to y at all. If we only use the traditional TE method, we may conclude that there is causal influence from z to y and therefore there is an information flow pathway from z to y, which is not true because they are both influenced by a common cause. Thus, the detection of direct and indirect causality is necessary for capturing the true process connectivity.

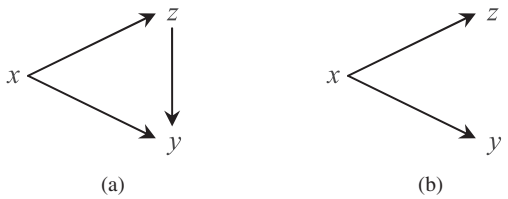


Fig. 2. Information flow pathways between x, y, and z with (a) true and direct causality from z to y, and (b) spurious causality from z to y (meaning that z and y have a common perturbing source, x, and therefore they may appear to be connected or correlated even when they are not connected physically).

An important application of causality analysis for capturing process connectivity is to find the fault propagation pathways and diagnose the root cause of certain disturbance or faults. If we only detect causality via the traditional TE approach, total causality and spurious causality would be detected to yield an overly complicated set of pathways from which root cause diagnosis of faults would be difficult if not erroneous. However, if we are able to differentiate between direct and indirect, true and spurious causality, then the derived causal map may be much simpler and more accurate to tell the fault propagation pathways and which variable is the likely root cause. This point is clearly illustrated by the industrial case study presented in Section IV.

B. Relationships Between DTE$_{\mathrm{diff}}$ and DTE$_{\mathrm{disc}}$

The $\mathrm{TE}_{\mathrm{diff}}$ and the $\mathrm{DTE}_{\mathrm{diff}}$ mentioned above are defined for continuous random variables. For continuous random variables, a widely used TE calculation procedure is to perform quantization first and then use the formula of $\mathrm{TE}_{\mathrm{disc}}$ [17]. Thus, we need to establish a connection between this quantization-based procedure and the $\mathrm{TE}_{\mathrm{diff}}$ procedure.

For the continuous random variables x, y, and z, let $\tilde{x}$, $\tilde{y}$ and $\tilde{z}$ denote the quantized x, y, and z, respectively. Assume that the supports of x, y, and z, i.e., $[x_{\min}, x_{\max}]$, $[y_{\min}, y_{\max}]$, and $[z_{\min}, z_{\max}]$, are classified into n_x, n_y, and n_z nonoverlapping intervals (bins), respectively, and the corresponding quantization bin sizes of x, y, and z are Δ_x, Δ_y, and Δ_z, respectively. Taking x for an example, if we choose a uniform quantizer, then we have

$$\Delta_x = \frac{x_{\max} - x_{\min}}{n_x - 1}.$$

We can see that the quantization bin size is related to the variable support and the number of quantization intervals (bin number). Given a variable support, the larger the bin number, the smaller is the quantization bin size.

After quantization, the TE from x to y can be approximated by the $\mathrm{TE}_{\mathrm{disc}}$ from $\tilde{x}$ to $\tilde{y}$

$$t_{\tilde{x}\to\tilde{y}} = \sum p(\tilde{y}_{i+h_1}, \tilde{\mathbf{y}}_i^{(k_1)}, \tilde{\mathbf{x}}_i^{(l_1)}) \cdot \log \frac{p(\tilde{y}_{i+h_1}|\tilde{\mathbf{y}}_i^{(k_1)}, \tilde{\mathbf{x}}_i^{(l_1)})}{p(\tilde{y}_{i+h_1}|\tilde{\mathbf{y}}_i^{(k_1)})} \quad (6)$$

where the sum symbol represents $k_1 + l_1 + 1$ sums over all amplitude bins of the joint probability distribution and conditional probabilities; $\tilde{\mathbf{y}}_i^{(k_1)} = [\tilde{y}_i, \tilde{y}_{i-\tau_1}, \ldots, \tilde{y}_{i-(k_1-1)\tau_1}]$ and $\tilde{\mathbf{x}}_i^{(l_1)} = [\tilde{x}_i, \tilde{x}_{i-\tau_1}, \ldots, \tilde{x}_{i-(l_1-1)\tau_1}]$ denote embedding vectors;

$p(\tilde{y}_{i+h_1}, \tilde{\mathbf{y}}_i^{(k_1)}, \tilde{\mathbf{x}}_i^{(l_1)})$ denotes the joint probability distribution; and $p(\cdot|\cdot)$ denotes the conditional probabilities. The meaning of other parameters remains unchanged.

From (6) we can express $\mathrm{TE}_{\mathrm{disc}}$ using conditional Shannon entropies [26] by expanding the logarithm, as

$$\begin{aligned} t_{\tilde{x}\to\tilde{y}} &= \sum p(\tilde{y}_{i+h_1}, \tilde{\mathbf{y}}_i^{(k_1)}, \tilde{\mathbf{x}}_i^{(l_1)}) \log \frac{p(\tilde{y}_{i+h_1}, \tilde{\mathbf{y}}_i^{(k_1)}, \tilde{\mathbf{x}}_i^{(l_1)})}{p(\tilde{\mathbf{y}}_i^{(k_1)}, \tilde{\mathbf{x}}_i^{(l_1)})} \\ &\quad - \sum p(\tilde{y}_{i+h_1}, \tilde{\mathbf{y}}_i^{(k_1)}) \log \frac{p(\tilde{y}_{i+h_1}, \tilde{\mathbf{y}}_i^{(k_1)})}{p(\tilde{\mathbf{y}}_i^{(k_1)})} \\ &= H(\tilde{y}_{i+h_1}|\tilde{\mathbf{y}}_i^{(k_1)}) - H(\tilde{y}_{i+h_1}|\tilde{\mathbf{y}}_i^{(k_1)}, \tilde{\mathbf{x}}_i^{(l_1)}) \end{aligned} \tag{7}$$

where

$$H(\tilde{y}_{i+h_1}|\tilde{\mathbf{y}}_i^{(k_1)}) = -\sum p(\tilde{y}_{i+h_1}, \tilde{\mathbf{y}}_i^{(k_1)}) \log p(\tilde{y}_{i+h_1}|\tilde{\mathbf{y}}_i^{(k_1)})$$

and

$$\begin{aligned} &H(\tilde{y}_{i+h_1}|\tilde{\mathbf{y}}_i^{(k_1)}, \tilde{\mathbf{x}}_i^{(l_1)}) \\ &\quad = -\sum p(\tilde{y}_{i+h_1}, \tilde{\mathbf{y}}_i^{(k_1)}, \tilde{\mathbf{x}}_i^{(l_1)}) \log p(\tilde{y}_{i+h_1}|\tilde{\mathbf{y}}_i^{(k_1)}, \tilde{\mathbf{x}}_i^{(l_1)}) \end{aligned}$$

are the conditional Shannon entropies.

Similar to $\mathrm{TE}_{\mathrm{disc}}$, we can express the $\mathrm{TE}_{\mathrm{diff}}$ using differential conditional entropies, as

$$\begin{aligned} T_{x\to y} &= \int f(y_{i+h_1}, \mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)}) \log f(y_{i+h_1}|\mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)}) d\mathbf{w} \\ &\quad - \int f(y_{i+h_1}, \mathbf{y}_i^{(k_1)}) \log f(y_{i+h_1}|\mathbf{y}_i^{(k_1)}) d\mathbf{u} \\ &= H^c(y_{i+h_1}|\mathbf{y}_i^{(k_1)}) - H^c(y_{i+h_1}|\mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)}) \end{aligned} \tag{8}$$

where $\mathbf{u}$ denotes the random vector $[y_{i+h_1}, \mathbf{y}_i^{(k_1)}]$, and $H^c(y_{i+h_1}|\mathbf{y}_i^{(k_1)})$ and $H^c(y_{i+h_1}|\mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)})$ are the differential conditional entropies.

Theoretically, as the bin sizes approach zero, the probability $p(\tilde{y}_{i+h_1}, \tilde{\mathbf{y}}_i^{(k_1)}, \tilde{\mathbf{x}}_i^{(l_1)})$ in (7) can be approximated by $\Delta_y \Delta_y^{k_1} \Delta_x^{l_1} f(y_{i+h_1}, \mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)})$. Then we have

$$\begin{aligned} &\lim_{\Delta_x, \Delta_y \to 0} t_{\tilde{x}\to\tilde{y}} \\ &\quad = \lim_{\Delta_x, \Delta_y \to 0} \Bigg\{ \sum \Delta_y \Delta_y^{k_1} \Delta_x^{l_1} f(y_{i+h_1}, \mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)}) \\ &\qquad \cdot \log \frac{\Delta_y \Delta_y^{k_1} \Delta_x^{l_1} f(y_{i+h_1}, \mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)})}{\Delta_y^{k_1} \Delta_x^{l_1} f(\mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)})} \\ &\qquad - \sum \Delta_y \Delta_y^{k_1} f(y_{i+h_1}, \mathbf{y}_i^{(k_1)}) \\ &\qquad \cdot \log \frac{\Delta_y \Delta_y^{k_1} f(y_{i+h_1}, \mathbf{y}_i^{(k_1)})}{\Delta_y^{k_1} f(\mathbf{y}_i^{(k_1)})} \Bigg\} \\ &\quad = \lim_{\Delta_x, \Delta_y \to 0} \Bigg\{ \sum \Delta_y \Delta_y^{k_1} \Delta_x^{l_1} f(y_{i+h_1}, \mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)}) \\ &\qquad \cdot \Big(\log \Delta_y + \log f(y_{i+h_1}|\mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)}) \Big) \\ &\qquad - \sum \Delta_y \Delta_y^{k_1} f(y_{i+h_1}, \mathbf{y}_i^{(k_1)}) \\ &\qquad \cdot \Big(\log \Delta_y + \log f(y_{i+h_1}|\mathbf{y}_i^{(k_1)}) \Big) \Bigg\}. \end{aligned} \tag{9}$$

As $\Delta_x, \Delta_y \to 0$, we have

$$\begin{aligned} &\sum \Delta_y \Delta_y^{k_1} \Delta_x^{l_1} f(y_{i+h_1}, \mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)}) \\ &\qquad \to \int f(y_{i+h_1}, \mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)}) d\mathbf{w} = 1, \end{aligned}$$

$$\sum \Delta_y \Delta_y^{k_1} f(y_{i+h_1}, \mathbf{y}_i^{(k_1)}) \to \int f(y_{i+h_1}, \mathbf{y}_i^{(k_1)}) d\mathbf{u} = 1$$

and the integral of the function $f(\cdot)\log f(\cdot)$ can be approximated in the Riemannian sense by

$$\begin{aligned} &\sum \Delta_y \Delta_y^{k_1} \Delta_x^{l_1} f(y_{i+h_1}, \mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)}) \log f(y_{i+h_1}|\mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)}) \\ &\quad \to \int f(y_{i+h_1}, \mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)}) \log f(y_{i+h_1}|\mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)}) d\mathbf{w} \\ &\sum \Delta_y \Delta_y^{k_1} f(y_{i+h_1}, \mathbf{y}_i^{(k_1)}) \log f(y_{i+h_1}|\mathbf{y}_i^{(k_1)}) \\ &\quad \to \int f(y_{i+h_1}, \mathbf{y}_i^{(k_1)}) \log f(y_{i+h_1}|\mathbf{y}_i^{(k_1)}) d\mathbf{u}. \end{aligned}$$

Thus

$$\begin{aligned} &\lim_{\Delta_x, \Delta_y \to 0} t_{\tilde{x}\to\tilde{y}} \\ &\quad = \lim_{\Delta_y \to 0} \log \Delta_y \\ &\qquad + \int f(y_{i+h_1}, \mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)}) \cdot \log f(y_{i+h_1}|\mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)}) d\mathbf{w} \\ &\qquad - \lim_{\Delta_y \to 0} \log \Delta_y \\ &\qquad - \int f(y_{i+h_1}, \mathbf{y}_i^{(k_1)}) \cdot \log f(y_{i+h_1}|\mathbf{y}_i^{(k_1)}) d\mathbf{u} \\ &\quad = \int f(y_{i+h_1}, \mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)}) \cdot \log f(y_{i+h_1}|\mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)}) d\mathbf{w} \\ &\qquad - \int f(y_{i+h_1}, \mathbf{y}_i^{(k_1)}) \cdot \log f(y_{i+h_1}|\mathbf{y}_i^{(k_1)}) d\mathbf{u} \\ &\quad = \int f(y_{i+h_1}, \mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)}) \cdot \log \frac{f(y_{i+h_1}|\mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)})}{f(y_{i+h_1}|\mathbf{y}_i^{(k_1)})} d\mathbf{w} \\ &\quad = T_{x\to y}. \end{aligned} \tag{10}$$

This means that the differential TE from x to y is the same as the discrete TE from quantized x to quantized y in the limit as the quantization bin sizes of both x and y approach zero.

Remark: From (9) and (10), we can see that the difference between the differential conditional entropy and the limiting value of the Shannon conditional entropy as $\Delta_x, \Delta_y \to 0$ is an infinite offset, $\lim_{\Delta_y \to 0} \log \Delta_y$. Thus, the differential conditional entropy can be negative.

Similar to TE, the DTE from x to y can be approximated by a discrete DTE ($\mathrm{DTE}_{\mathrm{disc}}$) from $\tilde{x}$ to $\tilde{y}$

$$\begin{aligned} d_{\tilde{x}\to\tilde{y}} &= \sum p(\tilde{y}_{i+h}, \tilde{\mathbf{y}}_i^{(k)}, \tilde{\mathbf{z}}_{i+h-h_3}^{(m_2)}, \tilde{\mathbf{x}}_{i+h-h_1}^{(l_1)}) \\ &\quad \cdot \log \frac{p(\tilde{y}_{i+h}|\tilde{\mathbf{y}}_i^{(k)}, \tilde{\mathbf{z}}_{i+h-h_3}^{(m_2)}, \tilde{\mathbf{x}}_{i+h-h_1}^{(l_1)})}{p(\tilde{y}_{i+h}|\tilde{\mathbf{y}}_i^{(k)}, \tilde{\mathbf{z}}_{i+h-h_3}^{(m_2)})} \end{aligned} \tag{11}$$

where $\tilde{\mathbf{y}}_i^{(k)}$, $\tilde{\mathbf{z}}_{i+h-h_3}^{(m_2)}$, and $\tilde{\mathbf{x}}_{i+h-h_1}^{(l_1)}$ are embedding vectors of $\tilde{y}$, $\tilde{z}$, and $\tilde{x}$, respectively. The definitions of the other quantities are similar to that in (4).

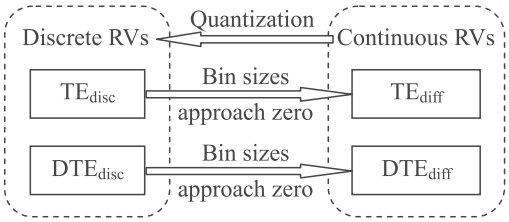


Fig. 3. Relationships between TEs and DTEs. RV means random variable.

For the DTE_{diff} and the DTE_{disc}, using the same proof procedure with the TE, we can obtain

$$\lim_{\Delta_x,\Delta_y,\Delta_z\to 0} d_{\tilde{x}\to\tilde{y}} = D_{x\to y}$$

which means that the DTE_{diff} from x to y is the same as the DTE_{disc} from quantized x to quantized y in the limit as the quantization bin sizes of x, y, and the intermediate variable z approach zero. Fig. 3 illustrates the relationships between TE_{diff} and TE_{disc}, and between DTE_{diff} and DTE_{disc}.

It should be noted that the smaller the bin size, the more accurate is the quantization and the closer are DTE_{disc} and DTE_{diff}. Note that the computational burden of the summation and the probability estimation in (6) and (11) will increase significantly with increasing quantization bin numbers, i.e., n_x, n_y, and n_z. Thus, for the choice of bin sizes, there is a tradeoff between the quantization accuracy and the computational burden in TE_{disc} and DTE_{disc} calculations. In practice, the conditions that the quantization bin sizes approach zero are difficult to satisfy. Thus, in order to avoid the roundoff error of quantization, we directly use TE_{diff} and DTE_{diff} to calculate TE and DTE, respectively.

C. Calculation Method

1) Required Assumptions for the DTE Calculation: Since the concept of DTE is an extension of TE, the required assumptions for DTE is exactly the same as TE; the collected sampled data must be stationary in a wide sense with a large data length, preferably no less than 2000 observations [17]. Stationarity requires that the dynamical properties of the system must not change during the observation period. Since in most cases we do not have direct access to the system and we cannot establish evidence that its parameters are indeed constant, we have to test for stationarity based on the available dataset.

For the purpose of testing for stationarity, the simplest and most widely used method is to measure the mean and the variance for several segments of the dataset (equivalent to an ergodicity test) and then use a standard statistical hypothesis test to check whether the mean and the variance change. More subtle quantities such as spectral components, correlations, or nonlinear statistics may be needed to detect less obvious nonstationarity [27]. In this paper, we use the mean and variance to test for stationarity.

We divide a given dataset, denoted by $x_i, i = 1, 2, \ldots, N$, into m consecutive segments, denoted by $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m$, each containing s data points. Let μ_j denote the mean value of $\mathbf{X}_j$, $j = 1, 2, \ldots, m$, and $\bar{\mu} = \sum_{j=1}^{m} \mu_j/m$; then the standard error of the estimated mean $\bar{\mu}$ is given by

$$\sigma = \sqrt{\frac{\sum_{j=1}^{m}(\mu_j - \bar{\mu})^2}{m(m-1)}}$$

where the standard deviation divided by an extra $\sqrt{m}$ is the error when estimating the mean value of Gaussian distributed uncorrelated numbers [27]. The null hypothesis for stationarity testing is that the dataset is stationary. The significance level for the mean testing is defined as

$$\frac{|\mu_j - \bar{\mu}|}{\sigma} > 6 \quad \text{for } j = 1, 2, \ldots, m. \tag{12}$$

A six-sigma threshold for the significance level is chosen here. Specifically, if there exists $\mu_j > \bar{\mu} + 6\sigma$ or $\mu_j < \bar{\mu} - 6\sigma$ for $j = 1, 2, \ldots, m$, then the null hypothesis that the dataset is stationary is rejected. If $\bar{\mu} - 6\sigma < \mu_j < \bar{\mu} + 6\sigma$ holds for all js, then the null hypothesis is accepted that the dataset is stationary.

For the variance test, let $\hat{\boldsymbol{x}}_i, i = 1, 2, \ldots, N$ denote the normalized dataset of x_i, and $\bar{\boldsymbol{x}}_1, \bar{\boldsymbol{x}}_2, \ldots, \bar{\boldsymbol{x}}_m$ denote the corresponding consecutive segments. Then we have $\bar{\boldsymbol{x}}_j = \hat{x}_{s(j-1)+1}, \hat{x}_{s(j-1)+2}, \ldots, \hat{x}_{sj}$ for $j = 1, 2, \ldots, m$. Since the sum of squares of the elements in each segment has the chi-squared distribution with s degrees of freedom $\hat{v}_j = \hat{x}^2_{s(j-1)+1} + \hat{x}^2_{s(j-1)+2} + \ldots + \hat{x}^2_{sj} \sim \chi^2_s$, we can check whether or not the dataset is stationary by comparing $\hat{v}_j$ with $\chi^2_s(\alpha)$. If there exists $\hat{v}_j > \chi^2_s(\alpha)$ for $j = 1, 2, \ldots, m$, then the null hypothesis that the dataset is stationary is rejected with $(1-\alpha) \times 100\%$ confidence. If $\hat{v}_j < \chi^2_s(\alpha)$ for all js, then the null hypothesis is accepted.

Multimodality is often encountered in industrial processes due to the normal operational changes as well as changes in the production strategy [28]. For such multimodal processes, a dataset with a large number of samples is most likely to be nonstationary as the data would reflect transitions from one mode to another, whereas a key assumption of the TE/DTE method is stationarity of the sampled data. In order to handle the process multimodality, one would have to partition the data into different segments corresponding to different modes. A few time-series analysis methods [29], [30] have been proposed for segmentation of time series to determine when the process mode has changed. As long as the segments corresponding to different modes are obtained, we can detect (direct) causality for each mode of the process using the appropriate segment. Note that the causal relationships may change with mode switching of the process.

2) Estimation of the TE_{diff} and the DTE_{diff}: For the TE from x to y, since (1) can be written as

$$T_{x\to y} = E\left\{\log \frac{f(y_{i+h_1}|\mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)})}{f(y_{i+h_1}|\mathbf{y}_i^{(k_1)})}\right\}$$

it can be approximated by

$$T_{x\to y} = \frac{1}{N - h_1 - r + 1}\sum_{i=r}^{N-h_1} \log \frac{f(y_{i+h_1}|\mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)})}{f(y_{i+h_1}|\mathbf{y}_i^{(k_1)})} \tag{13}$$

where N is the number of samples and $r = \max\{(k_1 - 1)\tau_1 + 1, (l_1 - 1)\tau_1 + 1\}$.

Just as with TE_{diff}, the DTE_{diff} (4) can be written as

$$D_{x\to y} = E\left\{\log \frac{f(y_{i+h}|\mathbf{y}_i^{(k)}, \mathbf{z}_{i+h-h_3}^{(m_2)}, \mathbf{x}_{i+h-h_1}^{(l_1)})}{f(y_{i+h}|\mathbf{y}_i^{(k)}, \mathbf{z}_{i+h-h_3}^{(m_2)})}\right\}$$

which can be approximated by

$$D_{x\to y} = \frac{1}{N-h-j+1} \cdot \sum_{i=j}^{N-h} \log \frac{f(y_{i+h}|\mathbf{y}_i^{(k)}, \mathbf{z}_{i+h-h_3}^{(m_2)}, \mathbf{x}_{i+h-h_1}^{(l_1)})}{f(y_{i+h}|\mathbf{y}_i^{(k)}, \mathbf{z}_{i+h-h_3}^{(m_2)})} \quad (14)$$

where $j = \max\{(k_1 - 1)\tau_1 + 1, (k_2 - 1)\tau_3 + 1, -h + h_3 + (m_2 - 1)\tau_3 + 1, -h + h_1 + (l_1 - 1)\tau_1 + 1\}$.

3) Kernel Estimation of pdfs: In (13) and (14), the conditional pdfs are expressed by the joint pdfs and then obtained by the kernel estimation method [31]. Here, the following Gaussian kernel function is used

$$k(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}.$$

Then a univariate pdf can be estimated by

$$\hat{f}(x) = \frac{1}{N\gamma} \sum_{i=1}^{N} k\left(\frac{x - X_i}{\gamma}\right) \quad (15)$$

where N is the number of samples, and γ is the bandwidth chosen to minimize the mean integrated squared error of the pdf estimation and calculated by $\gamma = 1.06\sigma N^{-1/5}$ according to the "normal reference rule-of-thumb" [31], [32], where σ is the standard deviation of the sampled data $\{X_i\}_{i=1}^N$.

For q-dimensional multivariate data, we use the Fukunaga method [31] to estimate the joint pdf. Suppose that $\mathbf{X}_1, \ldots, \mathbf{X}_N$ constitute a q-dimensional vector ($\mathbf{X}_i \in \mathbb{R}^q$) with a common pdf $f(x_1, x_2, \ldots, x_q)$. Let $\mathbf{x}$ denote the q-dimensional vector $[x_1, x_2, \ldots, x_q]^{\mathrm{T}}$; then the kernel estimation of the joint pdf is

$$\hat{f}(\mathbf{x}) = \frac{(\det \mathbf{S})^{-1/2}}{N\Gamma^q} \sum_{i=1}^{N} K\left\{\Gamma^{-2}(\mathbf{x} - \mathbf{X}_i)^{\mathrm{T}}\mathbf{S}^{-1}(\mathbf{x} - \mathbf{X}_i)\right\} \quad (16)$$

where Γ is similar to the bandwidth γ in (15). The estimated joint pdf is smoother when Γ is larger. However, a substantially larger Γ is most likely to result in an inaccurate estimation. Thus, Γ is also chosen to minimize the mean integrated squared error of the joint pdf estimation and calculated by $\Gamma = 1.06N^{-1/(4+q)}$. $\mathbf{S}$ is the covariance matrix of the sampled data, and K is the Gaussian kernel satisfying

$$K(u) = (2\pi)^{-q/2} e^{-\frac{1}{2}u}.$$

Note that when $q = 1$, (16) is simplified into (15).

For the TE, the estimation of the computational complexity is divided into two parts: the kernel estimation of the pdf using (16), and the calculation of the TE_{diff} using (13). For each joint pdf of dimension q, the computational complexity is $O(N^2q^2)$. Considering the conditional pdfs are estimated by the joint pdfs, the maximum dimension of the joint pdf is $k_1 + l_1 + 1$ and, thus, the computational complexity for the pdf estimation is $O(N^2(k_1 + l_1)^2)$. For calculation of the TE_{diff} in (13), approximately N summations are required. Thus, the total computational complexity for the TE_{diff} is $O(N^2(k_1 + l_1)^2)$. Similarly, we can obtain that the computational complexity for the DTE_{diff} using (14) is $O(N^2(k + m_2 + l_1)^2)$. It is obvious that the number of samples and the embedding dimensions determine the computing speed. Since the samples number is preferred to be no less than 2000 observations [17], we need to limit the choice of the embedding dimensions. The details on how to choose the embedding dimensions are given in the following subsection.

Note that the computational complexity is relatively large because of the kernel estimation of the (joint) pdfs, and that the computational complexity for the DTE increases with an increasing number of intermediate variables. Therefore, one would have to apply the method to smaller units with a smaller number of variables. A large-scale complex system can be broken down into smaller units and thereafter analyzed for causal relationships within each unit and between different units, and finally the information flow pathways of the whole process can be established.

4) Determination of the Parameters of the TE: In the use of the TE approach to detect causality, there are four undetermined parameters: 1) the prediction horizon (h_1); 2) the time interval (τ_1); 3) and the embedding dimensions (k_1 and l_1). Since these four parameters greatly affect the calculation results of the transfer entropies, we need to find a systematic method to determine them.

First, since $h_1 = \tau_1 \leq 4$ as a rule of thumb [17], we can further set initial values for h_1 and τ_1 according to *a priori* knowledge of the process. For example, we start by setting the initial values for $h_1 = \tau_1 = 1$.

Second, we can determine the embedding dimension of y, i.e., the window size of the historical y used for the future y prediction. The embedding dimension of y, i.e., k_1, can be determined as the minimum nonnegative integer above which the change rate of $H^c(y_{i+h_1}|\mathbf{y}_i^{(k_1)})$ decreases significantly. Considering a large k_1 can increase the dimension of the joint pdf and the difficulty in pdf estimation, if k_1 is greater than 3, we need to increase h_1 and τ_1 and repeat the calculation until a $k_1 \leq 3$ is found to make the change rate of $H^c(y_{i+h_1}|\mathbf{y}_i^{(k_1)})$ decrease significantly.

Finally, we can determine the embedding dimension of x, i.e., the window size of the historical x used for the future y prediction. Based on the values of $k_1, h_1,$, and τ_1, the embedding dimension of x, i.e., l_1, is determined as the minimum positive integer above which the change rate of the TE from x to y decreases significantly.

5) Normalization: It is easy to prove that both the TE and the DTE are conditional mutual information; thus they are always nonnegative. However, small values of the TE and the DTE suggest no causality or direct causality while large values do. In order to quantify the strength of the total causality and direct causality, normalization is necessary.

In [33], the normalized discrete TE ($\mathrm{NTE}_{\mathrm{disc}}$) is defined as

$$\mathrm{NTE}_{\tilde{x}\to\tilde{y}} = \frac{t_{\tilde{x}\to\tilde{y}} - t^{\mathrm{shuffled}}_{\tilde{x}\to\tilde{y}}}{H(\tilde{y}_{i+h_1}|\tilde{\mathbf{y}}_i^{(k_1)})} \in [0,1] \tag{17}$$

where $t^{\mathrm{shuffled}}_{\tilde{x}\to\tilde{y}}$ is an estimate of the same TE in shuffled data of $\tilde{x}$ and $\tilde{y}$. This $\mathrm{NTE}_{\mathrm{disc}}$ intuitively represents the fraction of information in $\tilde{y}$ not explained by its own past but explained by the past of $\tilde{x}$.

Eq. (17) is suitable for the normalization of the $\mathrm{TE}_{\mathrm{disc}}$. For $\mathrm{TE}_{\mathrm{diff}}$, we cannot just substitute $H(\tilde{y}_{i+h_1}|\tilde{\mathbf{y}}_i^{(k_1)})$ with the differential conditional entropy $H^c(y_{i+h_1}|\mathbf{y}_i^{(k_1)})$, since $H^c(y_{i+h_1}|\mathbf{y}_i^{(k_1)})$ could be negative. Moreover, using shuffled data to eliminate the calculation bias is not accurate because random shuffling may destroy the statistical properties of the time series. Also, $t^{\mathrm{shuffled}}_{\tilde{x}\to\tilde{y}}$ is an average of transfer entropies obtained on n trials. To obtain a better result, n should be large enough, which will increase the computational burden significantly. Thus, we need to propose a new normalization method for $\mathrm{TE}_{\mathrm{diff}}$.

In (17) the zero point is regarded as the origin and it represents a deterministic variable. For differential entropy, the value $-\infty$ instead of zero means that the variable is deterministic. The maximal differential entropy given a finite support is in the form of a uniform distribution [34]. So, we define the origin as the maximal differential entropy of y with the uniform distribution

$$\begin{aligned} H_0(y) &= -\int_{y_{\min}}^{y_{\max}} \frac{1}{y_{\max}-y_{\min}} \log \frac{1}{y_{\max}-y_{\min}} dy \\ &= \log(y_{\max}-y_{\min}) \end{aligned}$$

where $y_{\max}$ and $y_{\min}$ denote the maximum and minimum values of the variable y, respectively.

Considering that the $\mathrm{TE}_{\mathrm{diff}}$ is the difference between two differential conditional entropies, as shown in (8), we define the normalized differential TE ($\mathrm{NTE}_{\mathrm{diff}}$) as

$$\begin{aligned} \mathrm{NTE}^c_{x\to y} &= \frac{H^c(y_{i+h_1}|\mathbf{y}_i^{(k_1)}) - H^c(y_{i+h_1}|\mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)})}{H_0 - H^c(y_{i+h_1}|\mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)})} \\ &= \frac{T_{x\to y}}{H_0 - H^c(y_{i+h_1}|\mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)})} \\ &\in [0,1]. \end{aligned} \tag{18}$$

Intuitively, the numerator term represents the TE to capture the information about y not explained by its own history and yet explained by the history of x; the denominator term represents the information in y that is provided by the past values of both x and y. It is obvious that $\mathrm{NTE}^c_{x\to y} = 0$ if $T_{x\to y} = 0$. If y is uniformly distributed and the information about y explained by the history of both x and y is completely explained by the history of x, which means $H^c(y_{i+h_1}|\mathbf{y}_i^{(k_1)}) = H_0$, then according to (18) we obtain $\mathrm{NTE}^c_{x\to y} = 1$.

Since an entropy H represents the average number of bits needed to optimally encode independent draws of a random variable [16], the uncertain information contained in a signal is in fact proportional to 2^H. Here, a signal means a specific realization of the random variable. We extend the linear normalization function in (18) to a nonlinear function as follows:

$$\begin{aligned} \mathrm{NTE}^c_{x\to y} &= \frac{2^{H^c(y_{i+h_1}|\mathbf{y}_i^{(k_1)})} - 2^{H^c(y_{i+h_1}|\mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)})}}{2^{H_0} - 2^{H^c(y_{i+h_1}|\mathbf{y}_i^{(k_1)}, \mathbf{x}_i^{(l_1)})}} \\ &\in [0,1]. \end{aligned} \tag{19}$$

The meaning of (19) is the same as that in (18). This nonlinear normalization function (19) will be used later.

Since the $\mathrm{DTE}_{\mathrm{diff}}$ in (4) represents the information directly provided from the past x to the future y, a normalized differential DTE ($\mathrm{NDTE}_{\mathrm{diff}}$) is defined as

$$\begin{aligned} \mathrm{NDTE}^c_{x\to y} &= \frac{D_{x\to y}}{H^c(y_{i+h}|\mathbf{y}_i^{(k)}) - H^c(y_{i+h}|\mathbf{y}_i^{(k)}, \mathbf{z}_{i+h-h_3}^{(m_2)}, \mathbf{x}_{i+h-h_1}^{(l_1)})} \\ &\in [0,1) \end{aligned} \tag{20}$$

where $H^c(y_{i+h}|\mathbf{y}_i^{(k)})$ and $H^c(y_{i+h}|\mathbf{y}_i^{(k)}, \mathbf{z}_{i+h-h_3}^{(m_2)}, \mathbf{x}_{i+h-h_1}^{(l_1)})$ are the differential conditional entropies. Intuitively, this $\mathrm{NDTE}_{\mathrm{diff}}$ represents the percentage of direct causality from x to y in the total causality from both x and z to y.

D. Extension to Multiple Intermediate Variables

The definition of the $\mathrm{DTE}_{\mathrm{diff}}$ from x to y can be easily extended to multiple intermediate variables $z_1, z_2, \ldots, z_q$, as

$$\begin{aligned} D_{x\to y} = \int & f(y_{i+h}, \mathbf{y}_i^{(k)}, \mathbf{z}_{1,i_1}^{(s_1)}, \ldots, \mathbf{z}_{q,i_q}^{(s_q)}, \mathbf{x}_{i+h-h_1}^{(l_1)}) \\ & \cdot \log \frac{f(y_{i+h}|\mathbf{y}_i^{(k)}, \mathbf{z}_{1,i_1}^{(s_1)}, \ldots, \mathbf{z}_{q,i_q}^{(s_q)}, \mathbf{x}_{i+h-h_1}^{(l_1)})}{f(y_{i+h}|\mathbf{y}_i^{(k)}, \mathbf{z}_{1,i_1}^{(s_1)}, \ldots, \mathbf{z}_{q,i_q}^{(s_q)})} d\xi \end{aligned} \tag{21}$$

where $s_1, \ldots, s_q$ and $i_1, \ldots, i_q$ are the corresponding parameters determined by the calculations of the transfer entropies from $z_1, \ldots, z_q$ to y, and ξ denotes the random vector $[y_{i+h}, \mathbf{y}_i^{(k)}, \mathbf{z}_{1,i_1}^{(s_1)}, \ldots, \mathbf{z}_{q,i_q}^{(s_q)}, \mathbf{x}_{i+h-h_1}^{(l_1)}]$. If $d_{x\to y}$ is zero, then there is no direct causality from x to y, and the causal effects from x to y are all along the indirect pathways via the intermediate variables $z_1, z_2, \ldots, z_q$. If $d_{x\to y}$ is larger than zero, then there is direct causality from x to y.

The formulations of the proposed DTE and the partial TE in [25] are similar, but the basic ideas are quite different. The major difference is that for the partial TE, all the environmental variables are considered as intermediate variables, whereas for the DTE, the intermediate variables are chosen based on calculation results from the traditional TE. Specifically, the partial TE was proposed as a substitution of the traditional TE. We can choose either traditional TE to detect total causality or partial TE to detect partial causality only. However, the DTE was proposed here as an extension of the traditional TE and should be used after capturing the information flow pathways via the traditional TE method. The intermediate variables from x to y are determined as the variables within the information flow pathway from x to y (see Fig. 1) and common sources of both x and y (see Fig. 2).

More specific comparisons between the partial TE and the DTE are as follows:

1) The partial TE is defined such that all the environmental variables are considered as intermediate variables, which is not necessary in most cases and in any case this will increase the computational burden significantly. Moreover, this may even result in false causality detection. For example, given three variables x, y, and z, assume that x and y are independent and the true causal relationship between them is that x causes z and y also causes z, i.e., $x \rightarrow z \leftarrow y$. It is clear that, given the information of z, x and y are no longer independent. If we use the partial TE to detect the causality from x to y, given information of z, it is most likely to conclude that there is causality between x and y as long as there is a time delay between them, although in fact the detected causality between x and y does not exist. While using the DTE method, we need to first detect causality by using the traditional TE method, and then determine which variable is the intermediate variable. Since the information of z is unknown while using the traditional TE method in which only two variables x and y are considered, the causality between x and y cannot be detected, and then we conclude that there is no causal relationship between x and y, which is consistent with the fact.
2) For calculations of the traditional TE, the partial TE, and the DTE, it is important to determine the parameters for each variable, i.e., the prediction horizon, embedding dimensions, and the time interval. Much research has been done on how to determine these parameters for the traditional TE. For example, in [17], many simulations have been done to determine these parameters. In this paper, we also propose a parameter determination method for calculating the traditional TE in Section II-C. We can see that, for only two variables, it is not easy to determine the parameters. If all the environmental variables are considered as intermediate variables, the parameters for a large number of variables need to be chosen simultaneously and appropriately, which is nontrivial to achieve. But, for the DTE method, the parameters are determined based on the calculation of the traditional transfer entropies; thus, as long as the parameters of the traditional TE for each pair of the variables are determined, the parameters in the formulation of the DTE can be determined accordingly. Details can be found in the definition of the DTE in Section II-A.
3) From the application point of view, the utility of the partial TE is to detect unidirectional causalities [25]. The authors of [25] use the difference between the partial TE from x to y and the partial TE from y to x to quantify the causality from x to y, which is suitable in neurosciences; however, in industrial processes, feedback and bidirectional causalities are common due to recycle streams. If we still use the difference between TE from x to y and the TE from y to x to quantify the causality from x to y, it is most likely to lead to the conclusion that there is no causal relationship between x and y, although in fact there is bidirectional causality between x and y. Thus, we use the calculated TE and DTE to quantify total causality and direct causality, respectively. In addition, we propose normalization methods to quantify the strength of the total causality and direct causality, respectively.

III. Examples

In this section, we give three examples to show the usefulness of the proposed method. The first two examples use simple mathematical equations to represent causal relationships and the third example is a simulated 2×2 multiple-input multiple-output (MIMO) system.

Example 1: Assume three linear correlated continuous random variables x, y, and z satisfying

$$\begin{cases} z_{k+1} = 0.8x_k + 0.2z_k + v_{1k} \\ y_{k+1} = 0.6z_k + v_{2k} \end{cases}$$

where $x_k \sim N(0,1)$, $v_{1k}, v_{2k} \sim N(0, 0.1)$, and $z(0) = 3.2$. The simulation data consists of 6000 samples. To ensure stationarity, the initial 3000 data points were discarded.

To calculate the transfer entropies between x, z, and y, we need to determine the four design parameters. We take the TE from x to y in (1) as an example. First, we set initial values for h_1 and τ_1 as $h_1 = \tau_1 = 1$. Second, we calculate $H^c(y_{i+h_1}|\mathbf{y}_i^{(k_1)})$ with $k_1 = 0, 1, \ldots, 10$, as shown in the upper part of Fig. 4. The change rate of $H^c(y_{i+h_1}|\mathbf{y}_i^{(k_1)})$ with $k_1 = 0, 1, \ldots, 10$ is shown in the lower part of Fig. 4; we can see that as k_1 increases, the change rate of $H^c(y_{i+h_1}|\mathbf{y}_i^{(k_1)})$ does not vary sharply, which means that the history of y does not provide useful information for the future values of y. Therefore, we choose $k_1 = 0$. Finally, we calculate the TE $T_{x\rightarrow y}$ and its change rate with $l_1 = 1, \ldots, 10$, as shown in Fig. 5. Since the change rate of $T_{x\rightarrow y}$ decreases significantly after $l_1 = 2$, as shown in the lower part of Fig. 5, we choose $l_1 = 2$. Using the same procedure, the parameters for each pair of x, z, and y are determined as $h_1 = h_2 = h_3 = 1$, $\tau_1 = \tau_2 = \tau_3 = 1$, $k_1 = m_1 = k_2 = 0$, $l_1 = 2$, and $l_2 = m_2 = 1$. For the following example and case studies, the same procedure is used.

After the parameters are determined according to (19), the normalized transfer entropies between each pair of x, y, and z are shown in Table I. We can see that x causes z, z causes y, and x causes y because $\mathrm{NTE}^c_{x\rightarrow z} = 0.409$, $\mathrm{NTE}^c_{z\rightarrow y} = 0.393$, and $\mathrm{NTE}^c_{x\rightarrow y} = 0.348$ are relatively large. Thus we need to first determine whether there is direct causality from x to y. According to (4), we obtain $D_{x\rightarrow y} = 0.016$. According to (20), the normalized DTE from x to y is $\mathrm{NDTE}^c_{x\rightarrow y} = 0.016$, which is very small. Thus, we conclude that there is almost no direct causality from x to y. The information flow pathways for Example 1 are shown in Fig. 6(a).

This conclusion is consistent with the mathematical function, from which we can see that the information flow from x to y is through the intermediate variable z and there is no direct information flow pathway from x to y.

Example 2: Assume three nonlinear correlated continuous random variables x, y, and z satisfying

$$\begin{cases} z_{k+1} = 1 - 2 \mid 0.5 - (0.8x_k + 0.4\sqrt{z_k}) \mid + v_{1k} \\ y_{k+1} = 5(z_k + 7.2)^2 + 10\sqrt{\mid x_k \mid} + v_{2k} \end{cases}$$

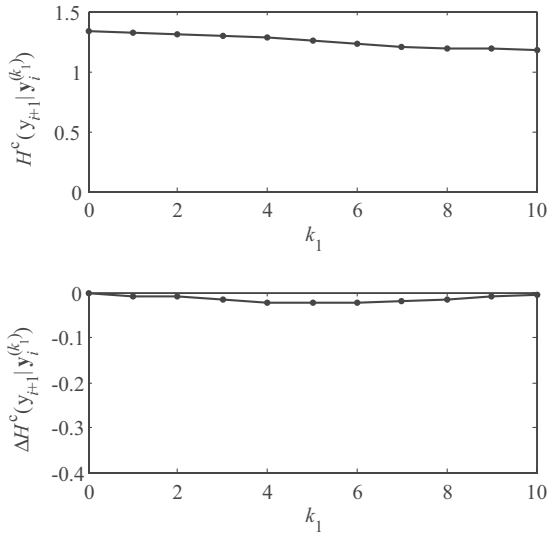


Fig. 4. Finding the embedding dimension of y for Example 1.

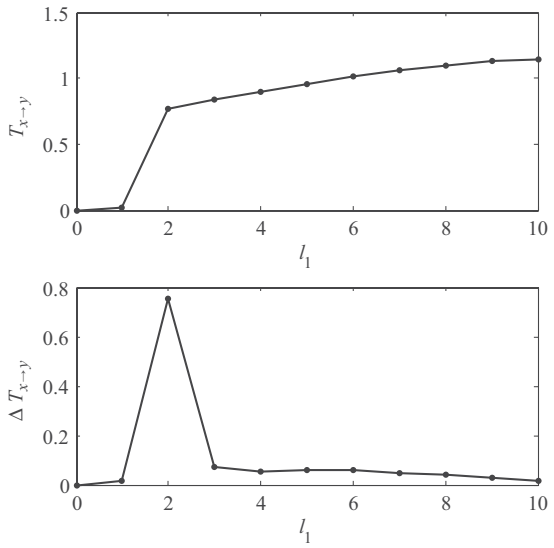


Fig. 5. Finding the embedding dimension of x for $T_{x\to y}$ of Example 1.

TABLE I
NORMALIZED TRANSFER ENTROPIES FOR EXAMPLE 1

$\text{NTE}^c_{\text{row}\to\text{column}}$	x	z	y
x	NA	0.409	0.348
z	0.058	NA	0.393
y	0.055	0.044	NA

where $x_k \in [4, 5]$ is a uniform distributed signal, $\upsilon_{1k}, \upsilon_{2k} \sim N(0, 0.05)$, and $z(0) = 0.2$. The simulation data consists of 6000 samples. To ensure stationarity, the initial 3000 data points were discarded.

The normalized transfer entropies between each pair of x, z, and y are shown in Table II. We can see that x causes z, z causes y, and x causes y because $\text{NTE}^c_{x\to z} = 0.623$, $\text{NTE}^c_{z\to y} = 0.308$, and $\text{NTE}^c_{x\to y} = 0.274$ are relatively large.

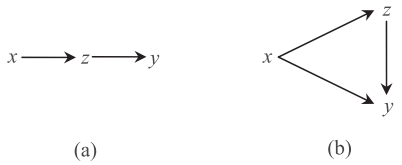


Fig. 6. Information flow pathways for (a) Example 1. (b) Example 2.

TABLE II
NORMALIZED TRANSFER ENTROPIES FOR EXAMPLE 2

$\text{NTE}^c_{\text{row}\to\text{column}}$	x	z	y
x	NA	0.623	0.274
z	0	NA	0.308
y	0	0.048	NA

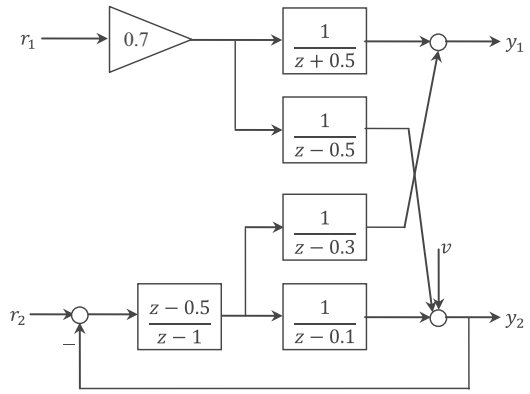


Fig. 7. System block diagram for Example 3.

Thus, we need to first determine whether there is direct causality from x to y. According to (4), we obtain $D_{x\to y} = 0.373$. According to (20), the normalized DTE from x to y is $\text{NDTE}^c_{x\to y} = 0.304$, which is much larger than zero. Thus, we conclude that there is direct causality from x to y. Second, we need to detect whether there is true and direct causality from z to y. According to (5), we obtain $D_{z\to y} = 0.538$, and thus the normalized DTE from z to y is $\text{NDTE}^c_{z\to y} = 0.438$, which is much larger than zero. Hence, we conclude that there is true and direct causality from z to y. The information flow pathways for Example 2 are shown in Fig. 6(b).

This conclusion is consistent with the mathematical function, from which we can see that there are direct information flow pathways both from x to y and from z to y.

Example 3: Fig. 7 shows a block diagram of a MIMO system with two inputs r_1 and r_2, and two outputs y_1 and y_2. Assume that $r_1 \sim N(0, 1)$ and $r_2 \sim N(0, 1)$ are independent, and $\upsilon \sim N(0, 0.1)$ is the sensor noise. The simulation data consists of 6000 samples. To ensure stationarity, the initial 3000 data points were discarded.

The normalized transfer entropies between each pair of r_1, r_2, y_1, and y_2 are shown in Table III. We can see that r_1 causes y_1 and y_2, r_2 also cause y_1 and y_2, and y_2 causes y_1. The corresponding information flow pathways are shown in Fig. 8.

TABLE III
NORMALIZED TRANSFER ENTROPIES FOR EXAMPLE 3

$\text{NTE}^c_{\text{row}\to\text{column}}$	r_1	r_2	y_1	y_2
r_1	NA	0.014	0.242	0.187
r_2	0.016	NA	0.212	0.259
y_1	0.018	0.016	NA	0.043
y_2	0.017	0.016	0.184	NA

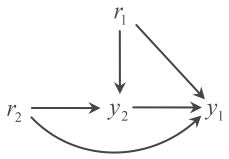


Fig. 8. Information flow pathways for Example 3.

As shown in Fig. 8, since y_1 and y_2 have common sources r_1 and r_2, we need to first detect whether the causality from y_2 to y_1 is true or spurious. According to (21), we obtain that the DTE from y_2 to y_1 with intermediate variables r_1 and r_2 is $D_{y_2\to y_1} = 0.474$. According to (20), the normalized DTE from y_2 to y_1 is $\text{NDTE}^c_{y_2\to y_1} = 0.366$, which is much larger than zero. Hence, we conclude that there is true and direct causality from y_2 to y_1.

Second, since r_1 causes y_2, y_2 causes y_1, and r_1 causes y_1, we need to further detect whether there is direct causality from r_1 to y_1. According to (4), we obtain that the DTE from r_1 to y_1 with the intermediate variable y_2 is $D_{r_1\to y_1} = 0.610$. According to (20), the normalized DTE from r_1 to y_1 is $\text{NDTE}^c_{r_1\to y_1} = 0.573$, which is much larger than zero. Thus, we conclude that there is direct causality from r_1 to y_1 in addition to the indirect causality through intermediate variable y_2. Similarly, we obtain that the DTE from r_2 to y_1 with the intermediate variable y_2 is $D_{r_2\to y_1} = 0.732$ and the normalized DTE from r_2 to y_1 is $\text{NDTE}^c_{r_2\to y_1} = 0.617$, which is also much larger than zero. Thus, we conclude that there is direct causality from r_2 to y_1. The information flow pathways are the same as those obtained from the results of calculated TEs, as shown in Fig. 8.

This conclusion is consistent with the block diagram, from which we can see that there are direct information flow pathways from r_1 to y_1, from r_2 to y_1, and from y_2 to y_1.

No matter whether the relationships of variables are linear or nonlinear, the DTE can detect direct causality and the normalized DTE can quantify the strength of direct causality.

IV. CASE STUDIES

In this section, an experimental and an industrial case studies are illustrated to validate the proposed direct causality detection method.

A. Experimental Case Study

In order to show the effectiveness of the proposed methods, a three-tank experiment was conducted. The schematic of the three-tank system is shown in Fig. 9. Water is drawn from a reservoir and pumped to tanks 1 and 2 by a gear pump and a three way valve. The water in tank 2 can flow down into tank 3. The water in tanks 1 and 3 eventually flows down into the reservoir. The experiment is conducted under open-loop conditions.

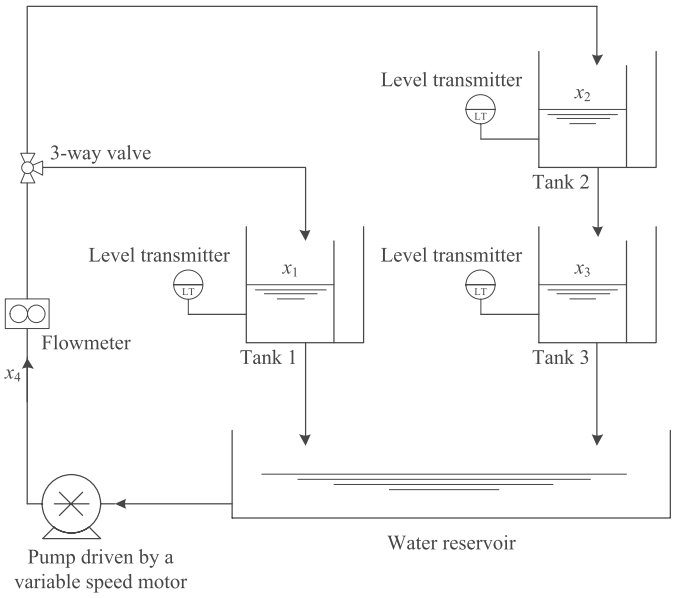


Fig. 9. Schematic of the three-tank system.

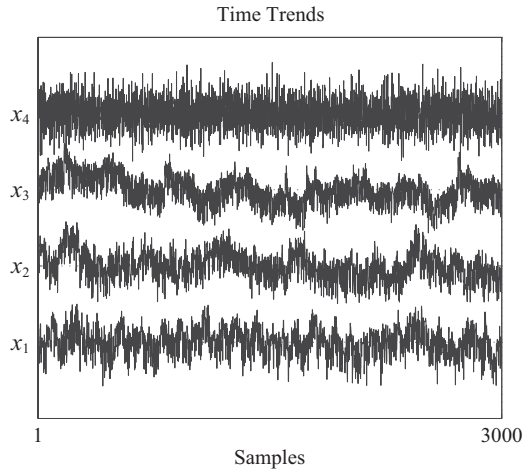


Fig. 10. Time trends of measurements of the three-tank system.

The water levels are measured by level transmitters. We denote the water levels of tanks 1–3 by x_1, x_2, and x_3, respectively. The flow rate of the water out of the pump is measured by a flow meter; we denote this flow rate by x_4. In this experiment, the normal flow rate of the water out of the pump is 10 l/min. However, the flow rate varies randomly with a mean value of 10 l/min because of the noise in the sensor and minor fluctuations in the pump. The sampled data of 3000 observations are analyzed. Fig. 10 shows the normalized time trends of the measurements. The sampling time is 1 s.

In order to detect the causality and direct causality using TE and DTE, we need to first test the stationarity of the dataset. Taking x_1 as an example, we divide the 3000 data points into

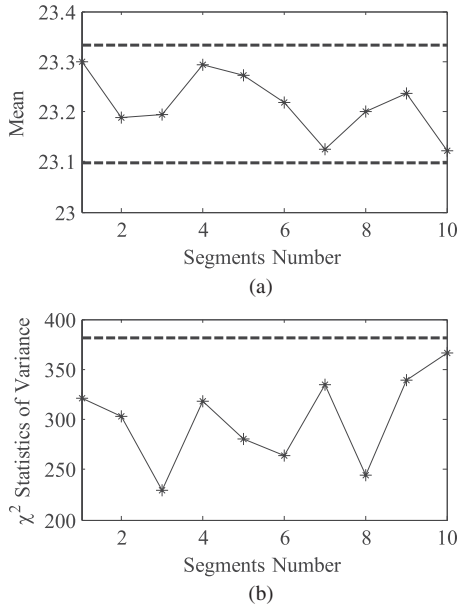


Fig. 11. Testing for stationarity. (a) Mean testing. (b) Variance testing. The dashed lines indicate the threshold.

TABLE IV
NORMALIZED TRANSFER ENTROPIES FOR THE THREE-TANK SYSTEM

$\text{NTE}^c_{\text{row}\to\text{column}}$	x_1	x_2	x_3	x_4
x_1	NA	0.024	0.010	0
x_2	0.012	NA	0.200	0
x_3	0.017	0.007	NA	0
x_4	0.199	0.171	0.152	NA

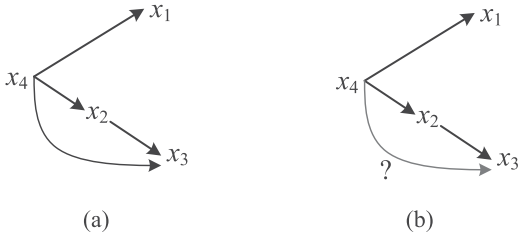


Fig. 12. Information flow pathways for the three-tank system based on the calculation results of normalized transfer entropies.

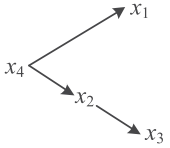


Fig. 13. Information flow pathways for three-tank system based on calculation results of normalized DTE.

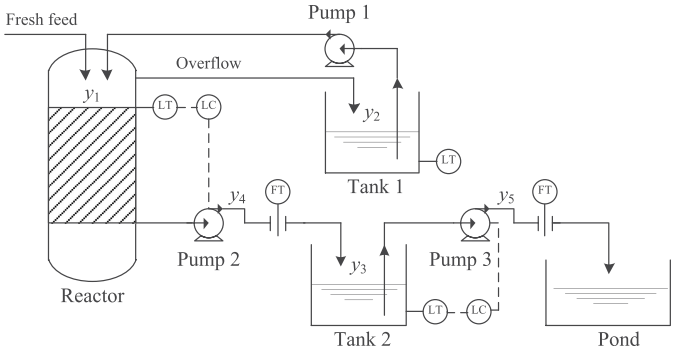


Fig. 14. Schematic of part of the FGD process.

10 consecutive segments, each containing 300 data points. The threshold of the mean values for each segment is determined by (12). In Fig. 11(a), the solid line shows the mean for each segment and the dashed line represents the threshold. Since all the mean values are within the threshold, we may conclude that the data is stationary according to the mean testing. Next, we test the properties of the variance. Here we choose $\alpha = 0.001$; thus, the threshold is $\chi^2_{300}(0.001) = 381.43$ with a 99.9% confidence. The χ^2 statistics of the variance for each segment is shown in Fig. 11(b), where the solid line shows the sum of squares of the elements for each segment after normalization and the dashed line represents the threshold. We can see that all the variance values are smaller than the threshold, and therefore we conclude that the dataset of x_1 is stationary. Using the same procedure, the stationary properties of other variables are tested. For the following industrial case study, the same procedure is used.

The normalized transfer entropies between each pair of x_1, x_2, x_3, and x_4 are shown in Table IV. We can see that x_2 causes x_3, and x_4 causes x_1, x_2, and x_3. The corresponding information flow pathways are shown in Fig. 12(a). As shown in Fig. 12(b), since x_4 causes x_2, x_2 causes x_3, and x_4 causes x_3, we need to first detect whether there is direct causality from x_4 to x_3.

According to (4), we obtain $D_{x_4\to x_3} = 0.006$. According to (20), the normalized DTE from x_4 to x_3 is $\text{NDTE}^c_{x_4\to x_3} = 0.030$, which is very small. Thus, we conclude that there is almost no direct causality from x_4 to x_3. The corresponding information flow pathways according to these calculation results are shown in Fig. 13, which are consistent with the information and material flow pathways of the physical three-tank system (see Fig. 9).

B. Industrial Case Study

Another case study is a part of a flue gas desulfurization (FGD) process at an oil company in Alberta, Canada. The schematic of this part of the process is shown in Fig. 14, including a reactor, two tanks, and a pond. Tank 1 receives the overflow from the reactor if it overflows. The liquid in Tank 1 is drawn into the reactor by Pump 1; the liquid in the reactor is drawn into Tank 2 by Pump 2, and the liquid level of the reactor is controlled by adjusting the flow rate of the liquid out of Pump 2; the liquid in Tank 2 is drawn into the pond by Pump 3, and the liquid level of Tank 2 is controlled by adjusting the flow rate of the liquid out of Pump 3. These two level control loops imply that there is a bidirectional relationship between the levels and the flow out of the tank due to material as well as information (due to feedback) flow pathways.

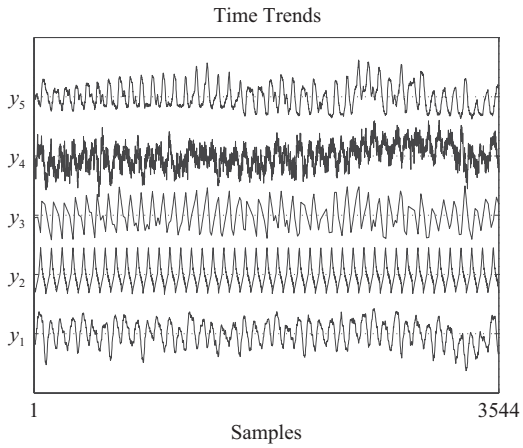


Fig. 15. Time trends of measurements of the FGD process.

TABLE V
NORMALIZED TRANSFER ENTROPIES FOR PART OF FGD PROCESS

$NTE^c_{row\to column}$	y_1	y_2	y_3	y_4	y_5
y_1	NA	0.001	0.089	0.177	0.014
y_2	0.131	NA	0.117	0.154	0.010
y_3	0.078	0.005	NA	0.008	0.105
y_4	0.128	0.005	0.095	NA	0.019
y_5	0.016	0.001	0.130	0.012	NA

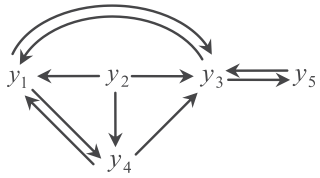


Fig. 16. Information flow pathways for part of FGD process based on calculation results of normalized transfer entropies.

We denote the liquid levels of the reactor, Tanks 1, and 2 by y_1, y_2, and y_3, respectively. There is no measurement of the flow rate of the liquid out of Pump 1. We denote the flow rates of the liquid out of pumps 2 and 3 by y_4 and y_5, respectively. The sampled data of 3544 observations are analyzed. Fig. 15 shows the normalized time trends of the measurements. The sampling time is 1 min.

The normalized transfer entropies between each pair of y_1, y_2, y_3, y_4, and y_5 are shown in Table V. We can choose the threshold as 0.02: if the normalized TE is less than 0.02, then there is almost no causality. The information flow pathways based on the normalized transfer entropies are shown in Fig. 16. We need to further determine whether the causality between y_1, y_2, y_3, y_4, and y_5 is true and direct.

Calculation steps of direct transfer entropies and corresponding simplified information flow pathways are shown in Fig. 17. We first determine whether the causality between y_1 and y_3 is true and direct by considering y_2 and y_4 as the possible intermediate variables (Steps 1 and 2). The calculation results of DTE and normalized DTE are shown in Table VI. Since the normalized DTEs between y_1 and y_3 are very small,

TABLE VI
CALCULATED AND NORMALIZED DTES FOR PART OF FGD PROCESS

	Intermediate Variable (s)	DTE	Normalized DTE
$y_1 \to y_3$	y_2, y_4	0.031	0.024
$y_3 \to y_1$	y_2, y_4	0.028	0.023
$y_2 \to y_1$	y_4	0.374	0.425
$y_2 \to y_4$	y_1	0.013	0.025
$y_2 \to y_3$	y_1, y_4	0.027	0.021

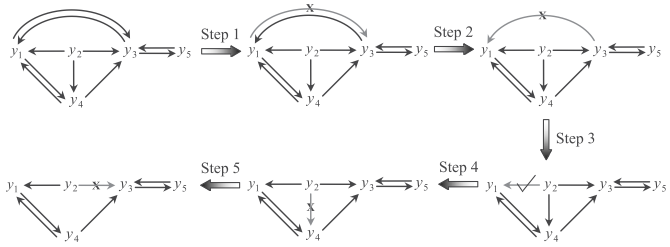


Fig. 17. Calculation steps of direct transfer entropies.

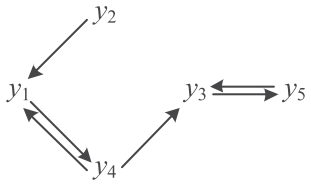


Fig. 18. Information flow pathways for part of FGD process based on calculation results of normalized DTE.

we conclude that there is almost no direct causality between them. Secondly, we determine whether the causality from y_2 to y_1 is direct by considering the possible intermediate variable y_4 (Step 3). Similarly, the causality from y_2 to y_4 can be determined by considering the possible intermediate variable y_1 (Step 4). Finally, we detect the direct causality from y_2 to y_3 with the possible intermediate variables y_1 and y_4 (Step 5). Based on the calculation results shown in Table VI, we conclude that, except for the causality from y_2 to y_1, the other detected causality is indirect or spurious. Note that here we do not need to further detect the direct causality between y_1 and y_4, from y_4 to y_3, and from y_5 to y_3 since there is no possible intermediate variable in their pathways. The information flow pathways based on calculated direct transfer entropies are shown in Fig. 18.

An overview of causality between process variables is shown in Fig. 19. Causal relationships from variables on the vertical axis to variables on the horizontal axis are represented by three different symbols. '.' means no causality; '▲' means direct causality; '△' means causality can be detected but it is indirect or spurious.

From Fig. 18, we can see that the spurious causality between the liquid levels of the reactor and Tank 2, i.e., between y_1 and y_3, is generated by the flow rate of the liquid out of Pump 2, i.e., y_4. Similarly, if we can obtain the measurement of the flow rate of the liquid out of Pump 1, the causality from the liquid level of Tank 1 to the liquid level of the reactor, i.e.,

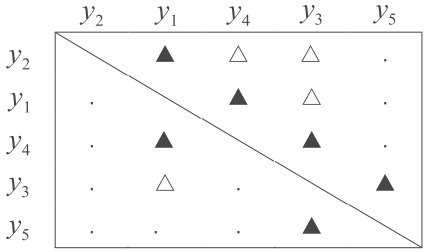


Fig. 19. Overview of causal relationships between FGD process variables. "." means no causality; "▲" means direct causality; "△" means causality can be detected, but it is indirect or spurious.

from y_2 to y_1, will also disappear. However, since the flow rate of the liquid out of Pump 1 is not measured, we still say that there is direct and true causality from y_2 to y_1. Thus, the connecting pathways shown in Fig. 18 are consistent with the information and material flow pathways of the physical process shown in Fig. 14, where the solid lines indicate material flow pathways and the dashed lines denote control loops. Note that, as mentioned earlier, the bidirectional causality between y_1 and y_4 and between y_3 and y_5 are due to the level feedback control loops.

V. Conclusion

In industrial processes, abnormalities often spread from one process variable to neighboring variables. It is important to determine the fault propagation pathways to find the root cause of the abnormalities and the corresponding fault propagation routes. TE can measure the causality between two process variables, i.e., the direction of the information flow. Furthermore, it is valuable to detect whether the influence is along direct or indirect pathways. A direct causality detection method based on the DTE has been proposed to detect whether there is a direct information and/or material flow pathway between each pair of variables. The DTE_{diff} for continuous random variables was defined based on an extension of the TE, which is suitable for both linear and nonlinear relationships. The TE_{diff} and the DTE_{diff} were shown, respecitively to be equivalent to the TE_{disc} and the DTE_{disc} in the limit as the quantization bin sizes approach zero. The NTE_{diff} and the $\text{NDTE}_{\text{diff}}$ were defined to measure the connectivity strength of causality and direct causality, respectively. The proposed methods were validated by two examples and two case studies.

Although the proposed method can be used to detect the direct causality between two process variables, there are still a number of unresolved questions.

1) The detection of direct information flow can be reformulated as a hypothesis test problem. Taking the direct causality from x to y with an intermediate variable z as an example, the null hypothesis should be that there is no direct causality from x to y and that the causality from x to y is indirect through z. In order to carry out this hypothesis testing, similar to the TE, we may use the bootstrap method [25] or the Monte Carlo method [17] by constructing resampling data or surrogate data (randomly shuffled data or by the iterative amplitude adjusted Fourier transform (iAAFT) method [35]). However, the constructed data must satisfy the null hypothesis that the direct information flow from x to y must be completely destroyed while the indirect pathway through z still exists. At the same time, the statistical properties of x, y, and z should not change. It is generally difficult to construct such surrogate or resampling data. Thus, our ongoing study is related to the confidence level determination of the DTE.
2) Similar to the TE, the calculation of the DTE needs to estimate the high-dimensional joint pdfs; for example, the dimension of $f(y_{i+h}, \mathbf{y}_i^{(k)}, \mathbf{z}_{i+h-h_3}^{(m_2)}, \mathbf{x}_{i+h-h_1}^{(l_1)})$ in (4) is $m_2 + l_1 + k + 1 \geq 3$. It is important to employ an accurate (less Type I and Type II errors) and efficient (less computational burden with a certain accuracy level) pdf estimation algorithm. Although the kernel estimation method is widely used, with increasing dimension of the variables, a more accurate and efficient pdf estimation algorithm needs to be developed.

Acknowledgment

The authors would like to thank Prof. N. F. Thornhill, Imperial College London, London, U.K., for comments and suggestions on this paper, Dr. V. Bavdekar, University of Alberta, Edmonton, AB, for discussions and suggestions, and Dr. Y. Shardt, University of Alberta, for help with the experiments.

References

[1] F. Yang, S. L. Shah, and D. Xiao, "Signed directed graph based modeling and its validation from process knowledge and process data," *Int. J. Appl. Math. Comput. Sci.*, vol. 22, no. 1, pp. 41–53, 2012.

[2] D. S. Nam, C. Han, C. W. Jeong, and E. S. Yoon, "Automatic construction of extended symptom-fault associations from the signed digraph," *Comput. Chem. Eng.*, vol. 20, no. 1, pp. S605–S610, 1996.

[3] M. Bauer, N. F. Thornhill, and A. Meaburn, "Specifying the directionality of fault propagation paths using transfer entropy," in *Proc. 7th Int. Symp. Dynamics Control Process Syst.*, Cambridge, MA, Jul. 2004, pp. 203–208.

[4] M. R. Maurya, R. Rengaswamy, and V. Venkatasubramanian, "A systematic framework for the development and analysis of signed digraphs for chemical processes. 1. Algorithms and analysis," *Ind. Eng. Chem. Res.*, vol. 42, no. 20, pp. 4789–4810, 2003.

[5] M. R. Maurya, R. Rengaswamy, and V. Venkatasubramanian, "A systematic framework for the development and analysis of signed digraphs for chemical processes. 2. Control loops and flowsheet analysis" *Ind. Eng. Chem. Res.*, vol. 42, no. 20, pp. 4811–4827, 2003.

[6] M. Winterhalder, B. Schelter, W. Hesse, K. Schwab, L. Leistriz, D. Klan, R. Bauer, J. Timmer, and H. Witte, "Comparison of linear signal processing techniques to infer directed interactions in multivariate neural systems," *Signal Process.*, vol. 85, no. 11, pp. 2137–2160, 2005.

[7] M. J. Kaminski and K. J. Blinowska, "A new method of the description of the information flow in the brain structures," *Biol. Cybern.*, vol. 65, no. 3, pp. 203–210, 1991.

[8] L. A. Baccala and K. Sameshima, "Partial directed coherence: A new concept in neural structure determination," *Biol. Cybern.*, vol. 84, no. 6, pp. 463–474, 2001.

[9] C. W. J. Granger, "Investigating causal relationships by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969.

[10] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1982.

[11] R. B. Govindan, J. Raethjen, F. Kopper, J. C. Claussen, and G. Deuschl, "Estimation of time delay by coherence analysis," *Phys. A, Stat. Mech. Appl.*, vol. 350, nos. 2–4, pp. 277–295, 2005.

[12] U. Feldmann and J. Bhattacharya, "Predictability improvement as an asymmetrical measure of interdependence in bivariate time series," *Int. J. Bifurcat. Chaos Appl. Sci. Eng.*, vol. 14, no. 2, pp. 505–514, 2004.

[13] M. Bauer and N. F. Thornhill, "Measuring cause and effect between process variables," in *Proc. Adv. Process Control Appl. Ind. Workshop*, Vancouver, BC, Canada, May 2005, pp. 1–6.

[14] M. Bauer, J. W. Cox, M. H. Caveness, J. J. Downs, and N. F. Thornhill, "Nearest neighbors methods for root cause analysis of plantwide disturbances," *Ind. Eng. Chem. Res.*, vol. 46, no. 18, pp. 5977–5984, 2007.

[15] K. Hlavackova-Schindler, M. Palus, M. Vejmelka, and J. Bhattacharya, "Causality detection based on information-theoretic approaches in time series analysis," *Phys. Rep.*, vol. 441, no. 1, pp. 1–46, 2007.

[16] T. Schreiber, "Measuring information transfer," *Phys. Rev. Lett.*, vol. 85, no. 2, pp. 461–464, 2000.

[17] 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 Trans. Control Syst. Technol.*, vol. 15, no. 1, pp. 12–21, Jan. 2007.

[18] R. Vicente, M. Wibral, M. Lindner, and G. Pipa, "Transfer entropy—a model-free measure of effective connectivity for the neurosciences," *J. Comput. Neurosci.*, vol. 30, no. 1, pp. 45–67, 2011.

[19] L. A. Overbey and M. D. Todd, "Dynamic system change detection using a modification of the transfer entropy," *J. Sound Vibrat.*, vol. 322, nos. 1–2, pp. 438–453, 2009.

[20] L. Barnett, A. B. Barrett, and A. K. Seth, "Granger causality and transfer entropy are equivalent for Gaussian variables," *Phys. Rev. Lett.*, vol. 103, no. 23, pp. 238701-1–238701-4, 2009.

[21] K. Hlavackova-Schindler, "Equivalence of granger causality and transfer entropy: A generalization," *Appl. Math. Sci.*, vol. 5, no. 73, pp. 3637–3648, 2011.

[22] M. Lungarella, K. Ishiguro, Y. Kuniyoshi, and N. Otsu, "Methods for quantifying the causal structure of bivariate time series," *Int. J. Bifurcat. Chaos*, vol. 17, no. 3, pp. 903–921, 2007.

[23] S. Gigi and A. K. Tangirala, "Quantitative analysis of directional strengths in jointly stationary linear multivariate processes," *Biol. Cybern.*, vol. 103, no. 2, pp. 119–133, 2010.

[24] B. Huang, N. F. Thornhill, S. L. Shah, and D. Shook, "Path analysis for process troubleshooting," in *Proc. Adv. Control Ind. Process.*, Kumamoto, Japan, Jun. 2002, pp. 149–154.

[25] V. A. Vakorin, O. A. Krakovska, and A. R. McIntosh, "Confounding effects of indirect connections on causality estimation," *J. Neurosci. Methods*, vol. 184, no. 1, pp. 152–160, 2009.

[26] R. W. Yeung, *Information Theory and Network Coding*. New York: Springer-Verlag, 2008, pp. 12–31.

[27] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1997.

[28] J. Yu, "A nonlinear kernel Gaussian mixture model based inferential monitoring approach for fault detection and diagnosis of chemical processes," *Chem. Eng. Sci.*, vol. 68, no. 1, pp. 506–519, 2012.

[29] E. J. Keogh, S. Chu, D. Hart, and M. Pazzani, "Segmenting time series: A survey and novel approach," in *Data Mining in Time Series Databases*, vol. 57. Singapore: World Scientific, 2004, pp. 1–22.

[30] A. Denis and F. Cremoux, "Using the entropy of curves to segment a time or spatial series," *Math. Geol.*, vol. 34, no. 8, pp. 899–914, 2002.

[31] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. New York: Chapman & Hall, 1986, pp. 34–48.

[32] Q. Li and J. S. Racine, *Nonparametric Econometrics: Theory and Practice*. Princeton, NJ: Princeton Univ. Press, 2007, pp. 4–15.

[33] B. Gourevitch and J. J. Eggermont, "Evaluating information transfer between auditory cortical neurons," *J. Neurophysiol.*, vol. 97, no. 3, pp. 2533–2543, 2007.

[34] J. Beirlant, E. J. Dudewicz, L. Gyorfi, and E. C. van der Meulen, "Nonparametric entropy estimation: An overview," *Int. J. Math. Stat. Sci.*, vol. 6, no. 1, pp. 17–39, 1997.

[35] T. Schreiber and A. Schmitz, "Surrogate time series," *Phys. D*, vol. 142, nos. 3–4, pp. 346–382, 2000.

Ping Duan received the B.Eng. degree in automation from Central South University, Changsha, China, in 2005. From September 2005 to August 2009, she was a Ph.D. student in the School of Information Science and Engineering at Central South University. Since September 2009, she has been a Ph.D. student in the Department of Electrical and Computer Engineering at the University of Alberta, Edmonton, AB, Canada

Her current research interests include process connectivity analysis, causality analysis, and detection and diagnosis of plant-wide disturbances.

Fan Yang (M'06) received the B.Eng. degree in automation and the Ph.D. degree in control science and engineering from Tsinghua University, Beijing, China, in 2002 and 2008, respectively.

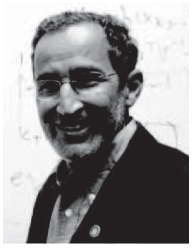
He was a PDF with Tsinghua University and the University of Alberta, Edmonton, AB, Canada. He joined the Department of Automation, Tsinghua University, as a Lecturer, in 2011. His current research interests include topology modeling of large-scale processes, abnormal events monitoring, process hazard analysis, and smart alarm management.

Dr. Yang received the Young Research Paper Award from the IEEE Control Systems Society Beijing Chapter in 2006, the Outstanding Graduate Award from Tsinghua University in 2008, and the Teaching Achievement Award from Tsinghua University in 2012.

Tongwen Chen (F'06) received the B.Eng. degree in automation and instrumentation from Tsinghua University, Beijing, China, in 1984, and the M.A.Sc. and Ph.D. degrees in electrical engineering from the University of Toronto, Toronto, ON, Canada, in 1988 and 1991, respectively.

He is currently a Professor of electrical and computer engineering with the University of Alberta, Edmonton, AB, Canada. His current research interests include computer and network based control systems, process safety and alarm systems, and their applications to the process and power industries.

Dr. Chen has served as an Associate Editor for several international journals, including the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, *Automatica*, and SYSTEMS AND CONTROL LETTERS.



Sirish L. Shah (M'76) is a Faculty with the University of Alberta, Edmonton, AB, Canada, where he held the NSERC-Matrikon-Suncor-iCORE Senior Industrial Research Chair in Computer Process Control from 2000 to 2012. He has held visiting appointments at Oxford University, Oxford, U.K., and Balliol College as a SERC Fellow from 1985 to 1986, Kumamoto University, Kumamoto, Japan, as a Senior Research Fellow of the Japan Society for the Promotion of Science in 1994, the University of Newcastle, Newcastle, Australia, in 2004, IIT-Madras, Chennai, India, in 2006, and the National University of Singapore, Singapore, in 2007. He has co-authored two books, entitled *Performance Assessment of Control Loops: Theory and Applications*, and *Diagnosis of Process Nonlinearities and Valve Stiction: Data Driven Approaches*. His current research interests include process and performance monitoring, system identification and design, and implementation of soft sensors.