Transfer Zero-Entropy and Its Application for Capturing Cause and Effect Relationship Between Variables

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

***Abstract*— Detection of causality is an important and challenging problem in root cause and hazard propagation analysis. It has been shown that the transfer entropy approach is a very useful tool in quantifying directional causal influence for both linear and nonlinear relationships. A key assumption for this method is that the sampled data should follow a well-defined probability distribution; yet this assumption may not hold for some industrial process data. In this paper, a new information theory-based measure, transfer 0-entropy (T0E), is proposed for causality analysis on the basis of the definitions of 0-entropy and 0-information without assuming a probability space. For the cases of more than two variables, a direct T0E (DT0E) concept is presented to detect whether there is a direct information and/or material flow pathway from one variable to another. Estimation methods for the T0E and the DT0E are addressed. The effectiveness of the proposed method is illustrated by two data sets, one based on data from a pilot scale process and a second evaluation based on data from a benchmark industrial case study.**

***Index Terms*— 0-entropy, causality analysis, direct transfer 0-entropy (DT0E), root cause diagnosis, transfer 0-entropy (T0E).**

I. Introduction

WHEN a disturbance is generated somewhere in a plant and propagates to the whole plant or specific units of the plant through information and/or material flow pathways, it is termed as a plant-wide disturbance [1]. Plant-wide disturbances are common in many processes because of highly integrated energy utilization streams as well as the presence of recycle streams. Their presence may impact the overall process performance and cause inferior quality products, larger rejection rates, excessive energy consumption, and even hazardous events. Thus, it is important to diagnose the root cause(s) of such disturbances to compensate for them.

Manuscript received April 21, 2014; accepted June 30, 2014. Date of publication August 26, 2014; date of current version April 14, 2015. Manuscript received in final form July 27, 2014. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada, in part by the Tsinghua University Initiative Scientific Research Program, in part by the 111 Project under Grant B08015, and in part by China Scholarship Council. 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 Tsinghua National Laboratory for Information Science and Technology, Department of Automation, 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).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCST.2014.2345095

Causality analysis provides an effective way to localize root cause of plant-wide abnormalities and disturbances since a causal map can represent the direction of disturbance propagation and allow investigation of fault propagation pathways [2], [3]. The basic idea of causality can be traced back to Wiener [4] who developed a mathematical definition for causality: given two random variables X and Y, X could be termed to cause Y if the predictability of Y is improved by incorporating information about X.

Wiener's idea lacked the machinery for practical implementation. It was Granger [5] who adapted this definition into the experimental practice, namely, analysis of data observed in consecutive time series. He formalized the prediction idea in the context of linear regression models [5]: X is said to have a causal influence on Y if the variance of the autoregressive prediction error of Y at the present time is reduced by inclusion of past measurements of X. From the definition, we can see that the flow of time or the notion of temporal asymmetry is a key point in causality analysis. Therefore, the interaction discovered by causality detection may be unidirectional or bidirectional. This directional interaction is the major difference between causal influence and relations reflected by the symmetric measures, such as ordinary coherence and mutual information.

Inspired by Granger's work, many advanced techniques for causality detection have been proposed, such as the extended and nonlinear Granger causality [6], directed transfer functions and partial directed coherence [7], predictability improvement [8], [9], nearest neighbors [10], and transfer entropy (TE) [11]–[13]. It has been pointed out that TE is a very useful tool in quantifying directional causal influence for both linear and nonlinear relationships; it has been successfully used in neurosciences [14] and chemical processes to find the direction of disturbance propagation and to diagnose the root cause [12]. Since the development of TE, there have been a lot of activities to extend ideas of causality detection. For example, partial TE [15] and direct TE [16] have been proposed to detect partial causality and direct causality, respectively. Recently, the concept of Rényian TE (RTE) was proposed in [17] as a measure of information that is transferred only between certain parts of underlying distributions. The authors have shown the usefulness of the RTE on stock market time series.

TE was proposed based on the key concept of Shannon's entropy that is defined stochastically as the averaged number of bits needed to optimally encode a source data set X with the

source probability distribution $P(X)$ [18]. Shannon's entropy represents the average unpredictability in a random variable. In other words, it is a measure of the uncertainty associated with a random variable. For a discrete-valued random variable X, assume X has n outcomes $\{x_1, \ldots, x_n\}$, then Shannon entropy is defined as [18]

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log p(x_i)$$

where $p(x_i)$ denotes the probability mass function of the outcome x_i, the base of the logarithm is two, and the unit is in bits. For example, a single toss of a fair coin has an entropy of 1 bit. A series of two fair coin tosses has an entropy of 2 bits. The number of fair coin tosses is its entropy in bits. The entropy rate for a fair coin toss is 1 bit per toss. However, if the coin is not fair, then the uncertainty (entropy rate) for each toss is lower. The reason is that if asked to predict the next outcome, we could choose the most frequent result and the prediction would be correct more often than wrong [19].

One reason for the definition of Shannon's entropy is that random variables in communication systems are generally prone to electronic circuit noises, which obey physical laws yielding well-defined distributions. In contrast, in industrial processes that contain a lot of mechanical and chemical components, the dominant disturbances may not follow a well-defined probability distribution since they may not necessarily arise from circuit noise [20]. Consequently, in process control, disturbances and uncertainties are sometimes treated as bounded unknowns or signals without *a priori* statistical structure.

In context of the aforementioned issue, a natural question to ask is: without assuming a probability space, is it possible to construct a useful analogue of the stochastic concept of the Shannon's entropy? Hartley entropy or 0-entropy H_0 [21] for discrete variables, and Rényi differential zeroth-order entropy or Rényi differential 0-entropy h_0 [22] for continuous variables provide answers to this question. If a random variable has a known range but an unknown distribution, then its uncertainty can be quantified by the logarithm of the cardinality (H_0) or the logarithm of the Lebesgue measure of its support (h_0). A related natural question is: without assuming a probability space, is it possible to construct a useful analogue of the TE for causality detection? This study is an attempt to provide an answer to this question.

The main contribution of this paper is a new information theory method to detect causal relationships between process variables of linear or nonlinear multivariate systems without assuming a probability space. The basic idea of this causality detection method was inspired by the concepts of the 0-entropy and 0-information described in [20].

The rest of this paper is organized as follows. In Section II, after introducing concepts of the (conditional) range, 0-entropy and 0-information, we define a transfer 0-entropy (T0E) to detect total causality and a direct T0E (DT0E) to detect and discriminate between direct and indirect causal relationships, respectively. The calculation method for T0E is proposed and the range estimation method for random variables is addressed in Section III. In Section IV, two numerical examples are described to illustrate the effectiveness of the proposed causality detection method. Two data sets, one from an experimental case study and a second from a benchmark industrial case study are introduced in the same section to demonstrate the utility of the proposed method for finding material and/or information flow pathways and fault propagation pathways for root cause diagnosis. In Section V, the conclusion is drawn.

II. Detection of Causality and Direct Causality

In this section, a T0E concept based on 0-entropy and 0-information is proposed to detect causality between two variables. In addition to this, a DT0E is proposed to detect whether there is direct causal influence from one variable to another.

A. Preliminaries

Before introducing the concept of the T0E, we describe the nonprobabilistic formulations of range, 0-entropy, and 0-information.

A random variable Y can be considered as a mapping from an underlying sample space $\mathbf{\Omega}$ to a set $\mathbf{Y}$ of interest. Each sample $\omega \in \mathbf{\Omega}$ can give rise to a realization $Y(\omega)$ denoted by $y \in \mathbf{Y}$. Then, the marginal range of Y is defined as [20]

$$[\![Y]\!] = \{Y(\omega) : \omega \in \mathbf{\Omega}\} \tag{1}$$

where $\{\cdot\}$ indicates a set. Given another random variable X taking values in $\mathbf{X}$, the conditional range of Y given $X(\omega) = x$ is defined as

$$[\![Y|x]\!] = \{Y(\omega) : X(\omega) = x, \omega \in \mathbf{\Omega}\}. \tag{2}$$

The relationship between the marginal range of Y and its conditional range given $X(\omega) = x$ satisfies that

$$\bigcup_{x \in [\![X]\!]} [\![Y|x]\!] = [\![Y]\!]. \tag{3}$$

The joint range of Y and X is defined as

$$[\![Y, X]\!] = \{(Y(\omega), X(\omega)) : \omega \in \mathbf{\Omega}\}. \tag{4}$$

The joint range is determined by the conditional and marginal ranges as follows:

$$[\![Y, X]\!] = \bigcup_{x \in [\![X]\!]} [\![Y|x]\!] \times \{x\} \tag{5}$$

where $\times$ represents the Cartesian product.

Variables Y and X are said to be unrelated iff the conditional range satisfies $[\![Y|x]\!] = [\![Y]\!]$, where $x \in [\![X]\!]$. Given another random variable Z that takes values in $\mathbf{Z}$, variables Y and X are said to be unrelated conditional on Z iff $[\![Y|x, z]\!] = [\![Y|z]\!]$, where $(x, z) \in [\![X, Z]\!]$ [20].

For example, Fig. 1(a) shows the case of two related variables Y and X. For a certain value $x \in [\![X]\!]$, the conditional range $[\![Y|x]\!]$ is strictly contained in the marginal range $[\![Y]\!]$. Note that in this case, the joint range $[\![Y, X]\!]$ is also strictly contained in the Cartesian product of marginal ranges, namely, $[\![Y]\!] \times [\![X]\!]$. Fig. 1(b) shows the ranges when Y and X are unrelated. For any $x \in [\![X]\!]$, the conditional range $[\![Y|x]\!]$

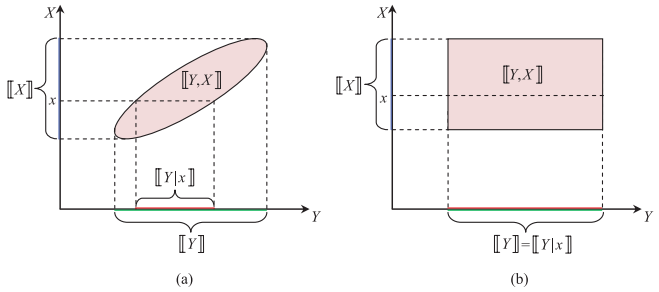


Fig. 1. Examples of marginal, conditional, and joint ranges for related and unrelated random variables (adapted from [20]). (a) Y and X are related. (b) Y and X are unrelated.

coincides with the marginal range $[\![Y]\!]$. Moreover, the joint range $[\![Y, X]\!]$ coincides with $[\![Y]\!] \times [\![X]\!]$.

Let $|\cdot|$ denotes set cardinality and μ denotes the Lebesgue measure. A function $\phi([\![Y]\!])$ is defined as

$$\phi([\![Y]\!]) = \begin{cases} |[\![Y]\!]| & \text{for discrete-valued } Y \\ \mu[\![Y]\!] & \text{for continuous-valued } Y \end{cases} \tag{6}$$

where $|[\![Y]\!]|$ indicates the set cardinality of $[\![Y]\!]$ for discrete-valued Y, and $\mu[\![Y]\!]$ can be understood as the length of the range $[\![Y]\!]$ for continuous-valued Y. The uncertainty associated with Y can be captured by the (marginal) 0-entropy defined as

$$H_0(Y) = \log \phi([\![Y]\!]) \tag{7}$$

where the base of the logarithm is 2, and the unit of H_0 is in bits. Note that if Y is a discrete-valued random variable, then $H_0(Y)$ represents the (marginal) Hartley entropy or 0-entropy [21] satisfying $H_0(Y) \in [0, \infty)$; if X is continuous valued, then $H_0(Y)$ indicates the (marginal) Rényi differential 0-entropy [22] that satisfies $H_0(Y) \in (-\infty, \infty)$.

A worst-case approach is taken to define the conditional 0-entropy of Y given X as follows [20], [23]:

$$H_0(Y|X) = \operatorname*{ess\,sup}_{x \in [\![X]\!]} \log \phi([\![Y|x]\!]) \tag{8}$$

where ess sup represents the essential supremum. Essential supremum means supremum for almost everywhere except a set of measure zero. In other words, when there are some outliers whose measure is zero, the calculation of essential supremum would ignore these outliers while the calculation of supremum would consider these outliers. $H_0(Y|X)$ can be understood as a measurement of the uncertainty that remains in Y after X is known.

To measure the information about Y gained from X, a nonprobabilistic 0-information metric from X to Y, $I_0(Y; X)$ is defined as follows [20], [23]:

$$I_0(Y; X) = H_0(Y) - H_0(Y|X) = \operatorname*{ess\,inf}_{x \in [\![X]\!]} \log \left(\frac{\phi([\![Y]\!])}{\phi([\![Y|x]\!])} \right) \tag{9}$$

where ess inf represents the essential infimum. From the definition, we can see that the 0-information is the worst-case log ratio of the prior to the posterior range set sizes/lengths, and it can be shown that $I_0(Y; X)$ is always nonnegative. $I_0(Y; X)$ represents the reduction in uncertainty about Y after X is known; thus, it can be understood as the information about Y provided by X. Note that the definition of the 0-information is asymmetric, that is, $I_0(Y; X) \neq I_0(X; Y)$.

B. Transfer 0-Entropy

The concept of 0-information provides an effective way to measure the information about Y provided by X. However, the time flow information is not considered in this definition. Since the time flow information is an important component in causality detection, 0-information cannot be directly used for causality analysis. To incorporate this, we propose a T0E concept for causality detection based on the concept of 0-information.

Before introducing the concept of T0E, a conditional 0-information from X to Y given Z is defined as follows:

$$I_0(Y; X|Z) = H_0(Y|Z) - H_0(Y|X, Z) \tag{10}$$

where $H_0(Y|Z)$ and $H_0(Y|X, Z)$ denote conditional 0-entropies defined in (8). The conditional 0-information measures the information about Y provided by X when Z is given.

Now consider two random variables X and Y with marginal ranges $[\![X]\!]$ and $[\![Y]\!]$ and joint range $[\![X, Y]\!]$, let them be sampled at time instant i to get X_i and Y_i with $i = 1, 2, \ldots, N$, where N is the number of samples.

Let Y_{i+h} denote the value of Y at time instant $i + h$, that is, h steps in the future from i, and h is referred to as the prediction horizon; $\mathbf{Y}_i^{(k)} = [Y_i, Y_{i-\tau}, \ldots, Y_{i-(k-1)\tau}]$ and $\mathbf{X}_i^{(l)} = [X_i, X_{i-\tau}, \ldots, X_{i-(l-1)\tau}]$ denote embedding vectors with elements from the past values of Y and X, respectively (k and l are the embedding dimensions of Y and X, respectively); τ is the time interval that allows the scaling in time of the embedded vector, which can be set to be $\tau = h$ as a rule of thumb [12]. Let $\mathbf{y}_i^{(k)} = [y_i, y_{i-\tau}, \ldots, y_{i-(k-1)\tau}]$ and $\mathbf{x}_i^{(l)} = [x_i, x_{i-\tau}, \ldots, x_{i-(l-1)\tau}]$ denote realizations of $\mathbf{Y}_i^{(k)}$ and $\mathbf{X}_i^{(l)}$, respectively. Thus, $[\![Y_{i+h}|\mathbf{x}_i^{(l)}, \mathbf{y}_i^{(k)}]\!]$ denotes the conditional range of Y_{i+h} given $\mathbf{X}_i^{(k)} = \mathbf{x}_i^{(l)}$ and $\mathbf{Y}_i^{(k)} = \mathbf{y}_i^{(k)}$, and $[\![Y_{i+h}|\mathbf{y}_i^{(k)}]\!]$ denotes the conditional range of Y_{i+h} given $\mathbf{Y}_i^{(k)} = \mathbf{y}_i^{(k)}$. The T0E from X to Y is then defined as follows:

$$T^0_{X \to Y} = I_0(Y_{i+h}; \mathbf{X}_i^{(l)} | \mathbf{Y}_i^{(k)}) \tag{11}$$

$$= H_0(Y_{i+h}|\mathbf{Y}_i^{(k)}) - H_0(Y_{i+h}|\mathbf{X}_i^{(l)}, \mathbf{Y}_i^{(k)}) \tag{12}$$

$$= \operatorname*{ess\,sup}_{\mathbf{y}_i^{(k)} \in [\![\mathbf{Y}_i^{(k)}]\!]} \log \phi([\![Y_{i+h}|\mathbf{y}_i^{(k)}]\!]) - \operatorname*{ess\,sup}_{(\mathbf{x}_i^{(l)}, \mathbf{y}_i^{(k)}) \in [\![\mathbf{X}_i^{(l)}, \mathbf{Y}_i^{(k)}]\!]} \log \phi([\![Y_{i+h}|\mathbf{x}_i^{(l)}, \mathbf{y}_i^{(k)}]\!])$$

$$= \log \frac{\operatorname*{ess\,sup}_{\mathbf{y}_i^{(k)} \in [\![\mathbf{Y}_i^{(k)}]\!]} \phi([\![Y_{i+h}|\mathbf{y}_i^{(k)}]\!])}{\operatorname*{ess\,sup}_{(\mathbf{x}_i^{(l)}, \mathbf{y}_i^{(k)}) \in [\![\mathbf{X}_i^{(l)}, \mathbf{Y}_i^{(k)}]\!]} \phi([\![Y_{i+h}|\mathbf{x}_i^{(l)}, \mathbf{y}_i^{(k)}]\!])} \tag{13}$$

where $[\![\mathbf{X}_i^{(l)}, \mathbf{Y}_i^{(k)}]\!]$ denotes the joint range of $\mathbf{X}_i^{(k)}$ and $\mathbf{Y}_i^{(k)}$; and $[\![\mathbf{Y}_i^{(k)}]\!]$ denotes the joint range of $\mathbf{Y}_i^{(k)}$.

Since $\bigcup_{\mathbf{x}_i^{(l)} \in [\![\mathbf{X}_i^{(l)}]\!]} [\![Y_{i+h}|\mathbf{x}_i^{(l)}, \mathbf{y}_i^{(k)}]\!] = [\![Y_{i+h}|\mathbf{y}_i^{(k)}]\!]$, we can infer that $[\![Y_{i+h}|\mathbf{x}_i^{(l)}, \mathbf{y}_i^{(k)}]\!]$ is contained in $[\![Y_{i+h}|\mathbf{y}_i^{(k)}]\!]$; thus,

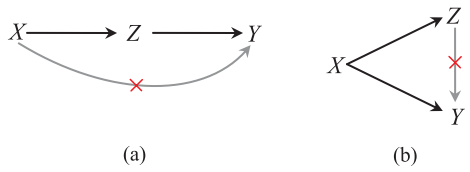


Fig. 2. Information flow pathways between X, Y, and Z. (a) Indirect causality from X to Y through the intermediate variable Z (meaning that there is no direct information flow from X to Y). (b) Spurious causality from Z to Y (meaning that Z and Y have a common perturbing source, X, and therefore they may appear to be connected or correlated even when they are not connected physically).

the T0E is always nonnegative. From the definition, we can see that the T0E from x to y is the conditional 0-information defined in (10). It measures the information transferred from X to Y given the past information of Y. In other words, the T0E represents the information about a future observation of variable Y obtained from simultaneous observations of past values of both X and Y, after discarding information about the future of Y obtained from past values of X alone. It is obvious that if T0E is greater than zero, then there is causality from X to Y; otherwise, there is no causal influence from X to Y.

From the definition of T0E shown in (13), we can see that the T0E is only related to ranges of the random variables and is independent of their probability distributions. Thus, we do not require a well-defined probability distribution of the data set. This means that the collected sampled data do not need to be stationary, which is a basic assumption for the traditional TE method.

C. Direct T0E

The T0E 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 without any intermediate variables or indirect pathways through some intermediate variables.

For example, given three random variables X, Y, and Z, if we find that $T^0_{X\to Y}$, $T^0_{X\to Z}$, and $T^0_{Z\to Y}$ are all greater than zero, then we can conclude that X causes Y, X causes Z, and Z causes Y. There are two possible connectivity realizations of these three variables. One case is that the causal influence from X to Y is only via the indirect pathway through the intermediate variable Z, as shown in Fig. 2(a). In this case, the causality from X to Y is indirect. The other case is that Z is not a cause of Y, yet the causality from Z to Y is generated by X, that is, X is the common source of both Z and Y [Fig. 2(b)]. In this case, the causality from Z to Y is spurious and the intermediate variable is X. Such cases are common in industrial processes, and thus, the detection of direct and indirect/spurious causality is necessary for capturing the true information/material flow pathways and faults propagation pathways [16].

To detect whether there is direct causality from X to Y or the causality is indirect through some intermediate variables, a DT0E from X to Y with intermediate variables $Z_1, Z_2, \ldots, Z_q$ is defined in (14), as shown at the top of the next page, where $\mathbf{Z}^{(s_j)}_{j,i_j} = [Z_{j,i_j}, Z_{j,i_j-\tau_j}, \ldots, Z_{j,i_j-(s_j-1)\tau_j}]$ denotes the embedding vector with elements from the past values of Z_j for $j = 1, \ldots, q$; $\mathbf{z}^{(s_j)}_{j,i_j}$ denotes a realization of $\mathbf{Z}^{(s_j)}_{j,i_j}$; and $(\mathbf{x}^{(l)}_i, \mathbf{y}^{(k)}_i, \mathbf{z}^{(s_1)}_{1,i_1}, \ldots, \mathbf{z}^{(s_q)}_{q,i_q}) \in [[\mathbf{X}^{(l)}_i, \mathbf{Y}^{(k)}_i, \mathbf{Z}^{(s_1)}_{1,i_1}, \ldots, \mathbf{Z}^{(s_q)}_{q,i_q}]]$. Note that the intermediate variables are chosen based on calculation results from the T0E [16]. Parameters $s_1, \ldots, s_q$, $i_1, \ldots, i_q$ and $\tau_1, \ldots, \tau_q$ in (14) are determined by the corresponding calculations of the T0E from $Z_1, \ldots, Z_q$ to Y.

The DT0E represents information transferred from X to Y given past information of both Y and intermediate variables $Z_1, \ldots, Z_q$. Similar to the T0E, it is conditional 0-information and always nonnegative. If this information is greater than zero, i.e., $D^0_{x\to y} > 0$, then we may conclude that there is direct causality (a direct information/material flow pathway) from X to Y. If $D^0_{x\to y} = 0$, then there is no direct causality from X to Y and the causal effect from X to Y is along indirect pathways via the intermediate variables $Z_1, \ldots, Z_q$.

III. Calculation Method

In this section, the calculation method for T0E and DT0E is proposed. Methods for determination of the parameters and confidence levels of T0E are also addressed.

A. Range Estimation

From the definition in (13), we can see that a key to T0E estimation is to estimate the joint and conditional ranges. For discrete-valued random variables, joint and conditional ranges can be estimated by finding all possible realizations of the variables. For example, $[[\mathbf{X}^{(l)}_i, \mathbf{Y}^{(k)}_i]]$ can be obtained by finding all possible realization sets of $(\mathbf{X}^{(l)}_i, \mathbf{Y}^{(k)}_i)$; and $[[Y_{i+h}|\mathbf{y}^{(k)}_i]]$ can be obtained by finding all possible realizations of Y_{i+h} given $\mathbf{Y}^{(k)}_i = \mathbf{y}^{(k)}_i$, and $\phi([[Y_{i+h}|\mathbf{y}^{(k)}_i]])$ is the count of these realizations. For continuous-valued random variables, the estimation of ranges is not as straightforward as the discrete-valued random variables since the realization sets of the continuous-valued random variables are not countable any more. Unfortunately, since most sampled data obtained from industrial processes are continuous valued, we need to figure out how to estimate the ranges for continuous-valued variables.

According to (3) and (5), it can be shown that the conditional ranges in (13) are fully determined by the joint range $[[Y_{i+h}, \mathbf{X}^{(l)}_i, \mathbf{Y}^{(k)}_i]]$. The joint range can be obtained by the well-developed support estimation method based on the concept of support vector machine (SVM) [24]. Here, we only give an algorithm for the joint range estimation, details on the support estimation method can be found in [25].

Let $\mathbf{v}$ denotes a p-dimensional random vector; and $\mathbf{v}_1, \ldots, \mathbf{v}_M \in \mathbf{V}$ denote M observations of $\mathbf{v}$, called training data, where $\mathbf{V}$ is a set of interest. Let ψ be a feature mapping from $\mathbf{V}$ into an inner product space $\mathbb{F}$ such that the inner product in the image of ψ can be computed by a kernel

$$k(\mathbf{v}_i, \mathbf{v}_j) = \langle \psi(\mathbf{v}_i), \psi(\mathbf{v}_j) \rangle \tag{15}$$

$$
\begin{aligned}
D^0_{X\to Y} &= I_0(Y_{i+h}; \mathbf{X}_i^{(l)} | \mathbf{Y}_i^{(k)}, \mathbf{Z}_{1,i_1}^{(s_1)}, \ldots, \mathbf{Z}_{q,i_q}^{(s_q)}) \\
&= H_0(Y_{i+h} | \mathbf{Y}_i^{(k)}, \mathbf{Z}_{1,i_1}^{(s_1)}, \ldots, \mathbf{Z}_{q,i_q}^{(s_q)}) - H_0(Y_{i+h} | \mathbf{X}_i^{(l)}, \mathbf{Y}_i^{(k)}, \mathbf{Z}_{1,i_1}^{(s_1)}, \ldots, \mathbf{Z}_{q,i_q}^{(s_q)}) \\
&= \log \frac{\underset{(\mathbf{y}_i^{(k)}, \mathbf{z}_{1,i_1}^{(s_1)}, \ldots, \mathbf{z}_{q,i_q}^{(s_q)})}{\text{ess}\quad\sup} \phi([\![Y_{i+h} | \mathbf{y}_i^{(k)}, \mathbf{z}_{1,i_1}^{(s_1)}, \ldots, \mathbf{z}_{q,i_q}^{(s_q)}]\!])}{\underset{(\mathbf{x}_i^{(l)}, \mathbf{y}_i^{(k)}, \mathbf{z}_{1,i_1}^{(s_1)}, \ldots, \mathbf{z}_{q,i_q}^{(s_q)})}{\text{ess}\quad\sup} \phi([\![Y_{i+h} | \mathbf{x}_i^{(l)}, \mathbf{y}_i^{(k)}, \mathbf{z}_{1,i_1}^{(s_1)}, \ldots, \mathbf{z}_{q,i_q}^{(s_q)}]\!])}.
\end{aligned} \tag{14}
$$

where $i, j = 1, \ldots, M$. Here, the following Gaussian kernel function is used:

$$k(\mathbf{v}_i, \mathbf{v}_j) = e^{-\gamma \|\mathbf{v}_i - \mathbf{v}_j\|^2} \tag{16}$$

where $\gamma \in \{2^{-15}, 2^{-14}, \ldots, 2^1, 2^2, 2^3\}$ and it is determined by the cross-validation approach [26], [27].

The basic idea of support estimation is to first map the training data into the feature space to separate them from the origin with the maximum margin, and then for a new data sample $\mathbf{v}_r$, the value of a decision function $f(\mathbf{v}_r)$ is determined by evaluating which side of the hyperplane it falls on in the feature space. In other words, the value of the decision function can tell whether the new sample $\mathbf{v}_r$ is within the support of $\mathbf{v}$.

Let $\alpha \in \mathbb{R}^M$ denote a vector with elements α_i for $i = 1, \ldots, M$. To separate the data set from the origin, we solve the following quadratic programming (QP) problem:

$$
\begin{aligned}
&\min_{\alpha} \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M} \alpha_i \alpha_j k(\mathbf{v}_i, \mathbf{v}_j), \\
&\text{s. t.} \quad 0 \le \alpha_i \le \frac{1}{\upsilon M}, \quad \sum_{i=1}^{M} \alpha_i = 1.
\end{aligned} \tag{17}
$$

where $\upsilon \in (0, 1]$ denotes an upper bound on the fraction of outliers, that is, training points outside the estimated region. This standard QP problem can be solved by the Quadratic Programming in C (QPC) toolbox [28].

For any α_i that satisfies $0 < \alpha_i < \frac{1}{\upsilon M}$, its corresponding data sample $\mathbf{v}_i$ satisfies

$$\rho = \sum_{j=1}^{M} \alpha_j k(\mathbf{v}_j, \mathbf{v}_i) \tag{18}$$

where ρ denotes the distance from the hyperplane to the origin.

Then, the decision function $f(\mathbf{v})$ is defined as follows:

$$f(\mathbf{v}) = \text{sgn}\left(\sum_{j=1}^{M} \alpha_j k(\mathbf{v}_j, \mathbf{v}) - \rho\right) \tag{19}$$

where

$$\text{sgn}(u) = \begin{cases} 1, & \text{for } u \ge 0 \\ -1, & \text{otherwise} \end{cases}$$

for a new data point $\mathbf{v}_r$, if $f(\mathbf{v}_r) = 1$, then the new data point $\mathbf{v}_r$ is within the support; if $f(\mathbf{v}_r) = -1$, then $\mathbf{v}_r$ is out of the support. By checking the value of the decision function for each data point, the joint range of the random vector $\mathbf{v}$ can be determined.

For simplicity, we take the range estimation of Y as an example to illustrate the usefulness of the range estimation method. Let $y_{\min}$ and $y_{\max}$ denote the minimum and maximum values of all the realizations denoted by y_i for $i = 1, 2, \ldots, M$ of Y. The calculation steps are described as follows.

Step 1: Calculate α according to (17) using the QPC toolbox [28].

Step 2: Calculate ρ according to (18).

Step 3: Determine the maximum range of Y, denoted by $[y_l, y_u]$, where y_l denotes the lower bound of Y, y_u denotes the upper bound of Y, $y_l = y_{\min} - \delta$, $y_u = y_{\max} + \delta$, and δ satisfies that $e^{-\gamma \delta^2} = \rho$. Note that this is obtained according to the following inequality:

$$\sum_{i=1}^{M} \alpha_i e^{-\gamma (y_l - y_i)^2} < \sum_{i=1}^{M} \alpha_i e^{-\gamma (y_l - y_{\min})^2} = e^{-\gamma \delta^2}.$$

To make $f(y_l) = -1$, which means that the points smaller or equal to y_l must be out of the range of Y, we may set that $e^{-\gamma \delta^2} = \rho$. Similarly, we can obtain that $y_u = y_{\max} + \delta$.

Step 4: Uniformly quantize the maximum range of Y, namely $[y_l, y_u]$, into n nonoverlapping intervals (bins) with bin size of δ_y. We set $n = 100$ in this paper.

Step 5: For the starting point within each bin, namely, y_l, $y_l + \delta_y$, $y_l + 2\delta_y$, $\ldots$, $y_u - \delta_y$, check the value of the decision function according to (19) and determine whether the point is within the range of Y.

Step 6: The Lebesgue measure of the marginal range of Y, namely, $\phi([\![Y]\!])$, is calculated by the number of points within the range times δ_y.

These steps can be easily extended for the joint range estimation. The difference is that we need to determine whether each data point within the maximum joint range, namely, the Cartesian product of the maximum range of each variable, is within the joint range or not using the decision function. Then, the Lebesgue measure of conditional ranges can be obtained using simple statistical calculations.

Similar to T0E, as long as the joint range $[\![Y_{i+h}, \mathbf{X}_i^{(l)}, \mathbf{Y}_i^{(k)}, \mathbf{Z}_{1,i_1}^{(s_1)}, \ldots, \mathbf{Z}_{q,i_q}^{(s_q)}]\!]$ is estimated, DT0E can be calculated according to (14).

B. Choice of Parameters

1) Determination of the Parameters of the T0E: Similar to the TE approach, there are four undetermined parameters in the definition of the T0E in (13): the prediction horizon (h), the time interval (τ), and the embedding dimensions (k and l). The basic idea for these parameters determination is similar to that for the TE method proposed in [16]. A systematic method to determine them is described below.

First, since $h = \tau$ as a rule of thumb [12], we can further set initial values for h and τ according to priori knowledge of the process dynamics. If the process dynamics are unknown, the small values of h and τ should give good results [12]; we may start by setting the initial values as $h = \tau = 1$.

Second, we can determine the embedding dimension of Y, namely, the window size of the historical Y used for the future Y prediction. The embedding dimension of Y, i.e., k, can be determined as the minimum nonnegative integer after which there is no significant change on $H_0(Y_{i+h}|\mathbf{Y}_i^{(k)})$. Considering that a large k can increase the dimension of the joint range and the difficulty in range estimation, if k is greater than three, we need to increase h and τ and repeat the calculation until a $k \leq 3$ is found.

Finally, we can determine the embedding dimension of X, namely, the window size of the historical X used for the prediction of future Y. Based on the values of k, h, and τ, the embedding dimension of X, i.e., l, is determined as the minimum positive integer after which there is no significant change on the TE from X to Y.

2) Confidence Level Determination of the T0E and DT0E: Small values of the T0E suggest no causality while large values do. The detection of causality can be reformulated as a hypothesis test problem. The null hypothesis is that the T0E measure, $T^0_{X\to Y}$, is small, that is, there is no causality from X to Y. If $T^0_{X\to Y}$ is large, then the null hypothesis can be rejected, which means there is causal influence from X to Y. To carry out this hypothesis testing, we may use the Monte Carlo method [12] by constructing a surrogate time series [29]. The constructed surrogate time series must satisfy the null hypothesis that the causal influence from X to Y is completely destroyed; at the same time, the statistical properties of X and Y should not change. To construct the surrogate time series that satisfies these two conditions, we propose a new surrogate time series construction method as follows.

Let X and Y be sampled at time instant i and denoted by X_i and Y_i with $i = 1, 2, \ldots, N$, where N is the number of samples; and M denotes the length of the training data set, namely, the data size for T0E and DT0E calculations. Then, a pair of surrogate time series for X and Y is constructed as

$$\begin{cases} X^{\text{surr}} = [X_i, X_{i+1}, \ldots, X_{i+M-1}] \\ Y^{\text{surr}} = [Y_j, X_{j+1}, \ldots, X_{j+M-1}] \end{cases} \tag{20}$$

where i and j are randomly chosen from $\{1, \ldots, N-M+1\}$ and $\|j-i\| \geq e$, where e is a sufficiently large integer (e is much larger than h) such that there is almost no correlation between X^{surr} and Y^{surr}.

By calculating the T0E from N_s surrogate time series such that $\lambda_n = T^0_{X^{\text{surr}}\to Y^{\text{surr}},\,n}$ for $n = 1, \ldots, N_s$, the significance level is then defined as

$$s_{X\to Y} = \frac{T^0_{X\to Y} - \mu_\lambda}{\sigma_\lambda} > 3 \tag{21}$$

where μ_λ and σ_λ are the mean and standard deviation of λ_n, respectively. Similarly, the value of $s_{X\to Y}$ can also be used as the significance level for the DT0E from X to Y.

C. Causality Detection Steps via the T0E Method

The causality detection steps via the T0E method (from X to Y) are described as follows.

- *Step 1:* Determine the four parameters in the definition of the T0E in (13) using the method described in Section III-B.
- *Step 2:* Obtain the Lebesgue measure of joint ranges appearing in (13) using the range estimation method described in Section III-A.
- *Step 3:* Calculate the T0E according to (13).
- *Step 4:* Calculate the confidence level for the T0E using the method described in Section III-B to determine whether there is significant causality from X to Y.

IV. Examples and Case Studies

The practicality and utility of the proposed method are illustrated by application to two numerical examples, an experimental data set and an industrial benchmark data set.

A. Examples

We use simple mathematical equations to represent causal relationships in the following two examples.

Example 1: Assume three linear correlated continuous random variables X, Y, and Z satisfying

$$\begin{cases} Y_{k+1} = 0.8X_k + 0.5Y_k + \upsilon_{1k} \\ Z_{k+1} = 0.6Y_k + \upsilon_{2k} \end{cases}$$

where $X_k \sim N(0,1)$; $\upsilon_{1k}, \upsilon_{2k} \sim N(0, 0.1)$; and $Y(0) = 3.2$. The simulation data set consists of 3000 samples. The initial 1000 data points are chosen as the training data and are used for causality analysis.

For range estimation, we set $\upsilon = 0.01$ since the fraction of outliers of the data is quite small. For determination of γ, the initial 1000 data points are used for training and the remaining 2000 samples are used for validation. Using the cross-validation approach, we find that $\gamma = 2^{-2}$ gives good results.

To calculate the T0Es between X, Y, and Z, we need to determine four design parameters. We take the T0E from X to Y as an example. First, initial values for h and τ are set as $h = \tau = 1$; second, we calculate $H_0(Y_{i+h}|\mathbf{Y}_i^{(k)})$ with $k = 0, 1, \ldots, 10$, as shown in the upper part of Fig. 3. The change rate of $H_0(Y_{i+h}|\mathbf{Y}_i^{(k)})$ with $k = 0, 1, \ldots, 10$ is shown in the lower part of Fig. 3, we can see that as k increases, there is no significant change in $H_0(Y_{i+h}|\mathbf{Y}_i^{(k)})$ after $k = 1$, which means that Y_i provides significantly useful information for Y_{i+1}, while $Y_{i-1}, Y_{i-2}, \ldots$ cannot provide any additional information for Y_{i+1} when Y_i is given. Thus, we choose $k = 1$. Finally, we calculate the TE $T^0_{X\to Y}$ and its change rate with

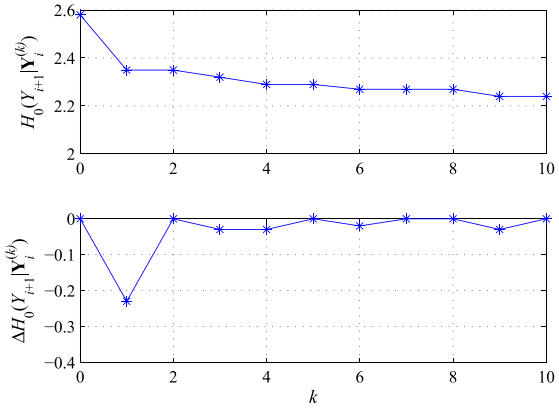


Fig. 3. Finding the embedding dimension of Y in Example 1.

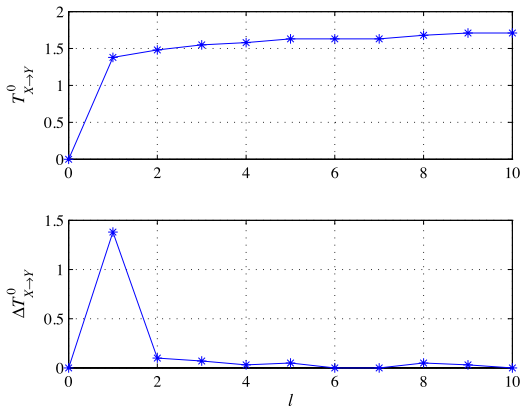


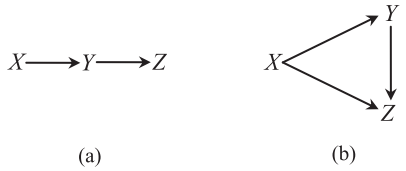
Fig. 4. Finding the embedding dimension of X for $T^0_{X\to Y}$ in Example 1.

$l = 1, \ldots, 10$, as shown in Fig. 4. Since there is no significant change in $T^0_{X\to Y}$ after $l = 1$, as shown in the lower part of Fig. 4, we choose $l = 1$. Using the same procedure, the parameters for each pair of X, Y, and Z can be determined. For the remaining example and case studies, the same procedure is used for parameters determination.

After the parameters are determined, according to (13) and (21), the T0Es between each pair of X, Y, and Z and the corresponding thresholds (see values within round brackets) obtained via the Monte Carlo method are shown in Table I. Note that the variables listed in column one are the cause variables and the corresponding effect variables appear in the first row. For surrogate time series construction, we set $e = 500$, i.e., $\|j - i\| \geq 500$ in (20), to ensure that there is almost no correlation between each pair of the surrogate data. For the remaining example and case studies, the same value of e is assigned. If the calculated T0E is greater than the corresponding threshold, then we may conclude that the causality is significant; otherwise there is almost no causal influence. Note that if the calculated T0E from one variable to another is zero, then we do not need to calculate the corresponding threshold since it is safe to accept the null hypothesis that there is no causality. From Table I, we can see that X causes Y, Y causes Z, and X causes Z because $T^0_{X\to Y} = 1.38$, $T^0_{Y\to Z} = 1.00$, and $T^0_{X\to Z} = 0.60$ are greater than the threshold. Next, we need to determine whether there

TABLE I
CALCULATED T0ES AND THRESHOLDS (VALUES IN ROUND BRACKETS) FOR EXAMPLE 1

$T^0_{\text{column 1}\to\text{row 1}}$	X	Y	Z
X	NA	1.38 (0.08)	0.60 (0.08)
Y	0.03 (0.07)	NA	1.00 (0.08)
Z	0	0	NA



(a) (b)

Fig. 5. Information flow pathways for (a) Example 1 and (b) Example 2.

TABLE II
CALCULATED T0ES AND THRESHOLDS (VALUES IN ROUND BRACKETS) FOR EXAMPLE 2

$T^0_{\text{column 1}\to\text{row 1}}$	X	Y	Z
X	NA	0.80 (0.07)	0.20 (0.06)
Y	0.03 (0.05)	NA	0.55 (0.08)
Z	0	0	NA

is direct causality from X to Z. According to (14), we obtain $D^0_{X\to Z} = 0$. Thus, we conclude that there is no direct causality from X to Z. The information flow pathways for Example 1 are shown in Fig. 5(a).

This conclusion is consistent with the mathematical function, from which we can see that there are information flow pathways both from X to Y and from Y to Z, and the information flow from X to Z is indirect through the intermediate variable Y.

Example 2: Assume three nonlinear correlated continuous random variables X, Y, and Z satisfying

$$\begin{cases} Y_{k+1} = 1 - 2(|\ 0.5 - (0.8X_k + 0.4\sqrt{|\ Y_k\ |})\ |) + \upsilon_{1k} \\ Z_{k+1} = 5(Y_k + 7.2)^2 + 10\sqrt{|\ X_k\ |} + \upsilon_{2k}. \end{cases}$$

where $X_k \in [4, 5]$ is a uniformly distributed signal; $\upsilon_{1k}, \upsilon_{2k} \sim N(0, 0.05)$; and $Y(0) = 0.2$. The simulation data consists of 3000 samples. The initial 1000 data points were chosen as the training data and were used for causality analysis.

The T0Es between each pair of X, Y, and Z are shown in Table II. The values within round brackets denote the corresponding thresholds. We may conclude that X causes Y, X causes Z, and Y causes Z because $T^0_{X\to Y} = 0.80$, $T^0_{X\to Z} = 0.20$, and $T^0_{Y\to Z} = 0.55$ are larger than their thresholds.

Thus, we need to first determine whether there is direct causality from X to Z. According to (14), we calculate the DT0E from X to Z with the intermediate variable Y and obtain $D^0_{X\to Z} = 0.23$, which is larger than the threshold 0.06. Thus, we conclude that there is direct causality from X to Z. Second, we need to detect whether there is true and direct causality from Y to Z since X is the common source of

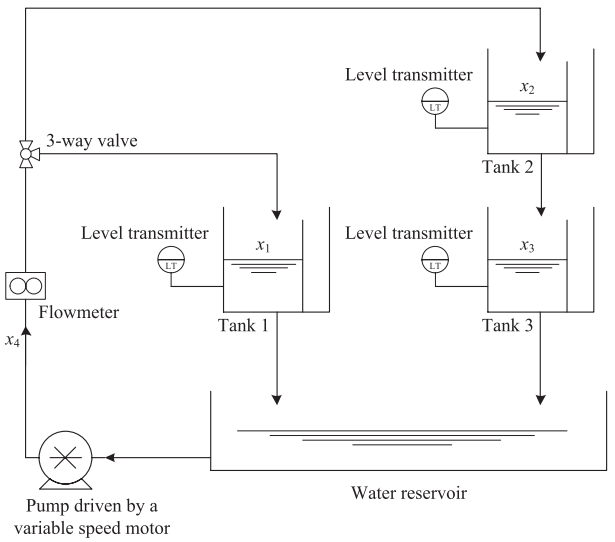


Fig. 6. Schematic of the three-tank system.

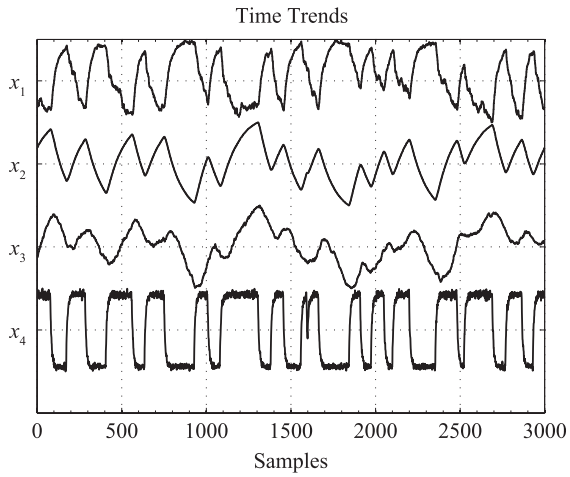


Fig. 7. Time trends of measurements of the three-tank system.

TABLE III
CALCULATED T0Es AND THRESHOLDS (VALUES IN ROUND BRACKETS) FOR THE THREE-TANK SYSTEM

$T^0_{\text{column 1}\to\text{row 1}}$	x_1	x_2	x_3	x_4
x_1	NA	0.05 (0.07)	0.14 (0.06)	0.04 (0.06)
x_2	0.05 (0.06)	NA	0.20 (0.07)	0
x_3	0.03 (0.06)	0.04 (0.07)	NA	0
x_4	0.17 (0.06)	0.16 (0.07)	0.06 (0.05)	NA

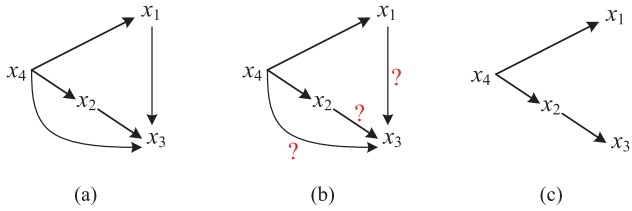


Fig. 8. Information flow pathways for the three-tank system based on (a) and (b) calculation results of T0Es which represent the total causality including both direct and indirect/spurious causality; (c) calculation results of DT0Es which correctly indicate the direct and true causality.

both Y and Z. We calculate the DT0E from Y to Z with the intermediate variable X and obtain $D^0_{Y\to Z} = 0.39$, which is also larger than the threshold 0.08. Thus, we conclude that there is true and direct causality from Y to Z. The information flow pathways for Example 2 are shown in Fig. 5(b). This conclusion is consistent with the mathematical function, from which we can see that there are direct information flow pathways from X to Y, from X to Z, and from Y to Z.

Compared with the traditional TE method, the T0E is defined without assuming a statistical space and the only issue is the (conditional) ranges of variables. This means that the time series does not need to be stationary, which is a basic assumption for the traditional TE method. The computational complexity for the T0E estimation is relatively small since we do not need to estimate the joint PDF. Another advantage of the T0E method is that the length of the data does not need to be very large. From the examples described above, we can see that 1000 samples are sufficient to give good results, while for the TE estimation, the sample number is preferred to be no less than 2000 observations [12].

B. Experimental Case Study

To illustrate the effectiveness of the proposed causality detection method for capturing information and/or material flow pathways, a three-tank experiment was conducted. The schematic of the three-tank system is shown in Fig. 6. 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 into the reservoir. The experiment is conducted under open-loop conditions.

The water levels are measured by level transmitters. We denote the water levels of tanks 1, 2, and 3 by x_1, x_2, and x_3, respectively. The flow rate of the pumped water is measured by a flow meter; we denote this flow rate by x_4. In this experiment, x_4 is set to be a pseudorandom binary sequence. The sampled data of 3000 observations were analyzed. Fig. 7 shows the normalized time trends of the measurements. The sampling time is 1 s. Note that this data set is not strictly stationary since it cannot pass the stationarity test described in [14]. The traditional TE method may not be suitable for this data set.

The initial 1000 data points are used as training data for γ determination and for causality analysis. The calculated T0Es between each pair of x_1, x_2, x_3, and x_4 are shown in Table III with the thresholds (see values within round brackets) obtained via the Monte Carlo method. If the calculated T0E is larger than the corresponding threshold, then we may conclude that the causality is significant; otherwise, there is no causality. We can see that x_1 and x_2 cause x_3, and x_4 causes x_1, x_2, and x_3. The corresponding connectivity realization is shown in Fig. 8(a).

Now, we need to determine whether the causality between x_1 and x_3 and between x_2 and x_3 is true or spurious, as shown in Fig. 8(b). To clarify this, we first calculate the DT0E from x_1 to x_3 with intermediate variables x_4 and x_2

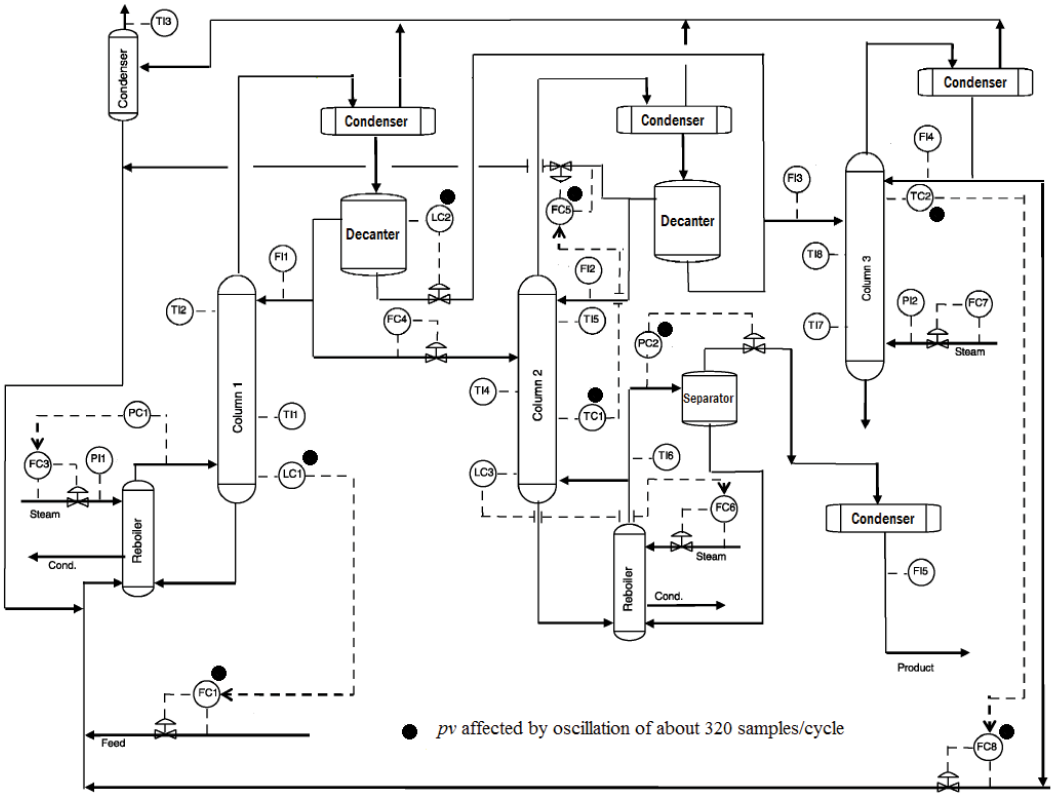


Fig. 9. Process schematic for the industrial case study. The oscillation process variables (*pv*'s) are marked by circle symbols.

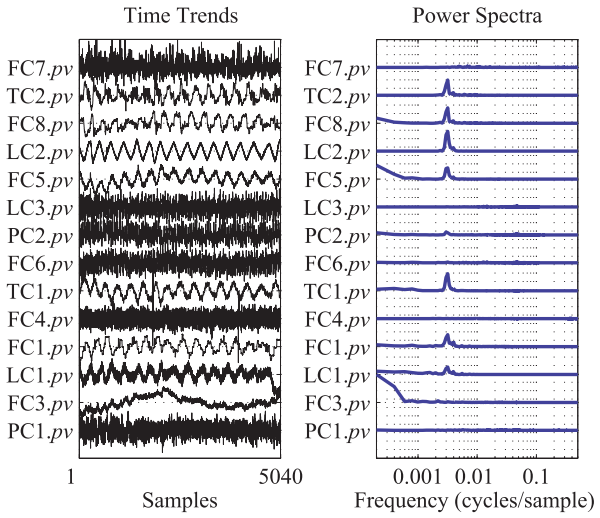


Fig. 10. Time trends and power spectra of measurements of process variables (*pv*'s).

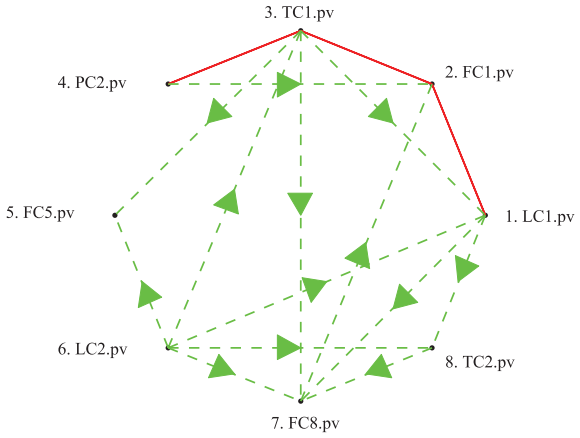


Fig. 11. Causal map of oscillation process variables based on calculation results of T0Es. A dashed line with an arrow indicates that there is unidirectional causality from one variable to the other, and a solid line connecting two variables without an arrow indicates there is bidirectional causality between the two variables.

and obtain $D^0_{x_1 \to x_3} = 0$, which means that there is no direct information/material flow pathway from x_1 to x_3 and the direct link should be eliminated. Next, we calculate the DT0E from x_2 to x_3 with intermediate variable x_4 and obtain $D^0_{x_2 \to x_3} = 0.18$, which is larger than the threshold 0.07. Thus, we conclude that there is true and direct causality from x_2 to x_3. As shown in Fig. 8(b), since x_4 causes x_2, x_2 causes x_3, and x_4 causes x_3, we need to further check whether there is direct causality from x_4 to x_3. According to (14), we calculate the DT0E from x_4 to x_3 with intermediate variable x_2 and obtain $D^0_{x_4 \to x_3} = 0$. Thus, we conclude that there is no direct causality from x_4 to x_3. The corresponding information flow pathways according to these calculation results are shown in Fig. 8(c), which are consistent with the information and material flow pathways of the physical three-tank system (Fig. 6).

C. Industrial Case Study

We use an industrial process data set [30] provided by the Advanced Controls Technology group of Eastman Chemical

TABLE IV
CALCULATED T0Es AND THRESHOLDS (VALUES IN ROUND BRACKETS) FOR THE INDUSTRIAL CASE STUDY

$T^0_{\text{column 1}\to\text{row 1}}$	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	NA	0.12 (0.06)	0	0	0	0	0.13 (0.07)	0.14 (0.06)
x_2	0.10 (0.07)	NA	0.16 (0.06)	0.04 (0.07)	0	0	0	0
x_3	0.13 (0.07)	0.15 (0.06)	NA	0.15 (0.07)	0.25 (0.07)	0	0.12 (0.06)	0
x_4	0	0.08 (0.06)	0.21 (0.07)	NA	0	0	0	0
x_5	0.06 (0.07)	0	0.05 (0.06)	0.04 (0.08)	NA	0	0	0.03 (0.07)
x_6	0.24 (0.07)	0	0.29 (0.06)	0	0.25 (0.07)	NA	0.13 (0.07)	0.25 (0.06)
x_7	0.04 (0.07)	0.13 (0.06)	0	0.04 (0.07)	0	0.03 (0.07)	NA	0.05 (0.06)
x_8	0	0	0	0.04 (0.07)	0	0	0.28 (0.06)	NA

TABLE V
CALCULATED DT0Es AND THRESHOLDS (VALUES IN ROUND BRACKETS) FOR THE INDUSTRIAL CASE STUDY

$D^0_{\text{column 1}\to\text{row 1}}$	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	NA	0.11 (0.06)	NA	NA	NA	NA	0 (0.07)	0 (0.06)
x_2	0.09 (0.07)	NA	0.03 (0.06)	NA	NA	NA	NA	NA
x_3	0 (0.07)	0.04 (0.06)	NA	0.15 (0.07)	0.23 (0.07)	NA	0 (0.06)	NA
x_4	NA	0 (0.06)	0.17 (0.07)	NA	NA	NA	NA	NA
x_5	NA	NA	NA	NA	NA	NA	NA	NA
x_6	0.20 (0.07)	NA	0.25 (0.06)	NA	0.05 (0.07)	NA	0.04 (0.07)	0.21 (0.06)
x_7	NA	0 (0.06)	NA	NA	NA	NA	NA	NA
x_8	NA	NA	NA	NA	NA	NA	0.24 (0.06)	NA

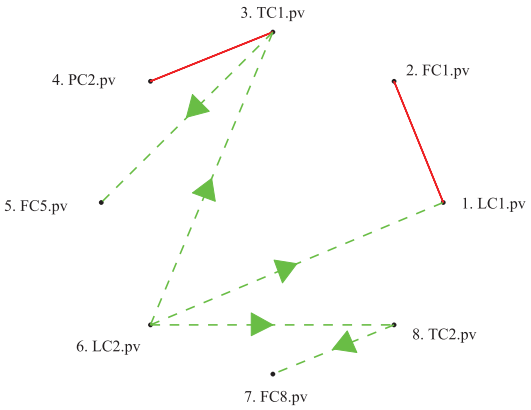


Fig. 12. Causal map of oscillation process variables based on calculation results of DT0Es. A dashed line with an arrow indicates that there is unidirectional causality from one variable to the other, and a solid line connecting two variables without an arrow indicates there is bidirectional causality between the two variables.

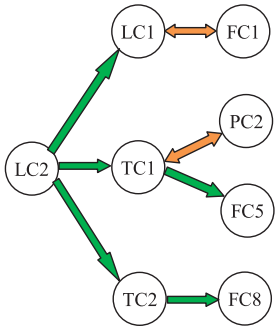


Fig. 13. Oscillation propagation pathways obtained via the T0E method.

Company, USA, to demonstrate the effectiveness of the proposed causality detection method. The Advanced Controls Technology group identified a need to diagnose a common oscillation of about 2 h (about 320 samples/cycle). It was assumed that this common oscillation is probably generated within a certain control loop. The process schematic is shown in Fig. 9. The process contains three distillation columns, two decanters, and several recycle streams.

Oscillations are present in the process variables (controlled variables), controller outputs, set points, controller errors (meaning errors between process variable measurements and their set points), or measurements from other sensors. The plant-wide oscillation detection and diagnosis methods can be used for any of these time trends [31], [32]. In this paper, we only use process variables for root cause analysis; and 14 controlled process variables corresponding to 14 PID controller loops are available. In total, 5040 sampled observations (from 28 h of data with the sampling interval 20 s) are analyzed. In this case study, FC, LC, PC, and TC represent flow, level, pressure, and temperature controllers, respectively. We denote the process variable by *pv*. Fig. 10 shows the normalized time trends and normalized power spectra of the 14 process variables. Note that this data set is not strictly stationary since it cannot pass the stationarity test described in [14]. The traditional TE method may not be suitable for this data set.

The power spectra in Fig. 10 indicate the presence of oscillation at the frequency of about 0.003 cycles/sample, corresponding to an approximate period of 2 h. This oscillation propagated throughout the interconnected units and affected

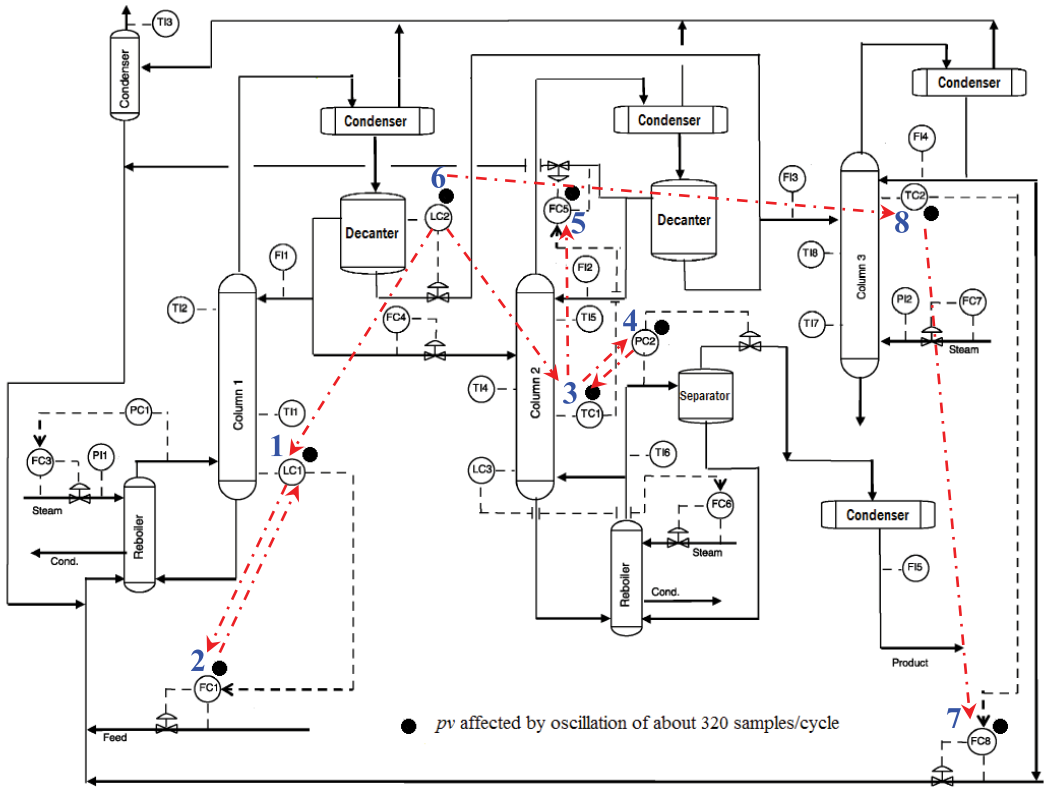


Fig. 14. Direct causal relationships between the oscillation process variables, namely, the oscillation propagation pathways, which are indicated by dash dot lines with arrows.

many variables in the process. Thus, our goal is to detect and diagnose the root cause of this oscillation.

For oscillation detection, the spectral envelope method was applied to determine which variables have oscillation at the frequency of 0.0032 cycles/sample. Details on the spectral envelope method can be found in [30]. Since this paper focuses on causality detection and its application to root cause diagnosis, we omit details of oscillation detection and only show the detection result, that is, the following eight process variables have common oscillations with 99.9% confidence level: LC1.*pv* (denoted by x_1), FC1.*pv* (denoted by x_2), TC1.*pv* (denoted by x_3), PC2.*pv* (denoted by x_4), FC5.*pv* (denoted by x_5), LC2.*pv* (denoted by x_6), FC8.*pv* (denoted by x_7), and TC2.*pv* (denoted by x_8). These variables are marked by dark circle symbols in Fig. 9. It is assumed that if a variable does not show significant power at the common oscillation frequency, then it does not belong to the group of likely root cause variables [30]. Therefore, we only need to find the information flow pathways among these variables that have oscillations at the common frequency.

The initial 1000 samples are used as training data for γ determination and also for causality analysis. Other samples are used as test data for γ determination. Using the cross-validation method, γ is set to be 2^{-2}. Table IV shows the T0Es and thresholds (values in brackets) between each pair of the process variables.

Based on Table IV, the causal relationships between the eight oscillation process variables are shown in Fig. 11, where a dashed line with an arrow indicates that there is unidirectional causality from one variable to the other, and a solid line connecting two variables without an arrow indicates there is bidirectional causality (also called causality feedback [33]) between the two variables.

The causal map in Fig. 11 shows a complicated set of pathways from which finding fault propagation pathways would be difficult. The reason is that by only calculating T0Es, both total and spurious causality can be detected. To derive a simpler and more accurate causal map, we need to differentiate between direct and indirect as well as true and spurious causality. Thus, we need to further calculate DT0Es between each pair of the variables that have causal relationship and have possible intermediate variable(s). For example, for the causal influence from x_6 to x_5, since x_6 causes x_3, and x_3 causes x_5, we need to calculate the DT0E from x_6 to x_5 with the intermediate variable x_3.

Table V shows the calculated DT0Es between each pair of the variables that have causal relationship and have possible intermediate variable(s). Note that if a pair of variables does not have significant causal relationship based on the calculation results of T0Es shown in Table IV, then we do not need to calculate its DT0E, and thus put NA in Table V. If the calculated DT0E is larger than the corresponding threshold, then we may conclude that the causality is direct and keep that information flow pathway; otherwise, there is no direct causality, and we can eliminate the information flow pathway in Fig. 11. The causal map based on calculation results of

DT0Es is shown in Fig. 12, which is much sparser than the previous causal map shown in Fig. 11.

The oscillation propagation pathways obtained from the causal map (Fig. 12) are shown in Fig. 13. They are also indicated by dash dot lines with arrows in the process schematic, as shown in Fig. 14. Note that the bidirectional propagation pathways between LC1.*pv* and FC1.*pv* and between TC1.*pv* and PC2.*pv* are generated by the cascade feedback control structure and recycle streams, respectively, which are consistent with the physical process. Figs. 13 and 14 show that LC2 can reach all the other loops but does not receive any significant causal effects from any other loops. Thus, we conclude that control loop LC2 is likely the root cause candidate. Figs. 13 and 14 also show that the oscillation in loop LC2 first propagates to loops LC1, TC1, and TC2. From Fig. 14, we can see that there are direct material flow pathways from the left-hand side decanter to columns 1, 2, and 3. Thus, the oscillation propagation pathways are validated by the physical process. It has been confirmed [30], [34] that the root cause of the plant wide oscillation was due to the valve stiction of control loop LC2; therefore, the causality analysis via the T0E method is indeed effective in finding the fault propagation pathways and determining the root cause candidate.

V. Conclusion

In industrial processes, fault diagnosis in a large-scale complex system is particularly challenging because of the high degree of integration between different units in the system as well as the presence of recycle streams and process control feedback loops. A simple fault may easily propagate along information and material flow pathways and affect many other parts of the system. It is important to determine the fault propagation pathways to find the root cause of the abnormalities and the corresponding fault propagation routes. Causality analysis can detect the causal influence between two process variables, including the direction of the information flow. However, in the case of more than two variables, it is valuable to detect whether the influence is along direct or indirect pathways. An information theory-based causality detection method based on the T0E has been proposed without assuming a probability space. Moreover, a direct causality detection method based on the DT0E has been presented to detect whether there is a direct information and/or material flow pathway between each pair of variables. The range estimation method for continuous-valued variables and the calculation method for both T0E and DT0E have been addressed. The practicality and utility of the proposed methods have been successfully illustrated by two numerical examples, plus an experimental data set and an industrial benchmark data set.

The outstanding advantage of the T0E method is that the data do not need to follow a well-defined probability distribution since the T0E is defined without assuming a statistical space and the only issue is its range. This means that the time series does not need to be stationary, which is a basic assumption for the traditional TE method. This point can also be seen from the range estimation point of view. According to the QP problem in (17), we can see that as long as each data point $\mathbf{v}_i$ is determined, the order of the data points will not affect the optimization results; this means that stationarity is not a necessary condition of the data. This point is clearly illustrated in the experimental three-tank case study and the industrial case study, as presented in Section IV. The analyzed data set is not strictly stationary. The T0E method can still find the information and/or material flow pathways using this data set. Another advantage of the T0E method compared with the traditional TE method is that the length of the data does not need to be very large. The reason is that the range estimation is based on the concept of SVM that can handle small sample data sets [35], [36]. Based on our experience, 500 samples are enough to give good results.

Although the proposed method has been validated by some examples, one limitation of this method is that the causality detection results may be conservative. For example, in the industrial case study in Section IV, there should be bidirectional causality between TC1 and FC5 and between TC2 and FC8 because of the temperature feedback control loops. However, we cannot find the causal influence from FC5 to TC1 and from FC8 to TC2. The possible reason for this is that the definition of 0-information in (9) is under the worst case that can be understood as the least information transferred from one variable to another. This is also the reason why we chose a three-sigma threshold for the significance level instead of six sigma. Our ongoing study is related to the conservative property of the T0E method.

References

[1] M. A. A. S. Choudhury, "Plantwide oscillations diagnosis—Current state and future directions," *Asia-Pacific J. Chem. Eng.*, vol. 6, no. 3, pp. 484–496, May/Jun. 2011.

[2] F. Yang and D. Xiao, "Progress in root cause and fault propagation analysis of large-scale industrial processes," *J. Control Sci. Eng.*, vol. 2012, pp. 478373-1–478373-10, Feb. 2012.

[3] P. Duan, T. Chen, S. L. Shah, and F. Yang, "Methods for root cause diagnosis of plant-wide oscillations," *AIChE J.*, vol. 60, no. 6, pp. 2019–2034, 2014.

[4] N. Wiener, "The theory of prediction," in *Modern Mathematics for Engineers*, E. F. Beckenbach, Ed. New York, NY, USA: McGraw-Hill, 1956.

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

[6] N. Ancona, D. Marinazzo, and S. Stramaglia, "Radial basis function approach to nonlinear granger causality of time series," *Phys. Rev. E*, vol. 70, no. 5, pp. 056221-1–056221-7, 2004.

[7] 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.

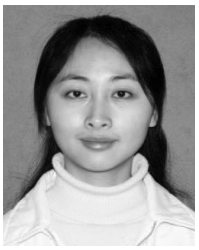
[8] U. Feldmann and J. Bhattacharya, "Predictability improvement as an asymmetrical measure of interdependence in bivariate time series," *Int. J. Bifurcation Chaos Appl. Sci. Eng.*, vol. 14, no. 2, pp. 505–514, 2004.

[9] M. Bauer, N. F. Thornhill, and J. W. Cox, "Measuring cause and effect of process variables," in *Proc. Adv. Process Control Appl. Ind. Workshop*, Vancouver, BC, Canada, 2005.

[10] 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.

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

[12] 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.
[13] 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.
[14] 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.
[15] 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.
[16] P. Duan, F. Yang, T. Chen, and S. L. Shah, "Direct causality detection via the transfer entropy approach," *IEEE Trans. Control Syst. Technol.*, vol. 21, no. 6, pp. 2052–2066, Nov. 2013.
[17] P. Jizba, H. Kleinert, and M. Shefaat, "Rényi's information transfer between financial time series," *Phys. A, Statist. Mech. Appl.*, vol. 391, no. 10, pp. 2971–2989, 2012.
[18] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Champaign, IL, USA: Univ. Illinois Press, 1949.
[19] *Entropy (Information Theory)*, Wikipedia. [Online]. Available: http://en.wikipedia.org/wiki/Shannon_entropy, accessed May 15, 2013.
[20] G. N. Nair, "A nonstochastic information theory for communication and state estimation," *IEEE Trans. Autom. Control*, vol. 58, no. 6, pp. 1497–1510, Jun. 2013.
[21] R. V. L. Hartley, "Transmission of information," *Bell Syst. Tech. J.*, vol. 7, no. 3, pp. 535–563, 1928.
[22] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Symp. Math. Statist. Probab.*, Berkeley, CA, USA, 1960, pp. 547–561.
[23] H. Shingin and Y. Ohta, "Disturbance rejection with information constraints: Performance limitations of a scalar system for bounded and Gaussian disturbances," *Automatica*, vol. 48, no. 6, pp. 1111–1116, 2012.
[24] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[25] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Comput.*, vol. 13, no. 7, pp. 1443–1471, 2001.
[26] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," Dept. Comput. Sci. Inf. Eng., National Taiwan Univ., Tech. Rep., 2010.
[27] G. McLachlan, K.-A. Do, and C. Ambroise, *Analyzing Microarray Gene Expression Data*. New York, NY, USA: Wiley, 2004.
[28] A. Wills. *Quadratic Programming in C (QPC) Toolbox (Version 2.0)*. [Online]. Available: http://sigpromu.org/quadprog/, accessed Mar. 10, 2013.
[29] T. Schreiber and A. Schmitz, "Surrogate time series," *Phys. D*, vol. 142, nos. 3–4, pp. 346–382, 2000.
[30] H. Jiang, M. A. A. S. Choudhury, and S. L. Shah, "Detection and diagnosis of plant-wide oscillations from industrial data using the spectral envelope method," *J. Process Control*, vol. 17, no. 2, pp. 143–155, 2007.
[31] N. F. Thornhill, B. Huang, and H. Zhang, "Detection of multiple oscillations in control loops," *J. Process Control*, vol. 13, no. 1, pp. 91–100, 2003.
[32] N. F. Thornhill, S. L. Shah, B. Huang, and A. Vishnubhotla, "Spectral principal component analysis of dynamic process data," *Control Eng. Pract.*, vol. 10, no. 8, pp. 833–846, 2002.
[33] P. E. Caines and C. Chan, "Feedback between stationary stochastic processes," *IEEE Trans. Autom. Control*, vol. 20, no. 4, pp. 498–508, Aug. 1975.
[34] N. F. Thornhill, J. W. Cox, and M. A. Paulonis, "Diagnosis of plant-wide oscillation through data-driven analysis and process understanding," *Control Eng. Pract.*, vol. 11, no. 12, pp. 1481–1490, 2003.
[35] W. Zheng *et al.*, "Support vector machine: Classifying and predicting mutagenicity of complex mixtures based on pollution profiles," *Toxicology*, vol. 313, nos. 2–3, pp. 151–159, 2013.
[36] L. Gao, Z. Ren, W. Tang, H. Wang, and P. Chen, "Intelligent gearbox diagnosis methods based on SVM, wavelet lifting and RBR," *Sensors*, vol. 10, no. 5, pp. 4602–4621, 2010.



Ping Duan received the B.Eng. degree in automation from Central South University, Changsha, China, in 2005, and the Ph.D. degree in electrical and computer engineering from the University of Alberta, Edmonton, AB, Canada, in 2014.

She has co-authored a brief entitled *Capturing Connectivity and Causality in Complex Industrial Processes* (Springer). 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 joined the Department of Automation at Tsinghua University, in 2011, after he was a Post-Doctoral Fellow with Tsinghua University and the University of Alberta, Edmonton, AB, Canada. He is currently an Associate Professor with Tsinghua University. He has co-authored a brief entitled *Capturing Connectivity and Causality in Complex Industrial Processes* (Springer). His current research interests include topology modeling of large-scale processes, abnormal events monitoring, process hazard analysis, and smart alarm management.

Dr. Yang was a recipient of 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.

Sirish L. Shah (M'76) has held visiting appointments with Balliol College, Oxford University, Oxford, U.K., 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, Callaghan, NSW, Australia, in 2004, IIT Madras, Chennai, India, in 2006, and the National University of Singapore, Singapore, in 2007. He is on the 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 co-authored books entitled *Performance Assessment of Control Loops: Theory and Applications*, *Diagnosis of Process Nonlinearities and Valve Stiction: Data Driven Approaches*, and more recently, *Capturing Connectivity and Causality in Complex Industrial Processes*. His current research interests include process and performance monitoring, system identification and design, and analysis and rationalization of alarm systems.

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.

Prof. Chen is a fellow of the International Federation of Automatic Control. He has served as an Associate Editor of several international journals, including the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, *Automatica*, and *Systems and Control Letters*.