

Causality analysis in process control based on denoising and periodicity-removing CCM

Causality
analysis in
process control

25

Qiming Chen, Xinyi Fei and Lie Xie

*State Key Laboratory of Industrial Control Technology, Zhejiang University,
Hangzhou, China, and*

Dongliu Li and Qibing Wang

Sicher Elevator Co., Ltd, Huzhou, China

Received 10 June 2020
Revised 10 August 2020
Accepted 1 September 2020

Abstract

Purpose – 1. To improve the causality analysis performance, a novel causality detector based on time-delayed convergent cross mapping (TD-CCM) is proposed in this work. 2. Identify the root cause of plant-wide oscillations in process control system.

Design/methodology/approach – A novel causality analysis framework is proposed based on denoising and periodicity-removing TD-CCM (time-delayed convergent cross mapping). We first point out that noise and periodicity have adverse effects on causality detection. Then, the empirical mode decomposition (EMD) and detrended fluctuation analysis (FDA) are combined to achieve denoising. The periodicities are effectively removed through singular spectrum analysis (SSA). Following, the TD-CCM can accurately capture the causalities and locate the root cause by analyzing the filtered signals.

Findings – 1. A novel causality detector based on denoising and periodicity-removing time-delayed convergent cross mapping (TD-CCM) is proposed. 2. Simulation studies show that the proposed method is able to improve the causality analysis performance. 3. Industrial case study shows the proposed method can be used to analyze the root cause of plant-wide oscillations in process control system.

Originality/value – 1. A novel causality detector based on denoising and periodicity-removing time-delayed convergent cross mapping (TD-CCM) is proposed. 2. The influences of noise and periodicity on causality analysis are investigated. 3. Simulations and industrial case shows that the proposed method can improve the causality analysis performance and can be used to identify the root cause of plant-wide oscillations in process control system.

Keywords Causality analysis, Root cause diagnosis, Oscillation diagnosis, Control performance monitoring, Convergent cross mapping

Paper type Research paper

1. Introduction

Oscillations are a common cause of performance degradation of process control system. They may generate in one control loop and propagate to other units (Lang *et al.*, 2018a). Then the whole plant may oscillate. These oscillations are likely to result in the waste of energy and materials, product quality fluctuation, and even poor safety (Chen *et al.*, 2020a). It is important for engineers and operators to correctly identify the root cause of oscillations as soon as possible (Yang *et al.*, 2014; Lang *et al.*, 2019). However, due to the disturbances of noises, periodicities, nonlinearities (Lang *et al.*, 2018b) and nonstationarities (Xie *et al.*, 2016), it is difficult to achieve the goal (Chen *et al.*, 2020b). Therefore, a lot of data-driven causality



© Qiming Chen, Xinyi Fei, Lie Xie, Dongliu Li and Qibing Wang. Published in *Journal of Intelligent Manufacturing and Special Equipment*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

Journal of Intelligent
Manufacturing and Special
Equipment
Vol. 1 No. 1, 2020
pp. 25-41
Emerald Publishing Limited
e-ISSN: 2633-660X
p-ISSN: 2633-6596
DOI 10.1108/JIMSE-06-2020-0003

analysis methods have been introduced to revealing the propagation paths and locating the root cause in the past decade (Lindner *et al.*, 2019a). A brief review is provided in the following.

Bauer and Thornhill (2008) analyzed the cross-correlation between variables to detect causality. In this method, the time delay corresponding to the maximum value of cross-correlation function (CCF) is regarded as the optimal delay. This CCF-based approach is simple and practical, but its results are not reliable in the case of nonlinearity. (Yuan and Qin, 2014) combined the principal component analysis (PCA) and the multivariate Granger causality to find the root cause of plant-wide oscillations. Granger causality is a popular method in various applications. However, it requires variables are independent and satisfy the linear stationary condition. Besides the causality methods in time domain, some related approaches are developed from the perspective of frequency domain, such as spectral Granger causality (SGC) (Xie *et al.*, 2019), partial directed coherence (PDC) (Zhang *et al.*, 2015) and directed transfer function (DTF) (Yang *et al.*, 2014), to name a few. Spectral Granger causality is based on the Fourier Transform, which is limited to processing the linear and stationary signals. Both PDC and DTF can describe the direction of cause and effect, but are not able to quantify the direct causality.

The above methods are restricted to analyzing linear relationships, which is not consistent with the fact that most industrial processes are nonlinear. Bayesian network learning (Yang and Xiao, 2006) utilized the graph structure with conditional probabilities to describe the causalities among various process variables. Nevertheless, Bayesian network is a directed acyclic graph, which is not proper to model the dynamic process. And the physical explanation of probabilities in Bayesian network is not straightforward, which cannot be easily accepted by engineers. Although (Richardson *et al.*, 1996) developed the cyclic causal discovery framework to allow the existence of cycles, it is difficult to collect massive industrial data. Lindner *et al.* (2019a) made a comprehensive analysis of Granger causality and transfer entropy in process control system and Lindner *et al.* (2019b) provided a systematic workflow for oscillation diagnosis using transfer entropy. However, Transfer entropy's performance depends on probability density function, which is not easy to be estimated.

In 2012, Sugihara *et al.* (2012) proposed a novel causality method called as convergent cross mapping (CCM) for analyzing relationships in complex systems. It is mainly based on the Takens' theorem (Takens, 1981) that if variable X is a cause of variable Y , then the historical information of X can be recovered from Y alone. CCM uses simplex projection to quantify the correspondence between the states of X and Y . And the Pearson correlation coefficients between estimated states recovered from Y and real states are adopted to measure the cross mapping ability, which indicates the causality degree. CCM has been widely used in various cases, but its criterion of convergence is too subjective. Recently, Ye *et al.* (2015) developed a time-delayed convergent cross mapping (TD-CCM) by assuming there is a time lag between the cause variable and the effect variable. It is reported that CCM shows the best performance when two variables are matched in accordance with proper time lag (Luo *et al.*, 2017). TD-CCM overcomes the subjectivity of the conclusion under the original CCM and can directly give the corresponding results. TD-CCM is the most promising progress of CCM in recent years. However, when it is applied to analyze oscillations in process control, the noise and periodicity degrade its performance. To eliminate the influences, this paper proposes to combine empirical mode decomposition (EMD) (Huang *et al.*, 1998) and detrended fluctuation analysis (DFA) (Kantelhardt *et al.*, 2002) to achieve denoising. The periodicities are removed by singular spectrum analysis (SSA) (Hassani and Thomakos, 2010). Following, the TD-CCM can accurately capture the causalities and locate the root cause by analyzing the filtered

signals. Simulations shows that the proposed denoising and periodicity-removing framework is able to improve the causality analysis performance of TD-CCM. In the end, the effectiveness and advantages of the proposed method are validated in Tennessee Eastman process.

The remainder of this paper is organized as follows. [Section 2](#) provides an overview of CCM and TD-CCM. The proposed causality analysis framework based on denoising and periodicity-removing TD-CCM is detailed in [Section 3](#). The industrial case is studied in [Section 4](#), followed by conclusions.

2. Preliminaries

2.1 Convergent cross mapping

([Sugihara et al., 2012](#)) proposed the convergent cross mapping (CCM) based on Taken's theorem ([Takens, 1981](#)) to analyze the causality between different variables in a nonlinear system. CCM assumes that if variable X has an effect on variable Y , denoted as $X \rightarrow Y$, Y will contain the information of X . The corresponding causality can be tested by measuring the correlation between reconstructed manifolds of X and Y . More specifically, for two time series X and Y with length N , the reconstructed states at time t can be expressed as

$$\begin{aligned} X_t &= [X(t), X(t - \tau), X(t - 2\tau), \dots, X(t - (E - 1)\tau)], \\ Y_t &= [Y(t), Y(t - \tau), Y(t - 2\tau), \dots, Y(t - (E - 1)\tau)], \\ \forall t &= (E - 1)\tau + 1, (E - 1)\tau + 2, \dots, N \end{aligned} \quad (1)$$

where E is the embedding dimension; τ is the time lag (default is 1). The embedding dimension is selected according to G-curve method ([Liu et al., 2008](#)). The sets of X_t and Y_t correspond to their shadow manifolds M_X and M_Y , respectively. According to Taken's theorem ([Takens, 1981](#)), if X and Y are coupled, M_X and M_Y are different observation forms of primitive manifold, i.e. diffeomorphism. Conversely, if there is no relationship between X and Y , the reconstructed states of M_X and M_Y will be far away. When X has an effect on Y , the information of M_X can be accurately estimated from M_Y , but M_Y cannot be recovered from M_X . CCM uses Pearson correlation coefficient to quantify the accuracy of X estimation from Y , shown as

$$\rho_{\hat{X}_i|M_Y} = \frac{\sum \hat{X}_i X_i - N \bar{\hat{X}_i} \bar{X_i}}{\sqrt{(\sum X_i^2 - N \bar{X_i}^2)} \sqrt{(\sum \hat{X}_i^2 - N \bar{\hat{X}_i}^2)}} \quad (2)$$

where \hat{X}_i is the estimation of X_i . The larger value of $\rho_{\hat{X}_i|M_Y}$, the stronger the influence of X on Y .

2.2 Time-delayed convergent cross mapping

It is reported that CCM is effective to detect causality from systems with weak to moderate coupling strength, but strong unidirectional forcing may lead to the phenomenon of generalized synchrony ([Sugihara et al., 2012](#)). Besides, [Ye et al. \(2015\)](#) pointed out that CCM suffered two main limitations: (1) This method only considers the causality between the variables at the same time (zero lag). However, there are a lot of causalities with time delay in the actual process, in which CCM would meet difficulties or failures; (2) CCM judges causality by observing whether the correlation coefficient curve converges, which is subjective. To tackle these issues, [Ye et al. \(2015\)](#) proposed a time-delayed CCM (TD-CCM).

The main difference between CCM and TD-CCM is that the latter utilizes M_Y to estimate $x_{t+\lambda}$, in which λ is the time lag; while in CCM, M_Y is used to calculate the estimation of x_t . In practice, different values of λ are tested to calculate the cross mapping index ρ . The λ corresponding to the maximum ρ value is the optimal time delay λ^* . $\lambda^* \leq 0$ corresponds to the relationship $X \rightarrow Y$, because $\lambda^* \leq 0$ means the past information of causal variable is contained in effect variable. If $\lambda^* > 0$, it reflects the future information of cause variable can predict the effect variable, which is ridiculous. Thus, $\lambda^* > 0$ indicates the causal relationship from X to Y is not tenable. In this way, TD-CCM can automatically and objectively determine causality.

Herein, a coupled Lorenz system (3) is constructed to show the performance of CCM and TD-CCM.

$$\begin{aligned}\dot{X}_i(t) &= 10(Y_i(t) - X_i(t)) + \xi_{X,i} \\ \dot{Y}_i(t) &= 28X_i(t) - Y_i(t) - X_i(t)Z_i(t) + \sum_{j \neq i} K_{ij}Y_j^2(t - \tau_{ij}) + \xi_{Y,i} \\ \dot{Z}_i(t) &= X_i(t)Y_i(t) - 8/3Z_i(t) + \xi_{Z,i} \\ & i, j = 1, 2; \\ & K_{12} = 1, K_{21} = 0; \\ & \tau_{12} = 1;\end{aligned}\tag{3}$$

where ξ is noise. X_1, Y_1 and Z_1 belong to system 1; X_2, Y_2 and Z_2 belong to system 2; Y_1 and Z_1, Y_2 and Z_2 are respectively coupled. Y_2 in system 2 affects Y_1 in system 1 in one direction. Z_1 and Z_2 are selected for causality analysis. Note that Z_2 is the cause of Z_1 . According to G -curve method (Liu *et al.*, 2008), the embedding dimension is set as $E^* = 3$. The causality results of CCM and TD-CCM are displayed in Figures 1a and b, respectively. It can be seen from Figure 1a that both red and blue curves tend to convergence, which indicates Z_1 and Z_2 are mutually coupled, i.e. two-way causality. This judgment contradicts the real situation of the system. From Figure 1b, It is observed that the optimal time lag of $Z_2 \rightarrow Z_1$ is $\lambda^* = -50$, which indicates $Z_2 \rightarrow Z_1$; while for $Z_1 \rightarrow Z_2$, the optimal time lag is positive, which is not consistent with the fact that the result event must happen after the cause event. It is concluded that Z_2 is the cause of Z_1 . Therefore, compared with the original CCM, the TD-CCM not only can automatically and objectively determine causality, but also solves the problem that the original CCM is not able to deal with strong coupling relationship.

3. Proposed framework

Although TD-CCM shows better performance than the original CCM, it meets difficulties in the case of processing signals from process control system. For data collected from industrial environments, they are contaminated by noise and periodicity (Chen *et al.*, 2019), which have an adverse influence on TD-CCM. Therefore, a denoising and periodicity-removing framework is proposed to improve TD-CCM's performance in this section.

3.1 Adverse influences of noise and periodicity on TD-CCM

A four-input and four-output system with correlated disturbances is taken from Wang *et al.* (2020) to show the influences of noise and periodicity on TD-CCM. Its transfer function matrix $G(q^{-1})$ is

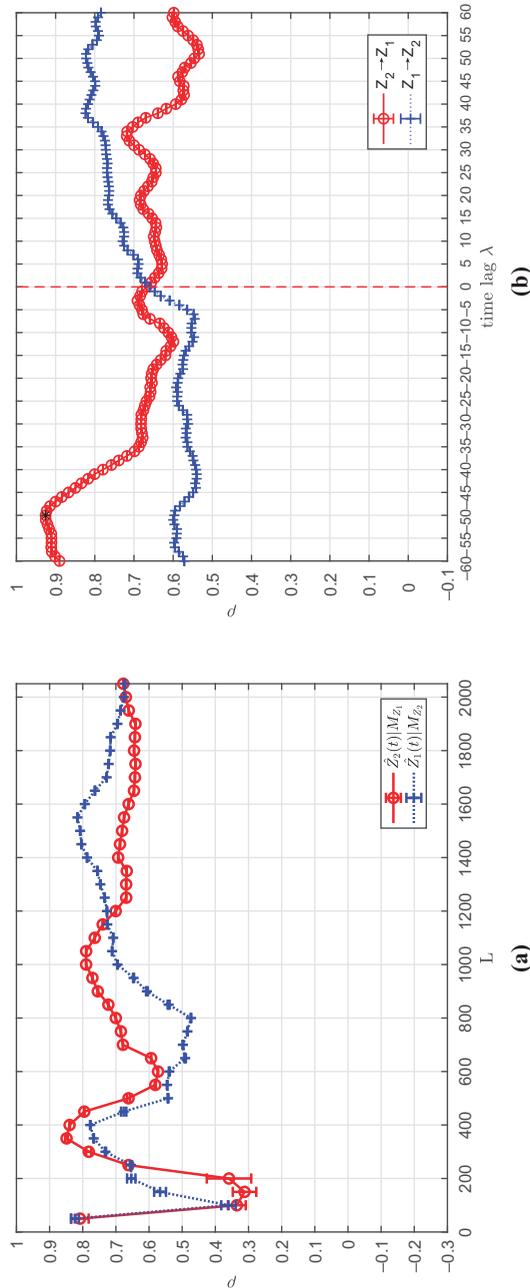


Figure 1. Causality results detected by (a) CCM and (b) TD-CCM. The red and blue curves represent the causality results in $Z_2 \rightarrow Z_1$ and $Z_1 \rightarrow Z_2$, respectively. For (a) CCM, the horizontal axis and vertical axis stand for the sample length and causality strength, respectively. Through the trends, it is judged that Z_1 and Z_2 are mutually coupled, i.e. two-way causality. For (b) TD-CCM, the horizontal axis and vertical axis stand for the time lag and causality strength, respectively. It is observed that the optimal time lag of $Z_2 \rightarrow Z_1$ is $\lambda^* = -50$, which indicates $Z_2 \rightarrow Z_1$; while for $Z_1 \rightarrow Z_2$, the optimal time lag is positive, which is not consistent with the fact that the result event must have happened after the cause event. Therefore, it is concluded $Z_2 \rightarrow Z_1$.

$$G(q^{-1}) = \begin{bmatrix} \frac{0.05q^{-3}}{1-0.95q^{-1}} & 0 & \frac{0.7q^{-3}}{1-0.3q^{-1}} & 0 \\ \frac{0.02966q^{-3}}{1-1.627q^{-1}+0.706q^{-2}} & \frac{0.0627q^{-6}}{1-0.937q^{-1}} & 0 & 0 \\ 0 & \frac{0.235q^{-5}}{1-0.765q^{-1}} & \frac{0.5q^{-2}}{1-q^{-1}+0.25q^{-2}} & 0 \\ \frac{0.5q^{-5}-0.4875q^{-6}}{1-1.395q^{-1}+0.455q^{-2}} & 0 & 0 & \frac{0.2q^{-6}}{1-0.8q^{-1}} \end{bmatrix}, \quad (4)$$

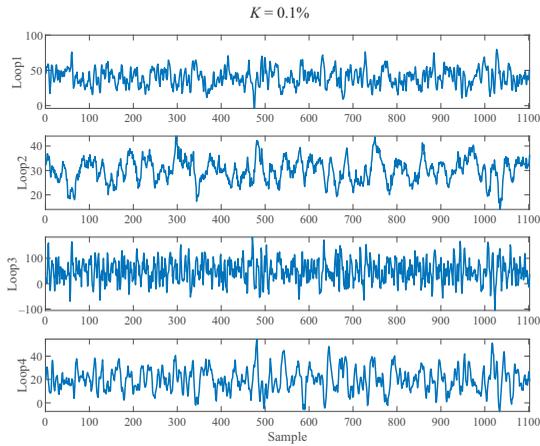
and the corresponding disturbance model is

$$N(q^{-1}) = \begin{bmatrix} \frac{1-0.1875q^{-1}}{1-0.9875q^{-1}} & 0 & 0 & 0 \\ 0 & \frac{1-0.1875q^{-1}}{1-0.9875q^{-1}} & 0 & 0 \\ 0 & 0 & \frac{1-0.1875q^{-1}}{1-0.9875q^{-1}} & 0 \\ 0 & 0 & 0 & \frac{1-0.1875q^{-1}}{1-0.9875q^{-1}} \end{bmatrix}. \quad (5)$$

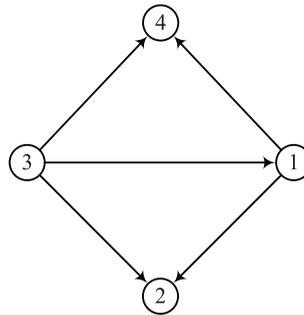
The disturbance source is generated by Gaussian white noise sequences. These loops are regulated by four PI controllers with $PI = K_C \left\{ 1 + \frac{1}{Tr} \times \frac{\Delta T}{1-q^{-1}} \right\}$, where ΔT is the controller sampling time. When the proportional gain and the integral time are $K_C = [0.816 \ 0.625 \ 0.184 \ 0.37]$ and $Tr = [20 \ 16 \ 2.86 \ 5]$ respectively, these loops are well controlled and the system works normally. A data-driven valve stiction model (Choudhury *et al.*, 2008) is embedded into the third loop to simulate the fault source.

First, we takes the noise level $K = 0.1\%$ as the benchmark, in which the noise level is rather low. The corresponding process variables are displayed in Figure 2a. TD-CCM tests the causal relationship of six pairs of variables, i.e. $L_1 - L_2$, $L_1 - L_3$, $L_1 - L_4$, $L_2 - L_3$, $L_2 - L_4$, and $L_3 - L_4$. The corresponding causality results are visualized in Figure 2b. It is observed that L_3 is the source because $L_3 \rightarrow L_1, L_2, L_4$ and $L_1 \rightarrow L_2, L_4$. These causality paths are completely consistent with the predefined system structure (4). However, when the noise level increases, such as $K = 10\%$, there will be many miscalculations in TD-CCM. As displayed in Figures 3a and b, when the noise level increases, TD-CCM thinks that there is a two-way causal relationship between loop 2 and loop 4 (marked with red lines), which contradicts the prior system structure. To enrich the experiments, we have counted the miscalculation of TD-CCM under different noise levels. The results listed in Table 1 indicate that with the increase of noise level, the number of causal misjudgments will go up. Therefore, it is necessary to denoise before using TD-CCM to analyze causality.

Second, we test the performance of TD-CCM when the signals are contaminated by various periodicities. The amplitudes and frequencies of periodicities vary from 12.5%

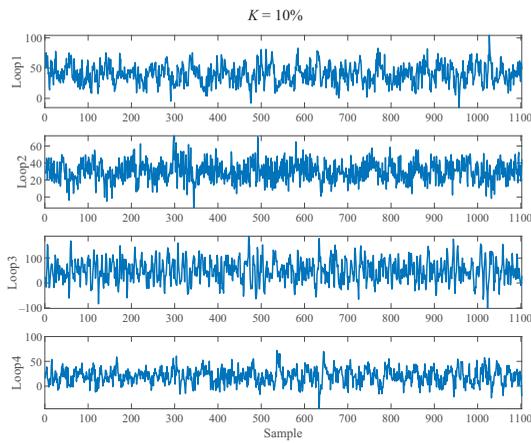


(a)

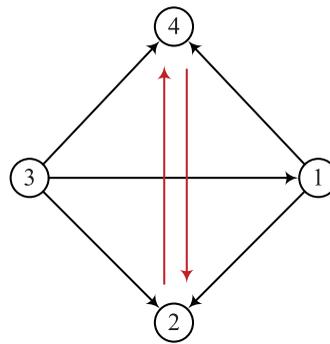


(b)

Figure 2. (a): Process variables in the case of $K = 0.1\%$ and (b): The corresponding causality results obtained from TD-CCM. The propagation paths are consistent with the predefined structure



(a)



(b)

Figure 3. (a): Process variables in the case of $K = 10\%$ and (b): The corresponding causality results obtained from TD-CCM. The red lines are wrong direction

K	0.1%	1%	5%	10%	50%	100%
Num	0	1	2	2	2	2

Table 1. With the increase of noise level, the number of causal misjudgments will increase

Period/Amp	12.5%	25%	50%	100%
30	0	0	1	1
60	2	2	2	3
100	0	1	3	3

Table 2. Number of TD-CCM misjudgments when signals are contaminated by periodicities

to 100% and 30 to 100 samples per cycle, respectively. The corresponding causality results obtained from TD-CCM are listed in Table 2. It is observed that due to the presence of periodicities, TD-CCM often misjudges the causalities in this system. Therefore, the periodicities are expected to be removed before TD-CCM are applied to detect causality.

Remark: The adverse effects of noise and periodicity on TD-CCM can be explained from the following two aspects: (1) TD-CCM involves searching the nearest neighbor. Noises will disturb the performance of nearest neighbor algorithm. Thus the causality results obtained from TD-CCM may make mistakes; (2) Periodicities strengthen the invalid coordination between variable, thus covering up the valuable information transfer flow, which may lead to two-way causality or missing relationship.

3.2 Denoising and periodicity-removing

The last section demonstrates the necessity of denoising and periodicity-removing procedure. Herein, a denoising and periodicity-removing framework is proposed to achieve this goal.

First, the signals are decomposed into a series of modes by empirical mode decomposition (EMD) (Huang *et al.*, 1998). EMD is a modern time-frequency analysis technique. The traditional signal processing methods, such as Fourier transform, are limited to processing linear and stationary signals (Lang *et al.*, 2020a). Although the wavelet decomposition can be used to deal with complex signals, its parameters, such as decomposition level and mother wavelet, should be provided in advance. On the contrary, EMD is totally adaptive and data-driven. It is capable to extract the intrinsic mode functions from nonlinear and nonstationary signals through a recursive sifting process that makes use of signal extrema. More specifically, by interpolating the extremum of the signal, the average values of the upper and lower envelopes are obtained, and then the local average values of the signal are obtained. These local average values are the low-frequency estimation of the data, and then they are removed from the input data recursively to generate the high-frequency mode in the signal. The process is repeated until all principal oscillatory modes present in the data are recovered. In EMD, the signal $x(t)$ can be expressed as

$$x(t) = \sum_{i=1}^K d_i(t) + r(t) \quad (6)$$

where $d_i(t)$ is the mode and $r(t)$ is trend.

After the signal is decomposed into a set of modes by EMD, the next task is to identify which modes are noise. There are many methods that can be adopted, such as permutation entropy (Lang *et al.*, 2020b), normalized correlation coefficient (Chen *et al.*, 2020c), etc. Herein, the detrended fluctuation analysis (DFA) (Bryce and Sprague, 2012) is utilized to identify whether the mode is noise. Specifically, a measurement factor α of each mode can be calculated by DFA. If $\alpha \leq 0.5$, the corresponding mode is regarded as noise. That is to say, only modes with $\alpha > 0.5$ are retained. The threshold 0.5 is recommended by Peng *et al.* (1995).

Now, we turn our attention to remove periodicities. First, the signal is decomposed by singular spectrum analysis (SSA) to obtain the eigenvalues and eigenvectors. Because the eigenvectors corresponding to periodicities are sine or cosine sequences with the same frequency and phase, the scatter diagram of a pair of eigenvectors of periodicities will form an approximate polygon. In this way, periodicities can be distinguished.

The proposed causality analysis framework is described in Figure 4. It consists of three parts: denoising, periodicity-removing and TD-CCM.

3.3 Simulation

In order to demonstrate the effectiveness and advantages of the proposed framework, the four-input and four-output control system (4) is used as a subject. Loop 3 is contaminated by noise and periodicity. Its original data, noise, periodicity, and filtered signal are displayed in Figure 5. It is observed that the noise and periodicity are accurately extracted through EMD-DFA-SSA procedures. Then, we apply TD-CCM to the original data, and denoising and periodicity-removing data, respectively. The causality results are reported in Figures 6a and b, respectively. It can be seen that the original TD-CCM misjudges the causality $L_4 \rightarrow L_2$ (red dotted line) and misses the causality between L_1 and

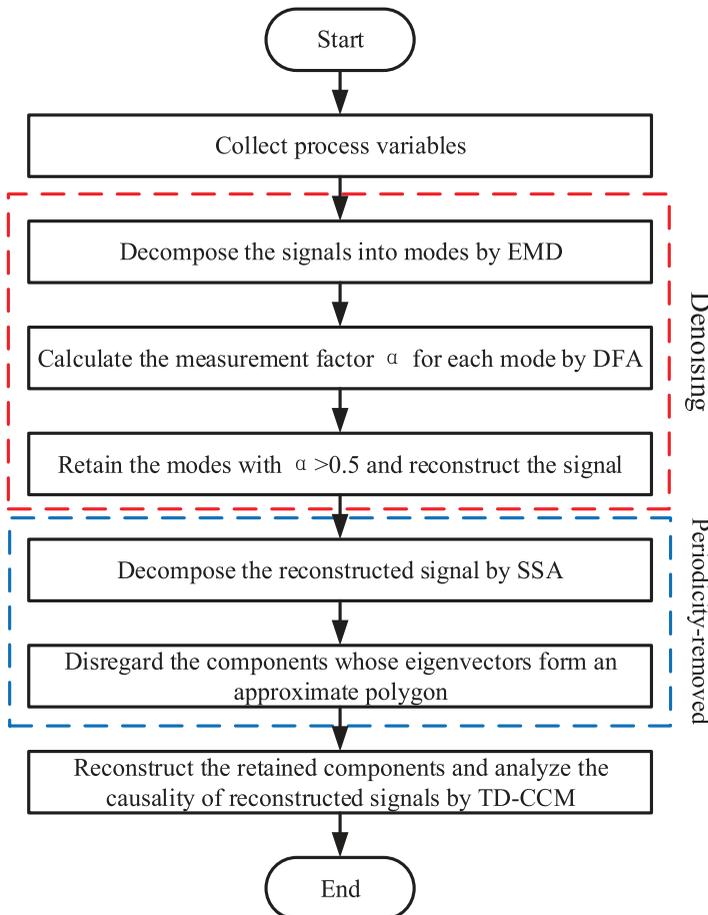


Figure 4. Flow chart of the proposed method

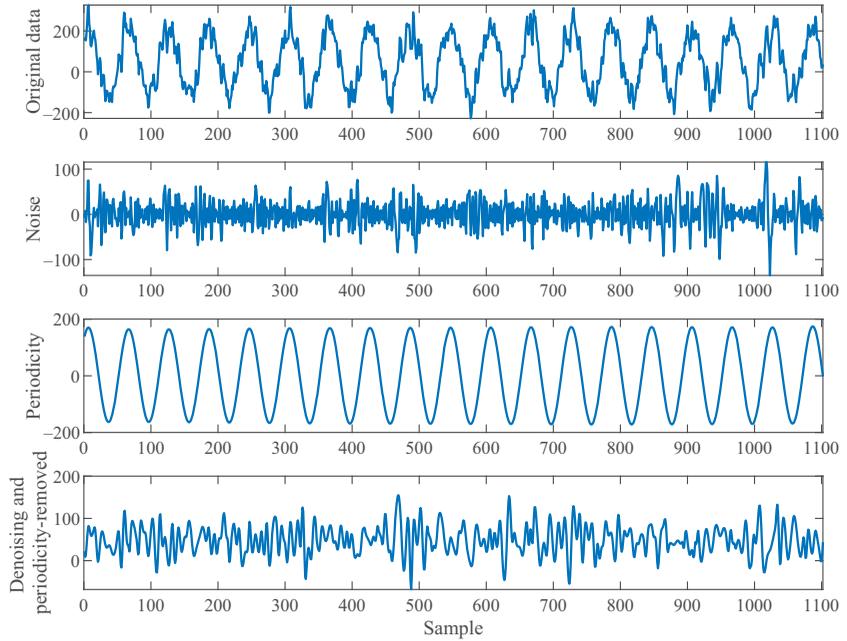


Figure 5. The first row is the original data of loop 3. The second row is the noises extracted by EMD and DFA; the third row is the periodicity obtained from SSA; the bottom is the denoising and periodicity-removing signal

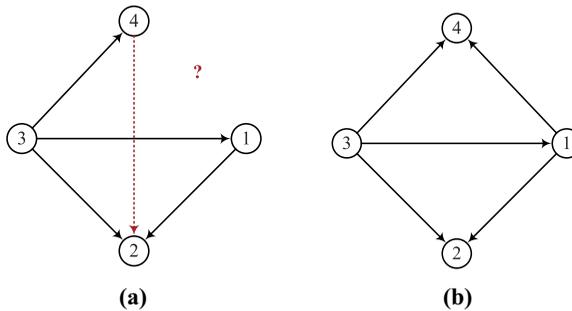


Figure 6. Causality results obtain from (a) original TD-CCM and (b) the proposed improved work

Note(s): It can be seen that the original TD-CCM misjudges the causality $L4 \rightarrow L2$ (red dotted line) and misses the causality between $L1$ and $L4$; while the proposed method accurately capture the causalities in the control system

$L4$; while the proposed method is able to correctly capture all causalities in the control system.

4. Industrial case study

In this section, the Tennessee Eastman process is used to demonstrate the utility of the proposed method in industrial situation (Zheng *et al.*, 2020). The corresponding

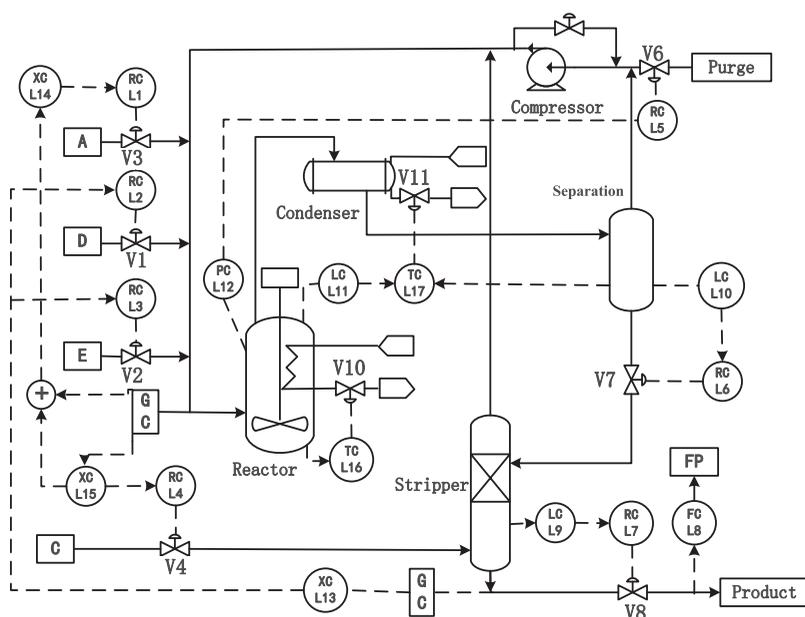


Figure 7.
Process schematic of a
plant from Eastman
chemical company

process schematic is shown in [Figure 7](#). It mainly consists of five parts: a reactor, a recycle compressor, a stripper, a product condenser, and a vapor separator. The process is regulated under a decentralized control strategy ([Ricker, 1996](#)). The predefined fault is added into loop 16. It can be seen from [Figure 8](#) that the predefined fault propagates to other loops and results in the plant-wide oscillations ([Lang et al., 2018c](#)). In this test, the sampling period is 0.1 h and the analyzed data are sampled from point 250–700. The causality detected by the proposed method is provided in [Figure 9](#). For comparison, multivariate Granger causality ([Lindner et al., 2019a](#)) is also tested and the corresponding results are plotted in [Figure 10](#). Because of the large number of process variables, the causal network 9 is too complex. It is not easy to distinguish the root cause with the naked eye. A net causal flow ([Yuan and Qin, 2014](#)) can be computed for each variable which is equal to the number of outgoing flows minus the number of incoming flows. According to [Yuan and Qin \(2014\)](#), a node with a high positive causal flow is likely to be a source; while a high negative causal flow represents a likely sink in the causality network. [Figures 11a](#) and [b](#) depict the causal flow obtained from the proposed method and multivariate Granger causality, respectively. The significance level of hypothesis test is 0.05 for multivariate Granger causality. In [Figure 11a](#), the variable 16 and 17 have the largest positive causal flow, and thus are regarded as the potential root causes. Because variable 16 has one-way causal effect on variable 17 in [Figure 9](#), it can be concluded that variable 16 is the real root cause. By contrast, the potential root causes of multivariate Granger causality are variable 6 and 15, which do not contain the predefined root cause. Therefore, the proposed method is successful to be applied to analyze the root cause in Tennessee Eastman process. And it shows better performance than the classical multivariate Granger causality.

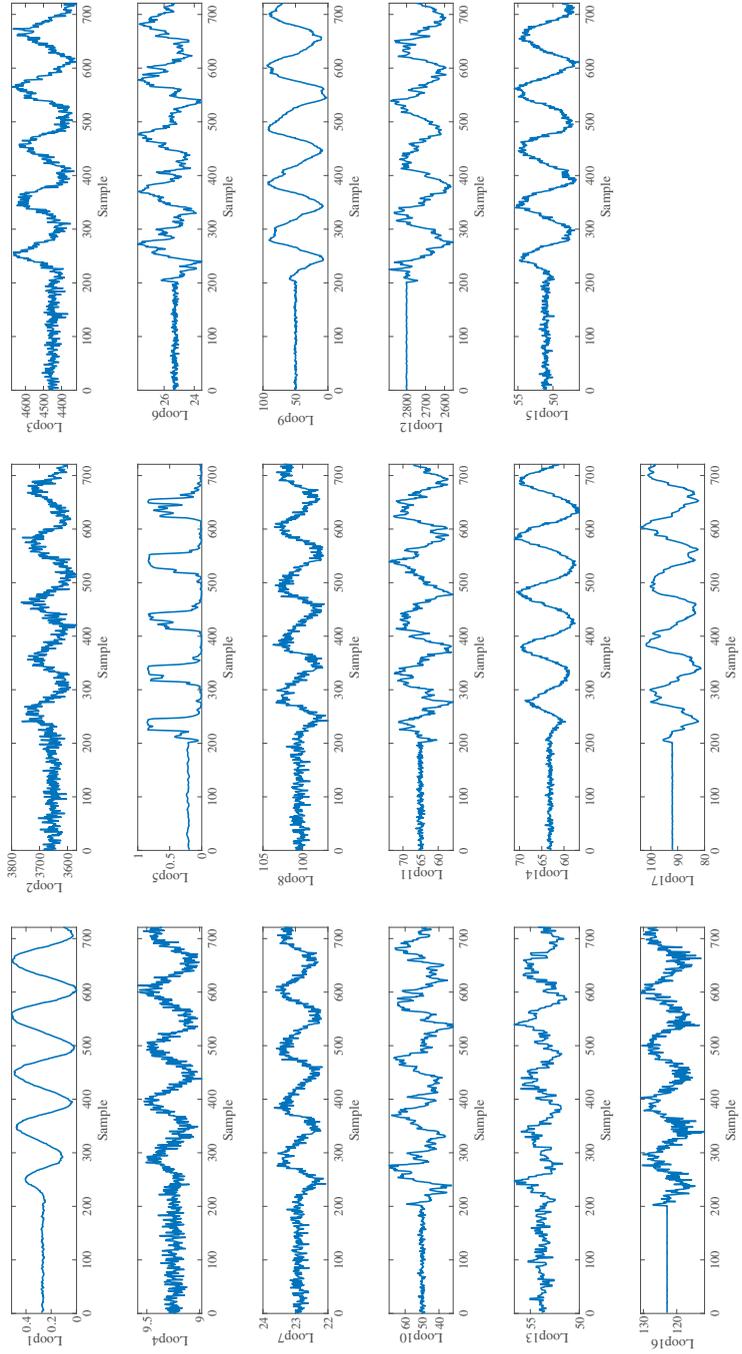


Figure 8.
Trends of process
variables of 17 loops in
Tennessee Eastman
process

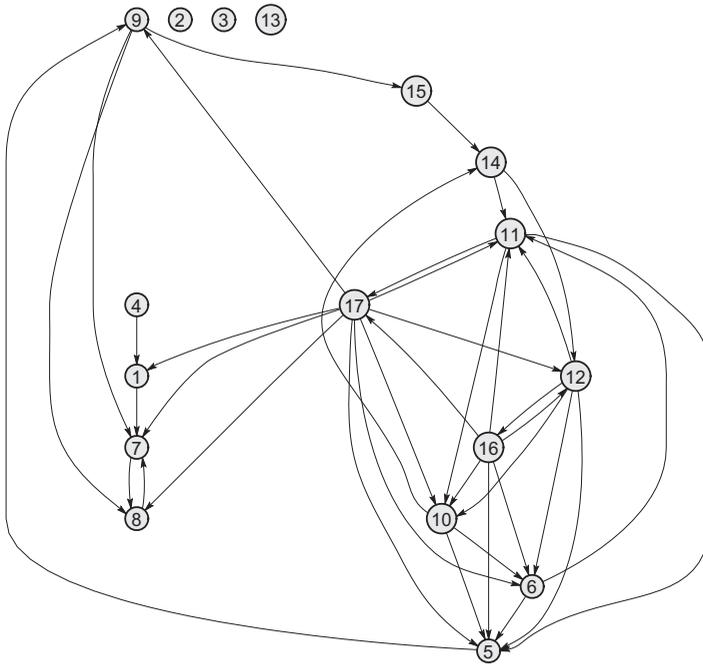


Figure 9.
Causality map in
Tennessee Eastman
process. This map is
obtained from the
proposed method

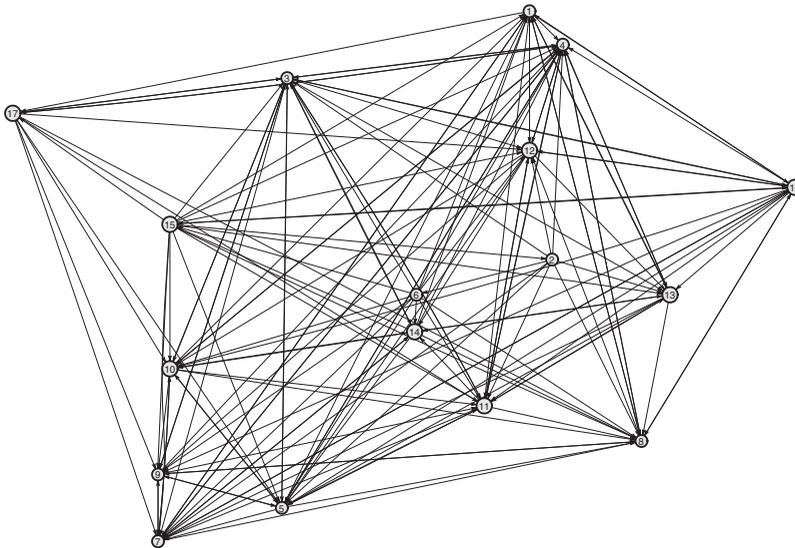


Figure 10.
Causality map in
Tennessee Eastman
process. This map is
obtained from the
multivariate Granger
causality

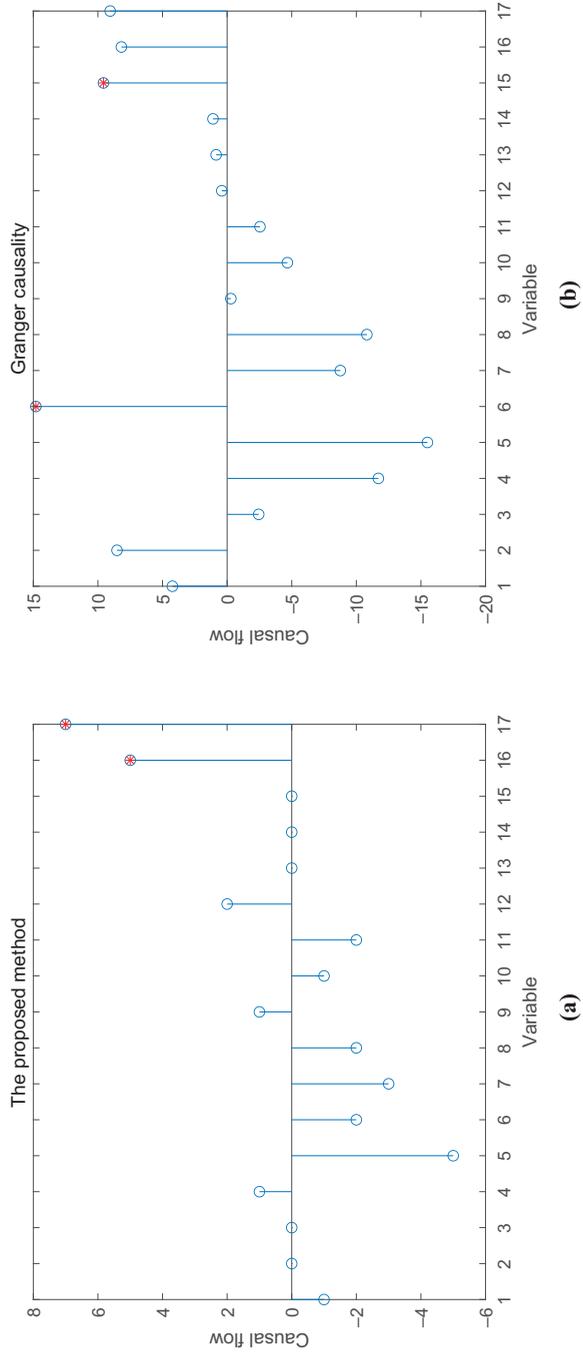


Figure 11. Causal flow for each node (variable) in Tennessee Eastman process. (a): The proposed method; (b) Multivariate Granger causality. The potential root causes are marked with red asterisk. It is observed that the potential root causes of the proposed method contain the real source; while that of Granger causality method excludes the true source

5. Conclusions

In this paper, a novel causality analysis framework is proposed based on denoising and periodicity-removing TD-CCM. First, the adverse effects of noise and periodicity are investigated. Then, the EMD and FDA are combined to achieve denoising. And SSA are used to remove the periodicities. Simulations demonstrate the proposed denoising and periodicity-removing procedure can effectively improve the performance of TD-CCM. It can reduce the miscalculation and omission of causality. In the end, the proposed causality analysis framework is applied to Tennessee Eastman process to identify the root cause of plant-wide oscillations. The application results show that the proposed method is effective and promising for root cause diagnosis in process control system.

References

- Bauer, M. and Thornhill, N.F. (2008), "A practical method for identifying the propagation path of plant-wide disturbances", *Journal of Process Control*, Vol. 18 Nos 7-8, pp. 707-719.
- Bryce, R. and Sprague, K. (2012), "Revisiting detrended fluctuation analysis", *Scientific Reports*, Vol. 2, p. 315.
- Chen, Q., Lang, X., Xie, L. and Su, H. (2019), "Detecting nonlinear oscillations in process control loop based on an improved VMD", *IEEE Access*, Vol. 7, pp. 91446-91462.
- Chen, Q., Chen, J., Lang, X., Xie, L., Lu, S. and Su, H. (2020), "Detection and diagnosis of oscillations in process control by fast adaptive chirp mode decomposition", *Control Engineering Practice*, Vol. 97, p. 104307.
- Chen, Q., Chen, J., Lang, X., Xie, L., Jiang, C. and Su, H. (2020), "Diagnosis of nonlinearity-induced oscillations in process control loops based on adaptive chirp mode decomposition", *2020 American Control Conference (ACC)*, IEEE, pp. 2772-2777.
- Chen, Q., Xie, L. and Su, H. (2020), "Multivariate nonlinear chirp mode decomposition", *Signal Processing*, p. 107667.
- Choudhury, A.A.S., Shah, S.L. and Thornhill, N.F. (2008), *Diagnosis of Process Nonlinearities and Valve Stiction: Data Driven Approaches*, Springer Science and Business Media.
- Hassani, H. and Thomakos, D. (2010), "A review on singular spectrum analysis for economic and financial time series", *Statistics and its Interface*, Vol. 3 No. 3, pp. 377-397.
- Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., *et al.* (1998), "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis", *Proceedings of the Royal Society of London Series A: Mathematical, Physical and Engineering Sciences*, Vol. 454 No. 1971, pp. 903-995.
- Kantelhardt, J.W., Zschiegner, S.A., Koscielny-Bunde, E., Havlin, S., Bunde, A. and Stanley, H.E. (2002), "Multifractal detrended fluctuation analysis of nonstationary time series", *Physica A: Statistical Mechanics and its Applications*, Vol. 316 Nos 1-4, pp. 87-114.
- Lang, X., Zhang, Z., Xie, L., Horch, A. and Su, H. (2018), "Time-frequency analysis of plant-wide oscillations using multivariate intrinsic time-scale decomposition", *Industrial and Engineering Chemistry Research*, Vol. 57 No. 3, pp. 954-966.
- Lang, X., Lu, S., Xie, L., Zakharov, A., Zhong, D. and Jämsä-Jounela, S.L. (2018), "Bihocurrence based industrial control loop nonlinearity detection and diagnosis in short nonstationary time series", *Journal of Process Control*, Vol. 63, pp. 15-28.
- Lang, X., Zheng, Q., Zhang, Z., Lu, S., Xie, L., Horch, A., *et al.* (2018), "Fast multivariate empirical mode decomposition", *IEEE Access*, Vol. 6, pp. 65521-65538.

- Lang, X., Zheng, Q., Xie, L., Horch, A. and Su, H. (2019), "Direct multivariate intrinsic time-scale decomposition for oscillation monitoring", *IEEE Transactions on Control Systems Technology*.
- Lang, X., ur Rehman, N., Zhang, Y., Xie, L. and Su, H. (2020), "Median ensemble empirical mode decomposition", *Signal Processing*, p. 107686.
- Lang, X., Zhang, Y., Xie, L., Jin, X., Horch, A. and Su, H. (2020), "The use of fast multivariate empirical mode decomposition for oscillation monitoring in noisy process plant", *Industrial and Engineering Chemistry Research*.
- Lindner, B., Auret, L., Bauer, M. and Groenewald, J.W. (2019), "Comparative analysis of Granger causality and transfer entropy to present a decision flow for the application of oscillation diagnosis", *Journal of Process Control*, Vol. 79, pp. 72-84.
- Lindner, B., Auret, L. and Bauer, M. (2019), "A systematic workflow for oscillation diagnosis using transfer entropy", *IEEE Transactions on Control Systems Technology*.
- Liu, S., Zhu, S. and Yu, X. (2008), "Determinating the embedding dimension in phase space reconstruction", *Journal of Harbin Engineering University*, Vol. 4, pp. 374-381.
- Luo, L., Cheng, F., Qiu, T. and Zhao, J. (2017), "Refined convergent cross-mapping for disturbance propagation analysis of chemical processes", *Computers and Chemical Engineering*, Vol. 106, pp. 1-16.
- Peng, C.K., Havlin, S., Stanley, H.E. and Goldberger, A.L. (1995), "Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series", *Chaos: An Interdisciplinary Journal of Nonlinear Science*, Vol. 5 No. 1, pp. 82-87.
- Richardson, T.S., Spirtes, P., *et al.* (1996), "Automated discovery of linear feedback models", *Carnegie Mellon [Department of Philosophy]*.
- Ricker, N.L. (1996), "Decentralized control of the Tennessee Eastman challenge process", *Journal of Process Control*, Vol. 6 No. 4, pp. 205-221.
- Sugihara, G., May, R., Ye, H., Hsieh, Ch., Deyle, E., Fogarty, M., *et al.* (2012), "Detecting causality in complex ecosystems", *Science*, Vol. 338 No. 6106, pp. 496-500.
- Takens, F. (1981), "Detecting strange attractors in turbulence", *Dynamical Systems and Turbulence, Warwick 1980*, Springer, pp. 366-381.
- Wang, K., Chen, J. and Song, Z. (2020), "A sparse loading-based contribution method for multivariate control performance diagnosis", *Journal of Process Control*, Vol. 85, pp. 199-213.
- Xie, L., Lang, X., Chen, J., Horch, A. and Su, H. (2016), "Time-varying oscillation detector based on improved LMD and robust Lempel-Ziv complexity", *Control Engineering Practice*, Vol. 51, pp. 48-57.
- Xie, T., Wang, J.G., Xie, Z.T., Yao, Y. and Xu, C. (2019), "Root cause diagnosis of plant-wide oscillation using spectral granger causality based on phase resampling test", *IEEE 8th Data Driven Control and Learning Systems Conference (DDCLS)*, IEEE, pp. 1230-1235.
- Yang, F. and Xiao, D. (2006), "Approach to fault diagnosis using SDG based on fault revealing time", *2006 6th World Congress on Intelligent Control and Automation*, IEEE, Vol. 2, pp. 5744-5747.
- Yang, F., Duan, P., Shah, S.L. and Chen, T. (2014), *Capturing Connectivity and Causality in Complex Industrial Processes*, Springer Science and Business Media.
- Ye, H., Deyle, E.R., Gilarranz, L.J. and Sugihara, G. (2015), "Distinguishing time-delayed causal interactions using convergent cross mapping", *Scientific Reports*, Vol. 5, p. 14750.
- Yuan, T. and Qin, S.J. (2014), "Root cause diagnosis of plant-wide oscillations using Granger causality", *Journal of Process Control*, Vol. 24 No. 2, pp. 450-459.
- Zhang, L., Zheng, J. and Xia, C. (2015), "Propagation analysis of plant-wide oscillations using partial directed coherence", *Journal of Chemical Engineering of Japan*, Vol. 48 No. 9, pp. 766-773.

Zheng, Y., Wang, Y., Yan, H., Wang, Y., Yang, W. and Tao, B. (2020), "Density peaks clustering-based steady/transition mode identification and monitoring of multimode processes", *The Canadian Journal of Chemical Engineering*.

Causality
analysis in
process control

Corresponding author

Lie Xie can be contacted at: leix@iipc.zju.edu.cn

41

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgroupublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com