University of Alberta

Information Theory-based Approaches for Causality Analysis with Industrial Applications

by

Ping Duan

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Control Systems

Department of Electrical & Computer Engineering

Spring 2014
Edmonton, Alberta

Dedicated to my parents and my husband for their support and love throughout my life.

Abstract

Detection and diagnosis of plant-wide abnormalities and disturbances are major problems in large-scale complex systems. To determine the root cause(s) of specific abnormalities, it is important to capture the process connectivity and investigate the fault propagation pathways, in which causality detection plays a significant and central role. This thesis focuses mainly on information theory-based approaches for causality analysis that are suitable for both linear and nonlinear process relationships.

Previous studies have shown that the transfer entropy approach is a very useful tool in quantifying causal influence by inferring material and information pathways in a system. However, the traditional transfer entropy method only determines whether there is causality from one variable to another; it cannot tell whether the causal influence is along a direct pathway or indirect pathways through some intermediate variables. In order to detect and discriminate between direct and indirect causality relationships, a direct transfer entropy concept is proposed in this thesis. Specifically, a differential direct transfer entropy concept is defined for continuous-valued random variables, and a normalization method for the differential direct transfer entropy is presented to determine the connectivity strength of direct causality.

A key assumption for the transfer entropy method is that the sampled data should follow a well-defined probability distribution; yet this assumption may not hold for all types of industrial process data. A new information theory-

based distribution-free measure, transfer 0-entropy, is proposed for causality analysis based on the definitions of 0-entropy and 0-information without assuming a probability space. For the cases of more than two variables, a direct transfer 0-entropy concept is presented to detect whether there is a direct information and/or material flow pathway from one variable to another. Additionally, estimation methods for the transfer 0-entropy and the direct transfer 0-entropy are also provided.

For root cause diagnosis of plant-wide oscillations, comparisons are given between the usefulness of these two information theory-based causality detection methods and other four widely used methods: the Granger causality analysis method, the spectral envelope method, the adjacency matrix method, and the Bayesian network inference method. All six methods are applied to a benchmark industrial data set and a set of guidelines and recommendations on how to deal with the root cause diagnosis problem is discussed.

Acknowledgements

I would like to use this opportunity to extend my sincere thanks to all the individuals without whom this thesis would not have been possible.

First of all, I would like to express my heartiest gratitude to my supervisor Dr. Tongwen Chen for his guidance during my study and research at the University of Alberta. He provided a lot of help to keep me on the right track. I learned a lot from him, including starting from simple numerical examples to more complex industrial processes, how to interpret the results theoretically, and how to dig out more information from simulation results. I am highly obliged to him for the time he spent in leading, supporting, and encouraging me. Dr. Chen's continuous support, encouragement, and help made the momentum of my research going.

I am highly grateful to my co-supervisor Dr. Sirish L. Shah for his guidance and valuable comments for the advancement of the research. Working under his supervision was one of the most positive experiences of my life. Dr. Shah gave me encouragement and support when I was confused. His constant and precious feedback, comments and guidelines played a significant role for this research. I learned from him how to write technical articles, how to revise papers and how to pay attention to details. I would like to thank him for always finding time for me in his busy schedule.

I am deeply thankful to Dr. Fan Yang at Tsinghua University (Beijing, China) for his unbounded support, valuable suggestions, insightful comments, and motivation on the research.

I would like to express my sincere thanks to Dr. Min Wu at Central South University (Changsha, China) for his long-term guidance and help.

My earnest thanks to Dr. Nina F. Thornhill from Imperial College London, UK, for comments and suggestions on direct causality analysis.

I also thank my PhD committee members, Dr. Qing Zhao, Dr. Ezra Kwok, and Dr. Masoud Ardakani for their valuable suggestions to improve the contents and presentation of this thesis.

Thanks are also due to my family; my parents, my husband, my sister, and my brother for bearing with me for all this time. I am particularly thankful to my husband for his encouragement and support at all times.

Last but not the least, I would like to thank all members of our research group for active discussions and sharing their research. My special thanks to Md Shahedul Amin, Yuri Shardt, Vinay Bavdekar, Sridhar Dasani, Salman Ahmed, and Jiadong Wang for their help in preparing this thesis.

Contents

List of Tables

List of Figures

List of Symbols

X, Y, Z	continuous random variables
$\tilde{X}, \tilde{Y}, \tilde{Z}$	quantized X, quantized Y, and quantized Z with quantization bin sizes $\Delta_X, \Delta_Y,$ and Δ_Z, respectively
$T_{X\to Y}$	differential Transfer Entropy ($\mathrm{TE}_{\mathrm{diff}}$) from $X$to Y
$D_{X\to Y}$	differential Direct Transfer Entropy ($\mathrm{DTE}_{\mathrm{diff}}$)from X to Y
$t_{\tilde{X}\to\tilde{Y}}$	discrete Transfer Entropy ($\mathrm{TE}_{\mathrm{disc}}$) from $\tilde{X}$ to $\tilde{Y}$
$d_{\tilde{X}\to\tilde{Y}}$	discrete Direct Transfer Entropy ($\mathrm{DTE}_{\mathrm{disc}}$)from $\tilde{X}$ to $\tilde{Y}$
$NTE_{\tilde{X}\to\tilde{Y}}$	normalized discrete Transfer Entropy ($\mathrm{NTE}_{\mathrm{disc}}$) from $\tilde{X}$ to $\tilde{Y}$
$NTE^{\mathrm{c}}_{X\to Y}$	normalized differential Transfer Entropy ($\mathrm{NTE}_{\mathrm{diff}}$) from X to Y
$NDTE^{\mathrm{c}}_{X\to Y}$	normalized differential direct Transfer Entropy ($\mathrm{NDTE}_{\mathrm{diff}}$) from X to Y
$[\![X]\!]$	marginal range of X
$[\![X, Y]\!]$	joint range of X and Y
$T^0_{X\to Y}$	transfer 0-entropy (T0E) from X to Y
$D^0_{X\to Y}$	direct transfer 0-entropy (T0E) from X to Y

List of Abbreviations

AIC	Akaike Information Criterion
AR	AutoRegressive
BIC	Bayesian Information Criterion
BN	Bayesian Network
DT0E	Direct Transfer 0-Entropy
DTE	Direct Transfer Entropy
DTF	Directed Transfer Functions
FGD	Flue Gas Desulfurization
GCCA	Granger Causal Connectivity Analysis
iAAFT	iterative Amplitude Adjusted Fourier Transform
MIMO	Multiple-Input and Multiple-Output
MLE	Maximum Likelihood Estimation
NLGC	Nonlinear Granger Causality
OCI	Oscillation Contribution Index
P&ID	Piping and Instrumentation Diagram/Drawing
PDC	Partial Directed Coherence
PDF	Probability Density Function
PFD	Process Flow Diagram
PRBS	Pseudo-Random Binary Sequence

PSD	Power Spectra Density
QP	Quadratic Programming
RTE	Rényian Transfer Entropy
RV	Random Variable
SDG	Signed Digraphs
SVM	Support Vector Machine
T0E	Transfer 0-Entropy
TE	Transfer Entropy

Chapter 1

Introduction

1.1 Motivation and Background

With the increase in scale and complexity of process operations in large industrial plants, faults may occur on any of the thousands of components and therefore result in unsatisfactory performance, failures or even hazardous accidents. When a disturbance is generated somewhere in a plant and propagates to the whole plant or some other units of the plant through information and/or material flow pathways, it is termed as a plant-wide disturbance [15]. Plant-wide disturbances are common in many processes because of interactions of units 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. Petrochemical plants on average suffer a major accident every three years that leads to devastating consequences [1]. Moreover, such abnormalities cost billions of dollars annually in industry due to unplanned shutdowns, equipment damages, performance degradations and operation failures. Thus, detection and diagnosis of plant-wide abnormalities and disturbances are major problems in the process industry.

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

In a complex industrial process, elements are not only connected to each

other, they are also mutually dependent. The concept of causality (also referred as causation) has been introduced to describe the cause-effect relationships between variables or events. To describe the causal relationships between all the variables, a network can be constructed with nodes denoting variables and arcs denoting their causal relationships; this network is usually called a causal map [7]. 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 along fault propagation pathways [7, 87].

Intuitively, causality is in our daily life; however, causality has not been accepted as a scientific concept until statisticians formulated a randomized experiment to test causal relations from data [24]. Wiener was one of the first mathematicians to develop a definition of causality between two random variables X and Y: X could be termed as to 'cause' Y if the predictability of Y is improved by incorporating information about X [81]. However, Wiener's idea lacked the machinery for practical implementation.

Granger adapted this definition into experimental practice by proposing a technique for analysis of data observed in consecutive time series. In his Nobel prize lecture [30], he identified two components of the statement about causality: (1) the cause occurs before the effect; and (2) the cause contains information about the effect that is unique, and is in no other variable. These statements intuitively mean that the causal variable can help to forecast the effect variable. It is said that a variable X 'Granger' causes another variable Y if the future of Y can be better predicted using the past information of both X and Y than only using the past information of Y. Granger formalized the prediction idea in the context of linear regression models [29]: X is said to have a causal influence on Y if the variance of the autoregressive (AR) prediction error of Y at the present time is reduced by inclusion of past measurements of X. This formalization has practical utility and thus has been widely used by the name of "Granger causality".

From the definition of causality, we can see that the flow of time or temporal direction is a key point in causality analysis. Therefore, the interaction discovered by causality detection may be unidirectional or bidirectional. In other words, causality is asymmetric: "X causes Y" does not imply "Y causes X" [2]. This directional interaction is the major difference between causal influence and relations reflected by the symmetric measures such as ordinary

coherence and mutual information. Correlation does not imply causality. One can say that X is correlated with Y, which implies that Y is correlated with X. Whereas if X causes Y, we cannot conclude that Y causes X. Inspired by Granger's work, many different kinds of techniques for causality detection have been proposed, especially in the area of neurosciences. One natural question to ask is: how to use and improve these techniques on historical process data to capture the causal relationships among process variables?

Although a causal relationship between two variables, X and Y, can be detected using causality analysis techniques, it is difficult to distinguish whether the causal influence is direct or indirect because it is possible that there is an intermediate variable or some intermediate variables which transfer information from X to Y. A direct causality from X to Y is defined as X directly causes Y, which means there is a direct information and/or material flow pathway from X to Y without any intermediate variables. Thus, we need to discriminate between direct and indirect causality between two variables.

The motivation for detection of direct and indirect causality based on measured process variables is as follows.

The purpose of process causality analysis is to investigate propagation of faults, alarms events, and signals through material and information flow pathways (for example via feedback control) and in this respect it is important to know if connection between variables of interest is direct or indirect. As shown in Fig. 1.1(a), we may conclude that there is causal influence from X to Y by using a certain causality detection method, but we cannot tell whether the causal influence is along a direct pathway or an indirect pathway through the intermediate variable Z. If direct causality from X to Y is detected, then there should be a direct information flow pathway from X to Y. Otherwise, there is no direct information flow pathway from X to Y and the direct link should be eliminated. This is clearly illustrated in the experimental 3-tank case study as presented in Chapters 2 and 3. Such cases are common in industrial processes because of the high degree of interconnections between elements. The traditional causality detection approaches will reveal a myriad of connections as it is not able to discriminate between direct and indirect causality; whereas once one is able to detect direct paths, the number of connecting pathways reduces significantly.

Another case to consider is if there is a common cause of both X and Y, and using some causality analysis techniques we can detect the causal relationships

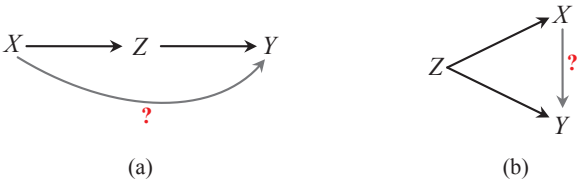


Figure 1.1: Detection of (a) direct or indirect causality and (b) true or spurious causality from X to Y.

between X and Y, for example, X causes Y (see Fig. 1.1(b)). However, it is possible that in fact X is not a cause of Y and the spurious causality is generated by the common source Z, a confounding variable that may affect both X and Y. Thus, in this case we need to further distinguish whether there is true causality between X and Y. In fact this can tell whether there is a direct information flow pathway from X to Y or there is no information flow pathway from X to Y at all. Thus, the detection of true/spurious causality is necessary for capturing the true process connectivity.

From an application point of view, one purpose of causality analysis is to find the fault propagation pathways and diagnose the root cause of certain disturbance or faults. If we only detect causality via the causality detection approach, total causality and spurious causality would be detected to yield an overly complicated set of pathways from which root cause diagnosis of faults and faults propagation pathways investigation would be difficult if not erroneous. However, if we are able to differentiate between direct and indirect, and true and spurious causality, then the derived causal map may be much simpler and more accurate to reveal the fault propagation pathways and which variable is the likely root cause. This point is clearly illustrated by the industrial case studies presented in Chapters 3 and 4. For example, for the benchmark industrial data set, if we only detect the total causality and do not distinguish direct/true and indirect/spurious causality, then the causal map of 8 oscillating variables obtained from the transfer entropy method is shown in Fig. 1.2, where a dashed line with an arrow indicates unidirectional causality and a solid line connecting two variables without an arrow indicates bidirectional causality. Fig. 1.2 shows a complicated set of pathways from which finding faults propagation pathways would be difficult. After detecting direct/true and indirect/spurious causality via the direct transfer entropy approach, the result is a simpler causal map as shown in Fig. 1.3, which correctly indicates

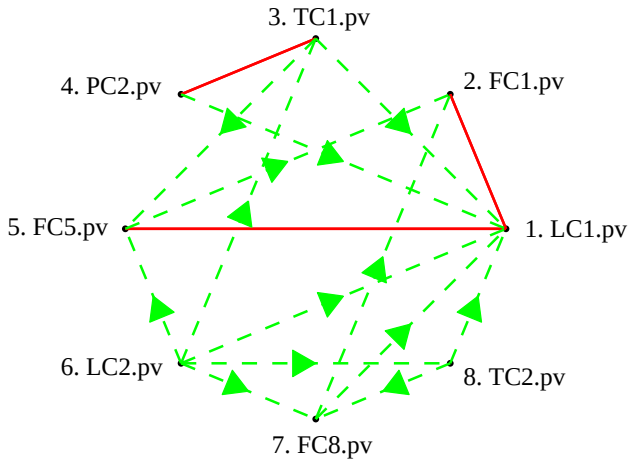


Figure 1.2: Causal map based on calculation results of transfer entropies which represent the total causality including both direct and indirect/spurious causality. A dashed line with an arrow indicates unidirectional causality and a solid line connecting two variables without an arrow indicates bidirectional causality.

direct and true causality. We can see that Fig. 1.3 is much sparser than Fig. 1.2 and it is much simpler to investigate the fault propagation pathways and determine the likely root cause from Fig. 1.3.

In summary, since information flow specifically means how variation propagates from one variable to another [26], the detection of direct/true and indirect/spurious causality is necessary for capturing the true process connectivity and finding faults propagation pathways.

An important application of causality analysis for capturing process connectivity is to find the fault propagation pathways and localize the likely root cause(s) of plant-wide abnormalities and disturbance. Take the plant-wide oscillations as an example, since various methods have already been proposed for diagnosis of plant-wide oscillations, there is no rule to determine which method to use when a plant-wide oscillation occurs. In order to give some suggestions on how to choose an appropriate method and provide some guidelines on how to deal with this common problem, it is necessary to illustrate the usefulness of the causality analysis approaches on a benchmark industrial data

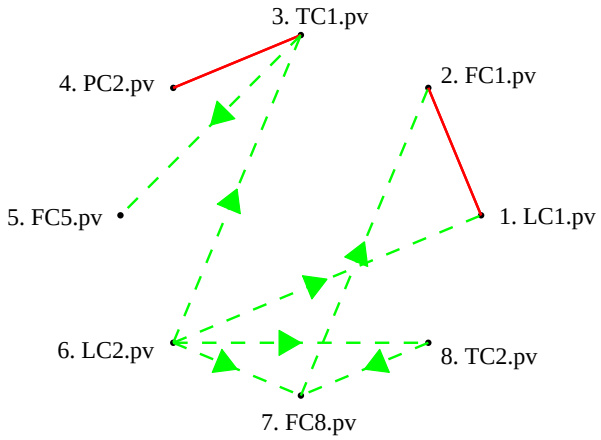


Figure 1.3: Causal map based on calculation results of direct transfer entropies which correctly indicate the direct and true causality. A dashed line with an arrow indicates unidirectional causality and a solid line connecting two variables without an arrow indicates bidirectional causality.

set and compare these approaches with several recently introduced methods for detection and/or root cause diagnosis of plant-wide oscillations.

1.2 Literature Review

The problem of focus is how to investigate systematic causality analysis techniques to capture process connectivity and find the material and information flow pathways of a process. The problem is divided into two parts. The first part is to detect causal relationships among process variables. The second part is to detect whether the causal influence between a pair of variables is along a direct pathway without any intermediate variables or indirect pathways through some intermediate variables. Corresponding to the two parts of the problem, the literature survey is also divided into the following two sections.

1.2.1 Detection of Causal Relationships

A qualitative process model in the form of a digraph has been widely used in root cause and hazard propagation analysis [86]. Digraph–based models usually express the causal relationships between faults and symptoms and

define the fault propagation pathways by incorporating expert knowledge of the process [54]. A drawback is that extracting expert knowledge is very time consuming and that knowledge is not always easily available. The modeling of digraphs can also be based on mathematical equations [49, 50], yet for large scale complex processes it is difficult to establish practical and precise mathematical models.

Data driven methods provide another way to find causal relationships between process variables. A few data-based methods are capable of detecting the causal relationships for linear processes [84]. In the frequency domain, directed transfer functions (DTF) [41] and partial directed coherence (PDC) [5] are widely used in brain connectivity analysis. Based on AR models of the process, Granger causality has its time-domain version and frequency-domain version (called spectral Granger causality) [21]. A MATLAB toolbox for Granger causal connectivity analysis (GCCA) has been developed [62]. The Granger causality method has been successfully used for root cause diagnosis of plant-wide oscillations in an industrial process [92]. Other methods such as path analysis [40] and cross-correlation analysis with lag-adjusted variables [28] are commonly used.

The predictability improvement based on nearest neighbors is proposed as an asymmetrical measure of interdependence in linear or nonlinear bivariate time series and applied to quantify the directional influences among physiological signals [23] and also industrial process variables [9]. The nearest neighbors method has been successfully used for root cause analysis of plant-wide disturbances [8]. Authors of [4] considered an extension of Granger causality to nonlinear bivariate time series and presented a nonlinear Granger causality (NLGC) approach with bivariate time series modeled by a generalization of radial basis functions. The usefulness of the NLGC approach was illustrated by some physiological examples [4].

Information theory provides a wide variety of approaches for measuring causal influence among multivariate time series [33]. Based on transition probabilities containing all information on causality between two variables, the transfer entropy (TE) approach was proposed to distinguish between driving and responding elements [59] and is suitable for both linear and nonlinear relationships; it has been successfully used in chemical processes [7] and neurosciences [79]. TE has two forms: discrete TE ($\mathrm{TE}_{\mathrm{disc}}$) for discrete-valued random variables [59] and differential TE ($\mathrm{TE}_{\mathrm{diff}}$) for continuous-valued ran-

dom variables [55]. Recently, a concept of Rényian transfer entropy (RTE) was proposed in [39] 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.

In [47], comparisons are given for several causality detection methods; these methods include TE, NLGC, and predictability improvement. The paper also includes a discussion on the usefulness of the methods for detecting asymmetric couplings and information flow directions in the deterministic chaotic systems. The authors conclude that, given a complex system with a priori unknown dynamics, the first method of choice might be TE. If a large number of samples are available, the alternative methods might be NLGC and predictability improvement.

It has been shown in [6] that, for Gaussian distributed variables with linear relationships, Granger causality and TE are equivalent. The equivalence of the two causality measures has been extended under certain conditions on probability density distributions of the data [32]. It has been shown that both the Granger causality method and the transfer entropy method are effective tools for causality detection. The Granger causality method is based on AR models of the process, which is suitable for linear multivariate processes. The problem of model misspecification may happen and thus the identified AR models may not be convincing. Compared to the Granger causality method, the TE method is an information-theoretic approach that does not need assumptions on the process model structure. It was proposed based on the concept of Shannon's entropy [63] and is suitable for both linear and nonlinear relationships. Similar to the Shannon's entropy, a key assumption of the TE method is that the sampled data should follow a well-defined probability distribution.

1.2.2 Direct or Indirect Causality Analysis

Since an important application of causality analysis is to capture the process connectivity and find the fault propagation pathways, it is necessary to detect whether the causal influence between a pair of variables is along a direct pathway without any intermediate variables or indirect pathways through some intermediate variables, and whether there is no information flow pathway between them at all and the spurious causality is generated by a common source.

In the frequency domain, a DTF/PDC-based method for quantification of direct and indirect energy flow in a multivariate process was recently proposed [26]. This method was based on vector auto-regressive or vector moving average model representations, which are suitable for linear multivariate processes. In the time domain, a path analysis method was used to calculate the direct effect coefficients [35]. The calculation was based on a regression model of the variables, which captures only linear relationships. In order to detect whether the interaction between two time series is direct or is mediated by other time series and whether the causal influence is simply due to differential time delays in their driving inputs, the bivariate Granger causality has already been generalized to the multivariate case [21]. All possible intermediate variables and confounding variables are included to construct the AR models of each process. Since the causality detection is based on residual analysis of AR models, this method also captures only linear relationships. Both time-domain and frequency-domain formulations of conditional Granger causality and their applications have been discussed in [21].

For both linear and nonlinear relationships, based on a multivariate version of TE, partial TE was proposed to quantify the total amount of indirect coupling mediated by the environment and was successfully used in neurosciences [78]. In [78] partial TE is defined such that all the environmental variables are considered as intermediate variables, which is not necessary in most cases; and in any case, this will increase the computational burden significantly. On the other hand, the utility of partial TE is to detect unidirectional causalities, which is suitable for neurosciences; however, in industrial processes, feedback and bidirectional causalities are common due to recycle streams and cascade control. Thus the partial TE method cannot be directly used for direct/indirect causality detection in the process industry.

1.2.3 Summary

After going through a survey of the existing results, we find that there are a number of useful results for the causality detection problem; however most of the detection methods are proposed and applied to detect causal relationships in the field of neurosciences and economics. It is an interesting topic to improve and apply these methods to detect causality among process variables, which can provide structural information about the process and can be further used to investigate the fault propagation pathways and determine the root cause(s)

of certain abnormality. As to the direct or indirect causality analysis problem, most of the current detection methods are suitable for linear multivariate processes; while for nonlinear relationships, the research on direct or indirect causality detection methods is limited.

It has been shown that the TE 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 all types of industrial process data. Thus, one natural question to ask is: without assuming a probability space, is it possible to construct a useful analogue of the TE for causality detection? Currently, this topic has not been studied yet.

Although comparisons are given for several causality detection methods in [47], the comparisons and discussions are based on applications of these methods to some numerical examples. As for the process industry, an important application of the causality analysis is to find the direction of disturbance propagation and determine the likely root cause of certain plant-wide disturbance. The research on comparisons and discussions of the usefulness of different causality detection methods for root cause and hazard propagation analysis is quite limited.

1.3 Thesis Contributions

The major contributions in this thesis that distinguish it from other work are listed below:

1. Proposed a transfer entropy based methodology to detect and discriminate between direct and indirect causality relationships between process variables of both linear and non-linear multivariate systems. Specifically this method is able to uncover explicit direct and indirect, as if through intermediate variables, connectivity pathways between variables.

2. Proposed a new information theory-based method—transfer 0-entropy (T0E) method—to detect causal relationships between process variables without assuming a probability space.

3. Illustrated the usefulness of the Granger causality method, the TE method, and the T0E method for root cause and hazard propagation analysis by using a benchmark data set.

4. Discussed and compared the causality analysis methods with another three widely used methods for root cause diagnosis of plant-wide oscillations: the spectral envelope method, the adjacency matrix method, and the Bayesian network inference method.

5. Provided guidelines and recommendations on how to choose an appropriate method for root cause diagnosis of plant-wide oscillations.

6. Provided a physical interpretation of the spectral envelope method for detection and diagnosis of plant-wide oscillations.

1.4 Thesis Outline

This thesis has been prepared, in 'paper' format, according to the guidelines from the Faculty of Graduate Studies and Research (FGSR) at the University of Alberta. The rest of the thesis is organized as follows.

In Chapter 2, we describe a direct causality detection approach suitable for both linear and nonlinear connections. Based on an extension of the transfer entropy approach, a direct transfer entropy (DTE) concept is proposed to detect whether there is a direct information flow pathway from one variable to another. Specifically, a differential direct transfer entropy concept is defined for continuous random variables, and a normalization method for the differential direct transfer entropy is presented to determine the connectivity strength of direct causality. The effectiveness of the proposed method is illustrated by several examples, including one experimental case study and one industrial case study.

In Chapter 3, we propose a new information theory-based approach for causality analysis. Without assuming a probability space, a transfer 0-entropy concept is proposed for causality detection on the basis of the definitions of 0-entropy and 0-information. For cases of more than two variables, a direct transfer 0-entropy (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 numerical examples, an experimental case study and an industrial case study.

Chapter 4 is concerned with the applications of the causality analysis methods for root cause and hazard propagation analysis. In this chapter, we com-

pare and discuss the usefulness of three causality analysis methods including the Granger causality method, the TE method, and the T0E method with another three widely used methods for root cause diagnosis of plant-wide oscillations: the spectral envelope method, the adjacency matrix method, and the Bayesian network inference method. All six methods are applied to an industrial benchmark data set and a set of guidelines on how to deal with this common problem is discussed. Moreover, the physical interpretation of the spectral envelope method is discussed. It turns out that for a given frequency, the magnitude of the optimal scaling is proportional to the amplitude of the Fourier transformation of the corresponding time series.

Chapter 5 highlights concluding remarks for this dissertation and presents some future work.

Chapter 2

Direct Causality Detection via the Transfer Entropy Approach*

2.1 Overview

As mentioned in Section 1.1, detection of direct causality, as opposed to indirect causality, is an important and challenging problem in root cause and hazard propagation analysis. Several methods provide effective solutions to this problem when linear relationships between variables are involved. For nonlinear relationships in industrial processes, currently only "overall" causality analysis can be conducted, direct causality cannot be identified.

In this chapter, we describe a direct causality detection approach suitable for both linear and nonlinear connections. An extension of the transfer entropy (TE)—direct transfer entropy (DTE)—is proposed to detect whether the causality between two variables is direct or indirect, and true or spurious. A discrete DTE (DTE_{disc}) and a differential DTE (DTE_{diff}) are defined for discrete-valued and continuous-valued random variables, respectively; and the relationship between them is discussed. Calculation methods and the normalization methods are also presented for the TE_{diff} and the DTE_{diff}. The effectiveness of the proposed method is illustrated by several numerical examples, an experimental case study and an industrial case study.

*A version of this chapter has been published as: P. Duan, F. Yang, T. Chen, and S.L. Shah. Direct causality detection via the transfer entropy approach. *IEEE Transactions on Control Systems Technology*, 21(6):2052–2066, 2013, and a short version has been published as: P. Duan, F. Yang, T. Chen, and S.L. Shah. Detection of direct causality based on process data. In *Proceedings of 2012 American Control Conference*, pages 3522–3527, Montreal, Canada, 2012.

2.2 Detection of Direct Causality

In this section, we apply the $\mathrm{TE}_{\mathrm{diff}}$ for continuous-valued random variables to detect total causality and define a differential DTE ($\mathrm{DTE}_{\mathrm{diff}}$)[1] to detect direct causality. The relationship between $\mathrm{DTE}_{\mathrm{diff}}$ and $\mathrm{DTE}_{\mathrm{disc}}$ is studied. Moreover, calculation methods and the normalization methods are proposed for both the $\mathrm{TE}_{\mathrm{diff}}$ and the $\mathrm{DTE}_{\mathrm{diff}}$.

2.2.1 Direct Transfer Entropy

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

The TE measures the amount of information transferred from one variable X to another variable Y. This extracted transfer information represents the total causal influence from X to Y. It is difficult to distinguish whether this influence is along a direct pathway or indirect pathways through some intermediate variables. In order to detect the direct and indirect pathways of the information transfer, the definition of a DTE is introduced as follows.

Since process variables take on values that vary continuously rather than a finite set of discrete values, we only consider continuous random variables in this thesis.

Given three continuous random variables X, Y, and Z, let them be sampled at time instants i and denoted by $X_i \in [X_{\min}, X_{\max}]$, $Y_i \in [Y_{\min}, Y_{\max}]$, and $Z_i \in [Z_{\min}, Z_{\max}]$ with $i = 1, 2, \ldots, N$, where N is the number of samples. The causal relationships between each pair of these variables can be estimated by calculating transfer entropies [59].

Let Y_{i+h_1} denote the value of Y at time instant $i + h_1$, that is, h_1 steps in the future from i, and h_1 is referred to as the prediction horizon; $\mathbf{Y}_i^{(k_1)} = [Y_i, Y_{i-\tau_1}, \ldots, Y_{i-(k_1-1)\tau_1}]$ and $\mathbf{X}_i^{(l_1)} = [X_i, X_{i-\tau_1}, \ldots, X_{i-(l_1-1)\tau_1}]$ denote embedding vectors with elements from the past values of Y and X, respectively (k_1 is the embedding dimension of Y and l_1 is the embedding dimension of X); τ_1 is the time interval that allows the scaling in time of the embedded vector,

[1] We caution the reader to be aware of the term: $\mathrm{DTE}_{\mathrm{diff}}$ for differential Direct Transfer Entropy and that it is different from the term discrete Direct Transfer Entropy ($\mathrm{DTE}_{\mathrm{disc}}$) as it applies to discrete-valued random variables.

which can be set to be $h_1 = \tau_1$ as a rule of thumb [7]; $f(Y_{i+h_1}, \mathbf{Y}_i^{(k_1)}, \mathbf{X}_i^{(l_1)})$ denotes the joint probability density function (PDF), and $f(\cdot|\cdot)$ denotes the conditional PDF, and thus $f(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)}, \mathbf{X}_i^{(l_1)})$ denotes the conditional PDF of Y_{i+h_1} given $\mathbf{Y}_i^{(k_1)}$ and $\mathbf{X}_i^{(l_1)}$ and $f(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)})$ denotes the conditional PDF of Y_{i+h_1} given $\mathbf{Y}_i^{(k_1)}$. The differential transfer entropy ($\mathrm{TE}_{\mathrm{diff}}$) from X to Y, for continuous variables, is then calculated as follows:

$$T_{X\to Y} = \int f(Y_{i+h_1}, \mathbf{Y}_i^{(k_1)}, \mathbf{X}_i^{(l_1)}) \cdot \log \frac{f(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)}, \mathbf{X}_i^{(l_1)})}{f(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)})} d\mathbf{w}, \tag{2.1}$$

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

Note that the time interval τ_1 in fact determines the sampling rate. A larger τ_1 indicates a lower sampling rate. If τ_1 is too large, i.e., the sampling rate is too low, then the historical data does not contain any information about the current Y even though there is causal relationship. Thus, it is important to determine a proper sampling rate (τ_1). The optimal parameter of τ_1 is a function of the process dynamics. Thus, for determination of τ_1, we need to keep the information of the system dynamics. If the process dynamics are known, then the sampling rate (τ_1) should be set accordingly. If the process dynamics are unknown, small values of τ_1 should give good results.

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

Similarly, the $\mathrm{TE}_{\mathrm{diff}}$ from X to Z is calculated as follows:

$$T_{X\to Z} = \int f(Z_{i+h_2}, \mathbf{Z}_i^{(m_1)}, \mathbf{X}_i^{(l_2)}) \cdot \log \frac{f(Z_{i+h_2}|\mathbf{Z}_i^{(m_1)}, \mathbf{X}_i^{(l_2)})}{f(Z_{i+h_2}|\mathbf{Z}_i^{(m_1)})} d\eta, \tag{2.2}$$

where h_2 is the prediction horizon, $\mathbf{Z}_i^{(m_1)} = [Z_i, Z_{i-\tau_2}, \ldots, Z_{i-(m_1-1)\tau_2}]$ and $\mathbf{X}_i^{(l_2)} = [X_i, X_{i-\tau_2}, \ldots, X_{i-(l_2-1)\tau_2}]$ are embedding vectors with time interval τ_2, and η denotes the random vector $[Z_{i+h_2}, \mathbf{Z}_i^{(m_1)}, \mathbf{X}_i^{(l_2)}]$.

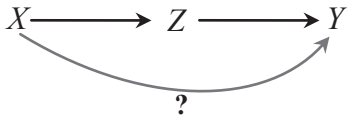


Figure 2.1: Detection of direct causality from X to Y.

The TE_{diff} from Z to Y is calculated as follows:

$$T_{Z\to Y} = \int f(Y_{i+h_3}, \mathbf{Y}_i^{(k_2)}, \mathbf{Z}_i^{(m_2)}) \cdot \log \frac{f(Y_{i+h_3}|\mathbf{Y}_i^{(k_2)}, \mathbf{Z}_i^{(m_2)})}{f(Y_{i+h_3}|\mathbf{Y}_i^{(k_2)})} d\zeta, \tag{2.3}$$

where h_3 is the prediction horizon, $\mathbf{Y}_i^{(k_2)} = [Y_i, Y_{i-\tau_3}, \ldots, Y_{i-(k_2-1)\tau_3}]$ and $\mathbf{Z}_i^{(m_2)} = [Z_i, Z_{i-\tau_3}, \ldots, Z_{i-(m_2-1)\tau_3}]$ are embedding vectors with time interval τ_3, and ζ denotes the random vector $[Y_{i+h_3}, \mathbf{Y}_i^{(k_2)}, \mathbf{Z}_i^{(m_2)}]$.

If $T_{X\to Y}$, $T_{X\to Z}$, and $T_{Z\to Y}$ are all larger than zero, then we conclude that X causes Y, X causes Z, and Z causes Y. We can also conclude that there is an indirect pathway from X to Y via the intermediate variable Z which transfers information from X to Y, as shown in Fig. 2.1. However, we cannot distinguish whether there is a direct pathway from X to Y, because it is possible that there exist both a direct pathway from X to Y and an indirect pathway via the intermediate variable Z. In this case, we need to distinguish whether the causal influence from X to Y is only via the indirect pathway through the intermediate variable Z, or in addition to this, there is another direct pathway from X to Y. We define a direct causality from X to Y as X directly causing Y, which means there is a direct information and/or material flow pathway from X to Y without any intermediate variables.

In order to detect whether there is a direct causality from X to Y, we define a differential direct transfer entropy (DTE_{diff}) from X to Y as follows:

$$\begin{aligned} D_{X\to Y} &= \int f(Y_{i+h}, \mathbf{Y}_i^{(k)}, \mathbf{Z}_{i+h-h_3}^{(m_2)}, \mathbf{X}_{i+h-h_1}^{(l_1)}) \\ &\quad \cdot \log \frac{f(Y_{i+h}|\mathbf{Y}_i^{(k)}, \mathbf{Z}_{i+h-h_3}^{(m_2)}, \mathbf{X}_{i+h-h_1}^{(l_1)})}{f(Y_{i+h}|\mathbf{Y}_i^{(k)}, \mathbf{Z}_{i+h-h_3}^{(m_2)})} d\mathbf{v}, \end{aligned} \tag{2.4}$$

where $\mathbf{v}$ denotes the random vector $[Y_{i+h}, \mathbf{Y}_i^{(k)}, \mathbf{Z}_{i+h-h_3}^{(m_2)}, \mathbf{X}_{i+h-h_1}^{(l_1)}]$; the prediction horizon h is set to be $h = \max(h_1, h_3)$; if $h = h_1$, then $\mathbf{Y}_i^{(k)} = \mathbf{Y}_i^{(k_1)}$, if $h = h_3$, then $\mathbf{Y}_i^{(k)} = \mathbf{Y}_i^{(k_2)}$; the embedding vector $\mathbf{Z}_{i+h-h_3}^{(m_2)} = [Z_{i+h-h_3}, Z_{i+h-h_3-\tau_3}, \ldots, Z_{i+h-h_3-(m_2-1)\tau_3}]$ denotes the past values of Z which can provide useful information for predicting the future Y at time instant $i+h$, where the

embedding dimension m_2 and the time interval τ_3 are determined by (2.3); the embedding vector $\mathbf{X}_{i+h-h_1}^{(l_1)} = [X_{i+h-h_1}, X_{i+h-h_1-\tau_1}, \ldots, X_{i+h-h_1-(l_1-1)\tau_1}]$ denotes the past values of X which can provide useful information to predict the future Y at time instant $i+h$, where the embedding dimension l_1 and the time interval τ_1 are determined by (2.1). Note that the parameters in $\mathrm{DTE}_{\mathrm{diff}}$ are all determined by the calculation of the transfer entropies for consistency.

The $\mathrm{DTE}_{\mathrm{diff}}$ represents the information about a future observation of Y obtained from the simultaneous observation of past values of both X and Z, after discarding the information about the future Y obtained from the past Z alone. This can be understood as follows: if the pathway from Z to Y is cut off, will the history of X still provide some helpful information to predict the future Y? Obviously, if this information is non-zero (greater than zero), then there is a direct pathway from X to Y. Otherwise there is no direct pathway from X to Y, and the causal influence from X to Y is all along the indirect pathway via the intermediate variable Z.

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

After the calculation of $D_{X\to Y}$, if there is direct causality from X to Y, we need to further judge whether the causality from Z to Y is true or spurious, because it is possible that Z is not a cause of Y and the spurious causality from Z to Y is generated by X, i.e., X is the common source of both Z and Y. As shown in Fig. 2.2, there are still two cases of the information flow pathways between X, Y, and Z, and the difference is whether there is true and direct causality from Z to Y.

Thus, $\mathrm{DTE}_{\mathrm{diff}}$ from Z to Y needs to be calculated:

$$\begin{aligned} D_{Z\to Y} &= \int p(Y_{i+h}, \mathbf{Y}_i^{(k)}, \mathbf{X}_{i+h-h_1}^{(l_1)}, \mathbf{Z}_{i+h-h_3}^{(m_2)}) \\ &\quad \cdot \log \frac{p(Y_{i+h}|\mathbf{Y}_i^{(k)}, \mathbf{X}_{i+h-h_1}^{(l_1)}, \mathbf{Z}_{i+h-h_3}^{(m_2)})}{p(Y_{i+h}|\mathbf{Y}_i^{(k)}, \mathbf{X}_{i+h-h_1}^{(l_1)})} d\mathbf{v}, \end{aligned} \tag{2.5}$$

where the parameters are the same as in (2.4). If $d_{Z\to Y} > 0$, then there is true and direct causality from Z to Y, as shown in Fig. 2.2(a). Otherwise, the

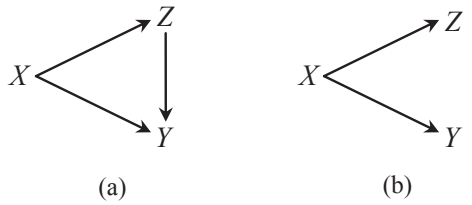


Figure 2.2: Information flow pathways between X, Y, and Z with (a) a true and direct causality from Z to Y and (b) a spurious causality from Z to Y (meaning that Z and Y have a common perturbing source, X, and therefore they may appear to be connected or 'correlated' even when they are not connected physically).

causality from Z to Y is spurious, which is generated by the common source X, as shown in Fig. 2.2(b).

2.2.2 Relationships Between $\mathrm{DTE_{diff}}$ and $\mathrm{DTE_{disc}}$

The $\mathrm{TE_{diff}}$ and the $\mathrm{DTE_{diff}}$ mentioned above are defined for continuous random variables. While for continuous random variables, a widely used TE calculation procedure is to perform quantization first and then use the formula of $\mathrm{TE_{disc}}$ [7]. Thus, we need to establish a connection between this quantization-based procedure and the $\mathrm{TE_{diff}}$ procedure.

For the continuous random variables X, Y, and Z, let $\tilde{X}$, $\tilde{Y}$, and $\tilde{Z}$ denote the quantized X, Y, and Z, respectively. Assume that the supports of X, Y, and Z, i.e., $[X_{\min}, X_{\max}]$, $[Y_{\min}, Y_{\max}]$, and $[Z_{\min}, Z_{\max}]$, are classified into n_X, n_Y, and n_Z non-overlapping intervals (bins), respectively, and the corresponding quantization bin sizes of X, Y, and Z are Δ_X, Δ_Y, and Δ_Z, respectively. Taking X for an example, if we choose a uniform quantizer, then we have

$$\Delta_X = \frac{X_{\max} - X_{\min}}{n_X - 1}.$$

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

After quantization, the TE from X to Y can be approximated by the $\mathrm{TE_{disc}}$ from $\tilde{X}$ to $\tilde{Y}$:

$$t_{\tilde{X}\to\tilde{Y}} = \sum p(\tilde{Y}_{i+h_1}, \tilde{\mathbf{Y}}_i^{(k_1)}, \tilde{\mathbf{X}}_i^{(l_1)}) \cdot \log \frac{p(\tilde{Y}_{i+h_1} | \tilde{\mathbf{Y}}_i^{(k_1)}, \tilde{\mathbf{X}}_i^{(l_1)})}{p(\tilde{Y}_{i+h_1} | \tilde{\mathbf{Y}}_i^{(k_1)})}, \tag{2.6}$$

where the sum symbol represents $k_1 + l_1 + 1$ sums over all amplitude bins of the joint probability distribution and conditional probabilities; $\tilde{\mathbf{Y}}_i^{(k_1)} = [\tilde{Y}_i, \tilde{Y}_{i-\tau_1}, \ldots, \tilde{Y}_{i-(k_1-1)\tau_1}]$ and $\tilde{\mathbf{X}}_i^{(l_1)} = [\tilde{X}_i, \tilde{X}_{i-\tau_1}, \ldots, \tilde{X}_{i-(l_1-1)\tau_1}]$ denote embedding vectors; $p(\tilde{Y}_{i+h_1}, \tilde{\mathbf{Y}}_i^{(k_1)}, \tilde{\mathbf{X}}_i^{(l_1)})$ denotes the joint probability distribution, $p(\cdot|\cdot)$ denotes the conditional probabilities. The meaning of other parameters remains unchanged.

From (2.6) we can express the $\mathrm{TE}_{\mathrm{disc}}$ using conditional Shannon entropies [89] by expanding the logarithm:

$$\begin{aligned} t_{\tilde{X}\to\tilde{Y}} &= \sum p(\tilde{Y}_{i+h_1}, \tilde{\mathbf{Y}}_i^{(k_1)}, \tilde{\mathbf{X}}_i^{(l_1)}) \log \frac{p(\tilde{Y}_{i+h_1}, \tilde{\mathbf{Y}}_i^{(k_1)}, \tilde{\mathbf{X}}_i^{(l_1)})}{p(\tilde{\mathbf{Y}}_i^{(k_1)}, \tilde{\mathbf{Y}}_i^{(l_1)})} \\ &\quad - \sum p(\tilde{Y}_{i+h_1}, \tilde{\mathbf{Y}}_i^{(k_1)}) \log \frac{p(\tilde{Y}_{i+h_1}, \tilde{\mathbf{Y}}_i^{(k_1)})}{p(\tilde{\mathbf{Y}}_i^{(k_1)})} \\ &= H(\tilde{Y}_{i+h_1}|\tilde{\mathbf{Y}}_i^{(k_1)}) - H(\tilde{Y}_{i+h_1}|\tilde{\mathbf{Y}}_i^{(k_1)}, \tilde{\mathbf{X}}_i^{(l_1)}), \end{aligned} \tag{2.7}$$

where

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

and

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

are the conditional Shannon entropies.

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

$$\begin{aligned} T_{X\to Y} &= \int f(Y_{i+h_1}, \mathbf{Y}_i^{(k_1)}, \mathbf{X}_i^{(l_1)}) \log f(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)}, \mathbf{X}_i^{(l_1)}) d\mathbf{w} \\ &\quad - \int f(Y_{i+h_1}, \mathbf{Y}_i^{(k_1)}) \log f(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)}) d\mathbf{u} \\ &= H^c(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)}) - H^c(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)}, \mathbf{X}_i^{(l_1)}), \end{aligned} \tag{2.8}$$

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

Theoretically, as the bin sizes approach zero, the probability $p(\tilde{Y}_{i+h_1}, \tilde{\mathbf{Y}}_i^{(k_1)}, \tilde{\mathbf{X}}_i^{(l_1)})$ in (2.7) is approximated by $\Delta_Y \Delta_Y^{k_1} \Delta_X^{l_1} f(Y_{i+h_1}, \mathbf{Y}_i^{(k_1)}, \mathbf{X}_i^{(l_1)})$. Then we

have

$$
\begin{aligned}
& \lim_{\Delta_X,\Delta_Y\to 0} t_{\tilde{X}\to\tilde{Y}} \\
= & \lim_{\Delta_X,\Delta_Y\to 0}\{\sum \Delta_Y \Delta_Y^{k_1}\Delta_X^{l_1} f(Y_{i+h_1}, \mathbf{Y}_i^{(k_1)}, \mathbf{X}_i^{(l_1)}) \\
& \cdot \log \frac{\Delta_Y \Delta_Y^{k_1}\Delta_X^{l_1} f(Y_{i+h_1}, \mathbf{Y}_i^{(k_1)}, \mathbf{X}_i^{(l_1)})}{\Delta_Y^{k_1}\Delta_X^{l_1} f(\mathbf{Y}_i^{(k_1)}, \mathbf{X}_i^{(l_1)})} \\
& - \sum \Delta_Y \Delta_Y^{k_1} f(Y_{i+h_1}, \mathbf{Y}_i^{(k_1)}) \\
& \cdot \log \frac{\Delta_Y \Delta_Y^{k_1} f(Y_{i+h_1}, \mathbf{Y}_i^{(k_1)})}{\Delta_Y^{k_1} f(\mathbf{Y}_i^{(k_1)})}\} \\
= & \lim_{\Delta_X,\Delta_Y\to 0}\{\sum \Delta_Y \Delta_Y^{k_1}\Delta_X^{l_1} f(Y_{i+h_1}, \mathbf{Y}_i^{(k_1)}, \mathbf{X}_i^{(l_1)}) \\
& \cdot \Big(\log \Delta_Y + \log f(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)}, \mathbf{X}_i^{(l_1)})\Big) \\
& - \sum \Delta_Y \Delta_Y^{k_1} f(Y_{i+h_1}, \mathbf{Y}_i^{(k_1)}) \\
& \cdot \Big(\log \Delta_Y + \log f(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)})\Big)\}. \qquad (2.9)
\end{aligned}
$$

As $\Delta_X, \Delta_Y \to 0$, we have

$$
\begin{aligned}
& \sum \Delta_Y \Delta_Y^{k_1}\Delta_X^{l_1} f(Y_{i+h_1}, \mathbf{Y}_i^{(k_1)}, \mathbf{X}_i^{(l_1)}) \\
& \to \int f(Y_{i+h_1}, \mathbf{Y}_i^{(k_1)}, \mathbf{X}_i^{(l_1)}) d\mathbf{w} = 1,
\end{aligned}
$$

$$
\sum \Delta_Y \Delta_Y^{k_1} f(Y_{i+h_1}, \mathbf{Y}_i^{(k_1)}) \to \int f(Y_{i+h_1}, \mathbf{Y}_i^{(k_1)}) d\mathbf{u} = 1,
$$

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

$$
\begin{aligned}
& \sum \Delta_Y \Delta_Y^{k_1}\Delta_X^{l_1} f(Y_{i+h_1}, \mathbf{Y}_i^{(k_1)}, \mathbf{X}_i^{(l_1)}) \log f(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)}, \mathbf{X}_i^{(l_1)}) \\
& \to \int f(Y_{i+h_1}, \mathbf{Y}_i^{(k_1)}, \mathbf{X}_i^{(l_1)}) \log f(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)}, \mathbf{X}_i^{(l_1)}) d\mathbf{w},
\end{aligned}
$$

$$
\begin{aligned}
& \sum \Delta_Y \Delta_Y^{k_1} f(Y_{i+h_1}, \mathbf{Y}_i^{(k_1)}) \log f(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)}) \\
& \to \int f(Y_{i+h_1}, \mathbf{Y}_i^{(k_1)}) \log f(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)}) d\mathbf{u}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
& \lim_{\Delta_X,\Delta_Y\to 0} t_{\tilde{X}\to\tilde{Y}} \\
= & \lim_{\Delta_Y\to 0} \log \Delta_Y \\
& + \int f(Y_{i+h_1}, \mathbf{Y}_i^{(k_1)}, \mathbf{X}_i^{(l_1)}) \cdot \log f(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)}, \mathbf{X}_i^{(l_1)}) d\mathbf{w} \\
& - \lim_{\Delta_Y\to 0} \log \Delta_Y \\
& - \int f(Y_{i+h_1}, \mathbf{Y}_i^{(k_1)}) \cdot \log f(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)}) d\mathbf{u} \\
= & \int f(Y_{i+h_1}, \mathbf{Y}_i^{(k_1)}, \mathbf{X}_i^{(l_1)}) \cdot \log f(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)}, \mathbf{X}_i^{(l_1)}) d\mathbf{w} \\
& - \int f(Y_{i+h_1}, \mathbf{Y}_i^{(k_1)}) \cdot \log f(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)}) d\mathbf{u} \\
= & \int f(Y_{i+h_1}, \mathbf{Y}_i^{(k_1)}, \mathbf{X}_i^{(l_1)}) \cdot \log \frac{f(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)}, \mathbf{X}_i^{(l_1)})}{f(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)})} d\mathbf{w} \\
= & \; T_{X\to Y}. \qquad (2.10)
\end{aligned}
$$

This means that the differential transfer entropy from X to Y is the same as the discrete transfer entropy from quantized X to quantized Y in the limit as the quantization bin sizes of both X and Y approach zero.

Remark: From (2.9) and (2.10) we can see that the difference between the differential conditional entropy and the limiting value of the Shannon conditional entropy as $\Delta_X, \Delta_Y \to 0$ is an infinite offset: $\lim_{\Delta_Y\to 0} \log \Delta_Y$. Thus, the differential conditional entropy can be negative.

Similar to TE, the DTE from X to Y can be approximated by a discrete direct transfer entropy ($\mathrm{DTE}_{\mathrm{disc}}$) from $\tilde{X}$ to $\tilde{Y}$:

$$
\begin{aligned}
d_{\tilde{X}\to\tilde{Y}} = & \sum p(\tilde{Y}_{i+h}, \tilde{\mathbf{Y}}_i^{(k)}, \tilde{\mathbf{Z}}_{i+h-h_3}^{(m_2)}, \tilde{\mathbf{X}}_{i+h-h_1}^{(l_1)}) \\
& \cdot \log \frac{p(\tilde{Y}_{i+h}|\tilde{\mathbf{Y}}_i^{(k)}, \tilde{\mathbf{Z}}_{i+h-h_3}^{(m_2)}, \tilde{\mathbf{X}}_{i+h-h_1}^{(l_1)})}{p(\tilde{Y}_{i+h}|\tilde{\mathbf{Y}}_i^{(k)}, \tilde{\mathbf{Z}}_{i+h-h_3}^{(m_2)})}, \qquad (2.11)
\end{aligned}
$$

where $\tilde{\mathbf{Y}}_i^{(k)}$, $\tilde{\mathbf{Z}}_{i+h-h_3}^{(m_2)}$, and $\tilde{\mathbf{X}}_{i+h-h_1}^{(l_1)}$ are embedding vectors of $\tilde{Y}$, $\tilde{Z}$, and $\tilde{X}$, respectively. The definitions of other quantities are similar to that in (2.4).

For the $\mathrm{DTE}_{\mathrm{diff}}$ and the $\mathrm{DTE}_{\mathrm{disc}}$, using the same proof procedure with the TE, we can obtain

$$
\lim_{\Delta_X,\Delta_Y,\Delta_Z\to 0} d_{\tilde{X}\to\tilde{Y}} = D_{X\to Y},
$$

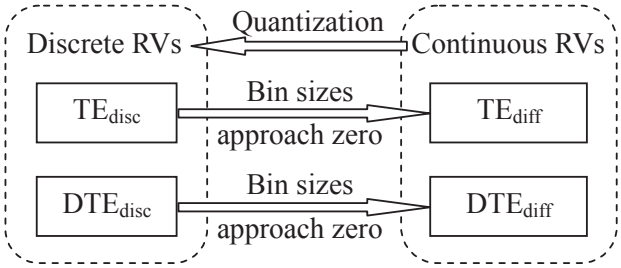


Figure 2.3: Relationships between TEs and DTEs. 'RVs' means random variables

which means that the $\mathrm{DTE}_{\mathrm{diff}}$ from X to Y is the same as the $\mathrm{DTE}_{\mathrm{disc}}$ from quantized X to quantized Y in the limit as the quantization bin sizes of X, Y, and the intermediate variable Z approach zero. Fig. 2.3 illustrates the relationships between $\mathrm{TE}_{\mathrm{diff}}$ and $\mathrm{TE}_{\mathrm{disc}}$, and between $\mathrm{DTE}_{\mathrm{diff}}$ and $\mathrm{DTE}_{\mathrm{disc}}$.

It should be noted that the smaller the bin size, the more accurate the quantization is and the closer the $\mathrm{DTE}_{\mathrm{disc}}$ and the $\mathrm{DTE}_{\mathrm{diff}}$ will be. Note that the computational burden of the summation and the probability estimation in (2.6) and (2.11) will increase significantly with increasing quantization bin numbers, i.e., n_X, n_Y, and n_Z. Thus, for the choice of bin sizes, there is a tradeoff between the quantization accuracy and the computational burden in $\mathrm{TE}_{\mathrm{disc}}$ and $\mathrm{DTE}_{\mathrm{disc}}$ calculations. In practice the conditions that the quantization bin sizes approach zero are difficult to satisfy. Thus, in order to avoid the roundoff error of quantization, we directly use $\mathrm{TE}_{\mathrm{diff}}$ and $\mathrm{DTE}_{\mathrm{diff}}$ to calculate TE and DTE, respectively.

2.2.3 Calculation Method

1) Required Assumptions for the DTE Calculation: Since the concept of DTE is an extension of TE, the required assumptions for DTE is exactly the same as TE: the collected sampled data must be wide-sense stationary with a large data length which is preferred to be no less than 2000 observations [7]. Stationarity requires that the dynamical properties of the system must not change during the observation period. Since in most cases we do not have direct access to the system and we cannot establish evidence that its parameters are indeed constant, we have to test for stationarity based on the available data set.

For the purpose of testing for stationarity, the simplest and most widely used method is to measure the mean and the variance for several segments

of the data set (equivalent to an ergodicity test), and using a standard statistical hypothesis test to check whether the mean and the variance change. More subtle quantities such as spectral components, correlations or nonlinear statistics may be needed to detect less obvious non-stationarity [42]. In this thesis, we use the mean and variance measurement to test for stationarity.

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

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

where the standard deviation divided by an extra $\sqrt{m}$ is the error when estimating the mean value of Gaussian distributed uncorrelated numbers [42]. The null hypothesis for stationarity testing is that the data set is stationary. The significance level for the mean testing is defined as

$$\frac{|\mu_j - \bar{\mu}|}{\sigma} > 6 \text{ for } j = 1, 2, \ldots, m. \tag{2.12}$$

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

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

Multimodality is often encountered in industrial processes due to the normal operational changes as well as changes in the production strategy [91]. For such multimodal processes, a data set with a large number of samples is most likely to be non-stationary as the data would reflect transitions from one mode

to others, whereas a key assumption of the TE/DTE method is stationarity of the sampled data. In order to handle the process multimodality, one would have to partition the data into different segments corresponding to different modes. A few time series analysis methods [20, 43] have been proposed for segmentation of time series to determine when the process mode has changed. As long as the segments corresponding to different modes are obtained, we can detect (direct) causality for each mode of the process using the appropriate segment. Note that the causal relationships may change with mode switching of the process.

2) Estimation of the $\mathrm{TE}_{\mathrm{diff}}$ *and the* $\mathrm{DTE}_{\mathrm{diff}}$*:* For the TE from X to Y, since (2.1) can be written as:

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

it can be approximated by

$$T_{X\to Y} = \frac{1}{N-h_1-r+1}\sum_{i=r}^{N-h_1}\log\frac{f(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)},\mathbf{X}_i^{(l_1)})}{f(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)})}, \tag{2.13}$$

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

Just as with $\mathrm{TE}_{\mathrm{diff}}$, the $\mathrm{DTE}_{\mathrm{diff}}$ (2.4) can be written as:

$$D_{X\to Y} = E\left\{\log\frac{f(Y_{i+h}|\mathbf{Y}_i^{(k)},\mathbf{Z}_{i+h-h_3}^{(m_2)},\mathbf{X}_{i+h-h_1}^{(l_1)})}{f(Y_{i+h}|\mathbf{Y}_i^{(k)},\mathbf{Z}_{i+h-h_3}^{(m_2)})}\right\},$$

which can be approximated by

$$D_{X\to Y} = \frac{1}{N-h-j+1}\sum_{i=j}^{N-h}\log\frac{f(Y_{i+h}|\mathbf{Y}_i^{(k)},\mathbf{Z}_{i+h-h_3}^{(m_2)},\mathbf{X}_{i+h-h_1}^{(l_1)})}{f(Y_{i+h}|\mathbf{Y}_i^{(k)},\mathbf{Z}_{i+h-h_3}^{(m_2)})}, \tag{2.14}$$

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

3) Kernel Estimation of PDFs: In (2.13) and (2.14), the conditional PDFs are expressed by joint PDFs and then obtained by the kernel estimation method [65]. Here the following Gaussian kernel function is used:

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

Then a univariate PDF can be estimated by

$$\hat{f}(x) = \frac{1}{N\gamma}\sum_{i=1}^{N}k\left(\frac{x-X_i}{\gamma}\right), \tag{2.15}$$

where N is the number of samples, and γ is the bandwidth chosen to minimize the mean integrated squared error of the PDF estimation and calculated by $\gamma = 1.06\sigma N^{-1/5}$ according to the "normal reference rule-of-thumb" [46, 65], where σ is the standard deviation of the sampled data $\{X_i\}_{i=1}^N$.

For q dimensional multivariate data, we use the Fukunaga method [65] to estimate the joint PDF. Suppose that $\mathbf{X}_1, \ldots, \mathbf{X}_N$ constitute a q dimensional vector ($\mathbf{X}_i \in \mathbb{R}^q$) with a common PDF $f(x_1, x_2, \cdots, x_q)$. Let $\mathbf{x}$ denote the q dimensional vector $[x_1, x_2, \cdots, x_q]^\intercal$, then the kernel estimation of the joint PDF is

$$\hat{f}(\mathbf{x}) = \frac{(\det \mathbf{S})^{-1/2}}{N\Gamma^q} \sum_{i=1}^{N} K\left\{\Gamma^{-2}(\mathbf{x} - \mathbf{X}_i)^\intercal \mathbf{S}^{-1}(\mathbf{x} - \mathbf{X}_i)\right\}, \tag{2.16}$$

where Γ is similar to the bandwidth γ in (2.15). The estimated joint PDF is smoother when Γ is larger. However, a substantially larger Γ is most likely to result in an inaccurate estimation. Thus, Γ is also chosen to minimize the mean integrated squared error of the joint PDF estimation and calculated by $\Gamma = 1.06N^{-1/(4+q)}$. $\mathbf{S}$ is the covariance matrix of the sampled data, and K is the Gaussian kernel satisfying

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

Note that when $q = 1$, (2.16) is simplified into (2.15).

For the TE, the estimation of the computational complexity is divided into two parts: the kernel estimation of the PDF using (2.16) and the calculation of the $\mathrm{TE}_{\mathrm{diff}}$ using (2.13). For each joint PDF of dimension q, the computational complexity is $O(N^2q^2)$. Considering the conditional PDFs are estimated by the joint PDFs, the maximum dimension of the joint PDF is $k_1 + l_1 + 1$, and thus, the computational complexity for the PDF estimation is $O(N^2(k_1+l_1)^2)$. For calculation of the $\mathrm{TE}_{\mathrm{diff}}$ in (2.13), about N summations are required. Thus, the total computational complexity for the $\mathrm{TE}_{\mathrm{diff}}$ is $O(N^2(k_1 + l_1)^2)$. Similarly, we can obtain that the computational complexity for the $\mathrm{DTE}_{\mathrm{diff}}$ using (2.14) is $O(N^2(k + m_2 + l_1)^2)$. It is obvious that the number of samples and the embedding dimensions determine the computing speed. Since the samples number is preferred to be no less than 2000 observations [7], we need to limit the choice of the embedding dimensions. The details on how to choose the embedding dimensions are given in the following subsection.

Note that the computational complexity is relatively large because of the kernal estimation of the (joint) PDFs, and the computational complexity for the

DTE increases with an increasing number of intermediate variables. Therefore one would have to apply the method to smaller units with smaller number of variables. A large-scale complex system can be broken down into smaller units and thereafter analyzed for causal relationships within each unit, between different units, and finally the information flow pathways of the whole process can be established.

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

Firstly, since $h_1 = \tau_1$ as a rule of thumb [7], we can further set initial values for h_1 and τ_1 according to a priori knowledge of the process. If the process dynamics are unknown, small values of h and τ_1 should give good results [7]; we may start by setting the initial values as $h_1 = \tau_1 = 1$.

Secondly, 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_1, can be determined as the minimum non-negative integer after which there is no significant change on $H^c(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)})$. Considering a large k_1 can increase the dimension of the joint PDF and the difficulty in PDF estimation, if k_1 is greater than 3, we need to increase h_1 and τ_1 and repeat the calculation until a $k_1 \leq 3$ is found.

Finally, we can determine the embedding dimension of X, namely, the window size of the historical X used for the future Y prediction. Based on the values of k_1, h_1, and τ_1, the embedding dimension of X, i.e. l_1, is determined as the minimum positive integer after which the change rate of the transfer entropy from X to Y decreases significantly.

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

In [27], the normalized discrete transfer entropy ($\mathrm{NTE}_{\mathrm{disc}}$) is defined as:

$$NTE_{\tilde{X}\to\tilde{Y}} = \frac{t_{\tilde{X}\to\tilde{Y}} - t^{\text{shuffled}}_{\tilde{X}\to\tilde{Y}}}{H(\tilde{Y}_{i+h_1}|\tilde{\mathbf{Y}}_i^{(k_1)})} \in [0,1], \tag{2.17}$$

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

Equation (2.17) is suitable for the normalization of the TE_{disc}. For TE_{diff}, we cannot just substitute $H(\tilde{Y}_{i+h_1}|\tilde{\mathbf{Y}}_i^{(k_1)})$ with the differential conditional entropy $H^c(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)})$, since $H^c(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)})$ could be negative. Moreover, using shuffled data to eliminate the calculation bias is not accurate, because random shuffling may destroy the statistical properties of the time series. Also, $t^{\text{shuffled}}_{\tilde{X}\to\tilde{Y}}$ is an average of transfer entropies obtained on n trials. To obtain a better result, n should be large enough, which will increase the computational burden significantly. Thus, we need to propose a new normalization method for TE_{diff}.

In (2.17) the zero point is regarded as the origin and it represents a deterministic variable. While for differential entropy, the value $-\infty$ instead of zero means that the variable is deterministic. The maximal differential entropy given a finite support is in the form of a uniform distribution [10]. So, we define the origin as the maximal differential entropy of Y with the uniform distribution:

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

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

Considering that the TE_{diff} is the difference between two differential conditional entropies, as shown in (2.8), we define the normalized differential transfer entropy (NTE_{diff}) as

$$\begin{aligned} NTE^c_{X\to Y} &= \frac{H^c(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)}) - H^c(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)},\mathbf{X}_i^{(l_1)})}{H_0 - H^c(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)},\mathbf{X}_i^{(l_1)})} \\ &= \frac{T_{X\to Y}}{H_0 - H^c(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)},\mathbf{X}_i^{(l_1)})} \\ &\in [0,1]. \end{aligned} \tag{2.18}$$

Intuitively, the numerator term represents the TE to capture the information about Y not explained by its own history and yet explained by the history of X; the denominator term represents the information in Y that is

provided by the past values of both X and Y. It is obvious that $NTE^c_{X\to Y} = 0$ if $T_{X\to Y} = 0$. If Y is uniform distributed and the information about Y explained by the history of both X and Y is completely explained by the history of X, which means $H^c(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)}) = H_0$, then according to (2.18) we obtain $NTE^c_{X\to Y} = 1$.

Since an entropy H represents the average number of bits needed to optimally encode independent draws of a random variable [59], the uncertain information contained in a signal is in fact proportional to 2^H. Here a signal means a specific realization of the random variable. We extend the linear normalization function in (2.18) to a nonlinear function as follows:

$$\begin{aligned} NTE^c_{X\to Y} &= \frac{2^{H^c(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)})} - 2^{H^c(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)},\mathbf{X}_i^{(l_1)})}}{2^{H_0} - 2^{H^c(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)},\mathbf{X}_i^{(l_1)})}} \\ &\in [0,1]. \end{aligned} \tag{2.19}$$

The meaning of (2.19) is the same with that in (2.18). This nonlinear normalization function (2.19) will be used later.

Since the $\mathrm{DTE}_{\mathrm{diff}}$ in (2.4) represents the information directly provided from the past X to the future Y, a normalized differential direct transfer entropy ($\mathrm{NDTE}_{\mathrm{diff}}$) is defined as

$$\begin{aligned} & NDTE^c_{X\to Y} \\ = \; & \frac{D_{X\to Y}}{H^c(Y_{i+h}|\mathbf{Y}_i^{(k)}) - H^c(Y_{i+h}|\mathbf{Y}_i^{(k)}, \mathbf{Z}_{i+h-h_3}^{(m_2)}, \mathbf{X}_{i+h-h_1}^{(l_1)})} \\ \in \; & [0,1), \end{aligned} \tag{2.20}$$

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

2.2.4 Extension to Multiple Intermediate Variables

The definition of the $\mathrm{DTE}_{\mathrm{diff}}$ from X to Y can be easily extended to multiple intermediate variables $Z_1, Z_2, \ldots, Z_q$:

$$\begin{aligned} D_{X\to Y} = & \int f(Y_{i+h}, \mathbf{Y}_i^{(k)}, \mathbf{Z}_{1,i_1}^{(s_1)}, \ldots, \mathbf{Z}_{q,i_q}^{(s_q)}, \mathbf{X}_{i+h-h_1}^{(l_1)}) \\ & \cdot \log \frac{f(Y_{i+h}|\mathbf{Y}_i^{(k)}, \mathbf{Z}_{1,i_1}^{(s_1)}, \ldots, \mathbf{Z}_{q,i_q}^{(s_q)}, \mathbf{X}_{i+h-h_1}^{(l_1)})}{f(Y_{i+h}|\mathbf{Y}_i^{(k)}, \mathbf{Z}_{1,i_1}^{(s_1)}, \ldots, \mathbf{Z}_{q,i_q}^{(s_q)})} d\xi, \end{aligned} \tag{2.21}$$

where $s_1, \ldots, s_q$ and $i_1, \ldots, i_q$ are the corresponding parameters determined by the calculations of the transfer entropies from $Z_1, \ldots, Z_q$ to Y, and ξ denotes the random vector $[Y_{i+h}, \mathbf{Y}_i^{(k)}, \mathbf{Z}_{1,i_1}^{(s_1)}, \ldots, \mathbf{Z}_{q,i_q}^{(s_q)}, \mathbf{X}_{i+h-h_1}^{(l_1)}]$. If $d_{X\to Y}$ is zero, then there is no direct causality from X to Y, and the causal effects from X to Y are all along the indirect pathways via the intermediate variables $Z_1, Z_2, \ldots, Z_q$. If $d_{X\to Y}$ is larger than zero, then there is direct causality from X to Y.

2.3 Examples

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

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

$$\begin{cases} Z_{k+1} = 0.8X_k + 0.2Z_k + v_{1k} \\ Y_{k+1} = 0.6Z_k + v_{2k}. \end{cases}$$

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

To calculate the transfer entropies between X, Z, and Y, we need to determine the four design parameters. We take the transfer entropy from X to Y in (2.1) as an example. First, we set initial values for h_1 and τ_1 as $h_1 = \tau_1 = 1$; Second, we calculate $H^c(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)})$ with $k_1 = 0, 1, \ldots, 10$, as shown in the upper part of Fig. 2.4. The change rate of $H^c(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)})$ with $k_1 = 0, 1, \ldots, 10$ is shown in the lower part of Fig. 2.4, we can see that as k_1 increases, the change rate of $H^c(Y_{i+h_1}|\mathbf{Y}_i^{(k_1)})$ does not vary sharply, which means that the history of Y does not provide useful information for the future values of Y. Thus, we choose $k_1 = 0$. Finally, we calculate the transfer entropy $T_{X\to Y}$ and its change rate with $l_1 = 1, \ldots, 10$, as shown in Fig. 2.5. Since the change rate of $T_{X\to Y}$ decreases significantly after $l_1 = 2$, as shown in the lower part of Fig. 2.5, we choose $l_1 = 2$. Using the same procedure, the parameters for each pair of X, Z, and Y are determined as $h_1 = h_2 = h_3 = 1$, $\tau_1 = \tau_2 = \tau_3 = 1$, $k_1 = m_1 = k_2 = 0$, $l_1 = 2$, and $l_2 = m_2 = 1$. For the following example and case studies, the same procedure is used.

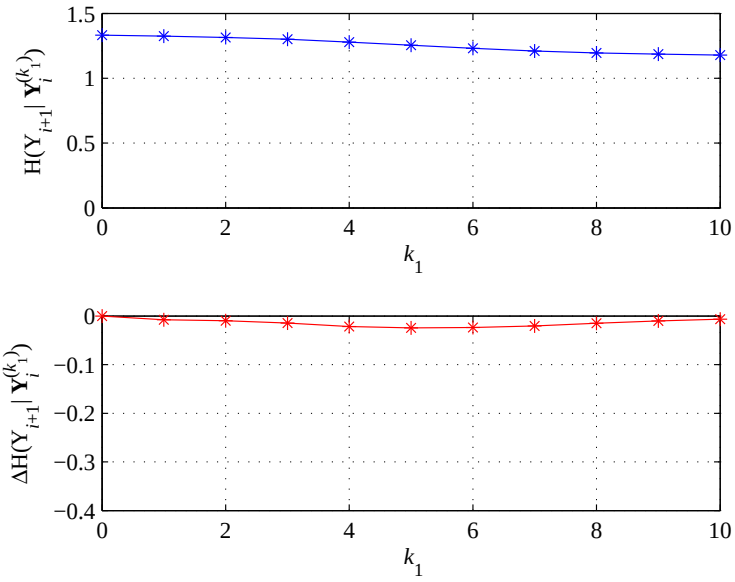


Figure 2.4: Finding the embedding dimension of Y for Example 1.

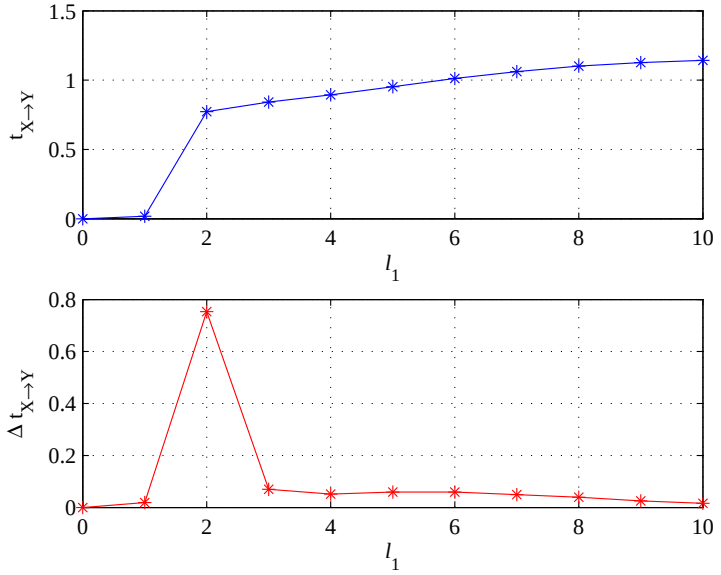


Figure 2.5: Finding the embedding dimension of X for $T_{X\to Y}$ of Example 1.

Table 2.1: Calculated transfer entropies for Example 1.

$T_{\texttt{column 1}\to\texttt{row 1}}$	X	Z	Y
X	NA	1.556	0.772
Z	0.071	NA	1.008
Y	0.067	0.065	NA

Table 2.2: Normalized transfer entropies for Example 1.

$NTE^c_{\texttt{column 1}\to\texttt{row 1}}$	X	Z	Y
X	NA	0.409	0.348
Z	0.058	NA	0.393
Y	0.055	0.044	NA

After the parameters are determined, according to (2.1) and (2.19), the calculated transfer entropies and normalized transfer entropies between each pair of X, Z, and Y are shown in Tables 2.1 and 2.2, respectively. Note that the variables listed in column one are the cause variables and the corresponding effect variables appear in the first row.

From the normalized transfer entropies in Table 2.2, We can see that X causes Z, Z causes Y, and X causes Y because $NTE^c_{X\to Z} = 0.409$, $NTE^c_{Z\to Y} = 0.393$, and $NTE^c_{X\to Y} = 0.348$ are relatively large. Thus we need to first determine whether there is direct causality from X to Y. According to (2.4), we obtain $D_{X\to Y} = 0.016$. According to (2.20), the normalized DTE from X to Y is $NDTE^c_{X\to Y} = 0.016$, which is very small. Thus, we conclude that there is almost no direct causality from X to Y. The information flow pathways for Example 1 are shown in Fig. 2.6(a).

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

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

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

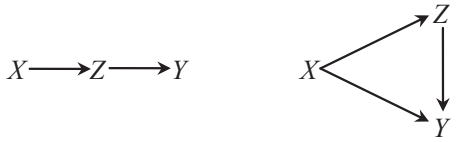


Figure 2.6: Information flow pathways for (a) Example 1 and (b) Example 2.

Table 2.3: Calculated transfer entropies for Example 2.

$T_{\mathtt{column}\,1\to\mathtt{row}\,1}$	X	Z	Y
X	NA	1.142	0.691
Z	0	NA	0.856
Y	0	0.025	NA

Table 2.4: Normalized transfer entropies for Example 2.

$NTE^c_{\mathtt{column}\,1\to\mathtt{row}\,1}$	X	Z	Y
X	NA	0.623	0.274
Z	0	NA	0.308
Y	0	0.048	NA

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

The calculated transfer entropies and normalized transfer entropies between each pair of X, Z, and Y are shown in Tables 2.3 and 2.4, respectively.

From the normalized transfer entropies in Table 2.4, we can see that X causes Z, Z causes Y, and X causes Y because $NTE^c_{X\to Z} = 0.623$, $NTE^c_{Z\to Y} = 0.308$, and $NTE^c_{X\to Y} = 0.274$ are relatively large.

Thus, we need to first determine whether there is direct causality from X to Y. According to (2.4), we obtain $D_{X\to Y} = 0.373$. According to (2.20), the normalized DTE from X to Y is $NDTE^c_{X\to Y} = 0.304$, which is much larger than zero. Thus, we conclude that there is direct causality from X to Y. Second, we need to detect whether there is true and direct causality from Z to Y. According to (2.5), we obtain $D_{Z\to Y} = 0.538$, and thus the normalized

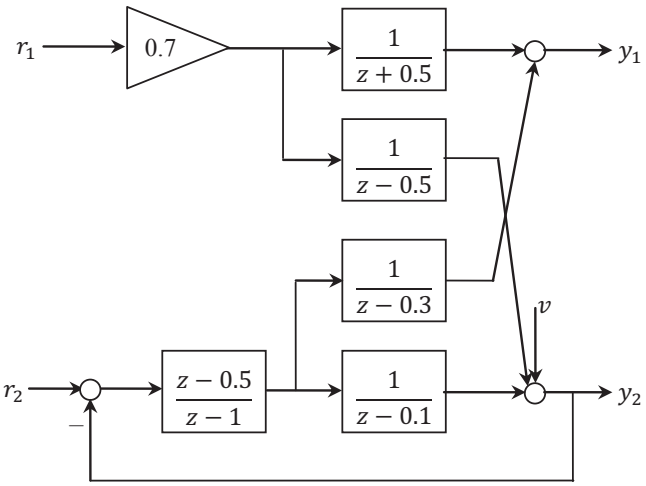


Figure 2.7: System block diagram for Example 3.

Table 2.5: Calculated transfer entropies for Example 3.

$T_{\texttt{column 1}\to\texttt{row 1}}$	r_1	r_2	y_1	y_2
r_1	NA	0.017	0.697	0.485
r_2	0.018	NA	0.562	0.831
y_1	0.020	0.020	NA	0.084
y_2	0.019	0.020	0.455	NA

DTE from Z to Y is $NDTE^{c}_{Z\to Y} = 0.438$, which is much larger than zero. Hence, we conclude that there is true and direct causality from Z to Y. The information flow pathways for Example 2 are shown in Fig. 2.6(b).

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

Example 3: Fig. 2.7 shows a block diagram of a MIMO system with two inputs r_1 and r_2, and two outputs y_1 and y_2. Assume that $r_1 \sim N(0,1)$ and $r_2 \sim N(0,1)$ are independent, and $v \sim N(0, 0.1)$ is the sensor noise. The simulation data consists of 6000 samples. To assure stationarity, the initial 3000 data points were discarded.

The calculated transfer entropies and normalized transfer entropies between each pair of r_1, r_2, y_1, and y_2 are shown in Tables 2.5 and 2.6, respectively.

Table 2.6: Normalized transfer entropies for Example 3.

$NTE^c_{\texttt{column 1}\to\texttt{row 1}}$	r_1	r_2	y_1	y_2
r_1	NA	0.014	0.242	0.187
r_2	0.016	NA	0.212	0.259
y_1	0.018	0.016	NA	0.043
y_2	0.017	0.016	0.184	NA

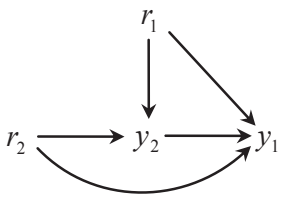


Figure 2.8: Information flow pathways for Example 3.

From the normalized transfer entropies in Table 2.6, we can see that r_1 causes y_1 and y_2, r_2 also causes y_1 and y_2, and y_2 causes y_1. The corresponding information flow pathways are shown in Fig. 2.8.

As shown in Fig. 2.8, since y_1 and y_2 have common sources r_1 and r_2, we need to first detect whether the causality from y_2 to y_1 is true or spurious. According to (2.21), we obtain that the DTE from y_2 to y_1 with intermediate variables r_1 and r_2 is $D_{y_2\to y_1} = 0.474$. According to (2.20), the normalized DTE from y_2 to y_1 is $NDTE^c_{y_2\to y_1} = 0.366$, which is much larger than zero. Hence, we conclude that there is true and direct causality from y_2 to y_1.

Second, since r_1 causes y_2, y_2 causes y_1, and r_1 causes y_1, we need to further detect whether there is direct causality from r_1 to y_1. According to (2.4), we obtain that the DTE from r_1 to y_1 with the intermediate variable y_2 is $D_{r_1\to y_1} = 0.610$. According to (2.20), the normalized DTE from r_1 to y_1 is $NDTE^c_{r_1\to y_1} = 0.573$, which is much larger than zero. Thus, we conclude that there is direct causality from r_1 to y_1 in addition to the indirect causality through intermediate variable y_2. Similarly, we obtain that the DTE from r_2 to y_1 with the intermediate variable y_2 is $D_{r_2\to y_1} = 0.732$ and the normalized DTE from r_2 to y_1 is $NDTE^c_{r_2\to y_1} = 0.617$, which is also much larger than zero. Thus, we conclude that there is direct causality from r_2 to y_1. The information flow pathways are the same as those obtained from the results of

calculated TEs, as shown in Fig. 2.8.

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

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

2.4 Case Studies

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

Experimental case study: In order to show the effectiveness of the proposed methods, a 3-tank experiment was conducted. The schematic of the 3-tank system is shown in Fig. 2.9. Water is drawn from a reservoir and pumped to tanks 1 and 2 by a gear pump and a three way valve. The water in tank 2 can flow down into tank 3. The water in tanks 1 and 3 eventually flows down into the reservoir. The experiment is conducted under open-loop conditions.

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, the normal flow rate of the water out of the pump is 10 L/min. However, the flow rate varies randomly with a mean value of 10 L/min because of the noise in the sensor and minor fluctuations in the pump. The sampled data of 3000 observations are analyzed. Fig. 2.10 shows the normalized time trends of the measurements. The sampling time is one second.

In order to detect the causality and direct causality using TE and DTE, we need to first test the stationarity of the data set. Taking x_1 as an example, we divide the 3000 data points into 10 consecutive segments, each containing 300 data points. The threshold of the mean values for each segment is determined by (2.12). As shown in Fig. 2.11(a), the solid line shows the mean for each segment and the dashed line represents the threshold. Since all the mean values are within the threshold, we may conclude that the data is stationary according to the mean testing. Next, we test the properties of the variance. Here we choose $\alpha = 0.001$; thus, the threshold is $\chi^2_{300}(0.001) = 381.43$ with

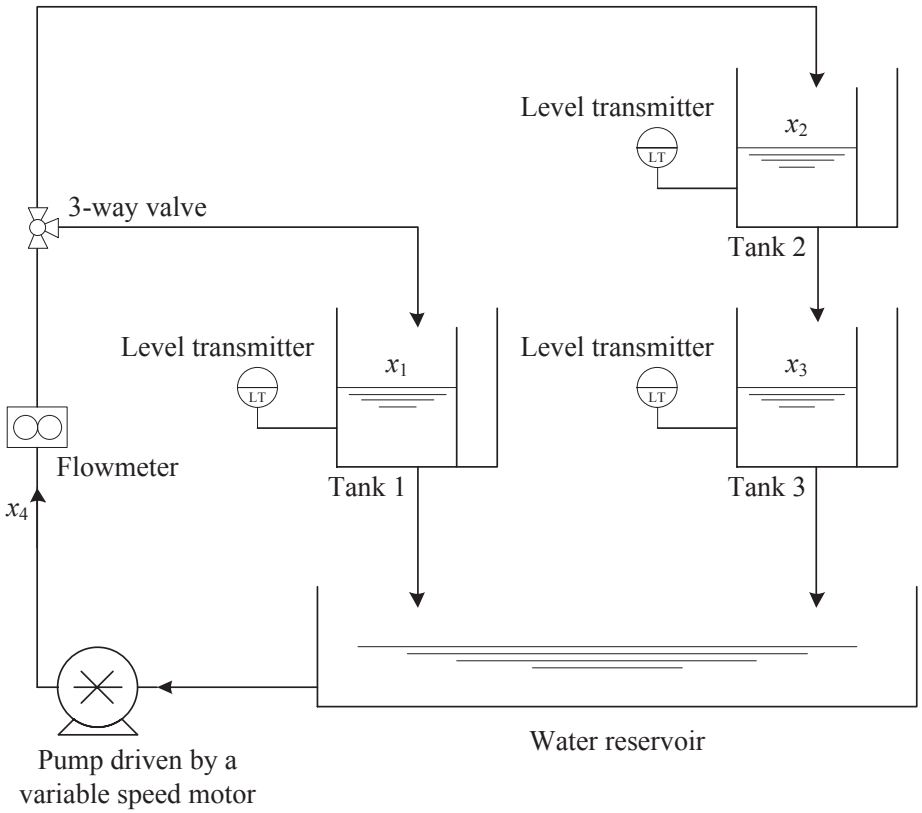


Figure 2.9: Schematic of the 3-tank system.

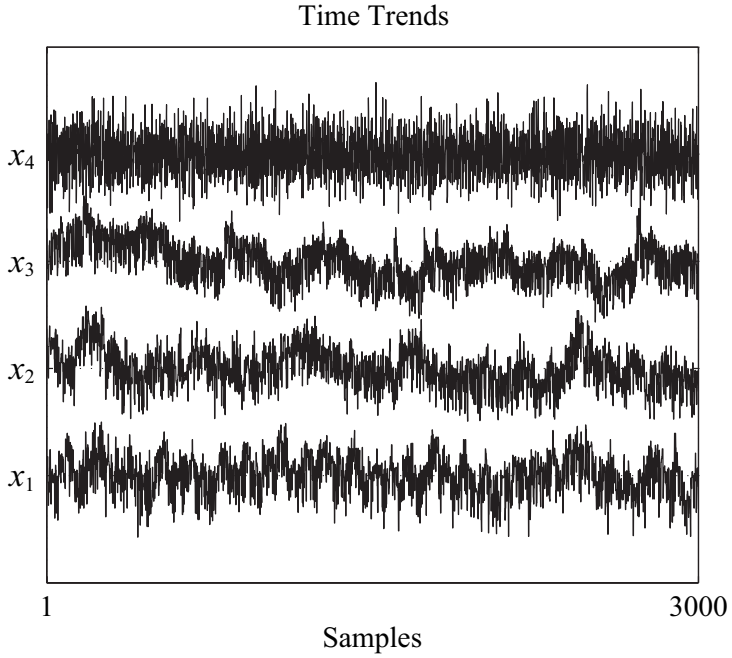


Figure 2.10: Time trends of measurements of the 3-tank system.

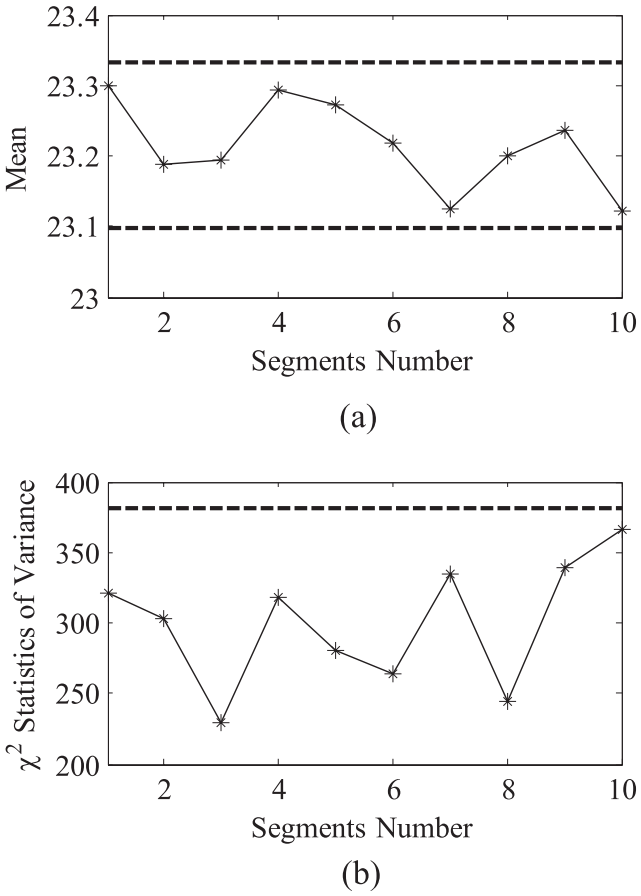


Figure 2.11: Testing for stationarity: (a) mean testing (b) variance testing. The dashed lines indicate the threshold.

a 99.9% confidence. The χ^2 statistics of the variance for each segment is shown in Fig. 2.11(b), where the solid line shows the sum of squares of the elements for each segment after normalization and the dashed line represents the threshold. We can see that all the variance values are smaller than the threshold, and therefore we conclude that the data set of x_1 is stationary. Using the same procedure, the stationary properties of other variables are tested. For the following industrial case study, the same procedure is used.

The calculated transfer entropies and normalized transfer entropies between each pair of x_1, x_2, x_3, and x_4 are shown in Tables 2.7 and 2.8, respectively.

From the normalized transfer entropies in Table 2.8, we can see that x_2 causes x_3, and x_4 causes x_1, x_2, and x_3. The corresponding information flow pathways are shown in Fig. 2.12(a). As shown in Fig. 2.12(b), since x_4 causes x_2, x_2 causes x_3, and x_4 causes x_3, we need to first detect whether there is

Table 2.7: Calculated transfer entropies for the 3-tank system.

$T_{\texttt{column 1}\to\texttt{row 1}}$	x_1	x_2	x_3	x_4
x_1	NA	0.016	0.007	0
x_2	0.006	NA	0.198	0
x_3	0.008	0.005	NA	0
x_4	0.126	0.144	0.139	NA

Table 2.8: Normalized transfer entropies for the 3-tank system.

$NTE^c_{\texttt{column 1}\to\texttt{row 1}}$	x_1	x_2	x_3	x_4
x_1	NA	0.024	0.010	0
x_2	0.012	NA	0.200	0
x_3	0.017	0.007	NA	0
x_4	0.199	0.171	0.152	NA

direct causality from x_4 to x_3.

According to (2.4), we obtain $D_{x_4\to x_3} = 0.006$. According to (2.20), the normalized DTE from x_4 to x_3 is $NDTE^c_{x_4\to x_3} = 0.030$, which is very small. Thus, we conclude that there is almost no direct causality from x_4 to x_3. The corresponding information flow pathways according to these calculation results are shown in Fig. 2.13, which are consistent with the information and material flow pathways of the physical 3-tank system (see Fig. 2.9).

Industrial case study: Another case study is a part of a flue gas desulfu-

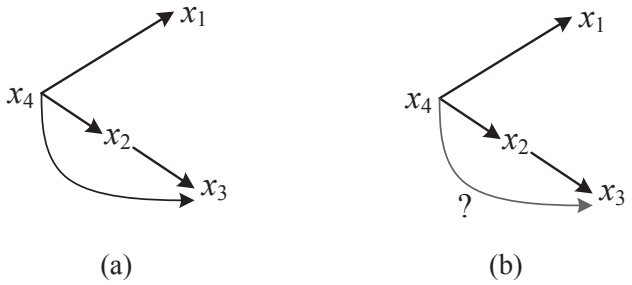


Figure 2.12: Information flow pathways for 3-tank system based on calculation results of normalized transfer entropies.

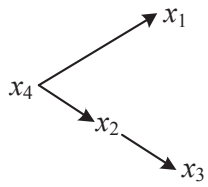


Figure 2.13: Information flow pathways for 3-tank system based on calculation results of normalized DTE.

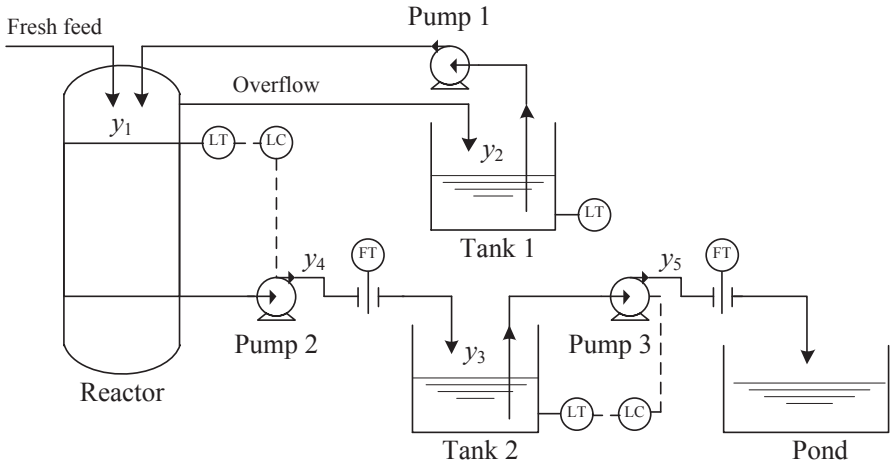


Figure 2.14: Schematic of part of the FGD process.

rization (FGD) process at an oil company in Alberta. The schematic of this part of the process is shown in Fig. 2.14, including a reactor, two tanks, and a pond. Tank 1 receives the overflow from the reactor if the reactor overflows. The liquid in Tank 1 is drawn into the reactor by Pump 1; the liquid in the reactor is drawn into Tank 2 by Pump 2, and the liquid level of the reactor is controlled by adjusting the flow rate of the liquid out of Pump 2; the liquid in Tank 2 is drawn into the pond by Pump 3, and the liquid level of Tank 2 is controlled by adjusting the flow rate of the liquid out of Pump 3. These two level control loops imply that there is bidirectional relationship between the levels and the flow out of the tank due to material as well as information (due to feedback) flow pathways.

We denote the liquid levels of the reactor, Tanks 1 and 2 by y_1, y_2, and y_3, respectively. There is no measurement of the flow rate of the liquid out of Pump 1. We denote the flow rates of the liquid out of Pumps 2 and 3 by y_4 and y_5, respectively. The sampled data of 3544 observations are analyzed. Fig. 2.15 shows the normalized time trends of the measurements. The sampling

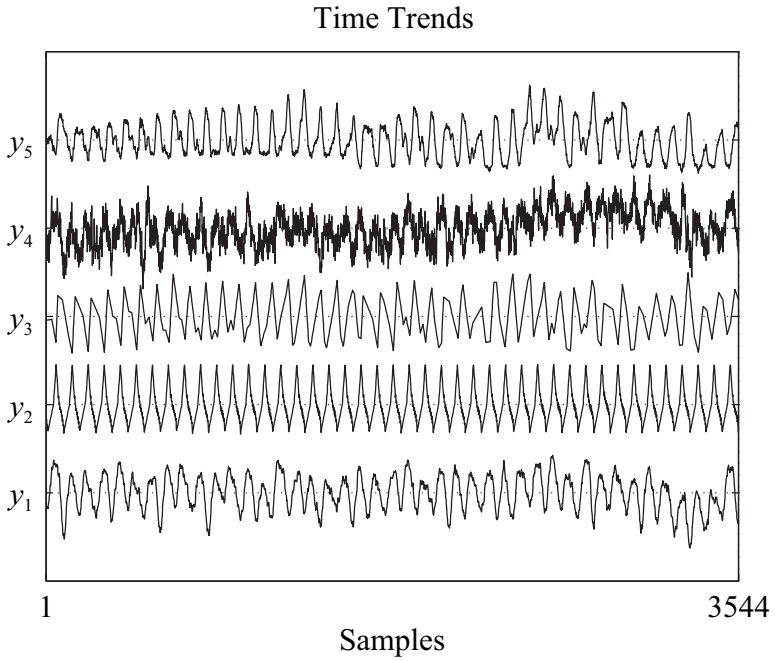


Figure 2.15: Time trends of measurements of the FGD process.

time is one minute.

The calculated transfer entropies and normalized transfer entropies between each pair of y_1, y_2, y_3, y_4 and y_5 are shown in Tables 2.9 and 2.10, respectively.

From the normalized transfer entropies in Table 2.10, we can choose the threshold as 0.05: if the normalized transfer entropy is less than or equal to 0.05, then there is almost no causality. The information flow pathways based on the normalized transfer entropies are shown in Fig. 2.16, we need to further determine whether the causality between y_1, y_2, y_3, y_4, and y_5 is true

Table 2.9: Calculated transfer entropies for part of the FGD process.

$T_{\texttt{column 1}\rightarrow\texttt{row 1}}$	y_1	y_2	y_3	y_4	y_5
y_1	NA	0.022	0.341	0.228	0.043
y_2	0.498	NA	0.488	0.190	0.032
y_3	0.256	0.087	NA	0.008	0.399
y_4	0.479	0.083	0.373	NA	0.057
y_5	0.046	0.011	0.564	0.012	NA

Table 2.10: Normalized transfer entropies for part of the FGD process.

$NTE^{c}_{\texttt{column}\,1\to\texttt{row}\,1}$	y_1	y_2	y_3	y_4	y_5
y_1	NA	0.001	0.089	0.177	0.014
y_2	0.131	NA	0.117	0.154	0.010
y_3	0.078	0.005	NA	0.008	0.105
y_4	0.128	0.005	0.095	NA	0.019
y_5	0.016	0.001	0.130	0.012	NA

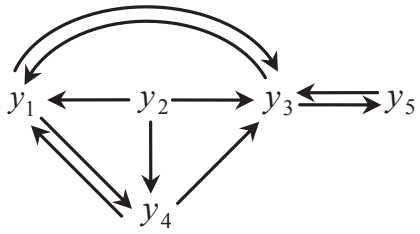


Figure 2.16: Information flow pathways for part of the FGD process based on calculation results of normalized transfer entropies.

and direct.

Calculation steps of direct transfer entropies and corresponding simplified information flow pathways are shown in Fig. 2.17. We first determine whether the causality between y_1 and y_3 is true and direct by considering y_2 and y_4 as the possible intermediate variables (Steps 1 and 2). The calculation results of DTE and normalized DTE are shown in Table 2.11. Since the normalized DTEs between y_1 and y_3 are very small, we conclude that there is almost no direct causality between them. Secondly, we determine whether the causality from y_2 to y_1 is direct by considering the possible intermediate variable y_4 (Step 3). Similarly, the causality from y_2 to y_4 can be determined by considering the possible intermediate variable y_1 (Step 4). Finally, we detect the direct causality from y_2 to y_3 with the possible intermediate variables y_1 and y_4 (Step 5). Based on the calculation results shown in Table 2.11, we conclude that except for the causality from y_2 to y_1, the other detected causality is indirect or spurious. Note that here we do not need to further detect the direct causality between y_1 and y_4, from y_4 to y_3, and from y_5 to y_3 since there is no possible intermediate variable in their pathways. The information flow pathways based

Table 2.11: Calculated and normalized DTEs for part of the FGD process.

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

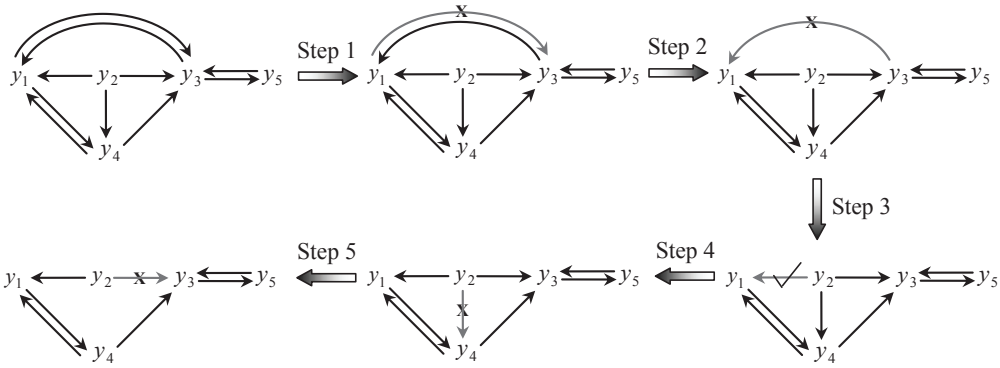


Figure 2.17: Calculation steps of direct transfer entropies.

on calculated direct transfer entropies are shown in Fig. 2.18.

An overview of causality between process variables is shown in Fig. 2.19. Causal relationships from variables on the vertical axis to variables on the horizontal axis are represented by three different symbols: '.' means no causality; '▲' means direct causality; '△' means causality can be detected, but it is indirect or spurious.

From Fig. 2.18, we can see that the spurious causality between the liquid

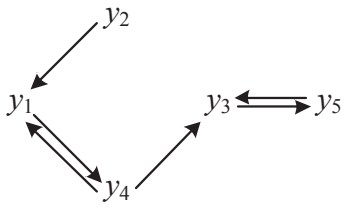


Figure 2.18: Information flow pathways for part of the FGD process based on calculation results of normalized DTE.

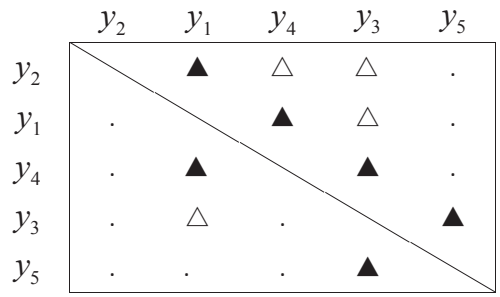


Figure 2.19: An overview of causal relationships between FGD process variables. '.' means no causality; '▲' means direct causality; '△' means causality can be detected, but it is indirect or spurious.

levels of the reactor and Tank 2, namely, between y_1 and y_3, is generated by the flow rate of the liquid out of Pump 2, namely, y_4. Similarly, if we can obtain the measurement of the flow rate of the liquid out of Pump 1, the causality from the liquid level of Tank 1 to the liquid level of the reactor, namely, from y_2 to y_1, will also disappear. However, since the flow rate of the liquid out of Pump 1 is not measured, we still say that there is direct and true causality from y_2 to y_1. Thus, the connecting pathways shown in Fig. 2.18 are consistent with the information and material flow pathways of the physical process shown in Fig. 2.14, where solid lines indicate material flow pathways and dashed lines denote control loops. Note that as mentioned earlier, the bidirectional causality between y_1 and y_4, and between y_3 and y_5 are because of the level feedback control loops.

2.5 Summary

In industrial processes, abnormalities often spread from one process variable to neighboring variables. It is important to find the fault propagation pathways to determine the likely root cause of the abnormalities. Transfer entropy can measure the causality between two process variables, i.e., the direction of the information flow. Furthermore, it is valuable to detect whether the causal influence is along direct or indirect pathways for root cause and hazard analysis. In this chapter, we proposed a direct causality detection method based on the DTE to detect whether there is a direct information and/or material flow pathway between a pair of variables. The DTE_{diff} for continuous random variables has been defined based on an extension of the transfer entropy, which is suitable for both linear and nonlinear relationships. The TE_{diff} and the

$\mathrm{DTE}_{\mathrm{diff}}$ have been respectively shown to be equivalent to the $\mathrm{TE}_{\mathrm{disc}}$ and the $\mathrm{DTE}_{\mathrm{disc}}$ in the limit as the quantization bin sizes approach zero. The $\mathrm{NTE}_{\mathrm{diff}}$ and the $\mathrm{NDTE}_{\mathrm{diff}}$ have been defined to measure the connectivity strength of causality and direct causality, respectively. The proposed methods have been validated by three examples and two case studies.

Chapter 3

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

3.1 Overview

In the previous chapter, we described a TE based methodology to detect direct causality relationships between a pair of process variables of either linear or non-linear multivariate systems. TE was proposed based on the key concept of Shannon's entropy which is defined stochastically as the averaged number of bits needed to optimally encode a source data set X with the source probability distribution $P(X)$ [63]. 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 $\{x_1, \ldots, x_n\}$, Shannon entropy is defined as [63]

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

where $p(x_i)$ denotes the probability mass function of the outcome x_i, the base of the logarithm is 2, and the unit is in bits. For example, a single toss of a fair coin has an entropy of one bit. A series of two fair coin tosses has an entropy of two bits. The number of fair coin tosses is its entropy in bits. The

*A version of this chapter has been submitted for publication as: P. Duan, F. Yang, S.L. Shah, and T. Chen. Transfer zero-entropy and its application for capturing cause and effect relationship between variables. *IEEE Transactions on Control Systems Technology*, 2013.

entropy rate for a fair coin toss is one 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 [82].

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 [53]. Consequently, in process control, disturbances and uncertainties are sometimes treated as bounded unknowns or signals without a priori statistical structure.

One 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 [31] for discrete variables, and Rényi differential 0th-order entropy or Rényi differential 0-entropy h_0 [57] for continuous variables provide an answer 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). Another 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.

In this chapter, we propose a new information theory method to detect causal relationships between a pair of variables without assuming a probability space. 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 transfer 0-entropy (DT0E) to detect and discriminate between direct and indirect causal relationships, respectively. The T0E approach is only related to the ranges of random variables instead of the PDFs. A calculation method for T0E is proposed and the range estimation method for random variables is addressed. Two numerical examples are described to show the effectiveness of the proposed causality detection method. An experimental case study and an industrial case study are introduced to show the usefulness of the proposed method for finding information flow pathways and faults propagation pathways, and for root cause diagnosis.

3.2 Detection of Causality and Direct Causality

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

3.2.1 Preliminaries

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

A random variable Y can be considered as a mapping from an underlying sample space Ω to a set $\mathbf{Y}$ of interest. Each sample $\omega \in \Omega$ can give rise to a realization $Y(\omega)$ denoted by $y \in \mathbf{Y}$. Then the marginal range of Y is defined as [53]

$$[\![Y]\!] = \{Y(\omega) : \omega \in \Omega\}, \tag{3.1}$$

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

$$[\![Y|x]\!] = \{Y(\omega) : X(\omega) = x, \omega \in \Omega\}. \tag{3.2}$$

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

$$\bigcup_{x \in [\![X]\!]} [\![Y|x]\!] = [\![Y]\!]. \tag{3.3}$$

The joint range of Y and X is defined as

$$[\![Y, X]\!] = \{(Y(\omega), X(\omega)) : \omega \in \Omega\}. \tag{3.4}$$

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

$$[\![Y, X]\!] = \bigcup_{x \in [\![X]\!]} [\![Y|x]\!] \times \{x\}, \tag{3.5}$$

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 taken values

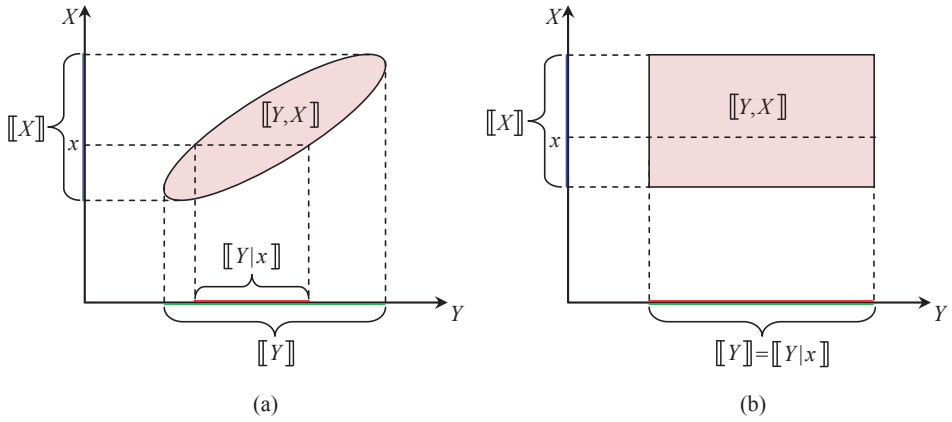


Figure 3.1: Examples of marginal, conditional, and joint ranges for related and unrelated random variables (adapted from [53]). (a) Y and X are related; (b) Y and X are unrelated.

in $\mathbf{Z}$, variables Y and X are said to be unrelated conditional on Z iff $[\![Y|x,z]\!] = [\![Y|z]\!]$, where $(x,z) \in [\![X,Z]\!]$ [53].

For example, Fig. 3.1(a) illustrates 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. 3.1(b) shows the ranges when Y and X are unrelated. For any $x \in [\![X]\!]$, the conditional range $[\![Y|x]\!]$ coincides with the marginal range $[\![Y]\!]$. Moreover, the joint range $[\![Y,X]\!]$ coincides with $[\![Y]\!] \times [\![X]\!]$.

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

$$\phi([\![Y]\!]) = \begin{cases} |[\![Y]\!]| & \text{for discrete-valued } Y, \\ \mu[\![Y]\!] & \text{for continuous-valued } Y, \end{cases} \tag{3.6}$$

where $|[\![Y]\!]|$ indicates the set size 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

$$H_0(Y) = \log \phi([\![Y]\!]), \tag{3.7}$$

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 [31] satisfying $H_0(Y) \in [0,\infty)$; if X

is continuous-valued, then $H_0(Y)$ indicates the (marginal) Rényi differential 0-entropy [57] which 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 [53, 64]:

$$H_0(Y|X) = \operatorname{ess\,sup}_{x \in [\![X]\!]} \log \phi([\![Y|x]\!]), \tag{3.8}$$

where ess sup represents the essential supremum. $H_0(Y|X)$ can be understood as a measurement of the uncertainty that remains in Y after X is known.

In order to measure the information about Y gained from X, a non-probabilistic 0-information metric, I_0, from X to Y is defined as follows [53, 64]:

$$I_0(Y;X) = H_0(Y) - H_0(Y|X) = \operatorname{ess\,inf}_{x \in [\![X]\!]} \log\left(\frac{\phi([\![Y]\!])}{\phi([\![Y|x]\!])}\right), \tag{3.9}$$

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 non-negative. $I_0(Y;X)$ 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, $I_0(Y;X) \neq I_0(X;Y)$.

3.2.2 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 transfer 0-entropy concept for causality detection based on the concept of 0-information.

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

$$I_0(Y;X|Z) = H_0(Y|Z) - H_0(Y|X,Z), \tag{3.10}$$

where $H_0(Y|Z)$ and $H_0(Y|X,Z)$ denote conditional 0-entropies defined in (3.8). 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 [7]. 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 transfer 0-entropy (T0E) from X to Y is then defined as follows:

$$
\begin{aligned}
& T^0_{X\to Y} \\
= \;& I_0(Y_{i+h}; \mathbf{X}_i^{(l)} | \mathbf{Y}_i^{(k)}) && (3.11)\\
= \;& H_0(Y_{i+h}|\mathbf{Y}_i^{(k)}) - H_0(Y_{i+h}|\mathbf{X}_i^{(l)}, \mathbf{Y}_i^{(k)}) && (3.12)\\
= \;& \operatorname{ess\,sup}_{\mathbf{y}_i^{(k)} \in [\![\mathbf{Y}_i^{(k)}]\!]} \log \phi([\![Y_{i+h}|\mathbf{y}_i^{(k)}]\!]) \\
& - \operatorname{ess\,sup}_{(\mathbf{x}_i^{(l)}, \mathbf{y}_i^{(k)}) \in [\![\mathbf{X}_i^{(l)}, \mathbf{Y}_i^{(k)}]\!]} \log \phi([\![Y_{i+h}|\mathbf{x}_i^{(l)}, \mathbf{y}_i^{(k)}]\!]) \\
= \;& \log \frac{\operatorname{ess\,sup}_{\mathbf{y}_i^{(k)} \in [\![\mathbf{Y}_i^{(k)}]\!]} \phi([\![Y_{i+h}|\mathbf{y}_i^{(k)}]\!])}{\operatorname{ess\,sup}_{(\mathbf{x}_i^{(l)}, \mathbf{y}_i^{(k)}) \in [\![\mathbf{X}_i^{(l)}, \mathbf{Y}_i^{(k)}]\!]} \phi([\![Y_{i+h}|\mathbf{x}_i^{(l)}, \mathbf{y}_i^{(k)}]\!])}, && (3.13)
\end{aligned}
$$

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

Since $\bigcup_{\mathbf{x}_i^{(l)} \in [\![\mathbf{X}_i^{(l)}]\!]} [\![Y_{i+h}|\mathbf{x}_i^{(l)}, \mathbf{y}_i^{(k)}]\!] = [\![Y_{i+h}|\mathbf{y}_i^{(k)}]\!]$, $[\![Y_{i+h}|\mathbf{x}_i^{(l)}, \mathbf{y}_i^{(k)}]\!]$ is contained in $[\![Y_{i+h}|\mathbf{y}_i^{(k)}]\!]$; thus, the T0E is always non-negative. From the definition, we can see that the T0E from x to y is the conditional 0-information defined in (3.10). 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.

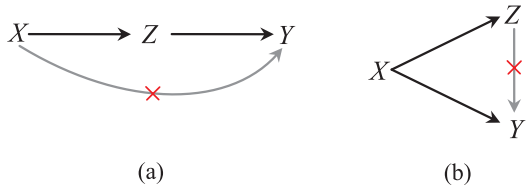
From the definition of T0E shown in (3.13), we can see that the T0E is only related to ranges of the random variables and is independent of their probability distributions. Thus, we do not require a well-defined probability distribution of the data set. This means that the collected sampled data does not need to be stationary, which is a basic assumption for the traditional transfer entropy method.

3.2.3 Direct Transfer 0-entropy

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. 3.2(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 (see Fig. 3.2(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 [22].

In order to detect whether there is direct causality from X to Y or the causality is indirect through some intermediate variables, a direct transfer 0-entropy (DT0E) from X to Y with intermediate variables $Z_1, Z_2, \ldots, Z_q$ is



(a) (b)

Figure 3.2: Information flow pathways between X, Y, and Z with (a) indirect causality from X to Y through the intermediate variable Z (meaning that there is no direct information flow from X to Y) and (b) spurious causality from Z to Y (meaning that Z and Y have a common perturbing source, X, and therefore they may appear to be connected or 'correlated' even when they are not connected physically).

defined as follows:

$$\begin{aligned} D^0_{X\to Y} &= I_0(Y_{i+h}; \mathbf{X}_i^{(l)} | \mathbf{Y}_i^{(k)}, \mathbf{Z}_{1,i_1}^{(s_1)}, \ldots, \mathbf{Z}_{q,i_q}^{(s_q)}) \\ &= H_0(Y_{i+h} | \mathbf{Y}_i^{(k)}, \mathbf{Z}_{1,i_1}^{(s_1)}, \ldots, \mathbf{Z}_{q,i_q}^{(s_q)}) - H_0(Y_{i+h} | \mathbf{X}_i^{(l)}, \mathbf{Y}_i^{(k)}, \mathbf{Z}_{1,i_1}^{(s_1)}, \ldots, \mathbf{Z}_{q,i_q}^{(s_q)}) \\ &= \log \frac{\underset{(\mathbf{y}_i^{(k)}, \mathbf{z}_{1,i_1}^{(s_1)}, \ldots, \mathbf{z}_{q,i_q}^{(s_q)})}{\text{ess sup}} \phi([\![Y_{i+h} | \mathbf{y}_i^{(k)}, \mathbf{z}_{1,i_1}^{(s_1)}, \ldots, \mathbf{z}_{q,i_q}^{(s_q)}]\!])}{\underset{(\mathbf{x}_i^{(l)}, \mathbf{y}_i^{(k)}, \mathbf{z}_{1,i_1}^{(s_1)}, \ldots, \mathbf{z}_{q,i_q}^{(s_q)})}{\text{ess sup}} \phi([\![Y_{i+h} | \mathbf{x}_i^{(l)}, \mathbf{y}_i^{(k)}, \mathbf{z}_{1,i_1}^{(s_1)}, \ldots, \mathbf{z}_{q,i_q}^{(s_q)}]\!])}, \end{aligned} \tag{3.14}$$

where $\mathbf{Z}_{j,i_j}^{(s_j)} = [Z_{j,i_j}, Z_{j,i_j-\tau_j}, \ldots, Z_{j,i_j-(s_j-1)\tau_j}]$ denotes the embedding vector with elements from the past values of Z_j for $j = 1, \ldots, q$; $\mathbf{z}_{j,i_j}^{(s_j)}$ denotes a realization of $\mathbf{Z}_{j,i_j}^{(s_j)}$; and $(\mathbf{x}_i^{(l)}, \mathbf{y}_i^{(k)}, \mathbf{z}_{1,i_1}^{(s_1)}, \ldots, \mathbf{z}_{q,i_q}^{(s_q)}) \in [\![\mathbf{X}_i^{(l)}, \mathbf{Y}_i^{(k)}, \mathbf{Z}_{1,i_1}^{(s_1)}, \ldots, \mathbf{Z}_{q,i_q}^{(s_q)}]\!]$. Note that the intermediate variables are chosen based on calculation results from the T0E [22]. Parameters $s_1, \ldots, s_q$, $i_1, \ldots, i_q$ and $\tau_1, \ldots, \tau_q$ in (3.14) are determined by the corresponding calculations of the T0E from $Z_1, \ldots, Z_q$ to Y.

The DT0E represents the information transferred from X to Y given the past information of both Y and intermediate variables $Z_1, \ldots, Z_q$. Similar to the T0E, it is a conditional 0-information and always non-negative. 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 $Z_1, \ldots, Z_q$.

3.3 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.

3.3.1 Range Estimation

From the definition in (3.13), 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, $[\![\mathbf{X}_i^{(l)}, \mathbf{Y}_i^{(k)}]\!]$ can be obtained by finding all possible realization sets of $(\mathbf{X}_i^{(l)}, \mathbf{Y}_i^{(k)})$; and $[\![Y_{i+h}|\mathbf{y}_i^{(k)}]\!]$ can be obtained by finding all possible realizations of Y_{i+h} given $\mathbf{Y}_i^{(k)} = \mathbf{y}_i^{(k)}$, and $\phi([\![Y_{i+h}|\mathbf{y}_i^{(k)}]\!])$ is the count of these realizations. For continuous-valued random variables, the estimation of ranges are 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 (3.3) and (3.5), it can be shown that the conditional ranges in (3.13) are fully determined by the joint range $[\![Y_{i+h}, \mathbf{X}_i^{(l)}, \mathbf{Y}_i^{(k)}]\!]$. The joint range can be obtained by the well-developed support estimation method based on the concept of support vector machine (SVM) [18]. Here we only give an algorithm for the joint range estimation, details on the support estimation method can be found in [58].

Let $\mathbf{v}$ denote 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:

$$k(\mathbf{v}_i, \mathbf{v}_j) = \langle \psi(\mathbf{v}_i), \psi(\mathbf{v}_j) \rangle, \tag{3.15}$$

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

$$k(\mathbf{v}_i, \mathbf{v}_j) = e^{-\gamma \|\mathbf{v}_i - \mathbf{v}_j\|^2}, \tag{3.16}$$

where $\gamma \in \{2^{-15}, 2^{-14}, \ldots, 2^1, 2^2, 2^3\}$ and it is determined by the cross validation approach [34, 52].

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 element α_i for $i = 1, \ldots, M$. In order to separate the data set from the origin, we solve the following quadratic programming (QP) problem:

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M} \alpha_i \alpha_j k(\mathbf{v}_i, \mathbf{v}_j), \tag{3.17}$$

$$\text{subject to} \quad 0 \leq \alpha_i \leq \frac{1}{vM}, \quad \sum_{i=1}^{M} \alpha_i = 1.$$

where $v \in (0, 1]$ 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 QPC toolbox [83].

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

$$\rho = \sum_{j=1}^{M} \alpha_j k(\mathbf{v}_j, \mathbf{v}_i), \tag{3.18}$$

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

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

$$f(\mathbf{v}) = \text{sgn}(\sum_{j=1}^{M} \alpha_j k(\mathbf{v}_j, \mathbf{v}) - \rho), \tag{3.19}$$

where

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

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.

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 (3.14).

3.3.2 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 (3.13): 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 [22]. A systematic method to determine them is described below.

Firstly, since $h = \tau$ as a rule of thumb [7], we can further set initial values for h and τ according to a priori knowledge of the process dynamics. If the process dynamics are unknown, small values of h and τ should give good results [7]; we may start by setting the initial values as $h = \tau = 1$.

Secondly, 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 non-negative 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 3, 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 T0E 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. In order to carry out this hypothesis testing, we may use the Monte Carlo method [7] by constructing a surrogate time series [60]. 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. In order to construct the surrogate time series that satisfy 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; 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

$$\begin{cases} X^{\text{surr}} = [X_i, X_{i+1}, \ldots, X_{i+M-1}], \\ Y^{\text{surr}} = [Y_j, X_{j+1}, \ldots, X_{j+M-1}], \end{cases} \tag{3.20}$$

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

$$s_{X \to Y} = \frac{T^0_{X \to Y} - \mu_\lambda}{\sigma_\lambda} > 3, \tag{3.21}$$

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.

3.4 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 a benchmark industrial case study.

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:

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

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

For ranges estimation, we set $v = 0.01$ since the fraction of outliers of the data is quite small. For determination of γ, the initial 1000 data points

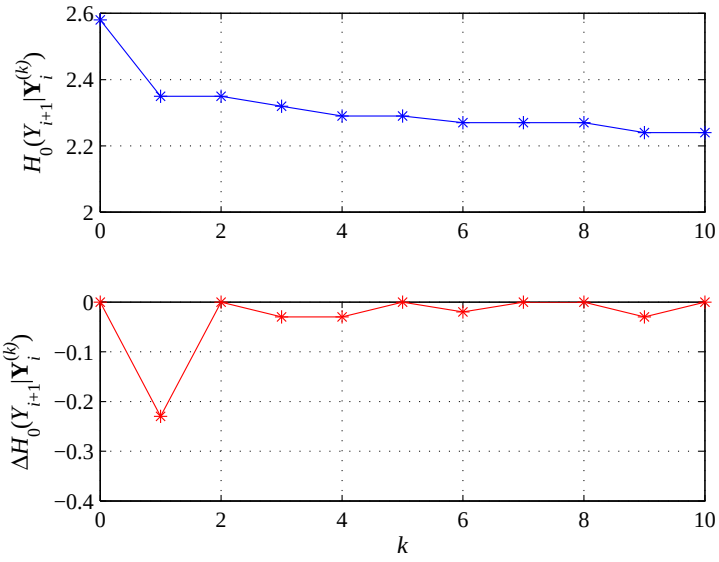


Figure 3.3: Finding the embedding dimension of Y in Example 1.

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 transfer 0-entropies between X, Y, and Z, we need to determine the four design parameters. We take the T0E from X to Y as an example. First, initial values for h and τ are set as $h = \tau = 1$; secondly, 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. 3.3. 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. 3.3, 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 more useful information for Y_{i+1} when Y_i is given. Thus, we choose $k = 1$. Finally, we calculate the transfer entropy $T^0_{X\to Y}$ and its change rate with $l = 1, \ldots, 10$, as shown in Fig. 3.4. Since there is no significant change in $T^0_{X\to Y}$ after $l = 1$, as shown in the lower part of Fig. 3.4, 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 (3.13) and (3.21), 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

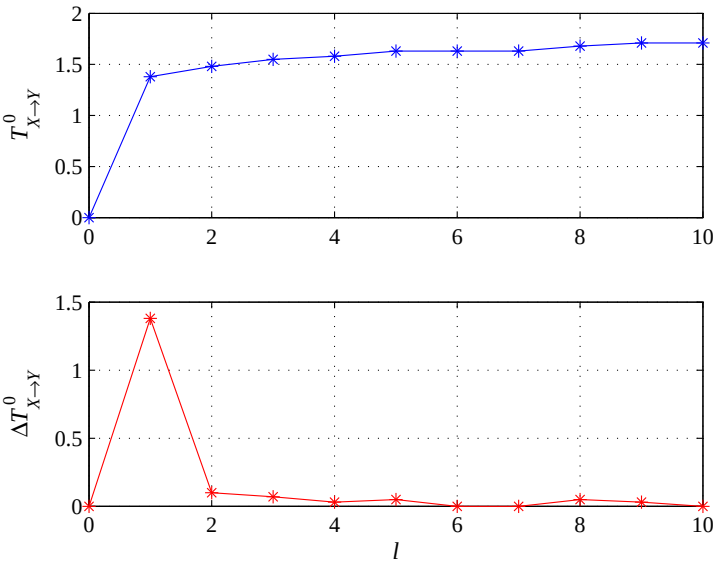


Figure 3.4: Finding the embedding dimension of X for $T^0_{X\to Y}$ in Example 1.

shown in Table 3.1. 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 (3.20), 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 3.1, 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 is direct causality from X to Z. According to (3.14), 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. 3.5(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.

Table 3.1: Calculated transfer 0-entropies and thresholds (values in round brackets) for Example 1.

$T^0_{\mathtt{column}\,1\to\mathtt{row}\,1}$	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)

Figure 3.5: Information flow pathways for (a) Example 1 and (b) Example 2.

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

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

where $X_k \in [4, 5]$ is a uniformly distributed signal; $v_{1k}, v_{2k} \sim N(0, 0.05)$; 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 3.2. 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 (3.14), 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. Secondly, we need to detect whether there is true and direct causality from Y to Z since X is the common source of 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

Table 3.2: Calculated transfer 0-entropies and thresholds (values in round brackets) for Example 2.

$T^0_{\mathtt{column}\,1\to\mathtt{row}\,1}$	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

and direct causality from Y to Z. The information flow pathways for Example 2 are shown in Fig. 3.5(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.

Experimental Case Study

In order to show the effectiveness of the proposed causality detection method for capturing information and/or material flow pathways, a 3-tank experiment was conducted. The schematic of the 3-tank system is shown in Fig. 3.6, which is the same as that in Section 2.4 of Chapter 2.

Similarly, we denote the water levels of tanks 1, 2, and 3, and the flow rate of the pumped water by x_1, x_2, x_3, and x_4, respectively. In this experiment, x_4 is set to be a pseudo-random binary sequence (PRBS). The sampled data of 3000 observations were analyzed. Fig. 3.7 shows the normalized time trends of the measurements. The sampling time is one second.

The initial 1000 data points were 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 3.3 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. 3.8(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. 3.8(b). To clarify this we first calculate the DT0E from x_1 to x_3 with intermediate variables x_4 and x_2 and obtain $D^0_{x_1\to x_3} = 0$, which means that there is no direct in-

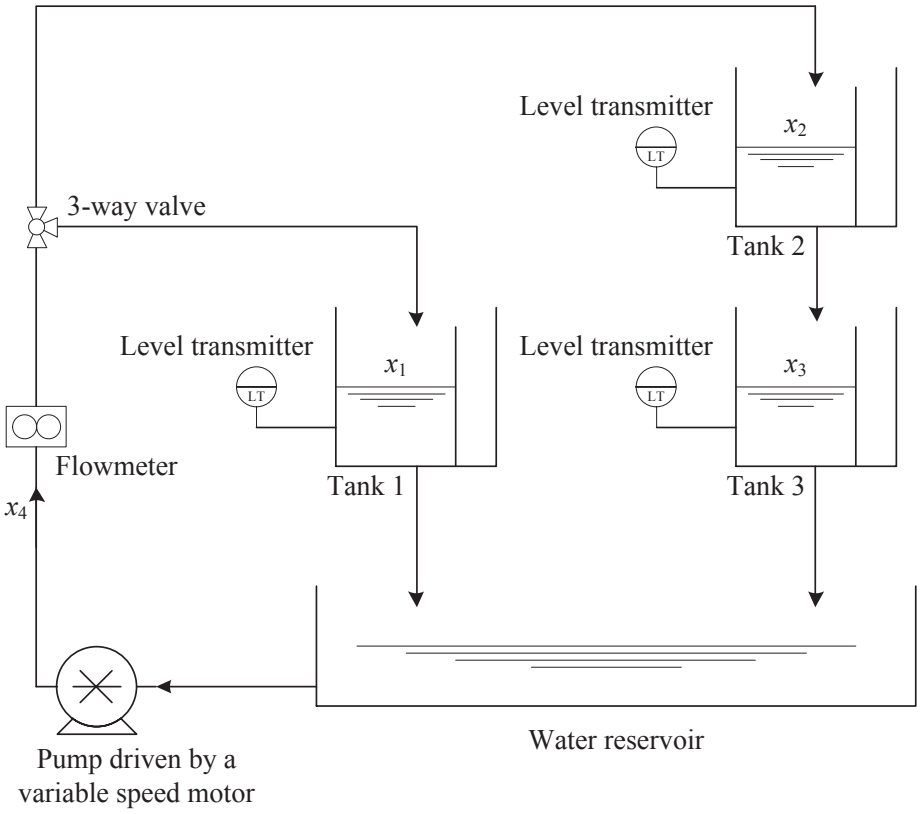


Figure 3.6: Schematic of the 3-tank system.

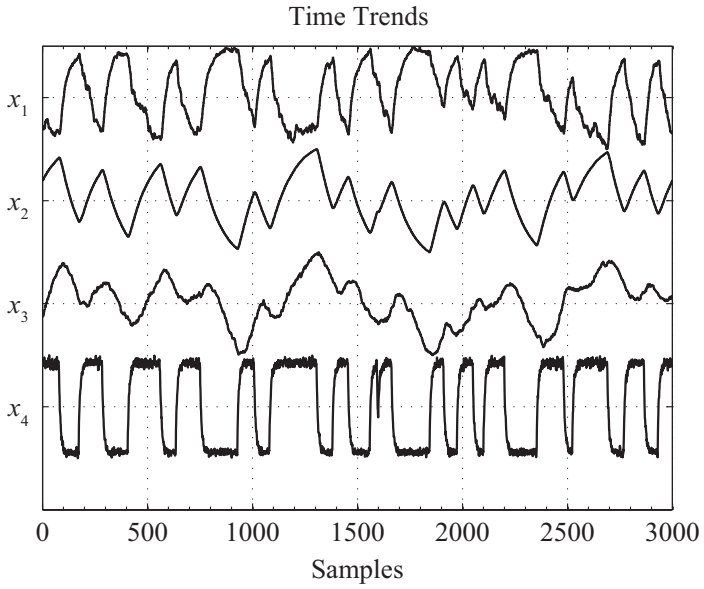


Figure 3.7: Time trends of measurements of the 3-tank system.

Table 3.3: Calculated transfer 0-entropies and thresholds (values in round brackets) for the 3-tank system.

$T^0_{\mathtt{column}\,1\to\mathtt{row}\,1}$	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

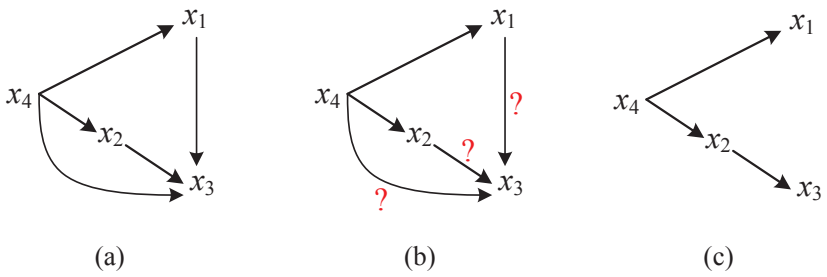


Figure 3.8: Information flow pathways for the 3-tank system based on (a) and (b) calculation results of T0Es which represent the total causality including both direct and indirect/spurious causality; (c) calculation results of DT0Es which correctly indicate the direct and true causality.

formation/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. 3.8(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 (3.14), 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. 3.8(c), which are consistent with the information and material flow pathways of the physical 3-tank system (see Fig. 3.6).

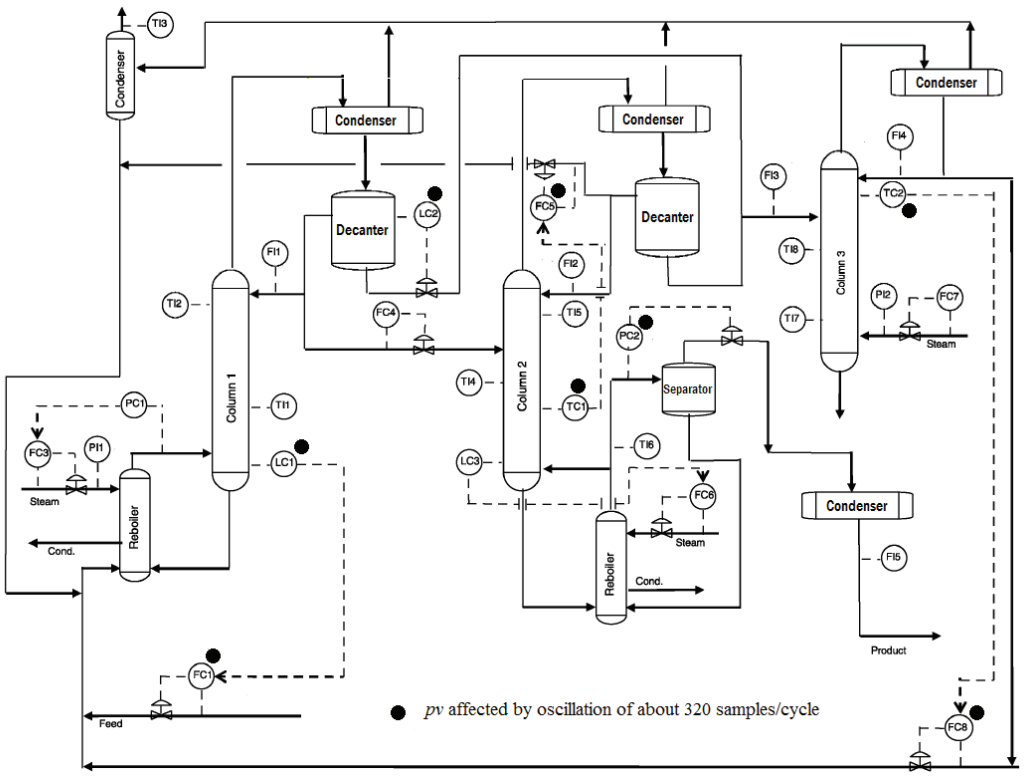


Figure 3.9: Process schematic. The oscillation process variables (pv) are marked by circle symbols.

Industrial Case Study

We use an industrial process data set [37, 73] provided by the Advanced Controls Technology group of Eastman Chemical Company, USA, to illustrate the effectiveness of the proposed causality detection method. Engineers at the Advanced Controls Technology group had identified a need to diagnose a common oscillation of about 2 hours (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. 3.9. 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 variables measurements and their set points) or in the measurements from other sensors. The plant-wide oscillation detection and diagnosis methods can be used for any of these time trends [75, 76]. In this thesis we only use process variables for root cause analysis; 14 controlled process variables corresponding to 14 PID controller loops are available. 5040 sampled observations (from 28

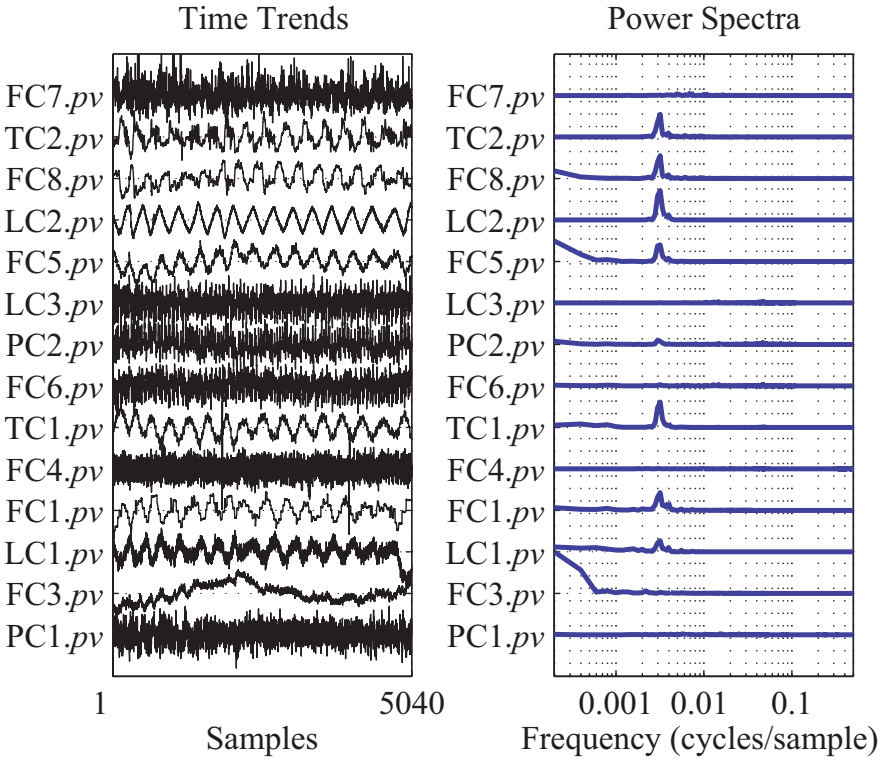


Figure 3.10: Time trends and power spectra of measurements of process variables (pvs).

hours of data with the sampling interval 20 seconds) 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. 3.10 shows the normalized time trends and power spectra of the 14 process variables.

The power spectra in Fig. 3.10 indicate the presence of oscillation at the frequency of about 0.003 cycles/sample, corresponding to an approximate period of 2 hours. It was shown that the control valve of loop LC2 suffered from a stiction problem which resulted in limit cycles in the corresponding process variable. [37]. It has been confirmed that the control valve caused the controlled variable LC2.pv to oscillate [73]. After that, this oscillation propagated throughout the inter-connected units and affected many other control loops in the process. More related information is provided in [73] and [37]. Thus, our goal is to detect and diagnose the root cause of this oscillation.

For oscillation detection, the spectral envelope method [37] was applied to determine which variables have oscillation at the frequency of 0.0032 cy-

Table 3.4: Calculated transfer 0-entropies and thresholds (values in round brackets) for the industrial case study.

$T^0_{\texttt{column 1}\to\texttt{row 1}}$	LC1	FC1	TC1	PC2	FC5	LC2	FC8	TC2
LC1	NA	0.12 (0.06)	0	0	0	0	0.13 (0.07)	0.14 (0.06)
FC1	0.10 (0.07)	NA	0.16 (0.06)	0.04 (0.07)	0	0	0	0
TC1	0.13 (0.07)	0.15 (0.06)	NA	0.15 (0.07)	0.25 (0.07)	0	0.12 (0.06)	0
PC2	0	0.08 (0.06)	0.21 (0.07)	NA	0	0	0	0
FC5	0.06 (0.07)	0	0.05 (0.06)	0.04 (0.08)	NA	0	0	0.03 (0.07)
LC2	0.24 (0.07)	0	0.29 (0.06)	0	0.25 (0.07)	NA	0.13 (0.07)	0.25 (0.06)
FC8	0.04 (0.07)	0.13 (0.06)	0	0.04 (0.07)	0	0.03 (0.07)	NA	0.05 (0.06)
TC2	0	0	0	0.04 (0.07)	0	0	0.28 (0.06)	NA

cles/sample. Details on the spectral envelope method are introduced in Chapter 4. Since this chapter 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 8 process variables have common oscillations with 99.9% confidence level: LC1.*pv*, FC1.*pv*, TC1.*pv*, PC2.*pv*, FC5.*pv*, LC2.*pv*, FC8.*pv*, and TC2.*pv*. These variables are marked by dark circle symbols in Fig. 3.9. 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 [37]. 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. By using the cross validation method, γ is set to be 2^{-2}. Table 3.4 shows the T0Es and thresholds (values in brackets) between each pair of the process variables, where we omit *.pv* in the tag names of variables for simplicity.

Based on Table 3.4, the causal relationships between the 8 oscillation pro-

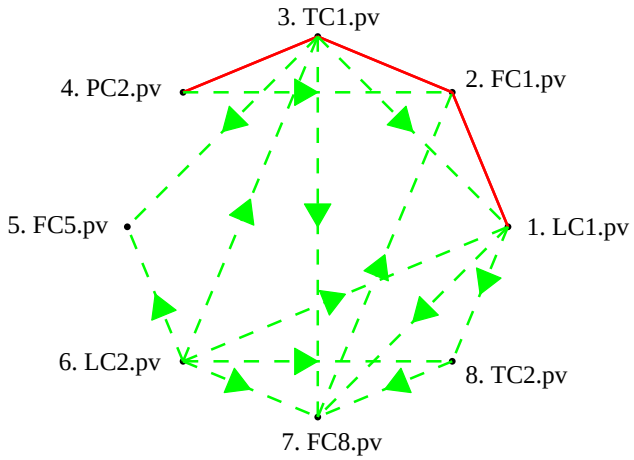


Figure 3.11: Causal map of oscillation process variables based on calculation results of T0Es. A dashed line with an arrow indicates that there is unidirectional causality from one variable to the other, and a solid line connecting two variables without an arrow indicates there is bidirectional causality between the two variables.

cess variables are shown in Fig. 3.11, 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 that there is bidirectional causality (also called causality feedback [13]) between the two variables.

The causal map in Fig. 3.11 shows a complicated set of pathways from which finding faults propagation pathways would be difficult. The reason is that by only calculating T0Es, both total and spurious causality can be detected. In order to derive a simpler and more accurate causal map, we need to differentiate between direct and indirect, 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 LC2.*pv* to FC5.*pv*, since LC2.*pv* causes TC1.*pv*, and TC1.*pv* causes FC5.*pv*, we need to calculate the DT0E from LC2.*pv* to FC5.*pv* with the intermediate variable TC1.*pv*.

Table 3.5 shows the calculated DT0Es between each pair of the variables

Table 3.5: Calculated direct transfer 0-entropies and thresholds (values in round brackets) for the industrial case study.

$D^0_{\texttt{column 1}\to\texttt{row 1}}$	LC1	FC1	TC1	PC2	FC5	LC2	FC8	TC2
LC1	NA	0.11 (0.06)	NA	NA	NA	NA	0 (0.07)	0 (0.06)
FC1	0.09 (0.07)	NA	0.03 (0.06)	NA	NA	NA	NA	NA
TC1	0 (0.07)	0.04 (0.06)	NA	0.15 (0.07)	0.23 (0.07)	NA	0 (0.06)	NA
PC2	NA	0 (0.06)	0.17 (0.07)	NA	NA	NA	NA	NA
FC5	NA	NA	NA	NA	NA	NA	NA	NA
LC2	0.20 (0.07)	NA	0.25 (0.06)	NA	0.05 (0.07)	NA	0.04 (0.07)	0.21 (0.06)
FC8	NA	0 (0.06)	NA	NA	NA	NA	NA	NA
TC2	NA	NA	NA	NA	NA	NA	0.24 (0.06)	NA

that has causal relationship and has possible intermediate variable(s). Note that if a pair of the variables does not have significant causal relationship based on the calculation results of T0Es shown in Table 3.4, then we do not need to calculate its DT0E and thus put 'NA' in Table 3.5. 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. 3.11. The causal map based on calculation results of DT0Es is shown in Fig. 3.12 which is much sparser than the previous causal map shown in Fig. 3.11.

The oscillation propagation pathways obtained from the causal map (see Fig. 3.12) are shown in Fig. 3.13, where One-headed arrows indicate unidirectional causality and double-headed arrows indicate bidirectional causality. The oscillation propagation pathways are also indicated by red lines with arrows in the process schematic, as shown in Fig. 3.14. 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. Figures

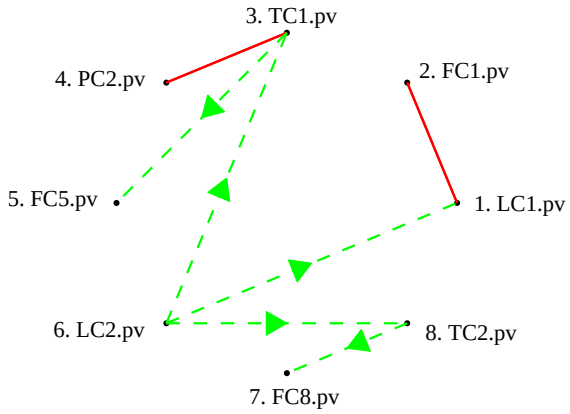


Figure 3.12: Causal map of oscillation process variables based on calculation results of DT0Es. A dashed line with an arrow indicates that there is unidirectional causality from one variable to the other, and a solid line connecting two variables without an arrow indicates there is bidirectional causality between the two variables.

3.13 and 3.14 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. Figures 3.13 and 3.14 also show that the oscillation in loop LC2 first propagates to loops LC1, TC1, and TC2. From Fig. 3.14, 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 [37, 73] that the root cause of the plant wide oscillation is the valve stiction of control loop LC2; therefore the causality analysis via the transfer 0-entropy method is indeed effective in finding the fault propagation pathways and determining the root cause candidate.

3.5 Summary

In industrial processes, fault detection and diagnosis in a large-scale complex system are 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 a-

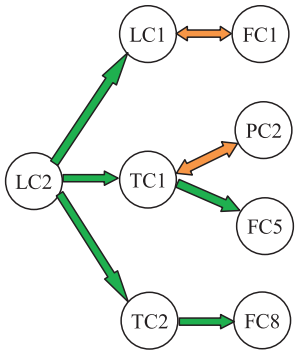


Figure 3.13: Oscillation propagation pathways obtained via the transfer 0-entropy method. One-headed arrows indicate unidirectional causality and double-headed arrows indicate bidirectional causality.

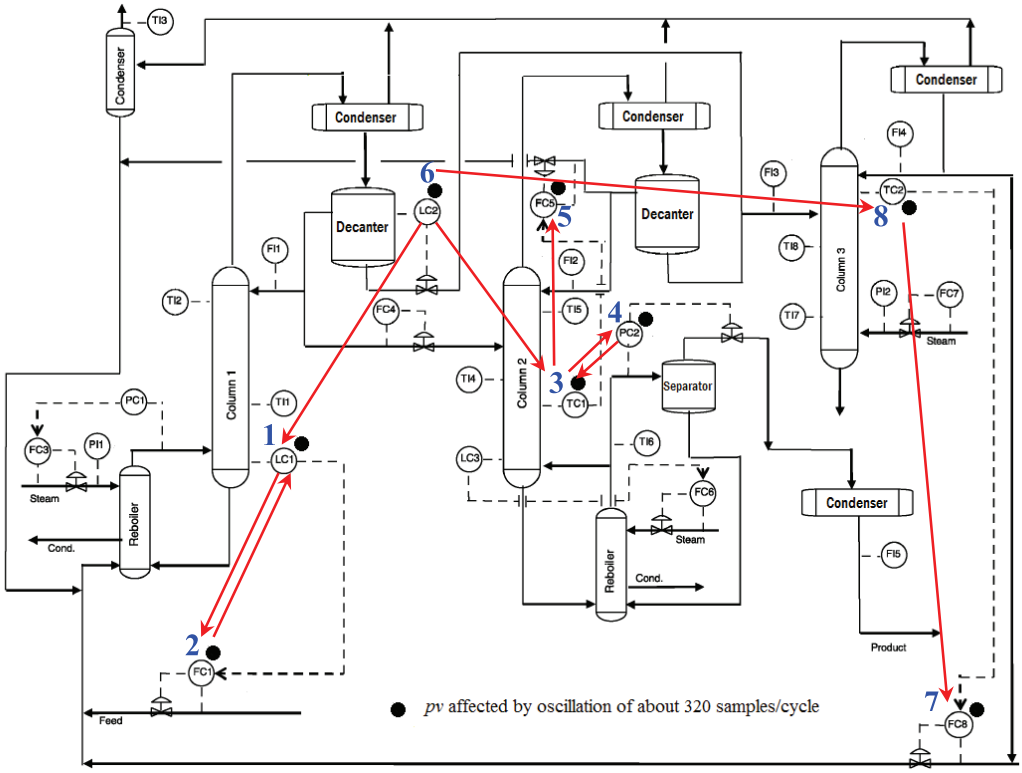


Figure 3.14: Direct causal relationships between the oscillation process variables, namely, the oscillation propagation pathways, which are indicated by red lines with arrows.

long information and material flow pathways and affect 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 application on two examples, an experimental data set and a benchmark industrial case study.

The outstanding advantage of the T0E method is that the data does 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 transfer entropy method. This point can also be seen from the range estimation point of view. According to the QP problem in (3.17), 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 3-tank case study as presented in Section 3.4. The analyzed data set is not strictly stationary. The T0E method can still find the information and/or material flow pathways by using this data set. Another advantage of the T0E method compared with the traditional transfer entropy method is that the length of the data does not need to be very large, for example larger than 2000. The reason is that the range estimation is based on the concept of SVM which can handle small sample data sets [25, 94]. Based on our experience, 500 samples are enough to give good results.

Chapter 4

Application of Causality Analysis for Root Cause Diagnosis of Plant-wide Oscillations*

4.1 Overview

Plant-wide oscillations are common in many industrial processes. They may impact the overall process performance and reduce profitability. It is important to detect and diagnose root causes of such oscillations. This chapter discusses applications of the causality analysis techniques to root cause diagnosis of plant-wide oscillations. Since various methods have already been proposed for identifying root causes of plant-wide oscillations, there is no rule to determine which method to use when a plant-wide oscillation occurs. In order to give some suggestions on how to choose appropriate methods, in this chapter, we review, discuss and recommend the usefulness of causality detection methods and several recently introduced methods for identifying possible root causes of plant-wide oscillations; these methods include three causality analysis methods using Granger causality, TE, and T0E, the spectral envelop method, the adjacency matrix method, and the Bayesian network inference method. All the six methods are evaluated by application to a benchmark industrial data set and recommendations and guidelines are given to help a practising engineer carry out the diagnostic process.

*A version of this chapter has been accepted for publication as: P. Duan, S.L. Shah, T. Chen, and F. Yang. Methods for detection and root cause diagnosis of plant-wide oscillations. *AIChE Journal*, 2014.

4.2 Introduction

Oscillations are periodic phenomena with well defined amplitudes and frequencies. When an oscillation is generated somewhere in a plant and propagates to the whole plant or some other units of the plant through information and/or material flow pathways, it is termed as a plant-wide oscillation [15]. Plant-wide oscillations are common in many processes because of the interacting material and information flow streams between units as well as the presence of recycle streams. Their presence may impact the overall process performance and cause inferior quality products, larger rejection rates and excessive energy consumption. Thus, it is important to detect and diagnose the causes of such oscillations in order to compensate for them. This includes two key requirements [74]: (1) detection of the presence of one or more periodic oscillations; and (2) determination of the locations of the various oscillations in the plant and their most likely root cause(s). Root causes of plant-wide oscillations can be poorly tuned controllers, process or actuator non-linearities caused by valve stiction, oscillatory disturbance(s), loop interaction, etc.

Plant-wide oscillation detection requires the recognition that an oscillation in one measurement is the same as that in another measurement. A high density plot provides an off-line visualization tool to describe temporal and spectral plots of all the concerned variables in a compact form [15]. Another visualization tool termed as the power spectral correlation map (PSCMAP) was developed to automatically collect and cluster variables with common spectral shapes in a multivariate process [70]. Disturbances that propagate throughout a plant due to recycle streams and/or heat integration stream or other means and can have an impact on product quality and running cost are termed as plant-wide disturbances. Not all plant-wide disturbances exhibit oscillatory behavior, but they do have the same effects on control loops of process units and do impact product quality. Such disturbances show up as having similar spectral shapes in all affected variables and therefore can be detected using the power spectral correlation criterion [70]. Based on the regularity of zero crossings of the filtered auto-correlation function (ACF), an automatic detection of clusters of similar oscillations was implemented in [75]. Spectral decomposition methods [69, 76, 85] can also be used to detect and classify spectral features in multivariate data sets. A spectral envelope method was used to detect and categorize process measurements with similar

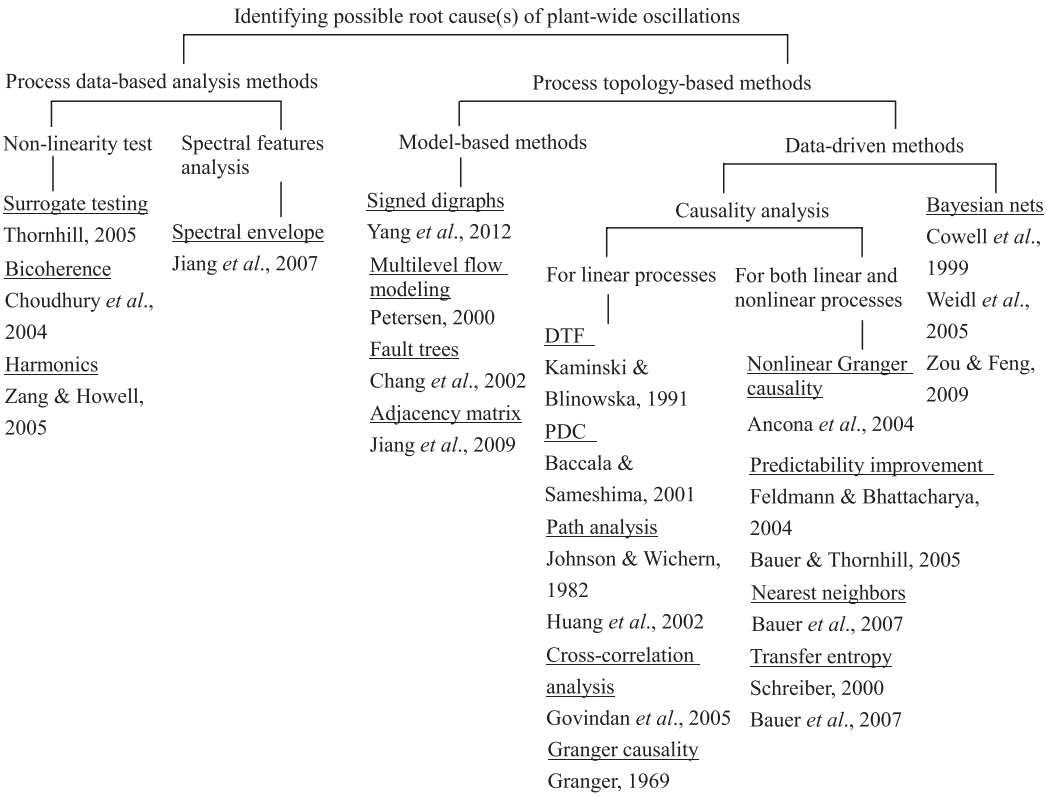


Figure 4.1: Family tree of methods for the first stage of root cause diagnosis of plant-wide oscillations.

spectral characteristics [37]; it makes use of the cross-spectra to find common frequency components in a multivariate data set.

For the plant-wide oscillation diagnosis problem, it is important to distinguish root causes of oscillations from the secondary propagated oscillations. Methods for identifying possible root causes can be divided into two main classes, namely, process data-based analysis methods and process topology-based methods. Fig. 4.1 is a family tree of methods for identifying root causes of plant-wide oscillations.

Process data-based analysis methods for oscillation diagnosis are widely used because of the availability of enormous amounts of process data. This class of methods diagnose root causes by analyzing characteristics of process data which mainly include non-linearity and power spectrum features. Root cause diagnosis based on non-linearity analysis has been reported on the assumption that the measurement with the highest non-linearity is closest to the root cause [74]. Surrogate testing [72], bispectrum and the related bi-

coherence [17], and harmonics [93] have been used to detect the presence of non-linearity in process data. In Ref. [37], based on oscillation detection results, the oscillation contribution index (OCI) was proposed to indicate how each measured variable contributes towards plant-wide oscillating frequency and then isolate the key variables as the potential root cause candidates of the common oscillation(s).

Process topology-based methods aim to capture process topology and find fault propagation pathways to determine the root cause(s) of certain plant-wide oscillations. This class of methods can be further divided into model-based methods and data-driven methods. For model-based methods, qualitative process information (piping and instrumentation diagram/drawing or P&ID) and expert knowledge of the process are implicitly used. It was shown that diagnosis would be enhanced if a qualitative model is available [45]. Qualitative models include signed digraphs (SDG) [86, 88], fault trees [14, 77], Multilevel Flow Modeling [56], and adjacency matrices [38]. A drawback of model-based methods is that extracting process information is very time consuming and that information is not always easily available. Often times the P&ID information is not accurately updated or is erroneous.

Data-driven methods provide another way to capture process connectivity. Bayesian networks [19, 80] have been introduced to describe dependency between multivariate data sets by incorporating probabilistic properties. As a hot and emerging topic, causality analysis provides an effective way to diagnose root cause of plant-wide oscillations since a causal map can capture process connectivity and allow investigation along fault propagation pathways [87].

After a potential root cause is identified using plant-wide oscillations diagnosis techniques, further analysis and field test(s) are usually carried out to confirm the root cause. If the root cause is a poorly tuned controller, the particular controller can be adjusted to remove oscillations. If the root cause is stiction in a control valve, there exist many techniques to detect valve stiction [36]. A common industrial practice is to put the loop in manual and observe the behavior of the oscillatory loop. The valve stiction can be confirmed if the limit cycle dies out in the manual control mode [15]. It is reported that both poorly tuned controller and valve stiction can be confirmed via a nonlinearity test and a controller gain change method [16, 90]. Note that just two oscillating variables having a common oscillation frequency component does

not mean that they are caused by the same root cause. Thus, a field test is needed to confirm that all the oscillations disappear when the potential root cause is isolated. However, it is very likely that the variables are somehow topologically connected and the root causes of the common oscillation are related.

Since various methods have already been proposed for identifying root causes of plant-wide oscillations, there is no rule to determine which method to use when a plant-wide oscillation occurs. In order to give some suggestions on how to choose appropriate methods, in this chapter, we discuss and recommend the usefulness of several recently introduced methods by application to a benchmark industrial data set (the Eastman data set described in Section 3.4 of Chapter 3). As shown in Fig. 4.1, since the non-linearity test methods under the first branch have been extensively discussed in [15] and [74], in this thesis we consider methods from other branches and choose one method from each branch as a representative for discussion and comparison; these methods include the spectral envelope method and four process topology analysis methods: adjacency matrix, two causality analysis methods using Granger causality and transfer entropy, and Bayesian networks. As such a wide spectrum of methods are compared and discussed. Specifically, the spectral envelope method is a spectral analysis method; the adjacency matrix is based on matrix algebra; the other three methods are data-based process topology exploration methods: the Granger causality is based on autoregressive (AR) models; the transfer entropy method is an information theory-based method; and the BN inference method is a probability-theoretic method. Since the T0E method was proposed in this thesis, we include it for discussion and comparison. Finally recommendations and guidelines based on the results of the benchmark case study are given to help a practising engineer carry out the diagnosis problem.

4.3 Methods

In this section, we describe four methods, namely, the spectral envelope method, the adjacency matrix method, the Granger causality method, and the Bayesian network inference method, including their mathematical backgrounds and how to use them for detection and/or diagnosis of plant-wide oscillations, and their applications to the Eastman data set described in Section 3.4 of Chapter 3.

Since the TE method has been studied in Chapter 2, in this section, we only introduce its application to the benchmark data set. Since the T0E method and its application to the benchmark data set have been extensively discussed in Chapter 3, we omit the T0E method in this section.

4.3.1 Spectral Envelope Method

The spectral envelope method is the only one among the six methods capable of both detection and root cause diagnosis of plant-wide oscillations. Since root cause diagnosis should be based on plant-wide oscillation detection results, that is, which variables have the common oscillation frequency component, we first need to detect these plant-wide oscillations. Although the high density plot (see Fig. 3.10) shows spectral peaks at the common oscillation frequency, these spectral plots in themselves cannot automatically provide a list of variables that have a common frequency component [15] since it is difficult to distinguish whether the spectral peak is significantly large or not. In contrast to this, the spectral envelope method provides an effective way to list all variables that have oscillations. Thus, after introducing the concept of the spectral envelope, this subsection describes the usefulness of the spectral envelope method for both detection and root cause diagnosis of plant-wide oscillations. The oscillation detection result will be used by all the six root cause diagnosis methods in their applications to the benchmark data set.

1) Concept of the Spectral Envelope: The concept of spectral envelope was first introduced by [67] as a statistical basis for frequency domain analysis of discrete symbolic data. The concept of spectral envelope was extended to continuous data [51] and applied to find optimal transformations for the analysis of time series and detect common signals in many time series [68]. The spectral envelope method was successfully used by [37] to detect and diagnose plant-wide oscillations from industrial data. The key idea of the spectral envelope is to select optimal transformations of a time series that emphasize any periodic occurrence in the frequency domain.

Let $\mathbf{x}(t) = [x_1(t), x_2(t), \cdots, x_n(t)]^{\mathrm{T}}$ be a vector-valued time series on $\mathbb{R}^n$. Each time series $x_i(t)$ is normalized by subtracting its mean value and dividing by its standard deviation. Thus, all n time series have identical power. Denote the covariance matrix of $\mathbf{x}(t)$ by $\mathbf{V_x}$ and the power spectral density (PSD) matrix of $\mathbf{x}(t)$ by $\mathbf{P_x}(\omega)$, where ω is the normalized frequency satisfying $-1/2 \leq \omega \leq 1/2$.

Let $y(t, \boldsymbol{\beta}) = \boldsymbol{\beta}^* \mathbf{x}(t)$ be a scaled series, where $\boldsymbol{\beta}$ is an n-dimensional vector which may be real or complex, and $*$ denotes the conjugate transpose. In fact, $y(t, \boldsymbol{\beta})$ is a linear combination of the elements of $\mathbf{x}(t)$. The variance of $y(t, \boldsymbol{\beta})$ can be expressed as $V_y(\boldsymbol{\beta}) = \boldsymbol{\beta}^* \mathbf{V}_\mathbf{x} \boldsymbol{\beta}$, and the power spectral density of $y(t, \boldsymbol{\beta})$ is $P_y(\omega, \boldsymbol{\beta}) = \boldsymbol{\beta}^* \mathbf{P}_\mathbf{x}(\omega) \boldsymbol{\beta}$.

The spectral envelope of $\mathbf{x}(t)$ is defined as

$$\lambda(\omega) = \sup_{\boldsymbol{\beta} \neq 0} \left\{ \frac{P_y(\omega, \boldsymbol{\beta})}{V_y(\boldsymbol{\beta})} \right\} = \sup_{\boldsymbol{\beta} \neq 0} \left\{ \frac{\boldsymbol{\beta}^* \mathbf{P}_\mathbf{x}(\omega) \boldsymbol{\beta}}{\boldsymbol{\beta}^* \mathbf{V}_\mathbf{x} \boldsymbol{\beta}} \right\}, \tag{4.1}$$

where the scaling vector $\boldsymbol{\beta}$ is usually constrained to $\boldsymbol{\beta}^* \mathbf{V}_\mathbf{x} \boldsymbol{\beta} = 1$. The scaling vector that results in the value $\lambda(\omega)$ is called the optimal scaling vector at frequency ω and denoted by $\boldsymbol{\beta}(\omega)$. The elements of the optimal scaling vector are called the optimal scalings. For a different frequency, the optimal scaling vector is different. Note that the relationship between $P_y(\omega, \boldsymbol{\beta})$ and $V_y(\boldsymbol{\beta})$ is

$$V_y(\boldsymbol{\beta}) = \int_{-1/2}^{1/2} P_y(\omega, \boldsymbol{\beta}) d\omega = 2 \int_0^{1/2} P_y(\omega, \boldsymbol{\beta}) d\omega.$$

Thus, The quantity $\lambda(\omega)$ represents the largest portion of power (or variance) that can be obtained at frequency ω for any scaled series.

With the notation that $\mathbf{V} = \text{diag}(\mathbf{V}_\mathbf{x})$, simulation studies showed that using $\mathbf{V}$ is superior to $\mathbf{V}_\mathbf{x}$ [68] in identifying oscillations in signals. By using $\mathbf{V}$ instead of $\mathbf{V}_\mathbf{x}$ in (4.1), a new expression for $\lambda(\omega)$ [66] is

$$\lambda(\omega) = \sup_{\boldsymbol{\beta} \neq 0} \left\{ \frac{\boldsymbol{\beta}^* \mathbf{P}_\mathbf{x}(\omega) \boldsymbol{\beta}}{\boldsymbol{\beta}^* \mathbf{V} \boldsymbol{\beta}} \right\}, \tag{4.2}$$

where $\boldsymbol{\beta}$ satisfies the constraint that $\boldsymbol{\beta}^* \mathbf{V} \boldsymbol{\beta} = 1$. Since the data has been normalized, we have $\mathbf{V} = I_{n \times n}$. Thus the constraint can be simplified as $\boldsymbol{\beta}^* \boldsymbol{\beta} = 1$. Then $\lambda(\omega)$ is the largest eigenvalue of $\mathbf{P}_\mathbf{x}(\omega)$, and $\boldsymbol{\beta}(\omega)$ is the corresponding eigenvector.

To calculate $\lambda(\omega)$ and $\boldsymbol{\beta}(\omega)$, we need to first estimate the PSD matrix $\mathbf{P}_\mathbf{x}(\omega)$. Assume the number of samples for $\mathbf{x}(t)$ is N, namely $t = 0, 1, \ldots, N-1$. The Fourier frequencies are $\omega_k = k/N$, for $k = 1, 2, \ldots, [N/2]$, where $[N/2]$ is the greatest integer less than or equal to $N/2$. If N is a large integer, using the fast Fourier transformation method, the periodogram can be estimated as

$$\hat{\mathbf{I}}_N(\omega_k) = \frac{1}{N} \left[\sum_{t=0}^{N-1} \mathbf{x}(t) \exp(-2\pi i t \omega_k) \right] \left[\sum_{t=0}^{N-1} \mathbf{x}(t) \exp(-2\pi i t \omega_k) \right]^*, \tag{4.3}$$

which provides a simple estimate of $\mathbf{P}_{\mathbf{x}}(\omega_k)$. Alternatively, a smoothed periodogram estimate can be used, that is

$$\hat{\mathbf{P}}_{\mathbf{x}}(\omega_k) = \sum_{j=-r}^{r} h_j \hat{\mathbf{I}}_N(\omega_{k+j}), \tag{4.4}$$

where h_j is symmetric positive weights satisfying $h_j = h_{-j}$ and $\sum_{j=-r}^{r} h_j = 1$. A simple average corresponds to the case where $h_j = 1/(2r+1)$ for $j = -r, \ldots, 0, \ldots, r$. Alternatively, h_j can be chosen as $h_j = (r - |j| + 1)/(r+1)^2$ for $j = -r, \ldots, 0, \ldots, r$. The number r is chosen to obtain a desired degree of smoothness. Details on how to determine r and h_j can be found in [68].

2) Usefulness of the Spectral Envelope Method: Since the spectral envelope method can be used for both detection and diagnosis of the plant-wide oscillation, we divide the description of its usefulness into two parts a) oscillation detection ; and b) root cause diagnosis.

a) Oscillation detection: The key idea of the spectral envelope method comes from the realization that a right linear combination of the original time series can enhance the signal to noise ratio [37, 66]. The spectral envelope plot represents the largest portion of power that can be obtained at each frequency for any linear combination of the original time series. In fact, this means that if the original time series has a common frequency component, the spectral envelope at that frequency will be larger than others. Thus common frequency components, that is one or more components, can be easily detected by seeking peaks in the spectral envelope plot.

For the benchmark data set, we first calculate the spectral envelope of the 14 controlled process variables by finding the largest eigenvalue of the PSD matrix $\mathbf{P}_{\mathbf{x}}(\omega)$. $\mathbf{P}_{\mathbf{x}}(\omega)$ is estimated using (4.4), here we choose $r = 1$ and weights $\{h_0 = 1/2, h_{\pm 1} = 1/4\}$. Fig. 4.2 shows the spectral envelope of the 14 variables. It is clear that there is a peak at a frequency of $\omega_{16} = 16/5040 \approx 0.0032$ cycles/sample, indicating an oscillation with a period of about 320 samples/cycle. This is the oscillation that the Advanced Controls Technology group of Eastman Chemical Company wanted to detect and diagnose. Then we need to identify all variables that have these oscillations.

Since the magnitude of the optimal scalings is a measure of the contribution of each time series to the spectral envelope $\lambda(\omega)$ at frequency ω, the time series having large optimal scaling magnitudes are the ones that contribute more to the spectral envelope, and thus are the ones having oscillations at

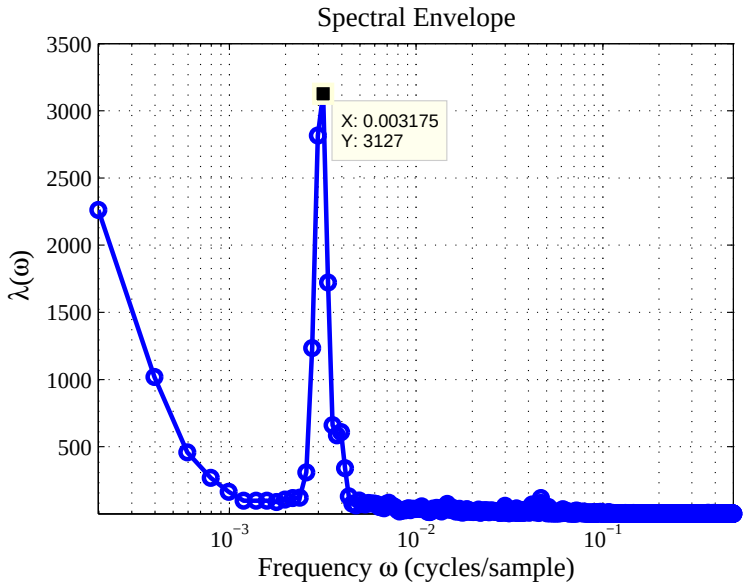


Figure 4.2: Spectral envelope of the 14 process variables.

that frequency. To identify the variables that have oscillations, a statistical hypothesis test can be performed to check whether a particular element of $\boldsymbol{\beta}(\omega)$ is zero. Details on the statistical hypothesis test can be found in [67] and [37], the computational aspects of the technique are described below.

Assume that $\lambda_1(\omega) = \lambda(\omega), \lambda_2(\omega), \ldots, \lambda_n(\omega)$ are the eigenvalues of $\hat{\mathbf{P}}_{\mathbf{x}}(\omega)$ arranged in decreasing order, and $\boldsymbol{\beta}_1(\omega) = \boldsymbol{\beta}(\omega), \boldsymbol{\beta}_2(\omega), \ldots, \boldsymbol{\beta}_n(\omega)$ are the corresponding eigenvectors. The asymptotic covariance matrix of the sample optimal scaling vector $\hat{\boldsymbol{\beta}}(\omega)$ is given by

$$\mathbf{V}_{\beta}(\omega) = v^{-2}\lambda_1(\omega)\sum_{l=2}^{n}\lambda_l(\omega)[\lambda_1(\omega) - \lambda_l(\omega)]^{-2}\boldsymbol{\beta}_l(\omega)\boldsymbol{\beta}_l^*(\omega), \qquad (4.5)$$

where $v = (\sum_{j=-r}^{r} h_j^2)^{-\frac{1}{2}}$.

The distribution of $2|\hat{\beta}_j(\omega) - \beta_j(\omega)|^2/\sigma_j(\omega)$ is approximately a Chi-square distribution with 2 degrees of freedom, where $\hat{\beta}_j(\omega), j = 1, \ldots, n$, is the jth element of the estimated optimal scaling vector, $\beta_j(\omega)$ is the jth element of the optimal scaling vector, and $\sigma_j(\omega)$ is the jth diagonal element of $\mathbf{V}_{\beta}(\omega)$. If $2|\hat{\beta}_j(\omega)|^2/\sigma_j(\omega) > \chi_2^2(\alpha)$, then the null hypothesis $\beta_j(\omega) = 0$ is rejected with $(1-\alpha)$ confidence, namely, the corresponding time series can be treated as having oscillation at frequency ω.

For the benchmark data set, in order to determine which variables have oscillation at the frequency of 0.0032 cycles/sample, we calculate the test

Table 4.1: *pv*s having oscillations at 320 samples/cycle.

Tag no.	Test statistic
LC1.*pv*	2120
FC1.*pv*	1760
TC1.*pv*	1140
PC2.*pv*	407
FC5.*pv*	527
LC2.*pv*	655
FC8.*pv*	253
TC2.*pv*	236

statistic $2|\hat{\beta}_j(\omega)|^2/\sigma_j(\omega)$ for $j = 1, 2, \ldots, 14$, where $\omega = 0.0032$, $\hat{\beta}_j(\omega)$ is the jth element of the calculated optimal scaling vector, namely, the jth element of the eigenvector of $\mathbf{P}_{\mathbf{x}}(\omega)$ corresponding to the largest eigenvalue, and $\sigma_j(\omega)$ is the jth diagonal element of $\mathbf{V}_{\beta}(\omega)$ calculated by (4.5). The variables that have the test statistic value bigger than $\chi^2_2(0.001) = 13.82$ at the oscillation frequency are shown in Table 4.1. We can conclude that the listed variables have common oscillations with 99.9% confidence level. These variables are marked by dark circle symbols in Fig. 3.9.

b) Root cause diagnosis: Since the magnitude of the optimal scalings represents the strength of the contribution of each oscillating variable to the plant-wide oscillations, an oscillation contribution index (OCI) of one oscillating variable $x_j(t)$ is defined to be [37]

$$\mathrm{OCI}_j(\omega) = \frac{|\hat{\beta}_j(\omega)|}{2\sigma_{\hat{\beta}}(\omega)}, \tag{4.6}$$

where $\sigma_{\hat{\beta}}(\omega)$ is the standard deviation of the magnitude of the optimal scalings of all the identified variables that have oscillations (see Table 4.1). OCI is used to isolate the key variables as the potential root cause candidates of the common oscillations. A general criterion is that the variables having $\mathrm{OCI}(\omega) > 1$ are the likely root cause variables at frequency ω because they contribute most to the spectral envelope peak at the plant-wide oscillation frequency.

For the benchmark data set, the OCI of each oscillating variable listed in Table 4.1 is calculated according to (4.6). Table 4.2 shows the variables that have OCI bigger than 1 at the oscillation frequency in a descending order. These variables are the root cause candidates. Among all the variables,

Table 4.2: Ranked list of *pv*s having OCI bigger than 1.

Tag no.	OCI
LC2.*pv*	1.41
TC1.*pv*	1.29
FC8.*pv*	1.16
TC2.*pv*	1.15
FC5.*pv*	1.07
FC1.*pv*	1.07

LC2.*pv* has the largest OCI at the oscillation frequency. This result indicates that the LC2 loop contributes most to the spectral envelope at the oscillation frequency; thus, we should take this loop as the first root cause candidate. This result is consistent with the fact that the root cause is due to the valve stiction in the LC2 control loop. Although the spectral envelope method provides an effective way to identify the likely root cause, it cannot tell the fault propagation pathways.

3) Physical Interpretation of the Spectral Envelope Method: Within the concept of spectral envelope, it is difficult to understand the physical interpretation of the linear combination $y(t, \boldsymbol{\beta}) = \boldsymbol{\beta}^* \mathbf{x}(t)$ when $\boldsymbol{\beta}$ is a complex vector. In this chapter, we give another physical interpretation of the spectral envelope method.

Since $\boldsymbol{\beta}$ is an n-dimentional column vector which may be real or complex, we assume

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} = \begin{bmatrix} \alpha_1 e^{\theta_1 i} \\ \alpha_2 e^{\theta_2 i} \\ \vdots \\ \alpha_n e^{\theta_n i} \end{bmatrix}; \tag{4.7}$$

it is obvious that $\alpha_j, j = 1, 2, \ldots, n$, is the magnitude of β_j and $\alpha_j \geq 0$. We also assume that the Fourier transformation of $\mathbf{x}(t)$ at frequency ω is

$$\mathbf{F}_{\mathbf{x}}(\omega) = \begin{bmatrix} f_1(\omega) \\ f_2(\omega) \\ \vdots \\ f_n(\omega) \end{bmatrix} = \begin{bmatrix} p_1 e^{\gamma_1 i} \\ p_2 e^{\gamma_2 i} \\ \vdots \\ p_n e^{\gamma_n i} \end{bmatrix}, \tag{4.8}$$

where $f_j(\omega)$ is the Fourier transformation of $x_j(t)$, p_j is the amplitude, and γ_j is the corresponding phase.

We have shown that (see Appendix) for a normalized vector-valued time series $\mathbf{x}(t)$, its optimal scaling vector satisfies that

$$(\gamma_1 - \theta_1) = (\gamma_2 - \theta_2) = \cdots = (\gamma_n - \theta_n).$$

This means that to make the power of the scaled time series at frequency ω maximum, the phase difference between each time series at frequency ω should be eliminated by introducing the optimal scaling vector. In other words, given a certain frequency ω, the physical interpretation of phases of the optimal scalings, namely, θ_j for $j = 1, 2, \ldots, n$, is to shift time series $x_j(t)$ by phase θ_j in the time domain such that all the shifted time series have the same phase at frequency ω. At the same time, the magnitudes of the elements of the scaling vector, namely, α_js, represent the scaling coefficients that scale amplitudes of these shifted time series. The power of the optimal scaled time series $y(t, \boldsymbol{\beta})$ at frequency ω is equivalent to the power of the summation of these shifted and scaled time series at frequency ω, namely,

$$\mathbf{P}_y(\omega) = (\alpha_1 p_1 + \alpha_2 p_2 + \cdots + \alpha_n p_n)^2.$$

Physically this means that the optimal scaled time series is the summation of these shifted and scaled time series. Since there is no phase difference between these shifted time series, the power of the summation of them must be maximum.

We have also shown (in the Appendix) that the optimal scaling vector and the spectral envelope respectively satisfy

$$\alpha_1 : \alpha_2 : \cdots : \alpha_n = p_1 : p_2 : \cdots : p_n, \tag{4.9}$$

$$\alpha_j = \frac{p_j}{\sqrt{(p_1^2 + p_2^2 + \cdots + p_n^2)}} \text{ for } j = 1, 2, \ldots, n, \tag{4.10}$$

and

$$\lambda(\omega) = p_1^2 + p_2^2 + \cdots + p_n^2. \tag{4.11}$$

Details on the proof are shown in the Appendix. From (4.9) and (4.10), we can see that for a given frequency ω, the magnitude of the optimal scaling, namely, $\alpha_j = |\beta_j|$, is proportional to the amplitude of the Fourier transformation of the corresponding time series $x_j(t)$, namely, p_j. From (4.11), we conclude that the spectral envelope at frequency ω (the power of the optimal scaled time series) is equivalent to the power summation of all the time series at frequency ω. It is obvious that for a certain frequency, the larger the magnitude of the optimal scaling, the more contribution of the corresponding time series to the spectral envelope.

4.3.2 Adjacency Matrix Method

The adjacency matrix method provides an effective way to capture a process topology. For diagnosis of plant-wide oscillations, it should be used together with another data-based method since the adjacency matrix method cannot tell if there is a plant-wide oscillation or which variables have oscillations. Thus, detection of plant-wide oscillations via a data-based method needs to be conducted when we use the adjacency matrix method for diagnosis of oscillations. In this thesis, the oscillation detection results from the spectral envelope method are used for oscillation diagnosis via the adjacency matrix method.

1) Concept of the Adjacency Matrix: A directed graph or a digraph represents the structural relationships between discrete objects [48]. The adjacency matrix is a common tool to represent digraphs, which provides one way to express the process topology. The concept of adjacency matrix was successfully developed and applied to root cause diagnosis of plant-wide oscillations in [38].

Digraphs are established by representing the process variables as graph nodes, representing the existence of relationship between two variables by edges. If a sense of direction is imparted to each edge of a graph, such a graph is called a directed graph. A directed graph can be easily converted into an adjacency matrix. In an adjacency matrix, both the rows and columns represent nodes. If there is a directed edge (an arc) from node i to node j, then the value of (i, j)th entry is set to be 1; otherwise it is 0.

Let $\mathbf{X} \in \mathbb{R}^{n\times n}$ denote the adjacency matrix of one digraph with n nodes. The (i, j)th element of $\mathbf{X}^k$ means the number of k-step edge sequences from node i to node j. We denote $\mathbf{A} = \mathbf{X} + \mathbf{X}^2 + \cdots + \mathbf{X}^n$. The reachability matrix of $\mathbf{X}$ is defined as the Boolean equivalence of matrix $\mathbf{A}$:

$$\begin{aligned} \mathbf{R} &= \mathbf{A}^{\#} \\ &= (\mathbf{X} + \mathbf{X}^2 + \cdots + \mathbf{X}^n)^{\#} \end{aligned} \tag{4.12}$$

where # denotes the Boolean operator, i.e.,

$$\mathbf{A}^{\#}(i, j) = \begin{cases} 1, & \text{if } \mathbf{A}(i, j) \neq 0, \\ 0, & \text{if } \mathbf{A}(i, j) = 0. \end{cases}$$

The reachability matrix represents process topology since its (i, j)th element indicates whether there is any path of any length from node i to node j.

2) Usefulness of the Adjacency Matrix Method: The (i, j)th element of the reachability matrix ($\mathbf{R}$) indicates whether there exists any directed path from

node i to node j. If the value of (i, j)th element of $\mathbf{R}$ is 1, then signals can propagate from node i to node j through one or more paths; otherwise, variation in node i cannot affect node j. Thus, among many oscillating variables (nodes), which can be detected using the spectral envelope method, if there exists one node that all elements of its corresponding row in $\mathbf{R}$ have values of 1 and all elements of its corresponding column have values of 0, then this node is most likely the root cause of the oscillations.

Because of feedback and/or feedforward control and other physical connections in a process, an oscillation often starts from a single control loop and propagates to other loops. In order to diagnose the control loop which causes plant-wide oscillations, a control loop digraph is defined in [38]. In this digraph, each controller in a process schematic is denoted as one node and an edge from node i to node j can be added if there is a direct interaction from node i to node j, i.e., if the controller output of controller $i(i.op)$ can directly affect the controlled variable of controller $j(j.pv)$ without going through controller output of any other nodes. The control loop digraph is constructed based on process information of the control structure and process flowsheet connections. It provides a measure of the domain of influence of each controller on other control loops. Then, the corresponding adjacency matrix and reachability matrix can be inferred. Based on the reachability matrix and oscillating variables detected by the spectral envelope method, the control loop(s) that may cause plant-wide oscillations can be determined.

For the benchmark case study, first we need to draw the control loop digraph of the Eastman chemical process as reported in [38]. There are 14 PID controllers, we take each controller as one node and add an edge from node i to node j if $i.op$ can directly affect $j.pv$. Fig. 4.3 [38] shows the control loop digraph of the Eastman chemical process. For example, node 1 and node 2 are the secondary and the master controllers in a cascade control loop. If the op of node 1 changes, then the pv of node 2 will be affected directly. Similarly, the op of node 2 has a direct influence on the pv of node 1. Therefore, we say that nodes 1 and 2 have direct interactions between them and we add edges between nodes 1 and 2. After a complete analysis of direct interactions between each pair of the nodes, the control loop digraph of this process is obtained as shown in Fig. 4.3.

Then, based on the control loop digraph, we construct the adjacency matrix as shown in Fig. 4.4(a). If there is an edge from node i to node j, then

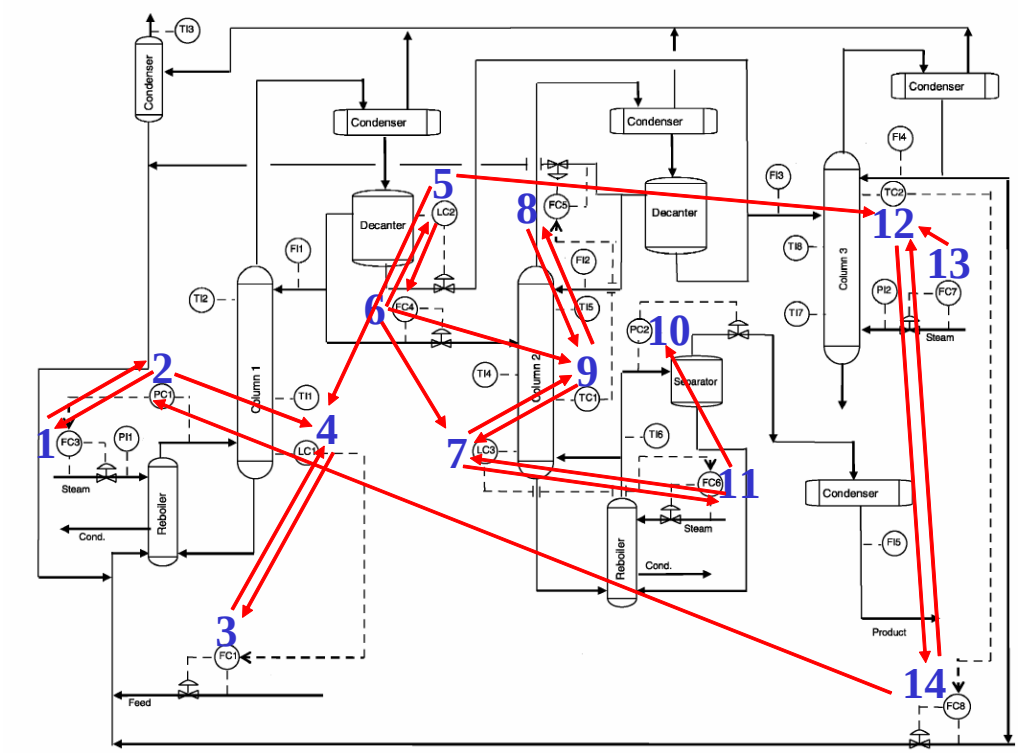


Figure 4.3: Control loop digraph of the process from Eastman Chemical Company [38].

		1	2	3	4	5	6	7	8	9	10	11	12	13	14
		FC3	PC1	FC1	LC1	LC2	FC4	LC3	FC5	TC1	PC2	FC6	TC2	FC7	FC8
1	FC3	1	1	0	0	0	0	0	0	0	0	0	0	0	0
2	PC1	1	1	0	1	0	0	0	0	0	0	0	0	0	0
3	FC1	0	0	1	1	0	0	0	0	0	0	0	0	0	0
4	LC1	0	0	1	1	0	0	0	0	0	0	0	0	0	0
5	LC2	0	0	0	1	1	1	0	0	0	0	0	1	0	0
6	FC4	0	0	0	0	1	1	1	0	1	0	0	0	0	0
7	LC3	0	0	0	0	0	0	1	0	1	0	1	0	0	0
8	FC5	0	0	0	0	0	0	0	1	1	0	0	0	0	0
9	TC1	0	0	0	0	0	0	1	1	1	0	0	0	0	0
10	PC2	0	0	0	0	0	0	0	0	0	1	0	0	0	0
11	FC6	0	0	0	0	0	0	1	0	0	1	1	0	0	0
12	TC2	0	0	0	0	0	0	0	0	0	0	0	1	0	1
13	FC7	0	0	0	0	0	0	0	0	0	0	0	1	1	0
14	FC8	0	1	0	0	0	0	0	0	0	0	0	1	0	1

(a) Adjacency matrix

		1	2	3	4	5	6	7	8	9	10	11	12	13	14
		FC3	PC1	FC1	LC1	LC2	FC4	LC3	FC5	TC1	PC2	FC6	TC2	FC7	FC8
1	FC3	1	1	1	1	0	0	0	0	0	0	0	0	0	0
2	PC1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
3	FC1	0	0	1	1	0	0	0	0	0	0	0	0	0	0
4	LC1	0	0	1	1	0	0	0	0	0	0	0	0	0	0
5	LC2	1	1	1	1	1	1	1	1	1	1	1	1	0	1
6	FC4	1	1	1	1	1	1	1	1	1	1	1	1	0	1
7	LC3	0	0	0	0	0	0	1	1	1	1	1	0	0	0
8	FC5	0	0	0	0	0	0	1	1	1	1	1	0	0	0
9	TC1	0	0	0	0	0	0	1	1	1	1	1	0	0	0
10	PC2	0	0	0	0	0	0	0	0	0	1	0	0	0	0
11	FC6	0	0	0	0	0	0	1	1	1	1	1	0	0	0
12	TC2	1	1	1	1	0	0	0	0	0	0	0	1	0	1
13	FC7	1	1	1	1	0	0	0	0	0	0	0	1	1	1
14	FC8	1	1	1	1	0	0	0	0	0	0	0	1	0	1

(b) Reachability matrix

Figure 4.4: Adjacency matrix and reachability matrix based on the control loop digraph.

the (i, j)th entry of the adjacency matrix is assigned a value of 1, otherwise it is assigned a value of 0. Note that the *op* of each node/controller has a direct interaction on the *pv* of itself; thus, all the diagonal elements of the table are set to be 1. The corresponding reachability matrix is shown in Fig. 4.4(b). The reachability matrix indicates the influence of a controller on another controller: where '1' denotes a link and '0' indicates no connection. From Fig. 4.4(b) we can see that nodes 5 and 6 can reach all the other nodes except node 13 and no other nodes can reach them.

By using the spectral envelope method for oscillation detection, we have already detected the oscillation frequency and isolated all the *pv*s that have

the common oscillation frequency as shown in Table 4.1. In Fig. 4.4(b), the controllers that have commonly oscillating *pv*s are highlighted in blue color. Based on the reachability matrix we conclude that: since loops 5 (LC2) and 6 (FC4) can reach all the detected oscillatory loops and they cannot be reached by any other oscillatory loops, the root cause should be either control loop 5 or loop 6. Based on this conclusion, we can further investigate these two loops and confirm possible root causes. Since it has been confirmed that valve stiction in loop 5 (LC2) is the root cause [38], we can see that the concept of adjacency matrix has also successfully suggested this loop as the potential root causes of plant-wide oscillation. The reason that both loops 5 and 6 are determined as root cause candidates is that there is interaction between these two loops. Similarly, if a plant-wide oscillation is generated within a reflow cycle, a group of all the loops within this reflow cycle will be regarded as root cause candidates by the adjacency matrix method.

4.3.3 Granger Causality Method

1) Concept of Granger causality: Granger causality is a measure of causal effect based on linear predictions of variables. According to Granger causality, we say that x_1 causes x_2 if the inclusion of past observations of x_1 reduces the variance of the prediction error of x_2 in a linear regression model of x_1 and x_2, as compared to a model which includes only previous observations of x_2 [29]. Granger causality has its time-domain version and frequency-domain version (called spectral Granger causality). Here we focus on the time-domain Granger causality method. Details on the spectral Granger causality can be found in [13, 21].

For two stationary time series $x_1(t)$ and $x_2(t)$ of length N, we can construct bivariate autoregressive (AR) models:

$$x_1(t) = \sum_{j=1}^{k} A_{11,j} x_1(t-j) + \sum_{j=1}^{k} A_{12,j} x_2(t-j) + \xi_{1|2}(t), \tag{4.13}$$

$$x_2(t) = \sum_{j=1}^{k} A_{21,j} x_1(t-j) + \sum_{j=1}^{k} A_{22,j} x_2(t-j) + \xi_{2|1}(t), \tag{4.14}$$

where k is the model order that defines the amount of lag considered, A's are the AR coefficients, and ξ's represent the prediction errors. The expressions in (4.13) and (4.14) are called a full model or unrestricted model. We can also

perform univariate AR modeling on each time series and obtain restricted models:

$$x_1(t) = \sum_{j=1}^{k} B_{1,j} x_1(t-j) + \xi_1(t), \tag{4.15}$$

$$x_2(t) = \sum_{j=1}^{k} B_{2,j} x_2(t-j) + \xi_2(t). \tag{4.16}$$

If variance of $\xi_{2|1}(t)$ is smaller than variance of $\xi_2(t)$, which means prediction of $x_2(t)$ is more accurate when including past values of x_1, then x_1 Granger causes x_2 and vice versa. The magnitude of interaction is measured by

$$F_{i\to j} = \ln \frac{\mathrm{var}(\xi_j)}{\mathrm{var}(\xi_{j|i})}, \tag{4.17}$$

where ξ_j is derived from the restricted model by only using $x_j(t)$ and $\xi_{j|i}$ is derived from the full model.

It is easy to generalize the bivariate Granger causality to multivariate case. For a system of n variables $(x_1, x_2, \ldots, x_n)$, x_i causes x_j if including x_i helps to predict x_j when all other variables are included in the regression models.
2) Usefulness of the Granger Causality Method: The AR coefficients can be calculated using the least square method and the model order k can determined by the Akaike Information Criterion (AIC) [3] or Bayesian Information Criterion (BIC) [61]. For a model of n variables, AIC and BIC are given as follows:

$$AIC(k) = \ln(det(\Sigma)) + \frac{2kn^2}{N}, \tag{4.18}$$

$$BIC(k) = \ln(det(\Sigma)) + \frac{\ln(N)kn^2}{N}, \tag{4.19}$$

where $\mathbf{\Sigma}$ is the residual covariance matrix of the full model and N is the number of observations.

A time domain Granger causality from x_i to x_j is significant if A_{ji}'s are jointly significant or large relative to zero. This is a hypothesis test problem. The null hypothesis is that A_{ji}'s are zero or there is no causality from x_i to x_j. Statistical significance can be determined via the F-statistical test [12]:

$$\frac{\frac{RSS_r - RSS_f}{k}}{\frac{RSS_f}{N-2k-1}} \sim F_{k,N-2k-1}, \tag{4.20}$$

where k is the model order, $RSS_r = \sum_{t=k+1}^{N} \xi_j^2(t)$ is the sum of squares of residual in the restricted model, $RSS_f = \sum_{t=k+1}^{N} \xi_{j|i}^2(t)$ is the sum of squares

of residual in the full model. The F-statistic approximately follows an F distribution with degrees of freedom k and $(N-2k-1)$. When the p-value is less than the significance level α (typically 0.01 or 0.05), the null hypothesis that there is no causality from x_i to x_j could be rejected.

For root cause diagnosis, 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 [37]. Therefore, we only need to find the information flow pathways among variables that have oscillations at the common frequency. For the benchmark data set, as shown in Table 4.1, the listed 8 process variables are the oscillating variables. As long as we capture the oscillation propagation pathways by using these variables, the possible root causes can be determined.

Granger causality is applied to capture the causality between each pair of the 8 oscillating variables. The BIC criterion is chosen to determine the model order. For the null hypothesis test, the significance level α is set to be 0.05. After calculation, the causal relationships between the 8 oscillating variables are shown in Fig. 4.5, 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 a bidirectional causality (also called causality feedback) between the two variables. Most of these causal relationships can be validated by the process schematic and the P&ID. For example, the bidirectional causality between LC1.pv and FC1.pv is generated by the cascade control strategy for the liquid level of Distillation Column 1.

The oscillation propagation pathways obtained from the causal map (see Fig. 4.5) are shown in Fig. 4.6, where One-headed arrows indicate unidirectional causality and double-headed arrows indicate bidirectional causality. Note that the bidirectional propagation pathways are generated by the cascade feedback control structure, which are consistent with the physical process. From Fig. 4.5 and Fig. 4.6, we can see that there are two process variables, LC2.pv and PC2.pv, that have causal effects to other 6 variables but do not receive any significant causal effects from any other variables. Thus, we may conclude that control loops LC2 and PC2 are the likely root cause candidates. Since the root cause of the plant wide oscillation is valve stiction in the actuator of control loop LC2, the causality analysis via the Granger causality method is effective in determining root cause candidates.

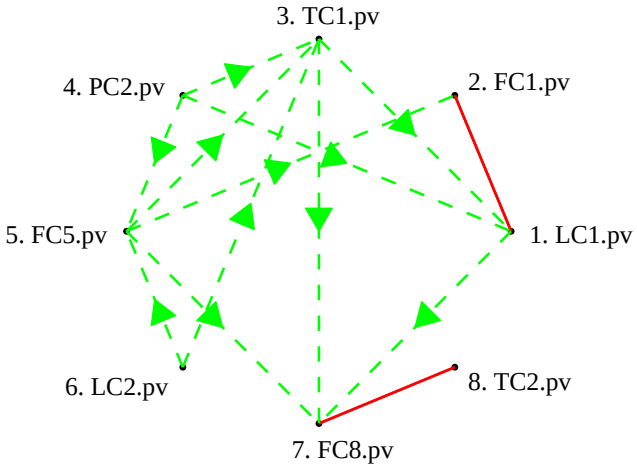


Figure 4.5: Causal map of 8 oscillating variables via the Granger causality method. A dashed line with an arrow indicates unidirectional causality and a solid line connecting two variables without an arrow indicates bidirectional causality.

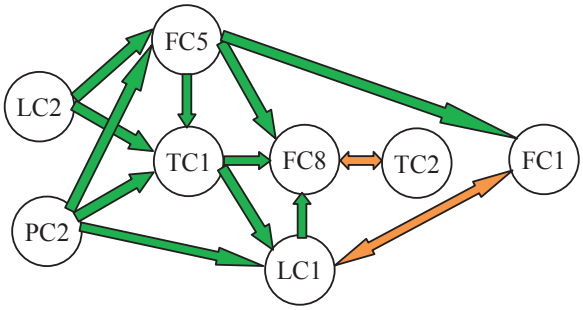


Figure 4.6: Oscillation propagation pathways obtained via the Granger causality method. One-headed arrows indicate unidirectional causality and double-headed arrows indicate bidirectional causality.

Table 4.3: Normalized transfer entropies for the benchmark data set.

$NTE^{c}_{\texttt{column 1}\rightarrow\texttt{row 1}}$	LC1	FC1	TC1	PC2	FC5	LC2	FC8	TC2
LC1	NA	0.23	0	0.05	0.09	0.04	0.01	0.01
FC1	0.20	NA	0	0.03	0.02	0.02	0	0
TC1	0.12	0.01	NA	0.13	0.34	0.03	0	0
PC2	0.08	0.02	0.16	NA	0.04	0.04	0.01	0.01
FC5	0.11	0.07	0.01	0.04	NA	0.03	0	0.01
LC2	0.22	0.02	0.17	0.04	0.12	NA	0.15	0.19
FC8	0.14	0.19	0	0.04	0.02	0.03	NA	0.01
TC2	0.13	0.01	0	0.04	0.03	0.03	0.35	NA

4.3.4 Transfer Entropy Method

Since the TE method and its usefulness have been extensively studied in Chapter 2, in this section, we omit description of this method and only introduce its application to the benchmark data set.

The TE approach described in Chapter 2 is used for root cause analysis of the plant-wide oscillations. First the causal relationships between the 8 oscillating variables are detected by calculating normalized differential transfer entropies according to (2.18). After calculation, the normalized transfer entropies between each pair of the 8 variables are shown in Table 4.3 where we omit *.pv* in the tag names of variables for simplicity. We can choose the threshold as 0.05: if the normalized transfer entropy is less than or equal to 0.05, then we conclude that there is almost no causality. The causal map based on the normalized transfer entropies is shown in Fig. 4.7, where we use the same symbols as in Fig. 4.5. The causal map shows a complicated set of pathways from which finding faults propagation pathways would be difficult, thus, we need to further determine whether the present causality is true and direct by calculating normalized direct transfer entropies according to (2.20).

Table 4.4 shows calculated normalized DTEs between each pair of the variables that have causal relationship and have possible intermediate variable(s). Note that if a pair of the variables does not have significant causal relationship based on the calculation results of TEs shown in Table 4.3, then we do not need to calculate its DTE and thus put 'NA' in Table 4.4. If a pair of the

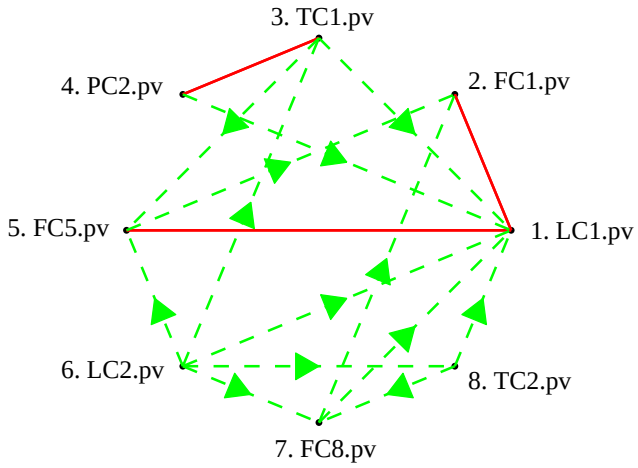


Figure 4.7: Causal map of 8 oscillating variables based on calculation results of normalized transfer entropies. A dashed line with an arrow indicates unidirectional causality and a solid line connecting two variables without an arrow indicates bidirectional causality.

variables has significant causal relationship but does not have any intermediate variable, then we also do not need to calculate its DTE since the causality must be true and direct, and thus assign a value of '1' in the table. If the calculated DTE is larger than 0.05, then we may conclude that the causality is true and direct and keep that information flow pathway; otherwise there is no direct causality, and we can eliminate the information flow pathway in Fig. 4.7. The causal map based on calculation results of normalized DTEs is shown in Fig. 4.8 which is much sparser than the previous causal map as shown in Fig. 4.7.

From the causal map in Fig. 4.8, we can see that LC2.*pv* has causal effects on all the other variables but does not receive any significant causal effects from any other process variables. Thus, we may conclude that control loop LC2 is likely the root cause candidate. The oscillation propagation pathways obtained from the causal map (see Fig. 4.8) are shown in Fig. 4.9. This figure shows that the control loop LC2 can reach all the other loops and the oscillation in loop LC2 first propagates to loops TC1, LC1, TC2 and FC8. From Fig. 3.9, we can see that there are direct material flow pathways from the left hand side

Table 4.4: Normalized DTEs for the benchmark data set.

$NDTE^c_{\texttt{column 1}\rightarrow\texttt{row 1}}$	LC1	FC1	TC1	PC2	FC5	LC2	FC8	TC2
LC1	NA	0.33	NA	NA	0.03	NA	NA	NA
FC1	0.27	NA	NA	NA	NA	NA	NA	NA
TC1	0.02	NA	NA	1	0.51	NA	NA	NA
PC2	0.02	NA	1	NA	NA	NA	NA	NA
FC5	0.04	0.01	NA	NA	NA	NA	NA	NA
LC2	0.10	NA	1	NA	0.03	NA	0.46	1
FC8	0.03	0.63	NA	NA	NA	NA	NA	NA
TC2	0.02	NA	NA	NA	NA	NA	0.74	NA

decanter to Columns 1, 2 and 3. Thus, the oscillation propagation pathways obtained from the transfer entropy method are consistent with the physical process.

We note that although the conclusion that the control loop LC2 is likely a root cause candidate is consistent with the Granger causality analysis results, there is an obvious difference between Figs 4.5 and 4.8. Especially, bidirectional causal relationships between control loops PC2 and TC1 are found via the transfer entropy method. That is why PC2 is no longer a root cause candidate according to Fig. 4.8. This conclusion is consistent with the fact that the root cause is valve stiction in the control loop LC2. Note that the bidirectional causality between PC2.pv and TC1.pv is also found via the T0E approach described in Chapter 3, and the conclusion drawn from the TE approach is consistent with that from the T0E method.

In summary, for the three causality analysis methods, although there is difference between the causal maps (see Figs 3.12, 4.5, and 4.8), and between the oscillation propagation pathways (see Figs 3.13, 4.6, and 4.9), and the conclusions on root cause candidates are not exactly the same, all three methods (the T0E method, the Granger causality method, and the transfer entropy method) provide an effective way to capture the fault propagation pathways and locate the likely root cause candidates.

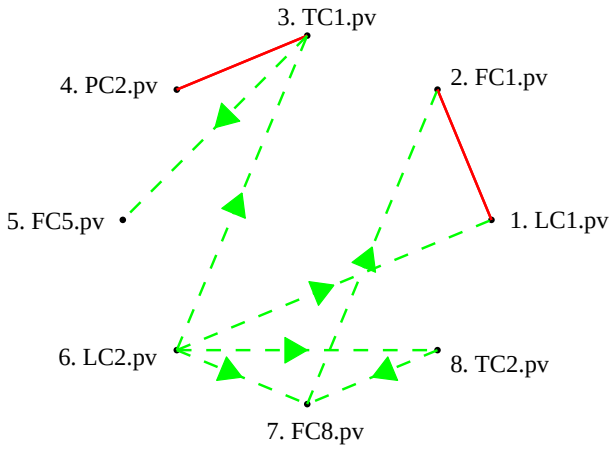


Figure 4.8: Causal map of 8 oscillating variables based on calculation results of normalized DTEs. A dashed line with an arrow indicates unidirectional causality and a solid line connecting two variables without an arrow indicates bidirectional causality.

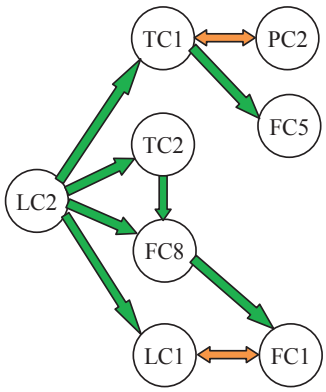


Figure 4.9: Oscillation propagation pathways obtained via the transfer entropy method. One-headed arrows indicate unidirectional causality and double-headed arrows indicate bidirectional causality.

4.3.5 Bayesian Network Structure Inference Method

1) Concept of Bayesian Networks: A Bayesian network is a specific type of graphical model which is a directed acyclic graph [11]. Each arc in the model is directed and acyclic. The set of nodes represents a set of random variables, and the arcs represent statistical dependence between the upstream variables and the downstream variables. The upstream variables are also called the parent variables of the downstream variables.

Bayesian network(s) (BN) provide an approach to dealing with uncertainty through the use of probability theory. Process topology can be captured by learning the BN structure. There are three broad classes of algorithms for BN structure inference: score-based approaches, constraint-based approaches and Bayesian model averaging approaches [44].

Score-based approaches are the most widely used methods for BN structure inference. These approaches address structure inference as a statistical model selection problem. We define a set of possible network structures (graphs) and a scoring function that measures how well each structure fits the observed data. The computational task is to search all possible structures and find the highest-scoring network structure. The key point of BN structure inference is the scoring function and the search algorithm.

There are two popular choices of the scoring function: likelihood score and Bayesian score. For a BN structure $\mathcal{G}$ with n nodes (variables) $x_1, x_2, \ldots, x_n$, given a particular observed data set $\mathcal{D} = \mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^N$ with a length of N and $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T$, let x_i^j denote the jth observation of the ith variable for $j = 1, 2, \ldots, N$ and $i = 1, 2, \ldots, n$. For this case, the likelihood of a parameter set $\theta_{\mathcal{G}}$ is

$$\begin{aligned} L(\theta_{\mathcal{G}} : \mathcal{D}) &= p(\mathcal{D}|\theta_{\mathcal{G}}) \\ &= \prod_{j=1}^{N} p(\mathbf{x}^j|\theta_{\mathcal{G}}), \\ &= \prod_{i=1}^{n}\prod_{j=1}^{N} p(x_i^j|Pa_{x_i^j}, \theta_{\mathcal{G}}), \end{aligned} \tag{4.21}$$

where $\theta_{\mathcal{G}}$ are the parameters that define the conditional probability of x_i given its parents, $Pa_{x_i^j}$ denotes parents of x_i^j and the parameters can be estimated by the maximum likelihood estimation (MLE) method [95]. Then, the likelihood score is defined as the logarithm of the likelihood function. Note that in

(4.21) there are two assumptions: one is that each observation is independent; the other is called the local Markov assumption [44], namely, each node x_i is independent of its non-descendants given its parents.

An alternative scoring function is the Bayesian information criterion (BIC) score defined as follows:

$$\text{score}_{BIC}(\mathcal{G} : \mathcal{D}) = \log p(\mathcal{D}|\theta_{\mathcal{G}}) - \frac{d}{2} \log N, \tag{4.22}$$

where d is the number of parameters, N is the length of the observed data set D, the expression of $p(\mathcal{D}|\theta_{\mathcal{G}})$ is the same as (4.21). The term of $\frac{d}{2} \log N$ is regarded as a penalty term in order to balance simplicity and accuracy of the model structure.

We now have a well-defined optimization problem. Our desired output is a network structure within the space of possible structures that maximizes the likelihood or the BIC score. Since the number of possible graphs increases exponentially with the number of nodes, some search algorithms are required. There are several search algorithms that can be applied; such as greedy structure search [44], annealing search, genetic algorithm search and $K2$ algorithm [95].

2) Usefulness of the Bayesian Network Structure Inference Method: Traditional Bayesian structure inference does not contain any time delay information, while one key point about causality is that "the cause occurs before the effect"; the temporal lag between the cause and the effect is also an important indicator of the direction of the signal propagation. In order to capture the time information, Zou and Feng (2009)[95] denoted a certain time lag for a specific variable by one node. Specifically, each variable $x_i, i = 1, 2, \ldots, n$ can be interpreted by a sequence of nodes $\{x_i^k, x_i^{k-1}, \ldots, x_i^{k-l}\}$, where $k, k-1, \ldots, k-l$ denote the time instants and $(l+1)$ nodes are used to denote the current and past information of x_i, in such way the time information is effectively captured. Therefore, there are totally $(l+1)n$ nodes in the Bayesian structure inference.

In this chapter, the Bayesian network inference is implemented via the following steps [95]:

1) Since there should be a time delay from the cause (parent node) to the effect (child node), the potential parent set for each node is determined to be all the nodes before it. The $K2$ algorithm [95] is used to deter-

mine the best parent nodes from the potential parent set for each node independently. In this way, a space of possible graphs is determined.

2) For each possible graph, the conditional probability of each node given its parents is estimated under an assumption that the data set has a Gaussian distribution [95]; and a corresponding score for this graph is obtained according to the BIC scoring function in (4.22).

3) The best network is determined as the graph with the highest score among all the possible graphs. Finally, causality is detected by checking parent nodes for each node according to the best network.

The BN inference method is applied to capture causality between the 8 oscillating variables for the benchmark data set. In order to include past information of variables, the lag node order for each variable is chosen to be 3, which means that each variable x_i for $i = 1, 2, \ldots, 8$ is represented by three lag-compensated nodes: x_i^k, x_i^{k-1}, and x_i^{k-2} representing the current information of x_i at the time instant k and its past information at time instants $k-1, k-2$, respectively. Note that increasing the lag order will increase the computational burden. The larger the order of lags within a certain range, the more accurate the obtained structure is. Here a certain range is similar to the embedding dimension of the embedding vectors with elements from the past values of each variable, which includes all the useful past information of each variable for forecasting other variables.

Fig. 4.10 shows the causal relationships between the 8 oscillating variables. The oscillation propagation pathways obtained from the causal map are shown in Fig. 4.11. From Figs 4.10 and 4.11, we can see that there are two variables LC2.pv and FC1.pv that do not receive causal effects from any other variables. LC2 can reach all the other loops except FC1, and FC1 can reach all the other loops except LC2 and TC2. Thus, we may conclude that loop LC2 is the first root cause candidate and loop FC1 is the second root cause candidate. Fig. 4.11 also shows that the oscillation of loop LC2 propagates to loops TC2 and FC5 first. By combining this information with the process schematic shown in Fig. 3.9, one can conclude that the oscillation of loop LC2 first propagates through material flow pathways from the left hand side decanter to columns 2 and 3, and then propagates to other loops. We can see that reasonable root cause candidates can also be found via the BN

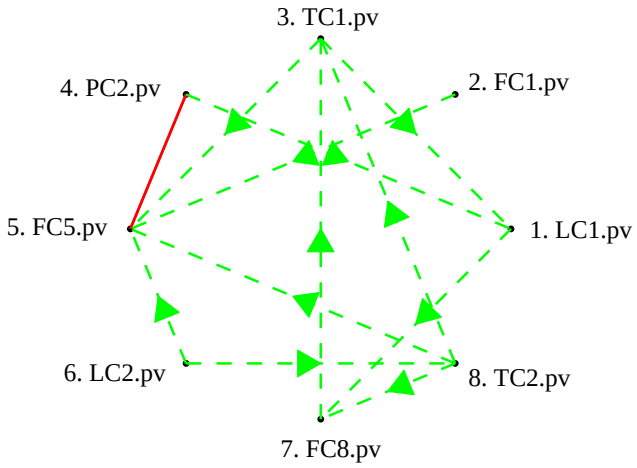


Figure 4.10: Causal map of 8 oscillating variables via the BN inference method. A dashed line with an arrow indicates unidirectional causality and a solid line connecting two variables without an arrow indicates bidirectional causality.

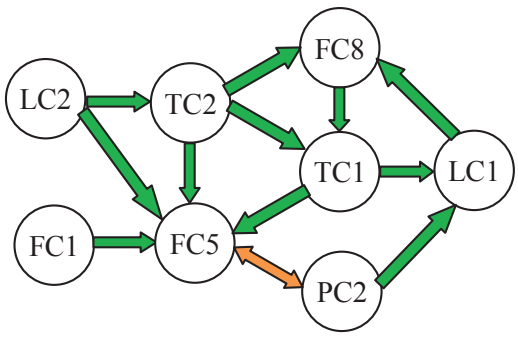


Figure 4.11: Oscillation propagation pathways obtained via the BN inference method. One-headed arrows indicate unidirectional causality and double-headed arrows indicate bidirectional causality.

structure inference method. However, some causal relationships are not captured whereas they are captured by the Granger causality method, the TE method, and the T0E method, for example, the causal relationships between loops LC1 and FC1.

4.4 Discussion

We can see that, to a certain extent, the industrial case study shows the effectiveness of all the methods we introduced for correctly identifying potential root causes of plant-wide oscillations. However, each method has its assumptions/conditions, advantages and disadvantages. The following remarks summarize the key findings in this comparative industrial case study.

4.4.1 Conditions and/or Assumptions

1) Spectral Envelope Method: This method is a frequency domain data-based method. The abnormal time series data with oscillations present is a necessary requirement. It can be used for both detection and diagnosis of plant wide oscillation(s). For root cause diagnosis, the main assumption of this method is that the root cause candidates are those variables that have relatively larger power at the specific oscillation frequency. Note that normalization of the data is necessary for the spectral envelope analysis.
2) Adjacency Matrix Method: It is a process knowledge-based method. One key condition of using this method is that the process knowledge must be given as would be available in a P&ID, process information of the control structure and process flowsheet connections. If a control loop digraph is developed to capture connectivity between the control loops, then the implicit assumption is that the root cause variables lie within these loops.
3) Causality Analysis Methods: The causality analysis methods: Granger causality, transfer entropy, and transfer 0-entropy are process data-based and the latter two are information-theory based methods. As long as the root cause of oscillations does not change causal relationships between the variables, both normal time series data and abnormal time series data can be used for root cause analysis. Both Granger causality and TE require that the collected sampled data must be wide-sense stationary with a large data length which is preferred to be no less than 2000 observations [7]. Stationarity requires that the dynamical properties of the system must not change during

the observation period. The T0E method does not need this requirement since the T0E is defined without assuming a probability space. For all of the three methods, the system should be sufficiently excited, especially for the Granger causality method since this method is based on the results of modeling. Since both the TE method and the T0E method are based on information theory, normalization of the data set is not necessary. Since the T0E is only related to a ratio of conditional ranges of one variable, random scalings (normalization) of the variable will not change this ratio. For the TE method, the key point is the (joint) PDFs of the variables, normalization of the data set will not change the ratio of conditional PDFs or the result of calculated transfer entropies.

Let us give a simple example to illustrate that normalization/scaling will not change the results of the transfer entropy method. Assume two correlated continuous random variables x and y satisfying:

$$y_{k+1} = ax_k + by_k + v_k, \tag{4.23}$$

where $x_k \sim N(0, \varepsilon^2)$ and $v_k \sim N(0, \sigma^2)$. Then, let $k_1 = l_1 = h_1 = \tau_1 = 1$, according to (2.1), it can be shown that

$$T_{x\to y} = \frac{1}{2}\log\left(1 + \frac{a^2\varepsilon^2}{\sigma^2}\right), \tag{4.24}$$

which means that the transfer entropy for this simple linear system is only related to the coefficient of x_k and the signal to noise ratio. Assume that $\hat{x}_k = cx_k$ and $\hat{y}_k = dy_k$ where c and d are any real values, then substituting $x_k = 1/c\hat{x}_k$ and $y_k = 1/d\hat{y}_k$ into (4.23), we obtain

$$\hat{y}_{k+1} = \frac{ad}{c}\hat{x}_k + b\hat{y}_k + dv_k.$$

According to (4.24), we have

$$\begin{aligned} T_{\hat{x}\to\hat{y}} &= \tfrac{1}{2}\log\left(1 + \frac{\frac{a^2d^2}{c^2}\cdot c^2\varepsilon^2}{d^2\sigma^2}\right) \\ &= \tfrac{1}{2}\log\left(1 + \frac{a^2\varepsilon^2}{\sigma^2}\right). \end{aligned}$$

Thus, we can see that the results of the transfer entropy will not change after random scalings (normalization) of x and y.

4) Bayesian Network Structure Inference Method: This method is also process data-based. Similar to the causality analysis methods, both normal and abnormal time series data can be used. The required assumption of the observed

data is that each observation is independent and identically distributed. This assumption is strict for industrial process data since most kinds of process data, including level, flow rate, temperature, etc., are generally autocorrelated. Note that under this assumption, the sampled data are stationary.

4.4.2 Advantages and Application Limitations

The advantage of both the spectral envelope method and the adjacency matrix method is that the computational burden is very small since we only need to calculate the eigenvalue and eigenvectors of the PSD matrix for the spectral envelope method and only simple matrix multiplications and summations are involved in the adjacency matrix method. Both of them are relatively simple to implement. The spectral envelope method is robust to parameters (r and h_j) and data selection changes. The detection and diagnosis results will not change no matter which part of the data is selected as long as the time series is abnormal data with oscillations present. A limitation of the application of the spectral envelope method is that the physical explanation of the spectral envelope is not straightforward, which is sometimes regarded as abstract and therefore unpractical by engineers. The main limitation of the adjacency matrix method based on the control loop digraph is requirement of apriori process knowledge about the connectivity between all the control loops. This is not always easily available and drawing the control loop digraph needs some time and careful consideration.

Causality analysis methods provide an effective way to capture faults propagation pathways. The major advantages of the Granger causality method are that its theoretical meaning is easy to understand; and its application techniques are well developed. For example, the null hypothesis test of causality is well defined. It is a relatively simple method to implement. A limitation of the application of the Granger causality method is that the Granger causality method is based on AR models, which is suitable for linear multivariate processes, the problem of model misspecification may happen and thus the identified AR models may not be convincing. If the model structure is incorrect, then the residuals can hardly give evidence of causality between the signals considered in the models. For the nonlinear Granger causality (NLGC) method, its computational burden is extremely large. The reason is that it is a kernel-based method for which for every sample we need to calculate a radial-based kernel function and the location of its center (e.g., using c-means

clustering). An additional drawback of NLGC is its heavy memory cost. In this chapter, we only used traditional Granger causality to detect causality for further comparisons with other methods.

For the transfer entropy method, the major advantage is that this method can be used for both linear and nonlinear multivariate processes. Application limitations of this method are: a good parameters determination procedure is needed since the TE is sensitive to the parameters (h_1, τ_1, k_1, and l_1) changes, and the computational burden is large since we need to estimate joint PDFs. The computational burden is related to the dimensions of embedding vectors and the number of samples. Moreover, unlike Granger causality, the distribution of the sample statistic is unknown, rendering significance testing to be difficult without recourse to computationally expensive bootstrap method [78] or the Monte Carlo method [7] by constructing resampling data [71] or surrogate data (randomly shuffled data or by the iterative amplitude adjusted Fourier transform (iAAFT) method [60]).

For the transfer 0-entropy method, the outstanding advantage 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 assumption that the time series should be stationary is not required, while this assumption is a basic one for the transfer entropy method. Similar to the transfer entropy method, an application limitation of the T0E method is that a good parameters determination procedure is needed since the T0E is sensitive to the parameters (h, τ, k, and l) changes. Another limitation of the T0E method is that the causality detection results may be conservative. The possible reason for this is that the definition of 0-information in (3.9) is in the worst case which can be understood as the least information transferred from one variable to another.

For the BN structure inference method, a major advantage is that it can handle the data with a short size, while both the Granger causality method and the transfer entropy method require large data lengths. Disadvantages of the BN structure inference method include the assumption that each observation is independent; this assumption is too strict for industrial process data, and the computational burden is large since we need to estimate the (conditional) PDFs of the data set. The computational burden is related to the number of samples and the nodes size in the BN structure. Moreover, the results are sensitive to the lags in the nodes and score-based approaches are in general not

Table 4.5: Comparisons of the introduced methods for diagnosis of plant-wide oscillations.

Methods	**Requirements**	**Samples length**	**Advantages**	**Application limitations**
Spectral envelope	Data-based (Abnormal data; frequency domain)	Small	Easiest to implement; robust to parameters and data selection changes; computational burden is small; suitable for both detection and diagnosis	Physical explanation is not straightforward
Adjacency matrix	Model-based (Process knowledge)	Plant data is not required except to identify oscillating variables	Easy to implement; can be used for other disturbances as long as the process structure is not changed; Computational burden is small	Process knowledge is not always available; it is time consuming to construct a control loop digraph
Granger causality (Linear)	Data-based (Normal or abnormal data)	Large	Easier to implement; robust to data selection changes; Computational burden is not large; application techniques are well developed	Only suitable for linear relationships between variables; model misspecification may happen
Transfer entropy	Data-based (Normal or abnormal data; time domain)	Large	Robust to data selection changes; Suitable for both linear and nonlinear relationships between variables	Sensitive to parameters changes; Computational burden is large
Transfer 0-entropy	Data-based (Normal or abnormal data; time domain)	Medium	Time series do not need to be stationary; Robust to data selection changes; Suitable for both linear and nonlinear relationships between variables	Causality detection results may be conservative; Sensitive to parameters changes
BN inference	Data-based (Normal or abnormal data; time domain)	Smallest	Suitable for data with small size; robust to data selection changes	Sensitive to the lags in the nodes; relatively difficult to implement; assumptions on the data for BN construction are difficult to satisfy; optimal structure is not guaranteed

guaranteed to find the optimal solution [44]. Thus, this method is relatively difficult to implement.

Simple comparisons are shown in Table. 4.5 for the six methods. Note that as long as the data requirements for each data-based method are satisfied, the results should be robust to the data selection changes.

From Table 4.5, we can see that there are advantages and application limitations for each method. To detect plant wide oscillations and categorize the variables with oscillations, the spectral envelope method is recommended. Based on the oscillation detection results, we need to choose appropriate method(s) to determine root cause candidates. For data-based methods, unless the data size is very small (usually less than 500), the Bayesian network structure inference method is the last choice since the data requirement that each observation is independent and identically distributed is difficult to meet. After root cause candidates are determined, we need to check the root cause candidates one by one by further analysis and field test(s) until the root cause of plant-wide oscillations is confirmed.

4.5 Summary

This chapter has presented a survey of root cause diagnosis methods for plant-wide oscillations. Several methods for identifying possible root cause(s) of plant-wide oscillations are discussed; these methods include the spectral envelope method, the adjacency matrix method, the Granger causality method, the TE method, the T0E method, and the Bayesian network structure inference method. Among these methods, the spectral envelope method is the only one that can be used to not only detect plant-wide oscillations and categorize the oscillating variables but also diagnose root cause(s) of plant-wide oscillations. We have discussed the physical interpretation of the spectral envelope method and established a relationship between the spectral envelope and Fourier transformation. Other introduced methods are used for identifying root cause candidates based on the detection results of which variables have oscillations via the spectral envelope method. The effectiveness of these methods have been shown by application to a benchmark industrial data set. Advantages and limitations for applications of each method have been discussed. In this way, readers can choose an appropriate method to detect and/or diagnose root cause(s) under certain conditions and assumptions. There are three main recommendations:

1. For detection of plant-wide oscillations, the spectral envelope method is recommended. Firstly, it is relatively simple to implement and its computational burden is small. Secondly, it is robust to parameters and data selection changes. Thirdly, all the common frequency components show up as peaks on a single plot, namely, the spectral envelope plot. Finally, one can also obtain a list of variables that have oscillations at a certain plant-wide oscillation frequency via the Chi-square statistical test. This method can be applied to linear or non-linear systems.

2. For a locally linear system, the Granger causality method is recommended to find the oscillation propagation pathways and determine the root cause candidate(s) because of its well-developed technique and its relatively easy implementation.

3. For a strongly nonlinear system, the TE method and the T0E method are recommended for root cause analysis; if the time series is not stationary, then the T0E method is recommended. For example, in multimodal

processes, a data set with a large number of samples is most likely to be non-stationary as the data would reflect transitions from one mode to another; in this case, the T0E method is recommended.

Chapter 5

Summary and Future Work

5.1 Summary of Contributions

Plant-wide disturbances and abnormalities are common in many industrial processes. The purpose of the work reported in this thesis is to improve and develop techniques for causality analysis that can be applied to investigation of disturbance propagation pathways and determination of likely root cause(s) of such abnormalities. We focus on information theory-based causality detection methods which are suitable for both linear and nonlinear relationships, and their applications to root cause and fault propagation analysis.

In Chapter 2, we proposed a direct causality detection method based on the DTE to detect whether there is a direct information and/or material flow pathway between process variables of both linear and non-linear multivariate systems. The $\mathrm{DTE}_{\mathrm{diff}}$ for continuous-valued random variables and the $\mathrm{DTE}_{\mathrm{disc}}$ for discrete-valued random variables have been defined based on an extension of the transfer entropy. The $\mathrm{NTE}_{\mathrm{diff}}$ and the $\mathrm{NDTE}_{\mathrm{diff}}$ have been defined to measure the connectivity strength of causality and direct causality, respectively. The direct causality detection method is able to uncover explicit direct and indirect connectivity pathways between variables, and plays an important role in capturing the true process connectivity and finding fault propagation pathways.

In Chapter 3, we presented a new information theory-based method to detect causal relationships between process variables without assuming a probability space. This offers a distribution-free approach for causality detection. A transfer 0-entropy concept and a direct transfer 0-entropy concept were proposed to detect causality and direct causality, respectively. Estimation methods for the transfer 0-entropy and the direct transfer 0-entropy were ad-

dressed. The effectiveness of the T0E method was illustrated by two numerical examples, one experimental case study and one industrial case study. Compared to the TE method, the outstanding advantage of the newly developed 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 (conditional) ranges of the time series.

Chapter 4 addressed the application of causality analysis techniques to root cause and fault propagation analysis. Discussions and comparisons were given for three causality detection methods and another three widely used methods for root cause diagnosis of plant-wide oscillations; these methods include the Granger causality method, the TE method, the T0E method, the spectral envelope method, the adjacency matrix method, and the Bayesian network inference method. All the six methods were applied to the Eastman benchmark data set and a set of guidelines was provided on how to deal with the root cause diagnosis problem when plant-wide oscillations occur. Moreover, a physical interpretation of the spectral envelope method was presented. It turns out that for a given frequency, the magnitude of the optimal scaling for each time series is proportional to the amplitude of its normalized Fourier transformation.

5.2 Future Work

As a hot and emerging topic, the research on causality analysis and its application to root cause and fault propagation analysis is an active area of research. There are some issues that have not been fully explored and merit further exploration and research.

Joint PDF Estimation

For the transfer entropy method, the calculations of both TE and DTE need to estimate the high-dimensional joint PDFs; for example, the dimension of $f(Y_{i+h}, \mathbf{Y}_i^{(k)}, \mathbf{Z}_{i+h-h_3}^{(m_2)}, \mathbf{X}_{i+h-h_1}^{(l_1)})$ in (2.4) is $m_2 + l_1 + k + 1 \geq 3$. It is important to employ an accurate (less Type I and Type II errors) and efficient (less computational burden with a certain accuracy level) PDF estimation algorithm. Although the kernel estimation method is widely used, with the increasing dimension of the variables, a more accurate and efficient PDF estimation algorithm needs to be developed.

Confidence Level Determination of the DTE and DT0E

The detection of direct information flow can be reformulated as a hypothesis testing problem. Taking the direct causality from X to Y with an intermediate variable Z as an example, the null hypothesis should be that there is no direct casuality from X to Y and the causality from X to Y is indirect through Z. In order to carry out this hypothesis testing, we may use the bootstrap method [78] or the Monte Carlo method [7] by constructing resampling data or surrogate data (randomly shuffled data or by the iterative amplitude adjusted Fourier transform (iAAFT) method [60]). However, the constructed data must satisfy the null hypothesis that the direct information flow from X to Y must be completely destroyed while the indirect pathway through Z still exists. At the same time, the statistical properties of X, Y, and Z should not change. It is generally difficult to construct such surrogate or resampling data. Although the normalized DTE can determine the connectivity strength of direct causality and the threshold of T0E works well for the DT0E, the significance level analysis of the DTE and DT0E is still necessary for the threshold setting.

Model Misspecification of the Granger Causality Method

The Granger causality method has been widely used. However, a limitation of the application of the Granger causality method is that the Granger causality method is based on AR models, the problem of model misspecification may happen and thus the identified AR models may not be convincing. If the model structure is incorrect, then the residuals can hardly give evidence of causality between the signals considered in the models. Thus, the research on model misspecification of the Granger causality method would be interesting and valuable.

Implementation

We note that the results from a data-based method should be combined with the qualitative process information in root cause diagnosis. For example, the results of the causality analysis should be validated by the P&IDs or process flow diagrams (PFDs) of the process. Plant-wide disturbances detection and diagnosis remain an off-line method so far and cannot utilize process information automatically. Qualitative models of processes will become almost as

readily available as historical data in the future. A future research direction can be related to integrating data-based causality analysis techniques with automatic information extraction from process models [74]. In this way, a powerful diagnostic tool for isolating the root causes of plant-wide abnormalities can be developed.

Appendix A

Proof for the Spectral Envelope Method in Chapter 4

The definition of the spectral envelope for the normalized time series $\mathbf{x}(t)$ in (4.2) can be rewritten as

$$\lambda(\omega) = \sup_{\boldsymbol{\beta}\neq 0} \left\{ \frac{\boldsymbol{\beta}^* \mathbf{P}_{\mathbf{x}}(\omega)\boldsymbol{\beta}}{\boldsymbol{\beta}^*\boldsymbol{\beta}} \right\}, \text{s.t. } \boldsymbol{\beta}^*\boldsymbol{\beta} = 1. \tag{A.1}$$

Since the PSD matrix of $\mathbf{x}(t)$ can be estimated by $\mathbf{P}_{\mathbf{x}}(\omega) = \mathbf{F}_{\mathbf{x}}(\omega)\mathbf{F}_{\mathbf{x}}(\omega)^*$, where $\mathbf{F}_{\mathbf{x}}(\omega)$ is given in (4.8). Then, the PSD of the scaled series $y(t, \boldsymbol{\beta})$ at frequency ω is written as

$$\begin{aligned}
\mathbf{P}_y(\omega) &= \boldsymbol{\beta}^*\mathbf{P}_{\mathbf{x}}(\omega)\boldsymbol{\beta} \\
&= \boldsymbol{\beta}^*\mathbf{F}_{\mathbf{x}}(\omega)\mathbf{F}_{\mathbf{x}}(\omega)^*\boldsymbol{\beta} \\
&= \begin{bmatrix} \alpha_1 e^{-\theta_1 i} & \alpha_2 e^{-\theta_2 i} & \cdots & \alpha_n e^{-\theta_n i} \end{bmatrix} \begin{bmatrix} p_1 e^{\gamma_1 i} \\ p_2 e^{\gamma_2 i} \\ \vdots \\ p_n e^{\gamma_n i} \end{bmatrix} \\
&\quad \cdot \begin{bmatrix} p_1 e^{-\gamma_1 i} & p_2 e^{-\gamma_2 i} & \cdots & p_n e^{-\gamma_n i} \end{bmatrix} \begin{bmatrix} \alpha_1 e^{\theta_1 i} \\ \alpha_2 e^{\theta_2 i} \\ \vdots \\ \alpha_n e^{\theta_n i} \end{bmatrix} \\
&= (\alpha_1 p_1 e^{(\gamma_1-\theta_1)i} + \cdots + \alpha_n p_n e^{(\gamma_n-\theta_n)i}) \\
&\quad \cdot (\alpha_1 p_1 e^{(\theta_1-\gamma_1)i} + \cdots + \alpha_n p_n e^{(\theta_n-\gamma_n)i}) \\
&\leq (\alpha_1 p_1 + \alpha_2 p_2 + \cdots + \alpha_n p_n)^2.
\end{aligned} \tag{A.2}$$

In (A.2), the last line can be easily shown since α_j and p_j for $j = 1, 2, \ldots, n$ are nonnegative values, and the equivalence holds if and only if $(\gamma_1 - \theta_1) =$

$(\gamma_2 - \theta_2) = \cdots = (\gamma_n - \theta_n)$. Since θ_j can be arbitrarily chosen, the equivalence can be easily achieved.

From the frequency domain point of view, the Fourier transform of the scaled time series $y(t, \boldsymbol{\beta})$ is

$$f_{\mathbf{y}}(\omega) = \begin{bmatrix} \alpha_1 e^{-\theta_1 i} & \alpha_2 e^{-\theta_2 i} & \cdots & \alpha_n e^{-\theta_n i} \end{bmatrix} \begin{bmatrix} p_1 e^{\gamma_1 i} \\ p_2 e^{\gamma_2 i} \\ \vdots \\ p_n e^{\gamma_n i} \end{bmatrix}.$$

Thus, the condition of $(\gamma_1 - \theta_1) = (\gamma_2 - \theta_2) = \cdots = (\gamma_n - \theta_n)$ represents that to make the power of the scaled time series at frequency ω maximum, the phase difference between each time series at frequency ω should be eliminated by introducing the optimal scaling vector. From (A.2), we can see that the power of time series $y(t, \boldsymbol{\beta})$ at frequency ω, i.e., $\mathbf{P}_y(\omega)$, is equivalent to the power of the summation of these shifted and scaled time series at frequency ω.

Now we consider the constraint in (A.1), i.e., $\boldsymbol{\beta}^*\boldsymbol{\beta} = 1$. According to (4.7), we have $\boldsymbol{\beta}^*\boldsymbol{\beta} = \alpha_1^2 + \alpha_2^2 + \cdots + \alpha_n^2 = 1$. Since α_j represents the scaling coefficient used to scale amplitudes of the normalized time series $x_j(t)$, the energy (variance) of the corresponding scaled time series is α_j^2. Thus, the physical interpretation of this constraint is to make the energy summation of these scaled time series equal to 1.

By combining with (A.2), (A.1) can be rewritten as

$$\begin{aligned} \lambda(\omega) \quad = \quad & \sup_{\prod_{j=1}^m \alpha_j \neq 0} \left\{ (\alpha_1 p_1 + \alpha_2 p_2 + \cdots + \alpha_n p_n)^2 \right\}, \\ & s.t. \;\; \alpha_1^2 + \alpha_2^2 + \cdots + \alpha_n^2 = 1. \end{aligned} \tag{A.3}$$

To solve this optimization problem, we assume

$$G = (\alpha_1 p_1 + \alpha_2 p_2 + \cdots + \alpha_n p_n)^2 - \eta(1 - \alpha_1^2 - \alpha_2^2 - \cdots - \alpha_n^2).$$

Taking derivative of G, we have

$$\begin{cases} \dfrac{dG}{d\alpha_1} = 2p_1(\alpha_1 p_1 + \alpha_2 p_2 + \cdots + \alpha_n p_n)^2 + 2\eta\alpha_1 = 0 \\ \dfrac{dG}{d\alpha_2} = 2p_2(\alpha_1 p_1 + \alpha_2 p_2 + \cdots + \alpha_n p_n)^2 + 2\eta\alpha_2 = 0 \\ \quad \vdots \\ \dfrac{dG}{d\alpha_n} = 2p_n(\alpha_1 p_1 + \alpha_2 p_2 + \cdots + \alpha_n p_n)^2 + 2\eta\alpha_n = 0. \end{cases} \tag{A.4}$$

(A.4) can be rewritten as

$$\begin{cases} \eta\alpha_1 = -p_1(\alpha_1 p_1 + \alpha_2 p_2 + \cdots + \alpha_n p_n)^2 \\ \eta\alpha_2 = -p_2(\alpha_1 p_1 + \alpha_2 p_2 + \cdots + \alpha_n p_n)^2 \\ \quad\vdots \\ \eta\alpha_n = -p_n(\alpha_1 p_1 + \alpha_2 p_2 + \cdots + \alpha_n p_n)^2. \end{cases} \tag{A.5}$$

Thus, we have (4.9). Combining (4.9) with the constraint $\alpha_1^2+\alpha_2^2+\cdots+\alpha_n^2 = 1$, the solution to (A.3) is (4.10) and (4.11).

Bibliography

[1] Abnormal situation management consortium definition: impact. `http://www.asmconsortium.net/defined/impact/Pages/default.aspx`.

[2] A. Agresti and B. Finlay. *Statistical Methods for the Social Sciences*. Pearson Prentice Hall, Upper Saddle River, New Jersey, 4th edition, 2009.

[3] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.

[4] N. Ancona, D. Marinazzo, and S. Stramaglia. Radial basis function approach to nonlinear granger causality of time series. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 70(52):056221–1–056221–7, 2004.

[5] L.A. Baccala and K. Sameshima. Partial directed coherence: a new concept in neural structure determination. *Biological Cybernetics*, 84(6):463–474, 2001.

[6] L. Barnett, A.B. Barrett, and A.K. Seth. Granger causality and transfer entropy are equivalent for gaussian variables. *Physical Review Letters*, 103(23):238701–1–238701–4, 2009.

[7] M. Bauer, J.W. Cox, M.H. Caveness, J.J. Downs, and N.F. Thornhill. Finding the direction of disturbance propagation in a chemical process using transfer entropy. *IEEE Transactions on Control Systems Technology*, 15(1):12–21, 2007.

[8] 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. *Industrial and Engineering Chemistry Research*, 46(18):5977–5984, 2007.

[9] M. Bauer, N.F. Thornhill, and J.W. Cox. Measuring cause and effect between process variables. In *Proceedings of Advanced Process Control Applications for Industry Workshop*, Vancouver, Canada, 2005.

[10] J. Beirlant, E.J. Dudewicz, L. Gyorfi, and E.C. van der Meulen. Nonparametric entropy estimation: An overview. *International Journal of Mathematical and Statistical Sciences*, 6(1):17–39, 1997.

[11] C.M. Bishop. *Pattern Recognition and Machine Learning*. Springer Press, New York, 2006.

[12] S.L. Bressler and A.K. Seth. Wiener-granger causality: a well established methodology. *NeuroImage*, 58(2):323–329, 2010.

[13] P.E. Caines and C.W. Chan. Feedback between stationary stochastic processes. *IEEE Transactions on Automatic Control*, 20(4):498–508, 1975.

[14] S. Chang, C. Lin, and C. Chang. A fuzzy diagnosis approach using dynamic fault trees. *Chemical Engineering Science*, 57(15):2971–2985, 2002.

[15] M.A.A.S. Choudhury. Plantwide oscillations diagnosis-current state and future directions. *Asia-Pacific Journal of Chemical Engineering*, 6:484–496, 2011.

[16] M.A.A.S. Choudhury, V. Kariwala, N.F. Thornhill, H. Douke, S.L. Shah, H. Takada, and J. Fraser Forbes. Detection and diagnosis of plant-wide oscillations. *Canadian Journal of Chemical Engineering*, 85(2):208–219, 2007.

[17] M.A.A.S. Choudhury, S.L. Shah, and N.F. Thornhill. Diagnosis of poor control loop performance using higher order statistics. *Automatica*, 40(10):1719–1728, 2004.

[18] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[19] R.G. Cowell, A.P. Dawid, S.L. Lauritzen, and D.J. Spiegelhalter. *Probabilistic Networks and Expert Systems*. Springer, New York, 1999.

[20] A. Denis and F. Cremoux. Using the entropy of curves to segment a time or spatial series. *Mathematical Geology*, 34(8):899–914, 2002.

[21] M. Ding, Y. Chen, and S.L. Bressler. Granger causality: basic theory and application to neuroscience. In B. Schelter, M. Winterhalder, and J. Timmer, editors, *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications*. Wiley-VCH, 2006.

[22] P. Duan, F. Yang, T. Chen, and S.L. Shah. Direct causality detection via the transfer entropy approach. *IEEE Transactions on Control Systems Technology*, 21(6):2052–2066, 2013.

[23] U. Feldmann and J. Bhattacharya. Predictability improvement as an asymmetrical measure of interdependence in bivariate time series. *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering*, 14(2):505–514, 2004.

[24] R.A. Fisher. *The Design of Experiments*. Oliver and Boyd, Edinburgh, UK, 1935.

[25] L. Gao, Z. Ren, W. Tang, H. Wang, and P. Chen. Intelligent gearbox diagnosis methods based on svm, wavelet lifting and rbr. *Sensors*, 10(5):4602–4621, 2010.

[26] S. Gigi and A.K. Tangirala. Quantitative analysis of directional strengths in jointly stationary linear multivariate processes. *Biological Cybernetics*, 103(2):119–133, 2010.

[27] B. Gourevitch and J.J. Eggermont. Evaluating information transfer between auditory cortical neurons. *Journal of Neurophysiology*, 97(3):2533–2543, 2007.

[28] R.B. Govindan, J. Raethjen, F. Kopper, J.C. Claussen, and G. Deuschl. Estimation of time delay by coherence analysis. *Physica A: Statistical Mechanics and its Applications*, 350(2-4):277–295, 2005.

[29] C.W.J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438, 1969.

[30] C.W.J. Granger. Time series analysis, cointegration, and applications. *The American Economic Review*, 94(3):421–425, 2004.

[31] R.V.L. Hartley. Transmission of information. *Bell System Technical Journal*, 7(3):535–563, 1928.

[32] K. Hlavackova-Schindler. Equivalence of granger causality and transfer entropy: a generalization. *Applied Mathematical Sciences*, 5(73):3637–3648, 2011.

[33] K. Hlavackova-Schindler, M. Palus, M. Vejmelka, and J. Bhattacharya. Causality detection based on information-theoretic approaches in time series analysis. *Physics Reports*, 441(1):1–46, 2007.

[34] C. Hsu, C. Chang, and C. Lin. A practical guide to support vector classification. Tech. rep., Department of Computer Science and Information Engineering, National Taiwan University, 2010.

[35] B. Huang, N.F. Thornhill, S.L. Shah, and D. Shook. Path analysis for process troubleshooting. In *Proceedings of Advanced Control of Industry Processes*, pages 149–154, Kumamoto, Japan, 2002.

[36] M. Jelali and B. Huang. *Detection and Diagnosis of Stiction in Control Loops: State of the Art and Advanced Methods*. Springer-Verlag, London, 2010.

[37] 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. *Journal of Process Control*, 17(2):143–155, 2007.

[38] H. Jiang, R. Patwardhan, and S.L. Shah. Root cause diagnosis of plant-wide oscillations using the concept of adjacency matrix. *Journal of Process Control*, 19(8):1347–1354, 2009.

[39] P. Jizba, H. Kleinert, and M. Shefaat. Rényi's information transfer between financial time series. *Physica A: Statistical Mechanics and its Applications*, 391(10):2971–2989, 2012.

[40] R.A. Johnson and D.W. Wichern. *Applied Multivariate Statistical Analysis*. Prentice Hall, New Jersey, 4th edition, 1998.

[41] M.J. Kaminski and K.J. Blinowska. A new method of the description of the information flow in the brain structures. *Biological Cybernetics*, 65(3):203–210, 1991.

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

[43] E.J. Keogh, S. Chu, D. Hart, and M. Pazzani. Segmenting time series: a survey and novel approach. In M. Last, A. Kandel, and H. Bunke, editors, *Data Mining in Time Series Databases*. World Scientific Publishing Co. Pte. Ltd, Singapore, 2004.

[44] D. Koller, N. Friedman, L. Getoor, and B. Taskar. Graphical models in a nutshell. In L. Getoor and B. Taskar, editors, *Introduction to Statistical Relational Learning*. MIT Press, 2007.

[45] G.B. Lee, S.O. Song, and E.S. Yoon. Multiple-fault diagnosis based on system decomposition and dynamic pls. *Industrial and Engineering Chemistry Research*, 42(24):6145–6154, 2003.

[46] Q. Li and J.S. Racine. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, Princeton, N.J., 2007.

[47] M. Lungarella, K. Ishiguro, Y. Kuniyoshi, and N. Otsu. Methods for quantifying the causal structure of bivariate time series. *International Journal of Bifurcation and Chaos*, 17(3):903–921, 2007.

[48] R.S.H. Mah. *Chemical Process Structures and Information Flows*. Butterworth-Heinemann, 1990.

[49] M.R. Maurya, R. Rengaswamy, and V. Venkatasubramanian. A systematic framework for the development and analysis of signed digraphs for chemical processes. 1. algorithms and analysis. *Industrial and Engineering Chemistry Research*, 42(20):4789–4810, 2003.

[50] M.R. Maurya, R. Rengaswamy, and V. Venkatasubramanian. A systematic framework for the development and analysis of signed digraphs for chemical processes. 2. control loops and flowsheet analysis. *Industrial and Engineering Chemistry Research*, 42(20):4811–4827, 2003.

[51] A.J. McDougall, D.S. Stoffer, and D.E. Tyler. Optimal transformations and the spectral envelope for real-valued time series. *Journal of Statistical Planning and Inference*, 57(2):195–214, 1997.

[52] G. McLachlan, K. Do, and C. Ambroise. *Analyzing Microarray Gene Expression Data*. Wiley-Interscience, 2004.

[53] G.N. Nair. A nonstochastic information theory for communication and state estimation. *IEEE Transactions on Automatic Control*, 58(6):1497–1510, 2013.

[54] D.S. Nam, C. Han, C.W. Jeong, and E.S. Yoon. Automatic construction of extended symptom-fault associations from the signed digraph. *Computers and Chemical Engineering*, 20(Supplement 1):S605–S610, 1996.

[55] L.A. Overbey and M.D. Todd. Dynamic system change detection using a modification of the transfer entropy. *Journal of Sound and Vibration*, 322(1-2):438–453, 2009.

[56] J. Petersen. Causal reasoning based on mfm. In *Proceedings of the Cognitive Systems Engineering in Process Control*, pages 36–43, Taejon, Korea, 2000.

[57] A. Rényi. On measures of entropy and information. In *Proceedings of 4th Berkeley Symposium on Mathematical Statistics and Probability*, pages 547–561, Berkeley, USA, 1960.

[58] B. Schölkopf, J.C. Platt, J. Shawe-taylor, A.J. Smola, and R.C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1471, 2001.

[59] T. Schreiber. Measuring information transfer. *Physical Review Letters*, 85(2):461–464, 2000.

[60] T. Schreiber and A. Schmitz. Surrogate time series. *Physica D*, 142(3–4):346–382, 2000.

[61] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

[62] A.K. Seth. A matlab toolbox for granger causal connectivity analysis. *Journal of Neuroscience Methods*, 186(2):262–273, 2010.

[63] C.E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, Champaign, Illinois, 1949.

[64] H. Shingin and Y. Ohta. Disturbance rejection with information constraints: Performance limitations of a scalar system for bounded and gaussian disturbances. *Automatica*, 48(6):1111–1116, 2012.

[65] B.W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London, 1986.

[66] D.S. Stoffer. Detecting common signals in multiple time series using the spectral envelope. *Journal of the American Statistical Association*, 94(448):1341–1356, 1998.

[67] D.S. Stoffer, D.E. Tyler, and A.J. McDougall. Spectral analysis for categorical time series: scaling and the spectral envelope. *Biometrika*, 80(3):611–622, 1993.

[68] D.S. Stoffer, D.E. Tyler, and D.A. Wendt. The spectral envelope and its applications. *Statistical Science*, 15(3):224–253, 2000.

[69] A.K. Tangirala, J. Kanodia, and S.L. Shah. Non-negative matrix factorization for detection and diagnosis of plantwide oscillations. *Industrial and Engineering Chemistry Research*, 46(3):801–817, 2007.

[70] A.K. Tangirala, S.L. Shah, and N.F. Thornhill. Pscmap: A new tool for plant-wide oscillation detection. *Journal of Process Control*, 15(8):931–941, 2005.

[71] J. Theiler, S. Eubank, A. Longtin, B. Galdrakian, and J.D. Farmer. Testing for nonlinearity in time series: the method of surrogate data. *Physica D: Nonlinear Phenomena*, 58(1–4):77–94, 1992.

[72] N.F. Thornhill. Finding the source of nonlinearity in a process with plant-wide oscillation. *IEEE Transactions on Control System Technology*, 13(3):434–443, 2005.

[73] N.F. Thornhill, J.W. Cox, and M.A. Paulonis. Diagnosis of plant-wide oscillation through data-driven analysis and process understanding. *Control Engineering Practice*, 11(12):1481–1490, 2003.

[74] N.F. Thornhill and A. Horch. Advances and new directions in plant-wide disturbance detection and diagnosis. *Control Engineering Practice*, 15(10):1196–1206, 2007.

[75] N.F. Thornhill, B. Huang, and H. Zhang. Detection of multiple oscillations in control loops. *Journal of Process Control*, 13(1):91–100, 2003.

[76] N.F. Thornhill, S.L. Shah, B. Huang, and B.A. Vishnubhotla. Spectral principal component analysis of dynamic process data. *Control Engineering Practice*, 10(8):833–846, 2002.

[77] N.H. Ulerich and G.J. Powers. On-line hazard aversion and fault diagnosis in chemical processes: the digraph + fault-tree method. *IEEE Transactions on Reliability*, 37(2):171–177, 1988.

[78] V.A. Vakorin, O.A. Krakovska, and A.R. McIntosh. Confounding effects of indirect connections on causality estimation. *Journal of Neuroscience Methods*, 184(1):152–160, 2009.

[79] R. Vicente, M. Wibral, M. Lindner, and G. Pipa. Transfer entropy-a model-free measure of effective connectivity for the neurosciences. *Journal of Computational Neuroscience*, 30(1):45–67, 2011.

[80] G. Weidl, A.L. Madsen, and S. Israelson. Applications of object-oriented bayesian networks for condition monitoring, root cause analysis and decision support on operation of complex continuous processes. *Computers and Chemical Engineering*, 29(9):1996–2009, 2005.

[81] N. Wiener. The theory of prediction. In E.F. Beckenbach, editor, *Modern Mathematics for Engineers*. McGraw-Hill, New York, 1956.

[82] Wikipedia. Entropy (information theory). `http://en.wikipedia.org/wiki/Shannon\_entropy`.

[83] A. Wills. Quadratic programming in c (qpc) toolbox (version 2.0). `http://sigpromu.org/quadprog/`.

[84] M. Winterhalder, B. Schelter, W. Hesse, K. Schwab, L. Leistriz, D. Klan, R. Bauer, J. Timmer, and H. Witte. Comparison of linear signal processing techniques to infer directed interactions in multivariate neural systems. *Signal Processing*, 85(11):2137–2160, 2005.

[85] C. Xia and J. Howell. Isolating multiple sources of plant-wide oscillations via spectral independent component analysis. *Control Engineering Practice*, 13(8):1027–1035, 2005.

[86] F. Yang, S.L. Shah, and D. Xiao. Signed directed graph based modeling and its validation from process knowledge and process data. *International Journal of Applied Mathematics and Computer Science*, 22(1):41–53, 2012.

[87] F. Yang and D. Xiao. Progress in root cause and fault propagation analysis of large-scale industrial processes. *Journal of Control Science and Engineering*, 2012:478373–1–478373–10, 2012.

[88] F. Yang, D. Xiao, and S.L. Shah. Signed directed graph-based hierarchical modelling and fault propagation analysis for large-scale systems. *IET Control Theory and Applications*, 7(4):537–550, 2013.

[89] R.W. Yeung. *Information Theory and Network Coding*. Springer, New York, USA, 2008.

[90] H. Yu, S. Lakshminarayanan, and V. Kariwala. Confirmation of control valve stiction in interacting systems. *Canadian Journal of Chemical Engineering*, 87(4):632–636, 2009.

[91] J. Yu. A nonlinear kernel gaussian mixture model based inferential monitoring approach for fault detection and diagnosis of chemical processes. *Chemical Engineering Science*, 68(1):506–519, 2012.

[92] T. Yuan and S.J. Qin. Root cause diagnosis of plant-wide oscillations using granger causality. In *Proceedings of the 8th IFAC Symposium on Advanced Control of Chemical Processes*, pages 160–165, Furama Riverfront, Singapore, 2012.

[93] X. Zang and J. Howell. Isolating the root cause of propagated oscillations in process plants. *International Journal of Adaptive Control Signal Processing*, 19(4):247–265, 2005.

[94] W. Zheng, D. Tian, X. Wang, W. Tian, H. Zhang, S. Jiang, G. He, Y. Zheng, and W. Qu. Support vector machine: Classifying and predicting mutagenicity of complex mixtures based on pollution profiles. *Toxicology*, 2013. Article in Press.

[95] C. Zou and J. Feng. Granger causality vs. dynamic bayesian network inference: a comparative study. *BMC Bioinformatics*, 10:122–1–122–17, 2009.